Extrinsic Calibration of an RGB Camera to a 3D Imaging Sonar

Aaron Marburg, Andy Stewart
Applied Physics Laboratory
University of Washington
Seattle, Washington 98105
Email: {amarburg,andy}@apl.washington.edu

***Abstract*—The introduction of low-cost RGB-depth (RGB-D) sensors have led to a diversity of algorithms for robust 3D scene reconstruction under controlled settings, but the underwater realization of such algorithms has been hampered by the constrained performance of most RGB-D sensors in water. We explore the possibility of fusing a point cloud generated from a high-frequency, mechanically scanned 3D imaging sonar with visual data from a camera to create a rich 3D representation of objects in the water column. A state-of-the-art algorithm for depth sensor-to-camera registration utilizing concurrent images of spherical targets is adapted, and the resulting alignment is used to combine sonar and visual imagery.**

I. Introduction

Low-cost sensors capable of simultaneously measuring photometric information and scene depth (so-called RGB-depth or RGB-D sensors) have revolutionized computer vision and robotics research in recent years. For applications ranging from object modeling and classification [1] to map-building and robot navigation [2], [3], the direct depth measurements provide a robust estimate of scene structure, while RGB imagery both reinforces those estimates and provide contextual information.

Underwater, the ability to fuse data from multiple sensors into a full and accurate representation of both the world around the robot and the robot's place in that world enables a broad range of applications from augmented situational awareness and haptic control for pilots of remotely-operated vehicles (ROVs) to fully autonmous manipulation and maneuvering. Unfortunately, the broad-field structured light approach used in many low-cost RGB-D sensors (e.g. the Microsoft Kinect) is ineffective underwater, where the water acts as both a scattering medium and strong attenuator of the projected light pattern [4].

Sonar systems have long been used in the ocean to acoustically measure time-of-flight or depth information. At lower frequencies, sonars can make measurements at ranges far exceeding the capabilities of visible light cameras, while high-frequency "imaging sonars" offer spatial ranges and resolutions comparable to vision. Such sonars offer an attractive set of properties orthogonal to those of a visible light camera. They are unaffected by turbidity and other causes of optical scattering in the water column, and do not require external lighting. In exchange, however, they typically have lower spatial resolution, and the accuracy of the resulting depth measurements are dependent on the acoustic characteristics of the local scene, the objects being imaged, and the intervening water column.

The fusion of video and depth data requires three accurate models: sensor models for both the camera and sonar which allows the projection of world points into each sensor's frame of reference, and a model for the extrinsic (geometric) relationship between the two sensors. This work adapts a state-of-the-art algorithm for the calibration and alignment of RGB-D sensors in air [5] to estimate the geometric alignment of a 3D imaging sonar to a video camera to create an opti-acoustic sensor pair.

II. Related Works

As in traditional camera intrinsic calibration, methods for extrinsic sensor calibration can be divided into explicit (or supervised) procedures which require a bespoke calibration target of known geometry, and unsupervised methods which use ad hoc features in the natural world (e.g., [6]). Explicit methods can be further differentiated by the proposed calibration target and the degree to which intrinsic sensor parameters are estimated in parallel with the extrinsic alignment. The simultaneous optimization of the intrinsic and extrinsic parameters requires relatively large data sets, which are readily available from the high-frame-rate video and depth sources (e.g. the Kinect) used in most research. As discussed below, however, the relatively low update rate of the sonar precludes this option.

A significant body of work has been built around extending the popular planar target or "checkerboard" calibration technique, presumably due both to the strong similarities between this problem and stereo camera extrinsic calibration, and also because it allows the intrinsic calibration of the camera from the same data set using an existing, field-tested algorithm [7]. Some rely on the detection of the outline of the calibration target [8]–[11], which necessitates a relatively precise depth sensor (e.g. a laser scanner) to accurately find the target edges. Others model the plane of the calibration target in the depth image [12]–[14] and match it to the estimated target location calculated as part of the camera's intrinsic calibration.

