

# 3D Occupancy Mapping Framework Based on Acoustic Camera in Underwater Environment

Yusheng Wang\* Yonghoon Ji\*\* Hanwool Woo\*  
Yusuke Tamura\* Atsushi Yamashita\* Asama Hajime\*

\* *Department of Precision Engineering, The University of Tokyo,  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan  
(e-mail: wang@robot.t.u-tokyo.ac.jp).*

\*\* *Department of Precision Mechanics, Chuo University,  
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan  
(e-mail: ji@mech.chuo-u.ac.jp)*

---

**Abstract:** In this paper, we present a novel probabilistic three-dimensional (3D) mapping framework that uses acoustic images captured in an underwater environment. Acoustic camera is a forward-looking imaging sonar that is commonly used in underwater inspection recently; however, the loss of elevation angle information makes it difficult to get a better understanding of underwater environment. To cope with this, we apply a probabilistic occupancy mapping framework with a novel inverse sensor model suitable for the acoustic camera in order to reconstruct the underwater environment in volumetric presentation. The simulations and experimental results demonstrate that our mapping framework for the acoustic camera can reconstruct dense 3D model of underwater targets successfully.

*Keywords:* 3D reconstruction, 3D occupancy mapping, underwater sensing, acoustic camera, inverse sensor model

---

## 1. INTRODUCTION

In recent years, underwater operations such as maintenance, ship hull inspection, and construction become much more important these days; however, hazards may prohibit human access (e.g., the Fukushima Daiichi nuclear power station, which has been the crisis since the 2011 earthquake off the Pacific coast of Tohoku in the east Japan). Moreover, the limited field of vision due to turbidity and lack of illumination makes it difficult for underwater operations. Therefore, in order to fulfill underwater tasks, reconstruction process of three-dimensional (3D) underwater environment by robots such as autonomous underwater vehicles (AUVs) or remotely operated underwater vehicles (ROVs) is necessary.

Sonars are useful for sensing an underwater environment since they are less affected by turbidity, illumination, and absorption as optical sensors or laser sensors. There exist a few studies on 3D underwater environment reconstruction using the sonars such as side scan sonars (SSSs) and mechanically scanned imaging sonars (MSISs). Sun et al. employed an approach based on Markov random field (MRF) that detects shadow using SSS to reconstruct 3D model of sea floor (Sun et al. (2008)). Kwon et al. applied occupancy mapping theory on MSISs to realize 3D reconstruction of underwater objects (Kwon et al. (2017)). However, these sonar sensors are facing with problems like

inflexibility or low resolution. SSSs are usually mounted on a ship and they cannot change their pose. Thus, the ship itself has to move forward to measure the surrounding environment. In other words, it is not suitable to be mounted on AUVs or ROVs with high degrees of freedom. MSISs are being used in relatively small AUVs or ROVs thanks to its small size, however, they have comparatively low resolutions and very low sampling rates.

Recently, the development of acoustic camera, such as dual frequency identification sonar (DIDSON) and adaptive resolution imaging sonar (ARIS) which can generate high-resolution and wide-range image, facilitates understanding of underwater situation (Belcher et al. (2002)). This kind of sensor is relatively small that is easy to be mounted on AUVs or ROVs and can gather information of a larger area much faster. However, wide fan-shape beam is emitted and information lost occurs during imaging, which makes 3D reconstruction difficult. For example, we are facing with the uncertainty of 14 deg in case of ARIS EXPLORER 3000 (Sound Metrics) which is the latest-model acoustic camera.

To the best of our knowledge, most previous studies that achieve 3D reconstruction using acoustic cameras are feature-based method because it is possible to calculate 3D coordinate values through matching of corresponding feature points extracted from multiple acoustic images, like stereo matching in optical images. Corners or lines are usually used as features for 3D reconstruction. Mai et al. achieved 3D reconstruction by tracing such features using

---

\* This work was in part supported by Tough Robotics Challenge, ImPACT Program (Impulsing Paradigm Change through Disruptive Technologies Program).

extended Kalman filter (EKF) (Mai et al. (2017a))(Mai et al. (2017b)). However, automatically detecting such features on the acoustic images is not as easy as the optical images. Moreover, this kind of sparse 3D model consisting of such low-level features is not suitable for representing complex objects.

