Acoustic Camera–based 3D Measurement of Underwater Objects through Automated Extraction and Association of Feature Points*

Yonghoon Ji[1], Seungchul Kwak[2], Atsushi Yamashita[1], and Hajime Asama[1], *Member, IEEE*

Abstract—This paper presents a novel scheme for the three-dimensional (3D) reconstruction of underwater objects by using multiple acoustic views based on geometric and image processing approaches. Underwater tasks such as maintenance, ship hull inspection, and harbor surveillance require accurate underwater information. In such cases, 3D reconstructed information would greatly contribute to a better understanding of the underwater environment. Acoustic cameras are the most suitable sensors because they provide acoustic images with more accurate details than other sensors, even in turbid water. In order to enable 3D measurement, feature points of each acoustic image should be extracted and associated in advance. In a previous study, we proposed a 3D measurement method, but it was limited by the assumption of complete correspondence information between feature points. This new methodology establishes a 3D measurement model by automatically determining correspondences between feature points through the application of geometric constraints and extracting these points. The result of the real experiment demonstrated that the proposed framework can automatically perform 3D measurement tasks of underwater objects.

I. INTRODUCTION

Numerous underwater tasks such as monitoring, investigation, and exploration require accurate underwater object recognition techniques [1–3]. However, there are many underwater environments where hazards prohibit human access (e.g., the Fukushima Daiichi nuclear power station, which has been in crisis since the 2011 earthquake off the Pacific coast of Tōhoku in east Japan), but accessibility is essential for investigation and exploration. In such cases, unmanned exploration by utilizing underwater robots such as autonomous underwater vehicles (AUVs) or remotely operated underwater vehicles (ROVs) is necessary in order to measure the 3D information of underwater objects for better understanding of the underwater environment.

Optical cameras have been used in numerous studies for underwater simultaneous localization and mapping (SLAM) [4,5] and underwater investigation [6,7] because they can provide high-resolution images with high accuracy. However, optical cameras are limited in turbid or deep water. Obtaining clear images by using optical cameras in turbid water is not possible.

[1]Y. Ji, A, Yamashita, and H. Asama are with the Department of Precision Engineering, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.{ji, yamashita, asama}@robot.t.u-tokyo.ac.jp.

[2]S. Kwak is with PLK Technologies, 57-5 Yangsan-ro, Yeongdeungpo-gu, Seoul, Korea. ksc5101@gmail.com.
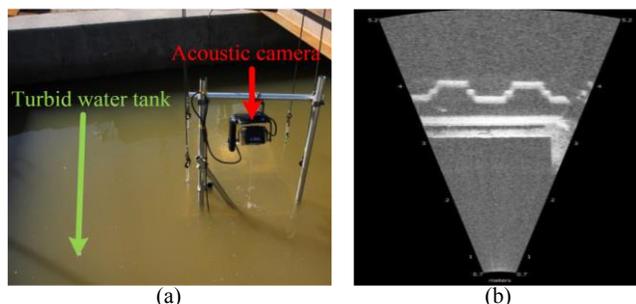
Figure 1. Because acoustic cameras provide acoustic images with accurate details, even in turbid water, they not only help with understanding the underwater situation but can also be applied to 3D measurement systems: (a) an acoustic camera system with ARIS in a turbid water tank and (b) an acoustic image of the turbid water tank bed.

Thus, acoustic cameras such as dual-frequency identification sonar (DIDSON) [8], adaptive resolution imaging sonar (ARIS) [9], and BlueView [10] are the most suitable sensors because they provide acoustic images with more accurate details even in turbid water than other sensors, as shown in Fig. 1. The high operating frequency of acoustic cameras (1.8–3.0 MHz in the case of ARIS EXPLORER 3000) allows them to provide high-resolution and wide-range images [11] to facilitate understanding of the underwater situation. However, there have been relatively few studies on establishing 3D measurement methodologies, despite their undeniable worth. Thus, there is a need to develop theoretical methodologies for the 3D measurement of underwater objects by using acoustic cameras.

Negahdaripour et al. and Majumder et al. conducted system calibration and 3D scene reconstruction by using opti-acoustic fusion systems [12–14]. Their approaches improve the accuracy of 3D reconstruction results for underwater objects by using a calibrated opti-acoustic stereo imaging system. However, their methods are still limited to clear water because they still rely on optical vision. Xu et al. mounted a concentrator lens on an acoustic camera [15] that allows acoustic camera beams to propagate further. Generally, an acoustic image from the acoustic camera does not include 3D information, whereas an acoustic camera with a concentrator lens can obtain 3D information directly within a narrow range by narrowing the field of view of the elevation angle (as described in the next section). However, this approach cannot sense a wide field range, which is a major benefit of acoustic cameras. A 3D measurement system for determining feature points on two acoustic images from different viewpoints was proposed in [16]; however, this system limits the acoustic cameras to vertical movement. A method for determining the 3D shape of a target from multiple acoustic images by using sequential carving of the non-target space was proposed in [17]. However, this method also limits
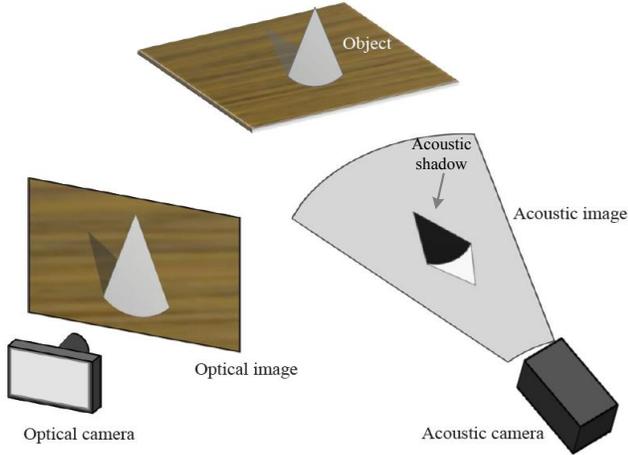
Figure 2. While optical cameras provide the same scenes as human sight, acoustic cameras provide significantly dissimilar images.

the movement of acoustic cameras to roll motions with only a few views. Consequently, these systems cannot determine the 3D shapes of underwater objects from arbitrary viewpoints.