Robustly identifying point cloud features from the planar calibration target can be challenging. While the plane itself is relatively easy to identify, it offers few strong points for robust localization. The outline/edge/corners are sensed as a strong discontinuity in the depth field, the accuracy of which is highly dependent on the resolution of the depth sensor.

978-0-933957-43-5 ©2015 MTS

Minor modifications to the planar pattern can make it far more visible in the depth image, assuming the new pattern remains visible and localizable in the camera image. Jung et al. [15] build on the camera pose estimation technique proposed by Vogiatzis and Hernańdez [16] which uses a planar target with a randomized hole pattern. In the visible light image the holes are readily identified using model-based circle extraction, while in the depth image the holes can be found by first identifying the plane, then searching for circular exclusions from the plane.

The method considered here is inspired by Staranowicz et al. [5] who use correspondences between the centers of spherical targets. Such targets are both straightforward to acquire (e.g. a ball) and the sphere/circle boundary is straightforward to extract from both image and point cloud data using robust model-based approaches. Moreover, calculation of the sphere center will tend to average out isotropic errors in the localization of the sphere boundary.

There have been a number of attempts to register a visible light camera to a so-called 2D imaging sonar, as distinct from the 3D sonar used here. A 3D imaging sonar produces an array of narrow cross-section "pencil beams." The sonar is pointed directly at the target, and a point cloud can be generated by measuring the time of flight for the first/strongest return from each of the beams (directly analogous to a multibeam echo sounder). A 2D imaging sonar also produces an array of beams, but each beam is itself fan-shaped orthogonal to the main array. The sonar is placed at a glancing angle to the target and each beam produces an independent time vs. return strength sequence as the ensonification passes over the target (analogous to a sidescan sonar). Negahdaripour et al. [17], [18] have been highly successful in fusing visible camera images to 2D imaging sonar output by obliquely imaging a planar calibration target containing a grid of holes. Each hole in the plate produces a strong, independent return regardless of the orientation of the plate to the sonar.

Hurtos et al. [19] successfully perform extrinsic calibration between a camera and a multibeam sonar mounted to an ROV by imaging a planar checkerboard calibration target raised above bottom of a pool, allowing the automatic segmentation of the planar target from the background in the multibeam imagery. The case the multibeam data is treated as individual line scans as the ROV maneuvers over the target, not as a complete point cloud.

There have also been a number of efforts to adapt non-sonar depth measurement technologies to underwater imaging. While off-the-shelf structured light approaches like the Kinect are not suitable, visible-pattern whole-field approaches have been demonstrated under controlled conditions [4], with severe degradation with increasing turbidity. Alternatively, narrow-beam or line-pattern approaches have also shown some promise [20], [21].

III. Camera–Sonar Alignment

The goal of camera–sonar calibration is to estimate the geometric relationship between the two sensor frames. In this case the calibration is estimated strictly from sensor data without external measurements. Combining the calibration with system models for each sensor allows the direct projection of data from one sensor into the frame of the other.

For a visible light camera, a projection function $\mathbf{P}(\cdot)$ maps a world point $\mathbf{X} = (X, Y, Z)$ to a pixel location $\mathbf{x}_C$ in the image

$$\mathbf{x}_C = \mathbf{P}\left(\mathbf{R}_C \mathbf{X} + \mathbf{t}_C\right), \tag{1}$$

where $\mathbf{R}_C, \mathbf{t}_C$ are the rigid body rotation and translation resp. which transform the point from the world frame into the camera frame. This work uses the pinhole projection model with five-coefficient radial lens distortion