To solve this problem, occupancy grid mapping theory can be considered as a way for building dense 3D volumetric representation. In 2005, Fairfield et al. successfully inspected a sinkhole in Mexico using the occupancy grid mapping theory with pencil beams which proves the effectiveness of such volumetric representation in underwater environment (Fairfield et al. (2005)). In 2016, Teixeira et al. accomplished 3D reconstruction of shiphull consisting of volumetric submaps using acoustic images (Teixeira et al. (2016)). However, when processing time of flight (TOF) data from acoustic images, either the first or the strongest return that is above a certain threshold is kept that too many data are discarded, which is not efficient. Guerneve et al. also applied grid-based method to acoustic camera, but their method is not fully probabilistic and robot can only be moved in  $z$ -axis (i.e., height direction) which is not useful for unmanned robot with arbitrary motion and not robust enough to noise (Guerneve and Petillot (2015)).

In this paper, we also use the volumetric 3D model for reconstruction of underwater environment, and Bayesian inference is used to update the probability of each voxel constituting the 3D model. Here, in order to make the occupancy mapping theory more suitable for acoustic camera, we design a novel inverse sensor model. By using occupancy mapping theory and applying a novel inverse sensor model, it is possible to reconstruct a dense underwater 3D environment robustly and efficiently from arbitrary acoustic views.

The remainder of this paper is organized as follows. Section 2 explains principles of acoustic camera. Section 3 describes 3D occupancy mapping with acoustic camera. The effectiveness of the proposed 3D mapping framework is evaluated with the experiment results in Section 4. Section 5 gives conclusions and future works of this paper.

## 2. PRINCIPLES OF ACOUSTIC CAMERA

An acoustic camera can sense a wide range of 3D area and generate images like an optical camera; however, the imaging principle is totally different. For optical camera, depth information is lost. On the contrary, for acoustic camera, elevation angle information is lost. This phenomenon can be explained by the model described in this section.

Acoustic cameras insonify a fan-shape acoustic wave in the forward direction with the azimuth angle  $\theta_{\max}$  and elevation angle  $\phi_{\max}$  within the scope of the maximum range of  $r_{\max}$  as shown in Fig. 1a. Once the acoustic wave hits underwater objects, it is reflected back to the sensor because of backscattering. As shown in Fig. 1b, acoustic camera is a multiple beam sonar that the 3D acoustic wave from the acoustic camera can be separated into 2D beam slices in and for each beam, only range information is acquirable. The result is that elevation angle information is lost.

