3D Reconstruction of Line Features Using Multi-view Acoustic Images in Underwater Environment

Ngoc Trung Mai, Hanwool Woo, Yonghoon Ji, Yusuke Tamura, Atsushi Yamashita, and Hajime Asama

Abstract— In order to understand the underwater environment, it is essential to use sensing methodologies able to perceive the three dimensional (3D) information of the explored site. Sonar sensors are commonly employed in underwater exploration. This paper presents a novel methodology able to retrieve 3D information of underwater objects. The proposed solution employs an acoustic camera, which represents the next generation of sonar sensors, to extract and track the line of the underwater objects which are used as visual features for the image processing algorithm. In this work, we concentrate on artificial underwater environments, such as dams and bridges. In these structured environments, the line segments are preferred over the points feature, as they can represent structure information more effectively. We also developed a method for automatic extraction and correspondences matching of line features. Our approach enables 3D measurement of underwater objects using arbitrary viewpoints based on an extended Kalman filter (EKF). The probabilistic method allows computing the 3D reconstruction of underwater objects even in presence of uncertainty in the control input of the camera's movements. Experiments have been performed in real environments. Results showed the effectiveness and accuracy of the proposed solution.

Fig. 1. Observation of underwater objects using an acoustic camera at multiple viewpoints.

I. INTRODUCTION

In recent years, the need for underwater sensing technology has increased. There are many problems in underwater environments such as inspection, survey, and exploration, etc [1][2]. Until now, many underwater sensing activities still rely on manpower like divers. However, in some cases, humans conducting these works directly in underwater environment face many potential risks. For these cases, it is desirable to apply a robot such as an autonomous underwater vehicle (AUV) and a remotely operated underwater vehicle (ROV) for underwater sensing instead of involving human directly.

Generally, in underwater environments, sonar sensors are commonly used. A sonar sensor can obtain reliable information even in dark water or muddy water. Therefore, it is the most suitable sensor for underwater sensing. In this regard, the measurement of the ocean bottom topography using seafloor mineral resources, fishery resources, and sonar sensors have been studied. Recently, the development of acoustic cameras such as dual frequency identification sonar (DIDSON) and adaptive resolution imaging sonar (ARIS) called
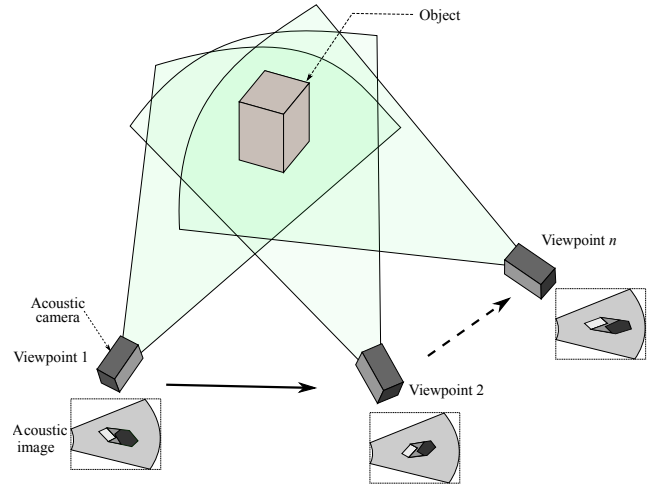
the next generation ultrasonic sensor has made information collection in the water environment more effective [3].

Thanks to the high resolution of acoustic images, a number of methodologies using an acoustic camera based systems for sensing underwater environments, such as mosaic processing [4], underwater object detection [5], position estimation of AUV [6], and underwater acoustic localization [7] have been proposed.