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \mathbf{R}_C \mathbf{X} + \mathbf{t}_C, \tag{2}$$

$$x' = X'/Z', \tag{3}$$
$$y' = Y'/Z', \tag{4}$$

$$r^2 = x'^2 + y'^2, \tag{5}$$
$$x'' = x'\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2p_1 x' y' + p_2(r^2 + 2x'^2), \tag{6}$$
$$y'' = y'\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_1(r^2 + 2y'^2) + 2p_2 x' y', \tag{7}$$

$$\mathbf{x}_C = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \begin{pmatrix} x'' \\ y'' \end{pmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}. \tag{8}$$

The camera intrinsic parameters include the focal lengths (f_x, f_y), the image center (c_x, c_y) and the lens distortion coefficients $(k_1, k_2, k_3, p_1, p_2)$. If these parameters are known *a priori*, the projection function can be inverted to find (x', y'), but the world 3-tuple (X', Y', Z') is only solvable to an unknown scale. This can be explicitly represented by decomposing $\mathbf{P}(\cdot)$ as

$$\mathbf{x}_C = \hat{\mathbf{P}}\left(\pi\left(\mathbf{R}_C \mathbf{X} + \mathbf{t}_C\right)\right), \tag{9}$$

with $\hat{\mathbf{P}}(\cdot)$ representing the invertible equations (5–8), while $\pi(\cdot)$ represents the projection given by equations (3–4).

For the results presented here, the sonar projection function was not explicitly modeled. Rather, sonar processing software was used to directly generate a metric point cloud, such that the sonar model is trivially

$$\mathbf{X}_S = \mathbf{R}_S \mathbf{X} + \mathbf{t}_S, \tag{10}$$

where $\mathbf{R}_S$ and $\mathbf{t}_S$ map the world point $\mathbf{X}$ into the sonar-aligned frame defined by the software.

A. Calibration by Point Alignment

If a world point is viewed by both the camera and the sonar, then the resulting camera/sonar imaged points will obey the equivalence:

$$\begin{aligned} \mathbf{x}_C &= \mathbf{P}\left(\mathbf{R}_C\mathbf{R}_S^{-1}\left(\mathbf{X}_S - \mathbf{t}_S\right) + \mathbf{t}_C\right), \\ &= \mathbf{P}\left({}_C\mathbf{R}_S\mathbf{X}_S + {}_C\mathbf{t}_S\right), \end{aligned} \tag{11}$$

with the sonar-to-camera rigid body transform

$${}_C\mathbf{R}_S = \mathbf{R}_C\mathbf{R}_S^{-1}, \tag{12}$$

$${}_C\mathbf{t}_S = -\mathbf{R}_C\mathbf{R}_S^{-1}\mathbf{t}_S + \mathbf{t}_C. \tag{13}$$

Note that when solved exclusively for ${}_C\mathbf{R}_S$ and ${}_C\mathbf{t}_S$ (i.e., $\mathbf{P}$ is known), equation (11) is equivalent to the perspective-n-point problem for determining camera pose from a set of known image-to-world point correspondences. As such, details of linear and minimal solutions, as well as pathological configurations can be drawn directly from the literature e.g. [22]–[24].

B. Non-linear optimization

Given a known sonar-to-image point correspondence $\mathbf{X}_S \sim \mathbf{x}_C$, the image reprojection error may be defined:

$$d(\mathbf{x}_C, \mathbf{X}_S) = \left\|\mathbf{x}_C - \mathbf{P}\left({}_C\mathbf{R}_S\mathbf{X}_S + {}_C\mathbf{t}_S\right)\right\|^2. \tag{14}$$

If the camera intrinsic coefficients are known, as is the case here, this is equivalent to:

$$\hat{d}(\mathbf{x}_C, \mathbf{X}_S) = \left\|\hat{\mathbf{P}}^{-1}(\mathbf{x}_C) - \pi\left({}_C\mathbf{R}_S\mathbf{X}_S + {}_C\mathbf{t}_S\right)\right\|^2. \tag{15}$$