Our research group previously proposed a novel method that reconstructs 3D feature points from multiple acoustic views of underwater objects [18]. When compared to previous research, our proposed methodology has the advantage in that it does not matter how far each viewpoint is from the others. In other words, the proposed methodology can successfully deal with the discontinuous frames of acoustic views, whereas the previous studies required obtaining successive frames of acoustic views. Huang et al. also recovered 3D feature points based on an optimization approach for the Bayesian SLAM [19]. However, both of the methods are restricted by the assumption of complete correspondence information between feature points because of difficulties with associating adjacent indistinguishable features. To overcome the limitation, this paper extends our previous work on using multiple acoustic views. In order to enable fully automated 3D measurement, the feature points of each acoustic image should be automatically extracted and associated in advance.

The contributions of this paper are as follows. Earlier approaches were limited by measuring 3D feature points under restrictive assumptions where operator instructions are required to define and associate feature points, as mentioned above. On the other hand, the proposed complete 3D measurement system in this paper requires no operator instructions for feature definition. In other words, the proposed 3D measurement methodology automatically determines correspondences between feature points by applying geometric constraints and extracting the points by using image processing techniques without any operator instructions.

The rest of this paper is organized as follows. Section II briefly introduces the principles of the acoustic cameras. Section III describes the 3D measurement model by using two acoustic images from different viewpoints. Section IV presents the feature extraction and association method for a fully automated system in detail. Section V discusses the effectiveness of the proposed 3D measurement framework based on the experimental results. Finally, section VI presents the conclusions and future works.

II. PRINCIPLES OF ACOUSTIC CAMERA

Acoustic cameras sense a 3D area in the same way as optical cameras; however, their output images are different, as shown in Fig. 2. This phenomenon can be explained by the models described in this section. These models elucidate the unique characteristics of acoustic cameras, including the signal processing and display mechanism. The principles of acoustic cameras are the basis for establishing the acoustic camera-imaging model that is represented in this section.

A. Acoustic projection model

As shown in Fig. 3 (a), acoustic cameras insonify acoustic waves in the forward direction to span the field of view with the azimuth angle θ_{cam} and elevation angle ϕ_{cam} within the scope of the maximum range r_{cam}. These parameters depend on the specifications of each acoustic camera. An insonified acoustic wave is propagated within the area of the determined sensing scope. After a forward-traveling acoustic wave hits underwater objects, it is reflected in different directions from the original propagation direction. The acoustic pressure of the acoustic wave diminishes with the travel distance and reflection.

As shown in Fig. 3 (b), acoustic cameras handle the sensing scope by using a few cross-sectional beams for signal processing. Each beam plane is perpendicular to the X_c-Y_c plane. The number of beams and the beam width also depend

(a)   (b)
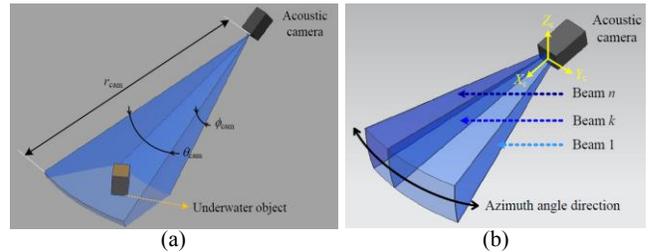
Figure 3. Acoustic projection model: (a) acoustic cameras insonify acoustic waves in the forward direction by spanning the field of view with the azimuth angle θ_{cam} and elevation angle ϕ_{cam} within the scope of the maximum range r_{cam}; (b) acoustic cameras handle the sensing scope by using a few cross-sectional beam slices (beam k, $k = 1, …, n$) for signal processing. Each beam plane is perpendicular to the X_c-Y_c plane, and the number of beams and the beam width depend on the specifications of the acoustic camera.

(a)   (b)
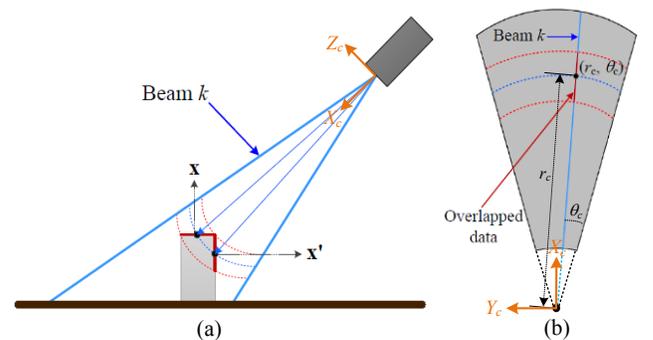
Figure 4. Overlapping phenomenon: (a) cross-sectional diagram of beam k and (b) pixels where acoustic pressure data overlap (red solid line). Because multiple acoustic pressure values overlap on the acoustic image when the detected points are located at the same distance from the acoustic camera in identical beam slices, the acoustic pressure values regarding the measurement points $\mathbf{x}$ and $\mathbf{x'}$ are output on the same pixel (r_c, θ_c).

on the specifications of the acoustic camera. Because the beam slices are arranged in the direction of the azimuth angle, the beam index is related to the value of the azimuth angle. This is clearly described in the next subsection.