Further, in other studies, some methodologies for restoring 3D feature points using acoustic images from multiple viewpoints have been proposed by Huang et al. [8], Aykin et al. [9], Kwak et al. [10], and Ji et al. [11]. In our previous research, we also proposed a methodology to obtain the 3D information of the feature points of underwater objects based on an extended Kalman filter (EKF) using acoustic camera images at multiple viewpoints [12]. By using the probabilistic method based on the EKF, 3D reconstruction of underwater objects is available even if the control input data for movements of the camera has uncertainty. In our previous research or other research related 3D reconstruction of underwater objects still based on the low–level feature points. However, with low-level features, the matching process for features can often fail given that the features are indistinguishable from each other, which decreases accuracy of 3D reconstruction. For extraction and association of feature points, it relies on a priori knowledge of the features which will be detected and then the manual
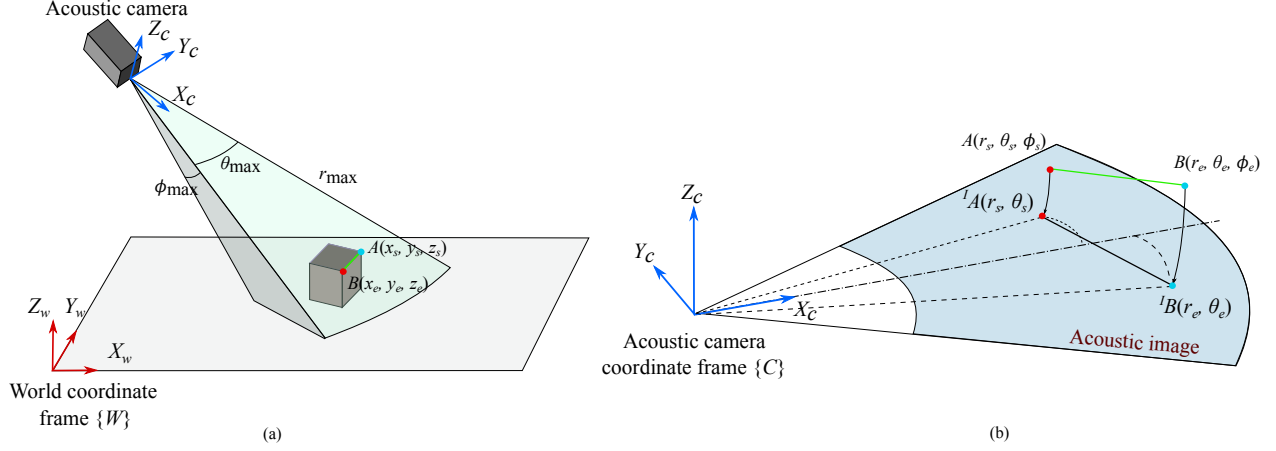
Fig. 2. Acoustic projection model and imaging sonar geometry: (a) geometrical model in the acoustic camera image generation and (b) imaging sonar geometry. The values of the range r and azimuth angle θ can be obtained from the pixel coordinate of the acoustic image, while the elevation angle ϕ is missing. Two coordinate systems are defined here. The world coordinate system (X_w, Y_w, Z_w) and the camera coordinate system (X_c, Y_c, Z_c). (x_s, y_s, z_s) and (x_e, y_e, z_e) represent respectively the 3D information of A and B in the world coordinate. ($^C x_s$, $^C y_s$, $^C z_s$) and ($^C x_e$, $^C y_e$, $^C z_e$) represent respectively the 3D information of A and B in the camera coordinate.

selection of features from the acoustic images.

In this paper, we use line segments instead of points as a landmark since there are several advantages to use line segments. We concentrate on the artificial environment. In this type of structured environment, the line segments are preferred over the points feature in order to represent structure information more effectively. For the typical environment, line features have been used as the landmarks in many studies using optical cameras for 3D measurement, simultaneous localization and mapping (SLAM), place recognition, etc [13]–[15].

Moreover, we also propose a method for extracting and matching of the line features automatically. A conceptual image of the proposed method based on multiple acoustic viewpoints is shown in Fig. 1. We use the EKF–based approach to obtain the 3D information of line features for underwater objects while estimating the pose of the acoustic camera.

The rest of the paper is organized as follows. Section II describes the principle of an acoustic camera. Section III presents a method for automatic line feature extraction on acoustic images. Section IV describes our proposed methodology. Section V presents the experimental results performed in a real underwater environment. Finally, conclusion is drawn in Section VI.

II. PRELIMINARIES OF ACOUSTIC CAMERA

A. Acoustic projection model

The acoustic camera transmits ultrasound waves to the 3D region space to generate an acoustic image. The sensing range is the maximum range $r_{\max}$, the maximum azimuth angle $\theta_{\max}$, and the maximum elevation angle $\phi_{\max}$, as shown in Fig. 2(a). The ultrasonic waves propagate forward, hit the object, and are reflected. The acoustic camera receives

the reflected waves from the object, calculates the power of the reflected waves, and reports it to the pixel corresponding to the direction of the reflected wave on the power map.

B. Imaging acoustic geometry

The output of an acoustic camera is 2D acoustic images. Each point (r, θ, ϕ) which is representation for polar coordinate system of (x, y, z) in the 3D sensing area is mapped to (r, θ) in the 2D image, as shown in Fig. 2(b). From the acoustic image, the range r and the azimuth angle θ are obtained; however, the elevation angle ϕ has been lost. For example, point $A(^C x_s, {}^C y_s, {}^C z_s)$ in the 3D sensing area is mapped at $^I A(r_s, \theta_s)$ in the 2D acoustic image, as shown in Fig. 2(b). Here, superscripts C and I respectively indicate camera coordinate system and an image coordinate system. This is due to the way the acoustic camera handles with the reflected acoustic waves. The acoustic waves reflected from the object are processed as a function of only the range r and the azimuth angle θ, not related to the elevation angle ϕ. In other words, different points in the 3D sensing area with same range r and same azimuth angle θ are mapped at same pixel on the 2D acoustic image.