If J such correspondences have been found, the total reprojection error is:

$$D(\mathbf{x}_{C1}, \ldots, \mathbf{x}_{CJ}, \mathbf{X}_{S1}, \ldots, \mathbf{X}_{SJ}) = \sum_{j=1}^{J} \hat{d}\left(\mathbf{x}_{Cj}, \mathbf{X}_{Sj}\right). \tag{16}$$

The optimal sonar-to-camera geometry will minimize this total error. Furthermore, if an estimate on the point localization error is available, (15) may be reformulated as a Mahalanobis distance, effectively solving the maximum-likelihood (MLE) estimate under the assumption of Gaussian noise. As this is a nonlinear least squares (NLSQ) cost function, this minimization can be addressed by a diversity of algorithms. In this case, an implementation was built around the Ceres nonlinear solver [25].

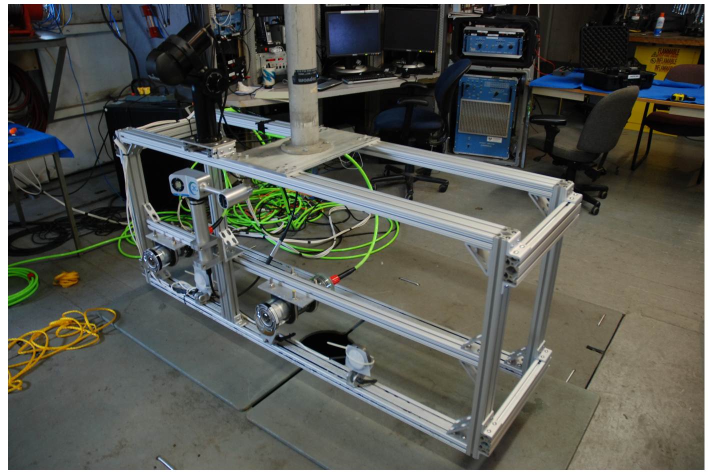


Fig. 1. Test frame with two HD cameras on bottom support and Blueview BV5000-2250 scanning sonar mounted on pan/tilt mechanism on top support.

Fig. 2. Sample image of calibration target. Note strong backlighting and low water visibility. The dark shapes in the upper corners are the pontoon hulls of the test vessel.

IV. Test Configuration

Two HD (1920×1080) streaming cameras (identified as *cam1* and *cam2*), and a Teledyne Blueview BV5000-2250 sonar were rigidly mounted to a test frame (Figure 1). This frame was lowered approx. two meters below the surface of the Lake Washington Ship Canal, a fresh water channel, with the sensors pointing horizontally. A work area of approximately three meters on a side was available in front of the sensors. The work area was clear of all solid objects other than the pontoon hulls of the test vessel. During testing the Secchi depth was estimated at approx 4.5 m.[1]

The BV5000-2250 3D imaging sonar is a mechanically scanned multibeam sonar operating at 2.25 Mhz. It produces 256 $1^\circ \times 1^\circ$ beams in a 90° fan. For each ping, the sonar detects one (or more) of the strongest returns and translates those points into a 3D location in a sonar-centered coordinate frame given a pre-specified local sound speed.

In the standard configuration (as used here), the sonar is mounted such that the sonar "looks" sideways with the

[1] From the King County Major Lakes Monitoring program [26].

fan oriented vertically. As such, a single ping of the sonar returns a single vertical profile of the scene. To generate a 3D point cloud, the sonar head is then mechanically panned at speeds from 0.5 – 2°/sec, with pan rate determining horizontal point spacing. There is no limitation on the horizontal field of view of the sonar, allowing full 360° point cloud generation, although in this case the horizontal sweep was limited to 90° to roughly encompass the same volume as the overlapping fields of view of the two cameras. Each point cloud requires 45–180 seconds to capture depending on the desired horizontal spatial resolution, due to the need to pan the head.