B. Imaging geometry model

Although an acoustic camera senses a 3D area determined by r_{cam}, θ_{cam}, and ϕ_{cam}, the output of the sensing process is a 2D acoustic image. This is because reflected acoustic waves are processed by an array of transducers as a function of the range and azimuth angle; the elevation angle is not included. In other words, the coordinate system of an acoustic image comprises the range and azimuth angle (r_c, θ_c). For instance, although the measurement point **x** is identified by the camera coordinate system (r_c, θ_c, ϕ_c), the elevation angle value ϕ_c does not matter when forming the acoustic image, as shown in Fig. 4.

This subsection focuses on the key aspects of the problem of omitting elevation angle values. Note that pixel values of an acoustic image originate from the acoustic pressure p. The diminished acoustic pressure value of acoustic waves detected at the measurement point **x** as they travel is output to the pixel corresponding to the range and azimuth angle coordinates (r_c, θ_c) of **x**. Thus, if there is another measurement point **x′** which has the same range and azimuth angle as **x** but a different elevation angle, the acoustic pressures at **x** and **x′** are output at the same pixel corresponding to (r_c, θ_c). This is shown in Fig. 4.

Thus, acoustic images are represented by three parameters: r, θ, and p. This means that pixels on acoustic images corresponding to the area where acoustic waves cannot travel (e.g., behind underwater objects) have no acoustic pressure values, so they rendered black in color (a black area in an acoustic image is called an acoustic shadow, as shown in Fig. 2). In addition, if multiple acoustic waves travel the same distance in identical beam slices, the numbers of acoustic pressures regarding those acoustic waves overlap at the same pixel on the acoustic image. For instance, in Fig. 4, because the upper side and part of the right side of the object are located at the same distance from the acoustic camera in identical beam slices, multiple acoustic pressure values overlap on the acoustic image. When the data overlap, the displayed pixel values are the aggregate of acoustic pressure values at each measurement point.

However, the pixel value is not a raw acoustic pressure value but a scale-converted color. Because the acoustic image is represented in grayscale, the applicable value mapped to each pixel is in grayscale. Each acoustic pressure value is rendered based on the correspondence between the range of all acoustic pressure values and the grayscale.

III. OVERVIEW OF 3D MEASUREMENT MODEL

Acoustic images are represented by three parameters: r, θ, and p. In other words, the values for the range r and azimuth angle θ are directly obtained from the pixel coordinate of acoustic image, but the elevation angle ϕ is missing. Therefore, it is generally impossible to recover the 3D information of objects by using an acoustic image. This section briefly explains the proposed theoretical methodology to recover 3D coordinates from two acoustic images from different viewpoints [18].

First, extracting feature points from an acoustic image in advance is an important task. Feature points indicate distinguishable points on acoustic images, such as corners or areas whose materials differ from each other because sound energy changes rapidly with such structures. The method for extracting and associating such feature points is described in detail in the next section.

It is impossible to obtain 3D coordinates of each feature point from an acoustic image, as mentioned above. Fortunately, however, candidates $^{Vn}\mathbf{P}^{Car}$ with respect to the global Cartesian coordinate frame can be predicted as follows:

$$^{Vn}\mathbf{P}^{Car} = \left[^{Vn}\mathbf{p}_i^{Car} \mid 1 < i < N \right]^{T} \tag{1}$$

$$^{Vn}\mathbf{p}_i^{Car} = \left[^{Vn}x_i \quad ^{Vn}y_i \quad ^{Vn}z_i \right]^{T} \tag{2}$$

$$^{Vn}\mathbf{p}_i = \left[^{Vn}r \quad ^{Vn}\theta \quad ^{Vn}\phi_i \right]^{T} \tag{3}$$

where N indicates the number of candidates for the viewpoint n. $^{Vn}\phi_1$ and $^{Vn}\phi_N$ are respectively zero and ϕ_{cam} (maximum elevation angle). Therefore, a larger number of candidates N improves the accuracy because the candidate points are denser given that it represents a discretization resolution. The superscript Vn represents a viewpoint (V1 or V2 in this case). Here, each $^{Vn}\mathbf{p}_i^{Car}$ is calculated from each $^{Vn}\mathbf{p}_i$ based on the relationship between the spherical and Cartesian coordinate frames as follows.

$$^{Vn}\mathbf{p}_i^{Car} = \begin{bmatrix} ^{Vn}r\cos(^{Vn}\phi_i)\sin(^{Vn}\theta) \\ ^{Vn}r\cos(^{Vn}\phi_i)\cos(^{Vn}\theta) \\ ^{Vn}r\sin(^{Vn}\phi_i) \end{bmatrix} \tag{4}$$

$$^{Vn}\mathbf{p}_i = \begin{bmatrix} \sqrt{(^{Vn}x_i)^2 + (^{Vn}y_i)^2 + (^{Vn}z_i)^2} \\ \arctan(^{Vn}y_i / ^{Vn}x_i) \\ \arctan(^{Vn}z_i / \sqrt{(^{Vn}x_i)^2 + (^{Vn}y_i)^2}) \end{bmatrix} \tag{5}$$