The conversion of Cartesian coordinates and Spherical coordinates is given as follows:

$$
\begin{bmatrix} {}^C x \\ {}^C y \\ {}^C z \end{bmatrix} = \begin{bmatrix} r \cos \phi \cos \theta \\ r \cos \phi \sin \theta \\ r \sin \phi \end{bmatrix}. \tag{1}
$$

The conversion of Image coordinates and Spherical coordinates is given as follows:

$$
\begin{bmatrix} {}^I x \\ {}^I y \end{bmatrix} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix}. \tag{2}
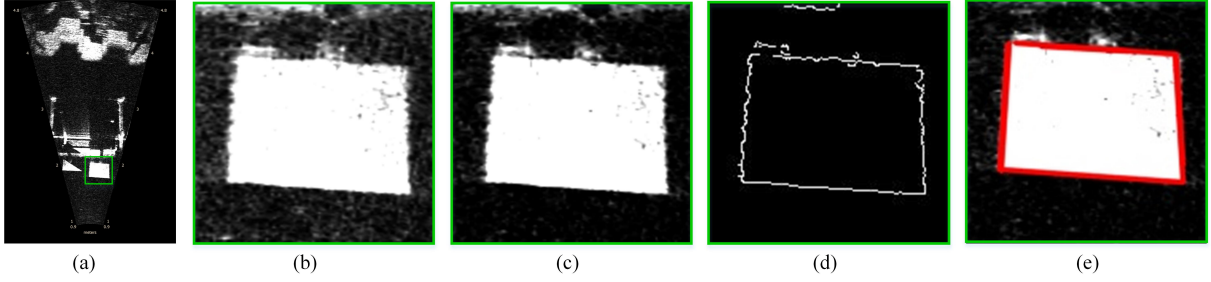$$

Fig. 3. Procedure of line feature extraction and results for extracting line feature of the square plate: (a) original acoustic image, (b) designation of ROI, (c) using Bilateral filter to smooth the image while preserving the edges, (d) using Canny edge detection to detect the edge of the image, and (e) extracting line segments by probabilistic Hough transform.

III. LINE SEGMENT EXTRACTION

In this study, the measurement information is defined as the 2D line segment of the object pictured in the acoustic image. In this section, we present a methodology for automated line feature extraction from the acoustic images. The 2D line segment model is defined as shown in Fig. 2(b), one line segment consists two endpoints $^IA(r_s, \theta_s)$ and $^IB(r_e, \theta_e)$ in the polar coordinate of the camera system.

For standard optical cameras, a straight line in 3D space is projected to a straight line in the 2D image. In the case of acoustic cameras, linear approximation can be used to project a straight line in 3D space into the 2D image. Justification of this assumption is reported below.

Assuming that the points $A(^Cx_s, {}^Cy_s, {}^Cz_s)$, $B(^Cx_e, {}^Cy_e, {}^Cz_e)$, and $M(^Cx_m, {}^Cy_m, {}^Cz_m)$ are in the same 3D straight line of the camera field of view (FoV), $\boldsymbol{v} = (a, b, c)$ is a vector parallel to this line, and the points are respectively mapped at $^IA(^Ix_s, {}^Iy_s)$, $^IB(^Ix_e, {}^Iy_e)$, and $^IM(^Ix_m, {}^Iy_m)$ in the 2D acoustic image. The equation of the straight line, which is passing through the point $M(^Cx_m, {}^Cy_m, {}^Cz_m)$ and parallel to the direction vector $\boldsymbol{v}$, can be written as follows:

$$\frac{x - {}^Cx_m}{a} = \frac{y - {}^Cy_m}{b} = \frac{z - {}^Cz_m}{c}. \tag{3}$$

Since the points $A(^Cx_s, {}^Cy_s, {}^Cz_s)$ and $B(^Cx_e, {}^Cy_e, {}^Cz_e)$ are on the straight line as Eq. (3), the following equations can be derived:

$$\frac{^Cx_s - {}^Cx_m}{a} = \frac{^Cy_s - {}^Cy_m}{b} = \frac{^Cz_s - {}^Cz_m}{c}, \tag{4}$$

$$\frac{^Cx_e - {}^Cx_m}{a} = \frac{^Cy_e - {}^Cy_m}{b} = \frac{^Cz_e - {}^Cz_m}{c}. \tag{5}$$

As shown in Table I, with respect to the elevation direction ϕ, the sensing range of the acoustic camera is very small, only about 15 deg (i.e., from -7 deg to 7 deg). This means that the value of $\cos\phi$ is almost 1. From Eqs. (1) and (2), the relationship between the values of points A, B, and M in the camera coordinate and in the image coordinate can be obtained as follows:

$$(^Ix_s, {}^Iy_s) \simeq (^Cx_s, {}^Cy_s), \tag{6}$$

$$(^Ix_s, {}^Iy_s) \simeq (^Cx_s, {}^Cy_s), \tag{7}$$

$$(^Ix_e, {}^Iy_e) \simeq (^Cx_e, {}^Cy_e). \tag{8}$$

Substituting Eqs. (6), (7) and (8) into Eqs. (4) and (5), the following equation can be obtained:

$$\frac{^Iy_s - {}^Iy_m}{^Ix_s - {}^Ix_m} \simeq \frac{^Iy_e - {}^Iy_m}{^Ix_e - {}^Ix_m} \simeq \frac{b}{a}. \tag{9}$$

Therefore, ^IA, ^IB, and ^IM appear to be on the same straight line. In conclusion, the straight line in the 3D space is mapped as almost straight line in the 2D acoustic image.