Timestamped video was recorded from both cameras concurrently. The target was moved within the scene, and held static to the best extent possible while scanning the sonar. While Staranowicz's original method proposed simultaneous estimation of the camera intrinsic properties and the camera-to-sonar registration, the large collection time for each sonar data point made this infeasible. Instead, the intrinsic coefficients for each camera was estimated independently using a standard planar target-based calibration routine. These images were also used to estimate the stereo baseline between the two cameras, again using standard techniques [24].

V. Generation of point correspondences from a spherical target

From [5], camera-to-sonar correspondences were generated by simultaneously viewing a spherical target with both sensors. In this case, Blueview had independently identified an open mesh sphere as an ideal target for repeatable and robust localization by their sonar. An approximately 22 cm diameter sphere was provided by Blueview for testing.

For each sonar scan, a single representative image was extracted from each camera's video stream. A background image was also generated for each camera by capturing the average pixel color for each pixel in a 30-second snippet of video without the sphere. An approximate sphere location was found in each image by identifying areas of high per-pixel difference relative to the averaged background image. The sphere outline itself was then refined within that region by Canny edge detection followed by Hough-voting circle detection. The center of the circle was used as the imaged sphere center $\mathbf{x}_C$.

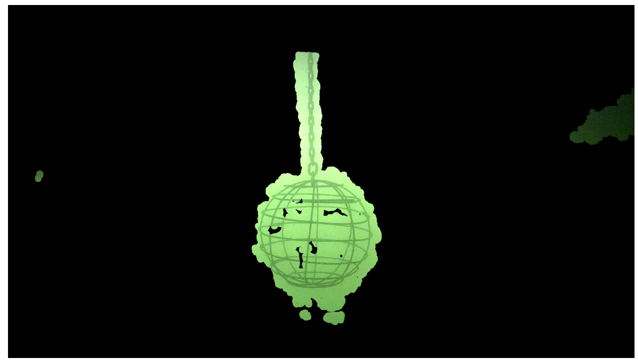
The sphere center location $\mathbf{X}_S$ was identified in the sonar point cloud data using a RANSAC-based robust model-matching scheme. As the test scene consisted primarily of open water, there was minimal background data other than the sphere and spurious noise — which might be symptomatic of the lack of a hard returning surface. Both the vision- and sonar-derived sphere locations were subsequently validated by hand.

VI. Results

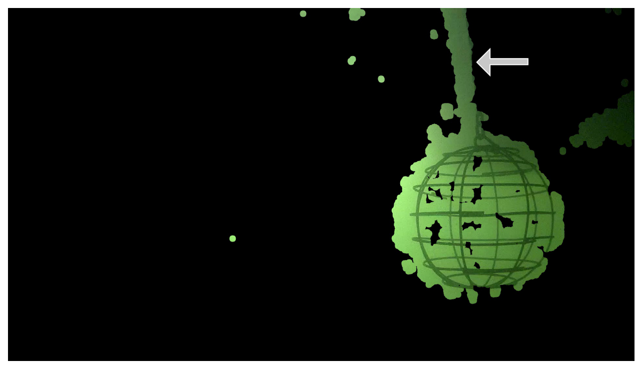
The sphere was imaged in nine distinct locations. For each, the sphere locations were calculated using the algorithms described in Section V, and a camera-to-sonar alignment was generated for each camera using the minimization described in Section III, while the camera-to-camera baseline was estimated using the standard planar-calibration-target method. As a ground truth, the length of the two camera-to-sonar baselines $|{}_C\mathbf{t}_S|$, and the camera-to-camera stereo baseline $|\mathbf{t}_{st}|$ were measured directly on the the test frame. The resulting baseline lengths are detailed in Table I. Precise measurement of the sensor-to-sensor baselines was challenging as the sensor origins did not necessarily correspond to a physical point on the sonar or camera housings.