After the above process is applied to acoustic images of each viewpoint, the 3D coordinates of the feature point $\hat{\mathbf{p}}^{Car}$ can be determined by finding an arc–arc intersection between the two candidates $^{V1}\mathbf{P}^{Car}$ and $^{V2}\mathbf{P}^{Car}$. Note that the set of candidates $^{Vn}\mathbf{P}^{Car}$ is in the shape of an arc (as shown in blue dotted line in Fig. 4 (a)). Because an intersection point may not exist due to various noises, the minimum distance between two arcs is calculated in actual implementation as follows:
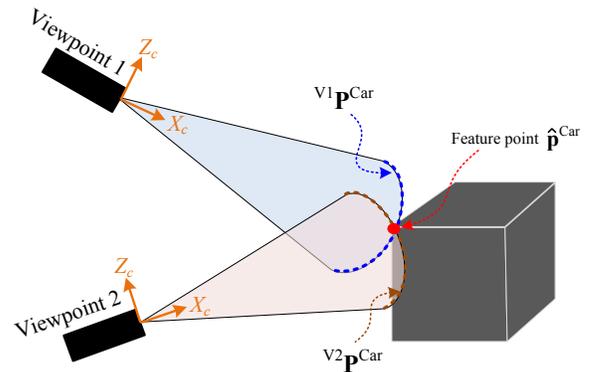
Figure 5. Conceptual image of the 3D measurement model using two acoustic viewpoints.
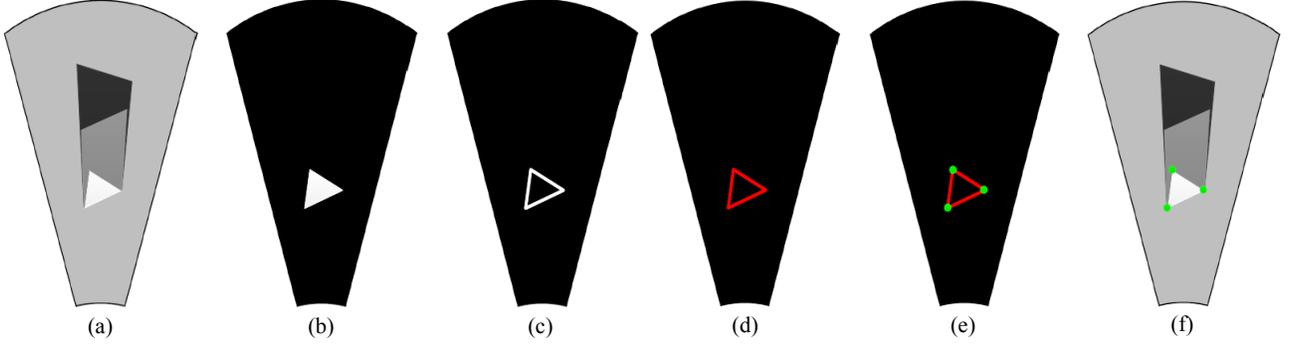
Figure 6. Process to extract feature points from an acoustic image: (a) original image, (b) designation of ROI, (c) contours of the image extracted by Canny edge detection, (d) extraction of lines with a high level of confidence by the probabilistic Hough transform, (e) extraction of endpoints of the lines as feature points, and (f) the feature points on the original image.

$$\hat{\mathbf{p}}^{Car} = \frac{1}{2}\left[{}^{V1}\mathbf{p}_{i*}^{Car} + {}^{V2}\mathbf{p}_{j*}^{Car}\right] \qquad (6)$$

$$(i*, j*) = \arg\min_{i,j}\left[\text{dist}({}^{V1}\mathbf{p}_i^{Car},\ {}^{V2}\mathbf{p}_j^{Car})\right] \qquad (7)$$

Here, Euclidian distance function dist(·) is applied to calculate minimum distance between two arcs. Figure 5 depicts a conceptual image of the 3D measurement model using the two acoustic viewpoints mentioned above.

IV. FEATURE EXTRACTION OF ACOUSTIC IMAGES

The previous section briefly presented a methodology for the 3D measurement of underwater objects by using two different acoustic views. However, the premise that underwater objects are known objects is indispensable not only to extract feature points from each acoustic image but also to determine the correspondences between them. To apply the 3D measurement methodology to unknown objects, an automated system for determining the correspondences between feature points from each acoustic image and extracting them is necessary. Therefore, a novel methodology is proposed in this section for automated feature extraction and association that is applicable to 3D measurement of unknown underwater objects. The approach assumes that extracted features in one viewpoint are visible from another viewpoint.

A. Extraction

In the proposed methodology, feature points indicate distinguishable points where acoustic pressure values change rapidly in the acoustic images. In the fields of computer vision, many studies about point feature extraction for the optical image, which includes SIFT [20] or SURF [21], have been proposed. However, these do not work well in the acoustic images given that the configuration of the acoustic image is completely different from the optical camera image as described in section II.

Our methodology limits the region of interest (ROI) to the area where acoustic pressure values are highest (white areas in acoustic images). Figure 6 shows the entire feature point extraction process. To improve the accuracy of the feature point extraction, the acoustic images are mapped in black except for the ROI (Fig. 6 (b)). Then, the contours of the images are extracted by Canny edge detection (Fig. 6 (c)). Not only the lines with a high level of confidence but also the endpoints of the line segments are extracted by probabilistic

Hough transform [22] after contour detection (Figs. 6 (d) and (e)). The endpoints of the extracted line segments are utilized as feature points because they must be theoretically distinguishable on acoustic images.