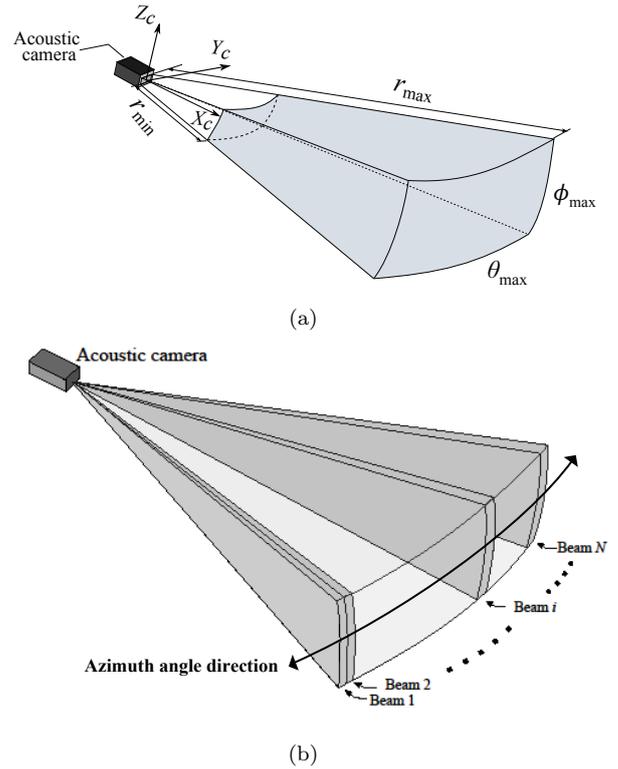


Fig. 1. Acoustic projection model:(a) shows the geometrical model of the acoustic image and (b) illustrates beam slices in the acoustic camera.

Imaging geometry of the acoustic camera is like different points in the 3D sensing area with the same range  $r$  and the same azimuth angle  $\theta$  are mapped at the same pixel on the 2D acoustic image. In other words, points can be represented by  $(r, \theta, \phi)$  in polar coordinate (i.e.,  $(x, y, z)$  in Cartesian coordinate) are mapped into  $(r, \theta)$  in the 2D acoustic image. Conversely, if the elevation angles  $\phi$  of each point are recovered, we can reconstruct 3D point cloud in the 3D sensing area for each measurement. More details on the principles of the acoustic camera can be found in (Kwak et al. (2015)).

## 3. 3D RECONSTRUCTION FRAMEWORK

Figure 2 shows overview of our 3D reconstruction framework. The input and output of our framework are:

*Input:*

- Acoustic images from multiple viewpoints
- Corresponding six degrees of freedom (6-DOF) camera poses

*Output:*

- Dense 3D reconstruction of an underwater environment

The process is divided into four steps: image segmentation, generation of input point cloud, 3D occupancy mapping and refinement. Detailed processes for each step are explained in the next subsections.

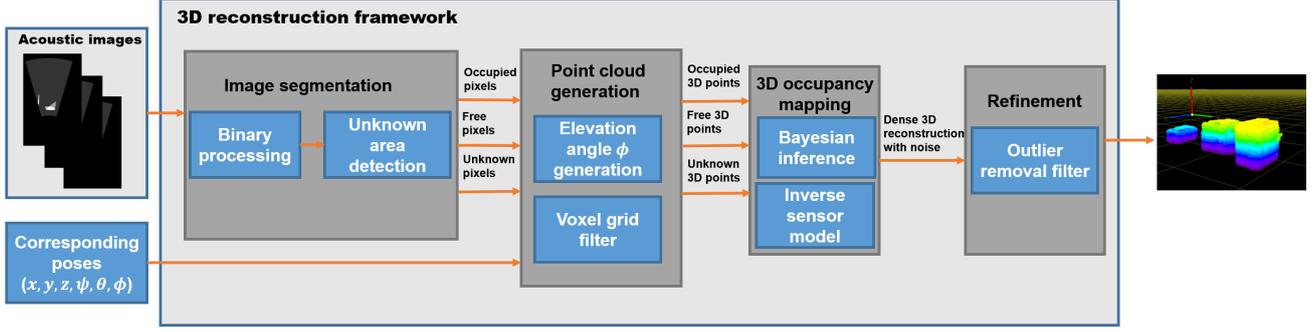


Fig. 2. Overview of 3D environment reconstruction framework based on occupancy mapping.

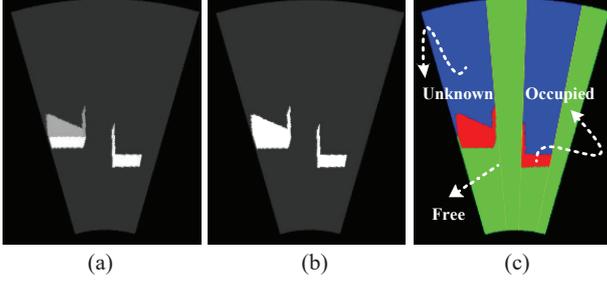


Fig. 3. Image segmentation process: (a) original acoustic image, (b) is the image after binarization and (c) shows segmentation process to classify pixels into occupied, free and unknown areas which are represented by red, green and blue respectively.