TABLE I. Measured and estimated baselines between camera and sonar reference frames, and between cameras. All lengths in mm.

	cam1–sonar	**cam2–sonar**	**cam1–cam2**
Estimated length $|{}_C\mathbf{t}_S|, |\mathbf{t}_{st}|$	844.79	896.03	707.50
Measured (ground truth) length	865.57	862.50	743.74
Error	2.4%	3.9%	4.9%
Assoc. length $|\hat{\mathbf{t}}|$	848.27	894.14	—
Error	0.41%	0.21%	—
Error length $|\mathbf{t} - \hat{\mathbf{t}}|$	42.84	66.74	—
Incl. angle $\angle(\mathbf{t}, \hat{\mathbf{t}})$	2.89°	4.27°	—



(a) Foreground segmentation of Figure 2 using the sonar point cloud.



(b) The segmentation of an alternative calibration image. Note that the image of the chain does not align with sonar projection (indicated by arrow).

Fig. 3. Results from the segmentation of the sphere from images by reprojection of the sonar point cloud.

If the three coordinate transforms (*cam1*-to-sonar, *cam2*-to-sonar, and *cam1*-to-*cam2*) are internally consistent, any single transform should be equal to the result of chaining the other two. That is, the transform $({}_1\mathbf{R}_S, {}_1\mathbf{t}_S)$, which gives the locations of the sonar in the frame of *cam1*, should be equal to the concatenation of $({}_2\mathbf{R}_S, {}_2\mathbf{t}_S)$, the sonar origin in the frame of *cam2*, with $(\mathbf{R}_{st}, \mathbf{t}_{st})$, the stereo baseline from *cam2* to *cam1*. Similarly, the transform from *cam2* to the sonar can also be generated by combining the *cam1*-to-sonar transform

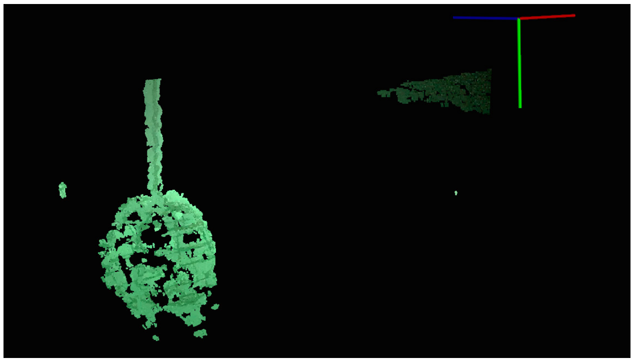


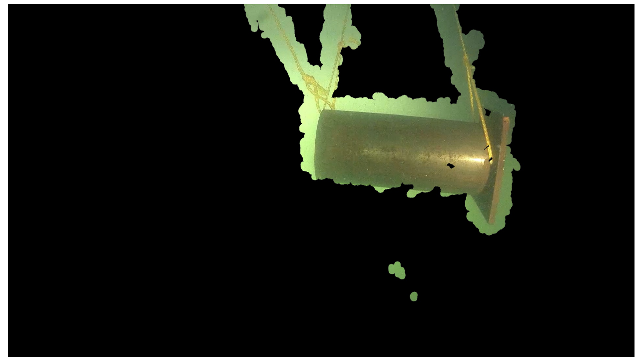
Fig. 4. A view of the point cloud corresponding to Figure 3(a). Points in the cloud are colored by sampling from the corresponding image.

and the stereo baseline. These translation vector $\hat{\mathbf{t}}$ for these *associative transforms* are included in Table I. As shown, the differences in magnitude between each sonar-to-camera baseline and its associative partner are quite small, indicated the three baseline lengths are internally consistent. However, examination of the length and included angle for the vector error between each baseline and its associative partner $|\mathbf{t} - \hat{\mathbf{t}}|$ reveals pointing errors between the two transforms.