B. Association

A geometric design is used to find correspondences between feature points on each acoustic image. As described in the previous section, feature point candidates, which presumably include the real coordinates, make part of an arc.
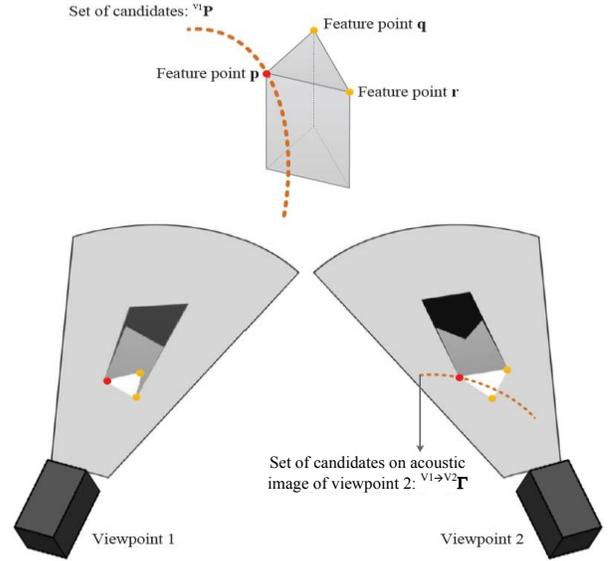
Figure 7. Conceptual image of the geometric design to find correspondences between feature points on each acoustic image.

TABLE I. SPECIFICATIONS OF ARIS EXPLORER 3000

Specification item [Unit]	Value
Identification frequency [MHz]	3.0
Identification range scope r_{cam} [m]	5.0
Azimuth angle θ_{cam} [deg]	32.0
Elevation angle ϕ_{cam} [deg]	14.0
Field of view [deg × deg]	32.0 × 14.0
Beam width [deg]	0.25
Number of transducer beams	128
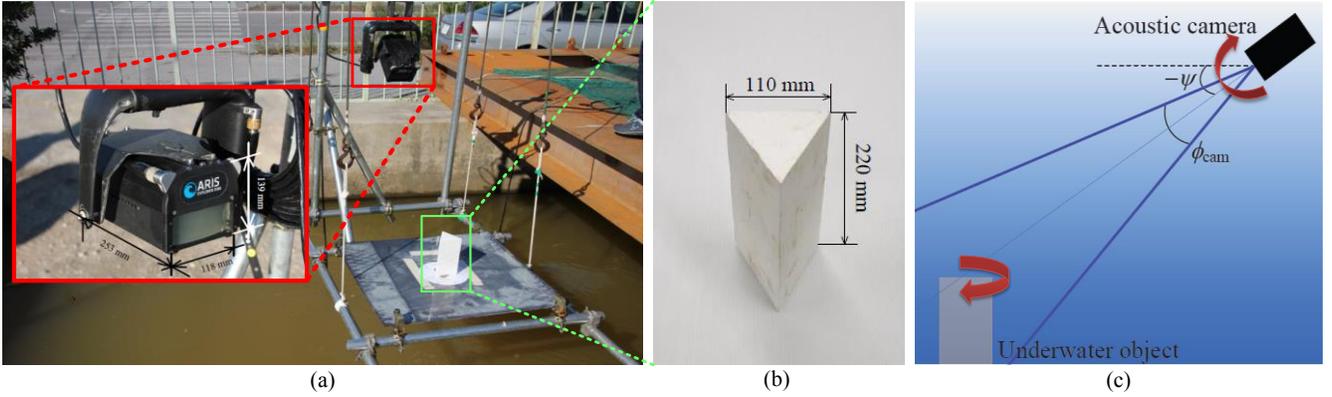
Figure 8. Experimental setup: (a) acoustic camera ARIS EXPLORER 3000 fixed to steel bar, (b) triangular prism used in the experiment, and (c) a conceptual image of environment settings for the experiment.
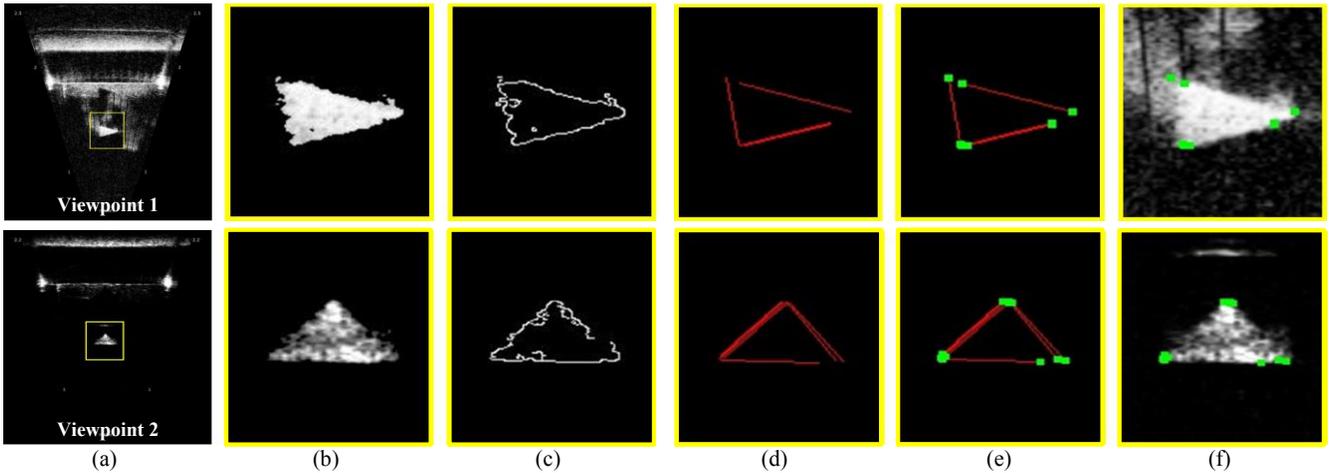
Figure 9. Feature extraction procedure and results from the triangular prism with respect to viewpoint 1 (top) and viewpoint 2 (bottom). Yellow squares represent the ROI: (a) original images, (b) designation of ROIs, (c) contours of the images extracted by Canny edge detection, (d) extraction of lines with a high level of confidence by probabilistic Hough transform, (e) extraction of endpoints of the lines as feature points, and (f) feature points on the original images.