Therefore, the line segment AB which consists of two endpoints $A(x_s, y_s, z_s)$ and $B(x_e, y_e, z_e)$ in the 3D camera FoV becomes the line segment $^IA^IB$ which consists of two endpoints $^IA(r_s, \theta_s)$ and $^IB(r_e, \theta_e)$ in the 2D acoustic image.

Figure 3 illustrates the whole process of line feature extraction. Firstly, the range to be analyzed is selected and limited by the region of interest (ROI), as shown in Fig. 3(b). Next, in order to improve line feature extraction accuracy, a Bilateral filter is used to eliminate noise (Fig. 3(c)). By using the Bilateral filter, the smoothing of the image preserves the edges. Then, the edge of the images are extracted by Canny edge detection (Fig. 3(d)). After edge detection, not only the lines but also the endpoints of the line segments are extracted by using probabilistic Hough transform [16] with a high reliability, as shown in Fig. 3(e).

IV. EKF-BASED 3D LINE ESTIMATION

A. Problem setting and formulation

A schematic representation of the method proposed in this research is shown in Fig. 4. The input of the algorithm is the 2D information of line segments extracted from acoustic images and control input data for movements of the camera. The output is the 3D information of line segments which are the bare bones of the underwater objects and the six degrees of freedom (6DoF) of camera poses. An EKF algorithm is used for the estimate. Euler angles are used for camera orientation.

State vectors $\boldsymbol{x}_c$, $\boldsymbol{x}_{li}$, and $\boldsymbol{x}_{l(1:n)}$ respectively represent the camera pose, the 3D information of each line segment, and the 3D information of all line segments, as follows:

$$\boldsymbol{x}_c = \begin{bmatrix} x_c & y_c & z_c & \psi_c & \theta_c & \varphi_c \end{bmatrix}^\top, \tag{10}$$

$$\boldsymbol{x}_{li} = \begin{bmatrix} x_{lis} & y_{lis} & z_{lis} & x_{lie} & y_{lie} & z_{lie} \end{bmatrix}^\top, \tag{11}$$
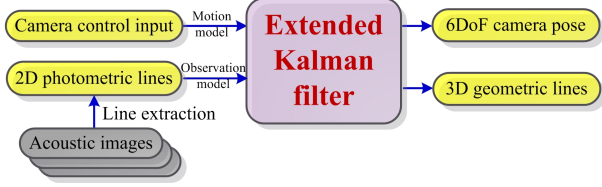
Fig. 4. Schematic representation of the proposed method. Extended Kalman filter algorithm is used to estimate the camera pose and the 3D information of line segments.

$$\boldsymbol{x}_{l(1:n)} = \begin{bmatrix} \boldsymbol{x}_{l1}^\top & \boldsymbol{x}_{l2}^\top & ... & \boldsymbol{x}_{li}^\top & ... & \boldsymbol{x}_{ln}^\top \end{bmatrix}^\top. \quad (12)$$

Here, the line segment is defined by start point $(x_{lis}, y_{lis}, z_{lis})$ and end point $(x_{lie}, y_{lie}, z_{lie})$. n is number of line segments. System state vector $\boldsymbol{X}$ indicating the 6DoF camera pose $\boldsymbol{x}_c$ and the 3D information of line segments $\boldsymbol{x}_{l(1:n)}$ are defined as follows:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_c^\top & \boldsymbol{x}_{l(1:n)}^\top \end{bmatrix}^\top. \quad (13)$$

Next, $\boldsymbol{P}$ is also defined as the covariance matrix, indicating the uncertainty of the state vector $\boldsymbol{X}$, as follows:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_c & \boldsymbol{P}_{c,l} \\ \boldsymbol{P}_{l,c} & \boldsymbol{P}_l \end{bmatrix}, \quad (14)$$

where the diagonal elements $\boldsymbol{P}_c$ and $\boldsymbol{P}_l$ are covariance matrices for the estimation of camera pose and line segments, respectively. The off-diagonal element $\boldsymbol{P}_{l,c} = \boldsymbol{P}_{c,l}$ is covariance matrices for the estimation of camera and line segments.

The estimation procedure consists of a prediction and an update step. In the prediction step, the system state $\boldsymbol{X}$ and corresponding covariance $\boldsymbol{P}$ are predicted based on the control input data for the camera movement. Next, in the update step, the system state $\boldsymbol{X}$ and corresponding covariance $\boldsymbol{P}$ are updated based on the observation data.