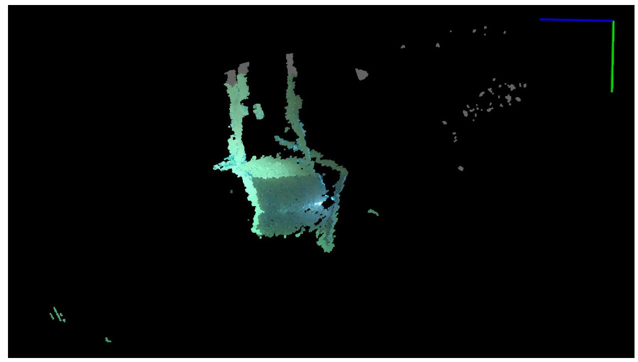
The resulting alignments can also be evaluated qualitatively. Camera-to-sonar registration allows any point in the sonar frame to be projected to a pixel location in the camera image by (11). Given the estimated registration, the sonar point cloud can be used to segment out the image pixels which correspond to the sphere and its chain. For example, Figure 3(a) is generated from Figure 2 by retaining pixels which lie within a fixed radius of a reprojected sonar point. Note however that the point cloud covers a larger physical extent than the sphere. This might be due to a scale error in the registration (arising from a consistent scale error in sphere extraction from either modality), or it might be due to the relatively coarse sonar resolution or the acoustic properties of the sphere itself. Figure 4 shows the converse representation, with points in the point cloud corresponding to Figure 2 colored by sampling the associated image pixels.

Unfortunately, the resulting registration is not perfect for all image-sonar pairs. Figure 3(b) shows another image from the calibration set segmented using the projection of the point cloud where the image of the sphere is offset slightly to the right within the point cloud. This bias causes the image of the chain to move outside of the reprojected sonar points (as indicated by the arrow). This mis-registration may be due to the relative paucity of sonar-camera calibration data, or systematic biases in the extraction of the sphere locations from the two data streams.

Once calculated, the image-sonar registration should remain valid for any sonar-camera image pair as long as the physical relationship between the two sensors is maintained. Figure 5(a) shows the segmentation of an image of a steel pipe with a square flange, while Figure 5(b) shows the corresponding point cloud.



(a) Segmentation of the flange assembly using the camera-to-sonar alignment generated in section VI.



(b) The point cloud for the flange assembly.

Fig. 5. Use of the calculated camera-sonar calibration to visualize a steel pipe-and-flange assembly. Note how the point cloud accurately represents the curvature of the pipe section and the square flange on the right end of the assembly.

VII. Conclusion

The ability to accurately register sensors from two different modalities enables the fusion of data from those sensors. Here, initial experiments demonstrate the merging of RGB images with point clouds generated by an 3D imaging sonar. In the examples shown here, the sonar data can be used to identify object pixels in the images, although the resulting areas are conservatively large due to the sonar's sensing behavior. Further processing will allow more tightly–integrated combinations of data, leading to more robust algorithms for underwater object and scene reconstruction.

The relatively slow update rate of the sonar places it at a severe disadvantage when observing dynamic scenes. By comparison, the camera provides high-bandwidth scene information but is unable to directly estimate scene depth and is susceptible to turbidity and other image distortions. In combination, the sonar might provide an initial metric scene structure while the video contents could then be used to track high-dynamics changes. Similarly, the sonar could provide infrequent or targeted "fill-in" depth updates to correct distortions in a dynamic world model.

Acknowledgments

The authors would like to thank Teledyne Blueview for the loan of the sonar and Bluhaptics Inc. for supporting the testing. They would also like to thank Benjamin Brand, Ryan Cox, and Paul Gibbs for their assistance with the preparation of the camera frame and in-water testing.

REFERENCES

[1] X. Ren, D. Fox, and K. Konolige, "Change Their Perception: RGB-D for 3-D Modeling and Recognition," *Robotics & Automation Magazine, IEEE*, vol. 20, no. 4, pp. 49–59, 2013.

[2] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pp. 127–136, 2011.

[3] T. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. J. Leonard, and J. McDonald, "Real-time large-scale dense RGB-D SLAM with volumetric fusion," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 598–626, Apr. 2015.