The proposed feature association methodology utilizes the imaginary arcs comprising the candidates. Figure 7 shows an example of finding correspondences between feature points. The set of candidates for the feature point $\mathbf{p}$ from viewpoint 1 is represented by $^{v1}\mathbf{P}^{Car}$, as given in Eq. (1). Next, the imaginary arcs made from viewpoint 1 (i.e., the set of candidates for the feature point $^{v1}\mathbf{P}$) are projected onto an acoustic image of viewpoint 2. The set of projected points on an acoustic image of viewpoint 2 is defined as $^{v1 \rightarrow v2}\Gamma$. With respect to an acoustic image from viewpoint 2, there must be at least one feature point which is on the curve $^{v1 \rightarrow v2}\Gamma$ because the feature point $\mathbf{p}$ that makes the arc is also taken from viewpoint 2.

However, more than one feature point (from viewpoint 2) is likely to be on the line $^{v1 \rightarrow v2}\Gamma$. In such cases, the index of the projected points that make up the line $^{v1 \rightarrow v2}\Gamma$ serves to distinguish whether or not the feature points from viewpoint 2 actually correspond with the feature point $\mathbf{p}$ from viewpoint 1. Here, measuring the real 3D coordinates of the candidate feature points greatly helps with the distinguishing because the calculation results of the minimum distances (described in the previous section) can be selection criteria. When the distance between both derived 3D coordinates is less than the threshold σ, the feature point on the line $^{v1 \rightarrow v2}\Gamma$ is determined as the

corresponding point with the feature point $\mathbf{p}$ from viewpoint 1. Here, the threshold σ is set to the maximum distance between feature points in a real environment when they have the same elevation angle. This is because of the uncertainty that 3D reconstructed feature points separated from vertices of underwater objects by more than that distance can be considered as vertices of underwater objects.

V. EXPERIMENTAL RESULTS

This section describes the results of a real data experiment performed with ARIS EXPLORER 3000, as shown in Fig. 8 (a), to validate the proposed methodologies. The detailed specifications of ARIS EXPLORER 300 used in this study are shown in Table I. A triangular prism (110L mm × 220H mm) was used as the underwater object, as shown in Fig. 8 (b). The acoustic camera was located higher than the objects to insonify acoustic waves at the objects diagonally. The distance from the prism to the acoustic camera was set to 1.35 m, and two acoustic images were acquired at different viewpoints by rotating the target object and the pitch angle of the acoustic camera ψ. Changing the acoustic camera poses allowed different incidence angles to be obtained at the same points on the underwater object. Figure 8 (c) shows conceptual images of the changing the viewpoint.
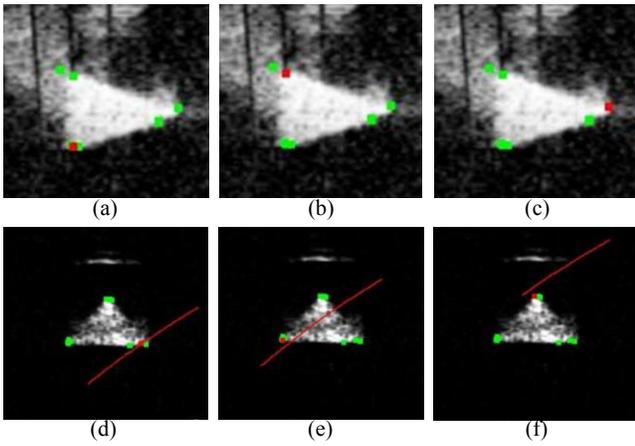
Figure 10. Finding correspondences between feature points: (a), (b), and (c) real acoustic images of the triangular prism with respect to viewpoint 1 and (d), (e), and (f) real acoustic images of the triangular prism with respect to viewpoint 2. The red point in (a) is a reference feature point to find the corresponding point in (d), and the red point in (d) is determined as corresponding by the red plotted line. Other correspondences can also be seen in (b), (e) and (c), (f).
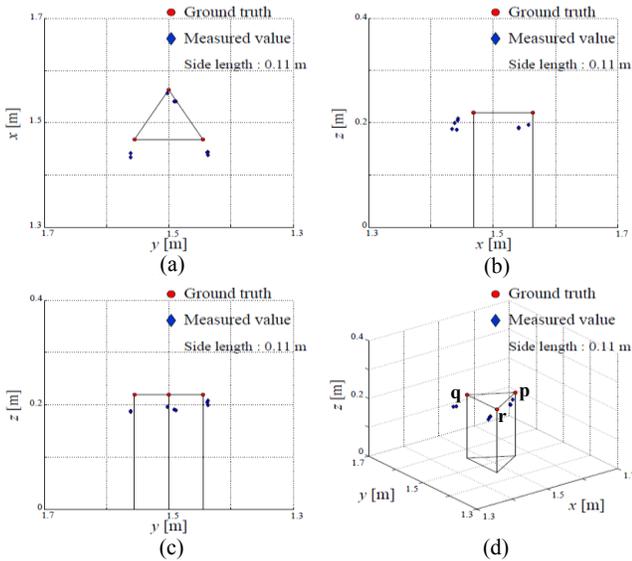
Figure 11. Experimental results for 3D measurement of the triangular prism. The 3D coordinates of the triangular prism's three vertices were measured from the real experiment: (a) 2D floorplan drawn on the *x-y* plane, (b) 2-D floorplan drawn on the *x-z* plane, (c) 2D floorplan drawn on the *y-z* plane, and (d) projection plan. The red circles indicate the ground truth, and the blue diamonds indicate the measured values.