B. Prediction

In the prediction step, the state vector $\bar{\boldsymbol{X}}_t$ and corresponding covariance matrix $\bar{\boldsymbol{P}}_t$ at time t are predicted as follows:

$$\bar{\boldsymbol{X}}_t = \boldsymbol{g}(\boldsymbol{u}_t, \boldsymbol{X}_{t-1}), \quad (15)$$

$$\bar{\boldsymbol{P}}_t = \boldsymbol{G}_t \boldsymbol{P}_{t-1} \boldsymbol{G}_t^\top + \boldsymbol{M}_t. \quad (16)$$

Here, function $\boldsymbol{g}(.)$ represents a motion model function of the system, $\boldsymbol{M}_t$ is the covariance of the control input noise, and $\boldsymbol{G}_t = \partial \boldsymbol{g} / \partial \boldsymbol{X}_t$ is the Jacobian for the motion model. The control input $\boldsymbol{u}_t$ is defined as follows:

$$\boldsymbol{u}_t = \begin{bmatrix} \Delta x_t & \Delta y_t & \Delta z_t & \Delta \psi_t & \Delta \theta_t & \Delta \varphi_t \end{bmatrix}^\top. \quad (17)$$

The motion model $\boldsymbol{g}(.)$ for predicting the next state vector $\bar{\boldsymbol{X}}_t$ is given as follows:

$$\bar{\boldsymbol{X}}_t = \boldsymbol{X}_{t-1} + \begin{bmatrix} \boldsymbol{I}_{(6)} \\ \boldsymbol{0}_{(6n \times 6)} \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{D}_t & \boldsymbol{0}_{(3 \times 3)} \\ \boldsymbol{0}_{(3 \times 3)} & \boldsymbol{E}_t \end{bmatrix} \boldsymbol{u}_t \right), \quad (18)$$

$$\boldsymbol{D}_t = \begin{bmatrix} \mathbf{c}_\theta \mathbf{c}_\varphi & \mathbf{c}_\varphi \mathbf{s}_\theta \mathbf{s}_\psi - \mathbf{c}_\psi \mathbf{s}_\varphi & \mathbf{s}_\psi \mathbf{s}_\varphi + \mathbf{c}_\psi \mathbf{c}_\varphi \mathbf{s}_\theta \\ \mathbf{c}_\theta \mathbf{s}_\varphi & \mathbf{c}_\psi \mathbf{c}_\varphi + \mathbf{s}_\theta \mathbf{s}_\psi \mathbf{s}_\varphi & \mathbf{c}_\psi \mathbf{s}_\theta \mathbf{s}_\varphi - \mathbf{c}_\varphi \mathbf{s}_\psi \\ -\mathbf{s}_\theta & \mathbf{c}_\theta \mathbf{s}_\psi & \mathbf{c}_\theta \mathbf{c}_\psi \end{bmatrix}, \quad (19)$$

$$\boldsymbol{E}_t = \begin{bmatrix} 1 & \mathbf{t}_\theta \mathbf{s}_\psi & -\mathbf{t}_\theta \mathbf{c}_\psi \\ 0 & \mathbf{c}_\psi & \mathbf{s}_\psi \\ 0 & \mathbf{s}_\psi / \mathbf{c}_\theta & \mathbf{c}_\psi / \mathbf{c}_\theta \end{bmatrix}, \quad (20)$$

where $\mathbf{c}, \mathbf{s}, \mathbf{t}$ represent sine, cosine and tangent funtions, respectively. Here, $\boldsymbol{I}_{(6)}$, $\boldsymbol{0}_{(6n \times 6)}$, and $\boldsymbol{0}_{(3 \times 3)}$ are respectively 6 rowed identity matrix, $6n \times 6$ zero matrix, and 3×3 zero matrix.

Using this approach the control input data $\boldsymbol{u}_t$ and the covariance of the control input noise $\boldsymbol{M}_t$ are used to predict the state vector $\bar{\boldsymbol{X}}_t$ and covariance matrix $\bar{\boldsymbol{P}}_t$.

C. Update

The updating phase uses the observation information from the acoustic images is used to update the state vector $\boldsymbol{X}_t$ and covariance matrix $\boldsymbol{P}_t$. In this study, the observation information is defined as line feature on the acoustic image. In this step, firstly, for each observation, the expected observation value is calculated. Next, by comparing the expected observation value with the observed value, the state vector and covariance will be corrected.

The expected observation value can be estimated from the predicted state vector. Here, it is necessary to define an observation model $\hat{z}_t = \boldsymbol{h}(\bar{\boldsymbol{X}}_t)$ for the line feature of the acoustic image. As shown in Fig. 2, besides the world coordinate system $\{W\}$, the camera coordinate system $\{C\}$ is also defined. In order to construct the observation model, it is necessary to first convert the 3D information of the line segment from world coordinate to camera coordinate and then estimate the expected observation value next.