[4] F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A. V. Razionale, "Experimentation of structured light and stereo vision for underwater 3D reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 4, pp. 508–518, Jul. 2011.

[5] A. Staranowicz, G. R. Brown, F. Morbidi, and G.-L. Mariottini, "Easy-to-Use and Accurate Calibration of RGB-D Cameras from Spheres," in *Image and Video Technology*. Berlin Heidelberg: Springer Berlin Heidelberg, 2014, pp. 265–278.

[6] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes," in *Intelligent Robots and Systems, 2007.IEEE/RSJ International Conference on.* IEEE, 2007, pp. 4164–4169.

[7] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*. IEEE, 1999, pp. 666–673 vol.1.

[8] I. V. Mikhelson, P. G. Lee, A. V. Sahakian, Y. Wu, and A. K. Katsaggelos, "Automatic, fast, online calibration between depth and color cameras," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 218–226, Jan. 2014.

[9] D. H. C, J. Kannala, and J. Heikkila, "Accurate and Practical Calibration of a Depth and Color Camera Pair," in *Computer Analysis of Images and Patterns*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 437–445.

[10] D. Herrera C, J. Kannala, and J. Heikkila, "Joint Depth and Color Camera Calibration with Distortion Correction," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 2058–2064, 2012.

[11] C. Raposo, J. P. Barreto, and U. Nunes, "Fast and Accurate Calibration of a Kinect Sensor," in *2013 International Conference on 3D Vision (3DV)*. IEEE, 2013, pp. 342–349.

[12] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Intelligent Robots and Systems, 2004. IEEE/RSJ International Conference on.* IEEE, 2004, pp. 2301–2306.

[13] R. Unnikrishnan and M. Hebert, "Fast Extrinsic Calibration of a Laser Rangefinder to a Camera," Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-05-09, Jul. 2005.

[14] C. Zhang and Z. Zhang, "Calibration Between Depth and Color Sensors for Commodity Depth Cameras," in *Computer Vision and Machine Learning with RGB-D Sensors*. Cham: Springer International Publishing, 2014, pp. 47–64.

[15] J. Jung, Y. Jeong, J. Park, H. Ha, J. D. Kim, and I.-S. Kweon, "A novel 2.5D pattern for extrinsic calibration of tof and camera fusion system," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 3290–3296.

[16] G. Vogiatzis and C. Hernández. Automatic camera pose estimation from dot pattern. [Online]. Available: http://george-vogiatzis.org/calib/

[17] S. Negahdaripour, "A new method for calibration of an opti-acoustic stereo imaging system," in *OCEANS 2010*. IEEE, 2010, pp. 1–7.

[18] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-Acoustic Stereo Imaging: On System Calibration and 3-D Target Reconstruction," *Image Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1203–1214, Jun. 2009.

[19] N. Hurtos, X. Cufí, and J. Salvi, "Calibration of optical camera coupled to acoustic multibeam for underwater 3D scene reconstruction," *OCEANS 2010*, pp. 1–7, 2010.

[20] C. Roman, G. Inglis, and J. Rutter, "Application of structured light imaging for high resolution mapping of underwater archaeological sites," *OCEANS 2010*, pp. 1–9, 2010.

[21] S. G. Narasimhan and S. K. Nayar, "Structured light methods for underwater imaging: light stripe scanning and photometric stereo," in *OCEANS 2005*. IEEE, 2005, pp. 2610–2617.

[22] L. Quan and Z. Lan, "Linear N-point camera pose determination," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 8, pp. 774–780, Aug. 1999.

[23] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.

[24] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press, 2003.

[25] S. Agarwal and K. Mierle. Ceres Solver. [Online]. Available: http://ceres-solver.org

[26] King County Major Lakes Monitoring. [Online]. Available: http://green2.kingcounty.gov/lakes/Default.aspx