Figure 9 (top) shows the feature extraction procedure and results for the triangular prism with respect to one viewpoint. Note that the corresponding conceptual images are illustrated in Fig. 6. Point features were extracted from the corners of the triangular prism. Similarly, the feature extraction procedure with respect to another viewpoint was also performed, as shown in Fig. 9 (bottom).

Figure 10 shows the results of finding correspondences between feature points using the method presented in subsection IV.B. In this experiment, two extracted feature points with the largest distance between them among all extracted feature points were selected to determine the threshold σ. The distance between the selected two feature

points in the real space when given the same elevation angle value was adopted as the value of threshold σ. The results (Figs. 9 and 10) showed that the proposed method described in section IV not only extracted feature points from each acoustic image but also determined the correspondences between them. Thus, it is applicable to 3D measurements, even for unknown objects.

Next, the 3D measurement of feature points was performed by using the proposed methodology described in section III. Figure 11 presents the results of the 3D reconstructed feature points with the ground truth. Table II compares the measured values with the ground truth and lists the root mean square errors. The results showed that the proposed methodologies are effective for the 3D measurement of underwater objects. However, as indicated in Table II, errors of approximately 0.014–0.027 m occurred.

The errors in the real data experiments can be explained as follows. The first problem was with the range measurement, which is strongly influenced by errors in acoustic speed. The possible error of the measured acoustic speed is known to be 10 m/s, which is approximately 0.667 % of the actual acoustic speed. As a result, an error of approximately 0.01 m in the range can occur when an object is located 1.5 m away from the acoustic cameras. The second problem was the extraction of feature points on acoustic images. Although acoustic cameras have outstanding visibility even in dark or turbid water, problems such as noise, multipath reflection, and scattering remains unsolved. Moreover, when bordering beam slices have a greater difference in acoustic pressure than 3 dB, the acoustic pressure values tend to be output on the neighboring beam. Therefore, areas where materials or the incidence angle of acoustic waves are greatly different from others tend to interfere with the output results of bordering beam slices, even though such areas are simultaneously considered to be distinguishable points for measuring 3D information. There were problems with extracting feature points from the acoustic image in the real experiment; for instance, the edge of the triangle prism was rounded at the end, as shown in Fig. 9 (b). The third problem was finding the exact location of the transmitter (i.e., origin point of the generated acoustic waves) which needs precise camera localization. The location of the transmitter could be calculated by using the floorplan. However, even though the location of the transmitter was guessed from the position relationships, the exact location of the transmitter was difficult to calculate. This also reduced the accuracy of the 3D measurement. Consequently, problems with the range measurement, exact extraction of feature points on acoustic images, and finding the exact location of the transmitter can be associated with the occurrence of errors.

TABLE II.   EXPERIMENTAL RESULTS FOR 3D MEASUREMENT

Vertex	Ground truth (*x*, *y*, *z*) [m]	Measured value (*x*, *y*, *z*) [m]	RMSE [m]
p	(1.564, 1.500, 0.220)	(1.541, 1.491, 0.191)	0.022
		(1.557, 1.502, 0.196)	0.014
		(1.541, 1.488, 0.190)	0.023
q	(1.468, 1.555, 0.220)	(1.442, 1.561, 0.187)	0.025
		(1.434, 1.561, 0.188)	0.027
r	(1.468, 1.445, 0.220)	(1.443, 1.438, 0.205)	0.017
		(1.438, 1.437, 0.200)	0.021
		(1.444, 1.437, 0.208)	0.016
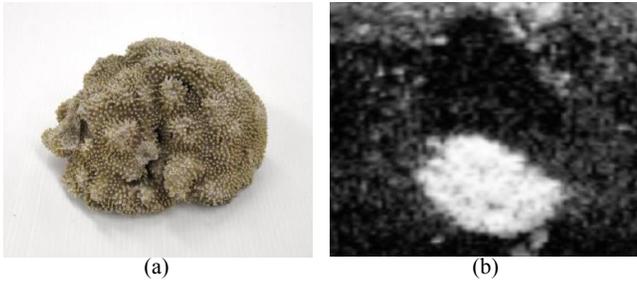
Figure 12. An example of difficult case to extract feature points from acoustic images: (a) an optical image of a coral and (b) a corresponding acoustic image.

VI. Conclusion

In this paper, a novel scheme for 3D measurement of the underwater object using two acoustic views is proposed. In order to achieve a complete 3D measurement system that can automatically extract feature points from acoustic images and determine correspondences between them, a geometric model and image processing techniques are adopted. These serve to prevent matching failures for feature points to obtain reliable 3D measurement results.