First, for converting coordinates, from the 3D information of two endpoints of the line segment $(x_{lis,t}, y_{lis,t}, z_{lis,t})$ and $(x_{lie,t}, y_{lie,t}, z_{lie,t})$ in the world coordinate system,

TABLE I

SPECIFICATIONS OF ARIS EXPLORER 3000

Identification frequency	3 MHz
Detection frequency	1.8 MHz
Depth rating	300 m
Identification range	5 m
Range resolution	Down to 3 mm
Field of view	30 deg × 15 deg
Beam width	0.25 deg
Number of transducer beams	128
Frame rate	Up to 15 frames / second

TABLE II

EXPERIMENTAL RESULTS

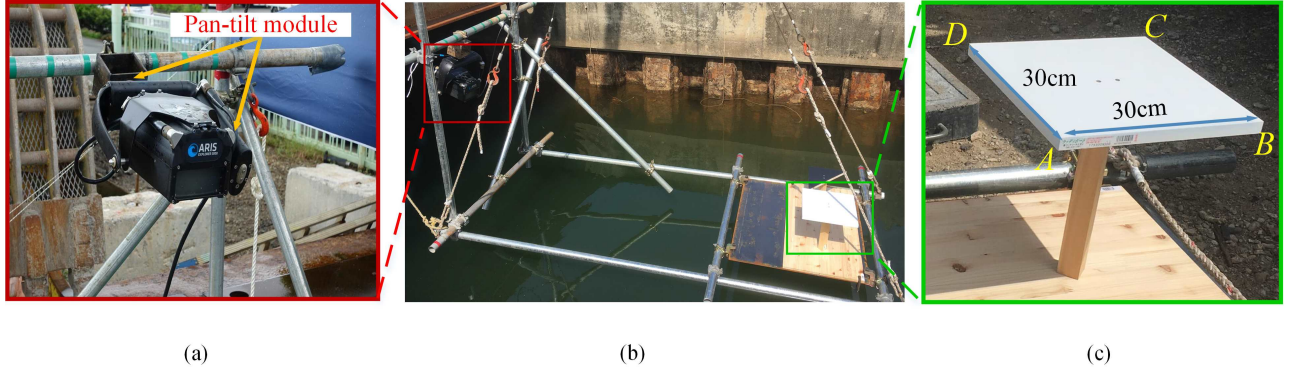| Line segment	True length [cm]	Estimated length [cm]	Error [cm]
AB	30	31.43	1.43
BC	30	30.53	0.53
CD	30	33.85	3.85
DA	30	29.40	0.60

Fig. 5. The acoustic camera and underwater object were used in the real experiment: (a) The acoustic camera (ARIS EXPLORER 3000) with the pan-tilt module, (b) The experimental system in turbid tank, and (c) The underwater object.

the 3D information of two endpoints of the line segment $(^C x_{lis,t}, {}^C y_{lis,t}, {}^C z_{lis,t})$ and $(^C x_{lie,t}, {}^C y_{lie,t}, {}^C z_{lie,t})$ in the camera coordinate system is obtained as follows:

$$\begin{bmatrix} ^C x_{lis,t} \\ ^C y_{lis,t} \\ ^C z_{lis,t} \end{bmatrix} = \boldsymbol{R}_t \begin{bmatrix} x_{lis,t} - x_{c,t} \\ y_{lis,t} - y_{c,t} \\ z_{lis,t} - z_{c,t} \end{bmatrix}, \qquad (21)$$

$$\begin{bmatrix} ^C x_{lie,t} \\ ^C y_{lie,t} \\ ^C z_{lie,t} \end{bmatrix} = \boldsymbol{R}_t \begin{bmatrix} x_{lie,t} - x_{c,t} \\ y_{lie,t} - y_{c,t} \\ z_{lie,t} - z_{c,t} \end{bmatrix}, \qquad (22)$$

where $\boldsymbol{R}_t$ is the rotation matrix for the coordinate transformation, which is defined as follows:

$$\boldsymbol{R}_t = \begin{bmatrix} \mathbf{c}_\theta \mathbf{c}_\varphi & \mathbf{c}_\psi \mathbf{s}_\varphi + \mathbf{c}_\varphi \mathbf{s}_\theta \mathbf{s}_\psi & \mathbf{s}_\psi \mathbf{s}_\varphi - \mathbf{c}_\psi \mathbf{c}_\varphi \mathbf{s}_\theta \\ -\mathbf{c}_\theta \mathbf{s}_\varphi & \mathbf{c}_\psi \mathbf{c}_\varphi - \mathbf{s}_\theta \mathbf{s}_\psi \mathbf{s}_\varphi & \mathbf{c}_\varphi \mathbf{s}_\psi + \mathbf{c}_\psi \mathbf{s}_\theta \mathbf{s}_\varphi \\ \mathbf{s}_\theta & -\mathbf{c}_\theta \mathbf{s}_\psi & \mathbf{c}_\theta \mathbf{c}_\psi \end{bmatrix}. \qquad (23)$$