An experiment was performed to demonstrate an application of the proposed approach in a real environment with an acoustic camera. Our system can measure the 3D coordinates of feature points on an underwater object even in turbid water. Consequently, the proposed methodology can deal with unknown objects and greatly contribute to performing underwater tasks in real environments.

Future works related to this paper will involve extending our methodologies to manage the various causes of errors described in the previous section. Especially, precise localization scheme for each viewpoint of the acoustic camera should be established. Furthermore, there still remains future work on more general structures with smoother and complex shapes (e.g., natural objects as shown in Fig. 12) because the proposed methodology in this paper could manage the cases that the target object has acute corners (i.e., distinguishable feature points on acoustic images).

References

[1] D. R. Yoerger, A. M. Bradley, B. B. Walden, M. H. Cormier, and W. B. F. Ryan, "Fine-Scale Seafloor Survey in Rugged Deep-Ocean Terrain with an Autonomous Robot," in *Proc. 2000 IEEE Int. Conf. Robotics and Automation*, vol. 2, pp. 1767–1774, 2000.

[2] S. Williams and I. Mahon, "Simultaneous Localisation and Mapping on the Great Barrier Reef," in *Proc. 2004 IEEE Int. Conf. Robotics and Automation*, vol. 2, pp. 1771–1776, 2004.

[3] P. C. Burns, R. C. Ewing, and A. Navrotsky, "Nuclear Fuel in a Reactor Accident," *Science (Washington)*, vol. 335, no. 6073, pp. 1184–1188, 2012.

[4] R. Eustice, H. Singh, J. Leonard, M. Walter, and R. Ballard, "Visually Navigating the RMS Titanic with SLAM Information Filters," in *Proc. Robotics: Science and Systems*, pp. 57–64, 2005.

[5] R. Eustice, H. Singh, J. Leonard, and M. Walter, "Visually Mapping the RMS Titanic: Conservative Covariance Estimates for SLAM Information Filters," *Int. J. Robotics Research*, vol. 25, no. 12, pp. 1223–1242, 2006.

[6] Y. Zhang, R. S. McEwen, J. P. Ryan, J. G. Bellingham, and H. Thomas, "A Peak-Capture Algorithm Used on an Autonomous Underwater Vehicle in the 2010 Gulf of Mexico Oil Spill Response Scientific Survey," *J. Field Robotics*, vol. 28, no. 4, pp. 484–496, 2011.

[7] A. Shibata, H. Fujii, A. Yamashita, and H. Asama, "Scale-Reconstructable Structure from Motion Using Refraction with a Single Camera," in *Proc. 2015 IEEE Int. Conf. Robotics and Automation*, pp. 5239–5244, 2015.

[8] E. Belcher, W. Hanot, and J. Burch, "Dual-Frequency Identification Sonar (DIDSON)," in *Proc. 2002 IEEE Int. Symposium Underwater Technology*, pp. 187–192, 2002.

[9] "ARIS," retrieved August 1, 2016, from http://www.soundmetrics.com /Products/ARIS-Sonars/ARIS-Explorer-3000.

[10] "Blueview," retrieved August 1, 2016, from http://www.blueview.com /products/2d-imaging-sonar/p900-2250-dual-frequency.

[11] N. O. Handegard and K. Williams, "Automated Tracking of Fish in Trawls Using the DIDSON (Dual Frequency IDentification SONar)," *ICES J. Marine Science: J. du Conseil*, vol. 65, no. 4, pp. 636–644, 2008.

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

[13] S. Majumder, S. Scheding, H. F. Durrant-Whyte, "Multisensor Data Fusion for Underwater Navigation," *Robotics and Autonomous Systems*, vol. 35, no. 2, pp. 97–108, 2001.

[14] H. Singh, C. Roman, L. Whitcomb, and D. Yoerger, "Advances in Fusion of High Resolution Underwater Optical and Acoustic Data," in *Proc. 2000 IEEE Int. Symposium Underwater Technology*, pp. 206–211, 2000.

[15] C. Xu, A. Asada, and K. Abukawa, "A Method of Generating 3D Views of Aquatic Plants with DIDSON," in *Proc. 2011 IEEE Symposium Underwater Technology and 2011 IEEE Workshop Scientific Use of Submarine Cables and Related Technologies*, pp. 1–5, 2011.

[16] H. Assalih, Y. Petillot, and J. Bell, "Acoustic Stereo Imaging (ASI) System, in *Proc. 2009 IEEE OCEANS*, pp. 1–6, 2009.

[17] M D. Aykin and S. Negahdaripour, "On 3-D Target Reconstruction from Multiple 2-D Forward-Scan Sonar Views," in *Proc. 2015 MTS/IEEE OCEANS*, pp. 1–6, 2015.

[18] S. Kwak, Y. Ji, A. Yamashita, and H. Asama, "3-D Reconstruction of Underwater Objects Using Arbitrary Acoustic Views," in *Proc. 2016 11th France-Japan Congr. Mechatronics 9th Europe-Asia Congr. Mechatronics 17th Int. Conf. Research and Education in Mechatronics*, pp. 74–79, 2016.

[19] T. A. Huang and M. Kaess, "Towards Acoustic Structure from Motion for Imaging Sonar," in *Proc. 2015 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 758–765, 2015.

[20] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[21] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[22] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A Probabilistic Hough Transform," *Pattern Recognition*, vol. 24, no. 4, pp. 303–316, 1991.