The expected observation value $\hat{\boldsymbol{z}}_t$ composed of the 2D information of the start point and the end point of the line segment $(\hat{r}_{is,t}, \hat{\theta}_{is,t}, \hat{r}_{ie,t}, \hat{\theta}_{ie,t})$ can be estimated as follows:

$$\hat{\boldsymbol{z}}_{i,t} = \begin{bmatrix} \hat{r}_{is,t} \\ \hat{\theta}_{is,t} \\ \hat{r}_{ie,t} \\ \hat{\theta}_{ie,t} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{^C x_{lis,t}{}^2 + {}^C y_{lis,t}{}^2 + {}^C z_{lis,t}{}^2} \\ \mathrm{atan2}(^C y_{lis,t}, {}^C x_{lis,t}) \\ \sqrt{^C x_{lie,t}{}^2 + {}^C y_{lie,t}{}^2 + {}^C z_{lie,t}{}^2} \\ \mathrm{atan2}(^C y_{lie,t}, {}^C x_{lie,t}) \end{bmatrix}. \qquad (24)$$

Next, the Kalman gain $\boldsymbol{K}_t$ is calculated as follows:

$$\boldsymbol{K}_t = \bar{\boldsymbol{P}}_t \boldsymbol{H}_t^\top (\boldsymbol{H}_t \bar{\boldsymbol{P}}_t \boldsymbol{H}_t^\top + \boldsymbol{Q}_t)^{-1}. \qquad (25)$$

Then, the state vector $\boldsymbol{X}_t$ and the covariance matrix $\boldsymbol{P}_t$ are updated based on the Kalman gain $\boldsymbol{K}_t$, as follows:

$$\boldsymbol{X}_t = \bar{\boldsymbol{X}}_t + \boldsymbol{K}_t (\boldsymbol{z}_t - \hat{\boldsymbol{z}}_t), \qquad (26)$$

$$\boldsymbol{P}_t = (\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{H}_t) \bar{\boldsymbol{P}}_t, \qquad (27)$$

where $\boldsymbol{Q}_t$ is the corvaricance of the observation noise, $\boldsymbol{H}_t$ is the Jacobian for the observation model, and $\boldsymbol{z}_t$ is actually observed value. $\boldsymbol{H}_t$ is defined as follows:

$$\boldsymbol{H}_t = \begin{bmatrix} \boldsymbol{H}_{c,t} & \boldsymbol{0}_{4 \times 6(i-1)} & \boldsymbol{H}_{l,t} & \boldsymbol{0}_{4 \times 6(n-1)} \end{bmatrix}. \qquad (28)$$

Here, $\boldsymbol{H}_{c,t} = \partial \boldsymbol{h}/\partial \boldsymbol{X}_{t-1}$ and $\boldsymbol{H}_{l,t} = \partial \boldsymbol{h}/\partial \boldsymbol{x}_{li,t-1}$ are respectively Jacobian for camera pose and Jacobian for i^{th} line segment.

By repeating the prediction and updating, the camera pose and the 3D information of each line segment are estimated at each time.

V. EXPERIMENT

In order to verify the proposed method, an experiment has been carried out by using the ARIS EXPLORER 3000 as shown in Fig. 5(a). The specification of the camera are reported in Table I. A square plate (30W × 30L × 2H cm) was used as an underwater object as shown in Fig. 5(c). An experimental system equipped with the acoustic camera and the underwater object was set up and placed into a turbid tank, as shown in Fig. 5(b). Underwater objects were set about 2 m away from the initial position of the acoustic camera. The acoustic camera was placed at a position higher than the object so that sound waves of the object were obliquely heard, as shown in Fig. 5(b). From this experiment, multiple acoustic images were acquired with various acoustic camera poses. We defined the world coordinate frame with reference to the first camera pose. In other words, the 6DoF of the initial camera pose was (0, 0, 0, 0, 0, 0). By rotating the roll axis of the acoustic camera (10 deg from 0 deg to 60 deg), acoustic images were acquired from several different viewpoints. Control input data for rolling was obtained from pan–tilt module as shown in Fig. 5(a).

Acoustic images of the object used for the 3D measurement is shown in Fig. 6(a). Even in turbid water, the acoustic camera was able to provide clear information of the object.

Figure 6(b) shows the results of reconstructed line segments in 3D space. Table II compares the estimated values with the true value and lists the estimated errors. Consequently, the results showed that the proposed methodology
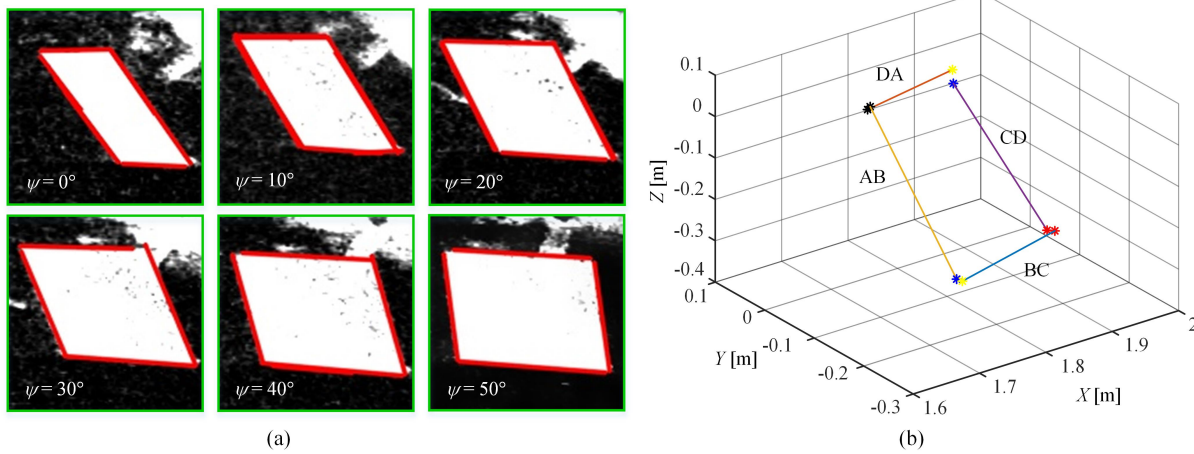
Fig. 6. The experiment results: (a) The real acoustic images of the square board with different camera poses and (b) The estimated result of line segments. The real acoustic images of the square board with different camera poses.

could accurately reconstruct 3D line segments with small errors about 0.53–3.85 cm.

VI. CONCLUSION

In this study, we proposed a novel methodology to reconstruct the 3D information of the lines feature for the underwater object based on the EKF using acoustic camera images captured from multi-viewpoints. By using the probabilistic method based on the EKF, 3D reconstruction of underwater objects can be conducted with the high accuracy. Moreover, we also developed a novel methodology for automatic line feature extraction. Experiments were performed in real environments in order to demonstrate the effectiveness of the proposed method.

In the future, 3D reconstruction methodology for more general structure with more complicated shape will be explored.

ACKNOWLEDGMENT

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, pp. 1767–1774, 2000.
[2] Y. W. Huang, Y. Sasaki, Y. Harakawa, E. F. Fukushima, and S. Hirose, "Operation of Underwater Rescue Robot Anchor Diver III During the 2011 Tohoku Earthquake and Tsunami," in Proc. 2011 MTS/IEEE OCEANS, pp. 1–6, 2011.
[3] E. Belcher, W. Hanot, and J. Burch, "Dual–Frequency Identification Sonar (DIDSON)," in Proc. 2002 International Symposium on Underwater Technology, pp. 187–192, 2002.
[4] N. Hurtos, S. Nagappa, N. Palomeras, and J. Salvi, "Real–Time Mosaicing with Two–Dimensional Forward–Looking Sonar," in Proc. 2014 IEEE Int. Conf. Robotics and Automation, pp. 601–606, 2014.
[5] H. Cho, J. Gu, H. Joe, A. Asada, and S. C. Yu, "Acoustic Beam Profile–Based Rapid Underwater Object Detection for an Imaging Sonar," Journal of Marine Science and Technology, vol. 20, no. 1, pp. 180–197, 2015.
[6] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar–aided navigation for autonomous underwater harbor surveillance," in Proc. 2010 IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 4396–4403, 2010.
[7] J. Li, P. Ozog, J. Abernethy, R.M. Eustice, and M. Johnson-Roberson, "Utilizing High-dimensional Features for Real-time Robotic Applications: Reducing the Curse of Dimensionality for Recursive Bayesian Estimation," in Proc. 2016 IEEE/RSJ Int. Conf. Intelligent Robots and Systems, pp. 1334–1341, 2016.
[8] 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.
[9] M. D. Aykin and S. Negahdaripour, "On 3D Target Reconstructionfrom Multiple 2D Forward–Scan Sonar Views," in Proc. 2015 MTS/IEEE OCEANS, pp. 1–6, 2015.
[10] S. Kwak, Y. Ji, A. Yamashita, and H. Asama, "3D 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.
[11] Y. Ji, S. Kwak, A. Yamashita, and H. Asama, "Acoustic Camera–based 3D Measurement of Underwater Objects through Automated Extraction and Association of Feature Points," in Proc. 2016 Int. Conf. Multisensor Fusion and Integration for Intelligent Systems, pp. 224–230, 2016.
[12] N. T. Mai, H. Woo, Y. Ji, Y. Tamura, A. Yamashita, and H. Asama, "3D Reconstruction of Underwater Object Based on Extended Kalman Filter by Using Acoustic Camera Images," Preprints of the 20th World Congress of the International Federation of Automatic Control, pp. 1066–1072, 2017.
[13] A. P. Gee and W. Mayol-Cuevas, "Real-time model-based SLAM using line segments," in Proc. 2006 IEEE Int. Symp. Visual Computing, pp. 354–363 , 2006.
[14] J. H. Lee, G. Zhang, J. Lim, and I. H. Suh, "Place recognition using straight lines for vision-based SLAM," in Proc. 2013 IEEE Int. Conf. Robotics and Automation, pp. 3799–3806, 2013.
[15] R. I. Hartley, "A Linear Method for Reconstruction from Lines and Points," in Proc. 1995 IEEE Int. Conf. Computer Vision, pp. 882–887, 1995.
[16] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A Probabilistic Hough Transform," Pattern Recognition, vol. 24, no. 4, pp. 303–316, 1991.