### 3.1 Image Segmentation

After acoustic images are inputted into the system, they are binarized by a threshold set preliminary. Figure 3b shows an acoustic image after binarization process. The binarized acoustic image is then segmented into three areas which are defined as occupied, free and unknown. As we mentioned above, the acoustic wave can be discretized into 128 2D beams in the azimuth angle direction. Considering one beam, every pixel in this radial direction is searched and labeled depending on its intensity and the relationship with several adjacent pixels. Algorithm 1 describes an algorithm for image segmentation process. Here,  $n$  denotes the pixel numbers in radial directions and  $j$  represents the pixel index in one radial direction from sensor origin.  $I_j$  is the intensity value of each pixel. If  $j$  is bigger than  $n$ ,  $I_j$  equals zero. Figure 3 shows the acoustic image after segmentation, where red area represents occupied, green area represents free and blue area represents unknown.

### 3.2 Point Cloud Generation

After processing image segmentation, pixel data of each group from all images are converted to input point clouds by generating elevation angles  $\phi$ . To begin with, each pixel in acoustic image is represented by  $(r, \theta)$ .  $\phi$  is generated as

$$\phi_j = \frac{\phi_{\max}}{2} - \frac{\phi_{\max} j}{M}. \quad (1)$$

Here, the elevation direction is equally separated by  $M$  radian intervals. Thus,  $j$  is integer zero to  $M$ .  $\phi_{\max}$  denotes

### Algorithm 1 Image segmentation

---

**Input:** Acoustic image  
**Output:** Classified pixels

```

while  $i < 128$  do
   $Flag = 0$ 
  while  $j < n$  do
    if  $I_j > 0$  then
      return occupied pixels
    else if  $I_j \leq 0$  and  $Flag = 0$  then
      return free pixels
    else
      return unknown pixels
    end if
    if  $I_j > 0$  and  $I_{j+1}, I_{j+2}, \dots, I_{j+k} \leq 0$  then
       $Flag = 1$ 
    end if
    if  $j + k > n$  then
       $Flag = 1$ 
    end if
     $j \leftarrow j + 1$ 
  end while
   $i \leftarrow i + 1$ 
end while

```

---

the sensing scope of the acoustic camera in elevation angle direction. Then, the generated points are converted from the camera spherical coordinates  $(r, \theta, \phi)$  to the camera Cartesian coordinates  $(x_c, y_c, z_c)$  as:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r \cos \phi \sin \theta \\ r \cos \phi \cos \theta \\ r \sin \phi \end{bmatrix}. \quad (2)$$

From the camera Cartesian coordinate, the points are then transformed to the World Cartesian coordinate  $(x_w, y_w, z_w)$  as:

$$\begin{bmatrix} x_w & y_w & z_w & 1 \end{bmatrix}^T = \begin{bmatrix} c_\theta c_\phi & c_\phi s_\theta s_\psi - c_\psi s_\phi & s_\psi s_\phi + c_\psi c_\phi s_\theta & x \\ c_\theta s_\phi & c_\psi c_\phi + s_\theta s_\psi s_\phi & c_\psi s_\theta s_\phi - c_\phi s_\psi & y \\ -s_\theta & c_\theta s_\psi & c_\theta c_\psi & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}, \quad (3)$$

where  $(x, y, z, \psi, \theta, \phi)$  indicates the 6-DOF pose of the acoustic camera (i.e., camera viewpoint). Here,  $s$  and  $c$  represent sine and cosine functions, respectively.

Figures 4a and 4b show an example of input acoustic image after segmentation process and corresponding 3D point cloud data through the generation process above. The point cloud is also divided into three groups by labeling.

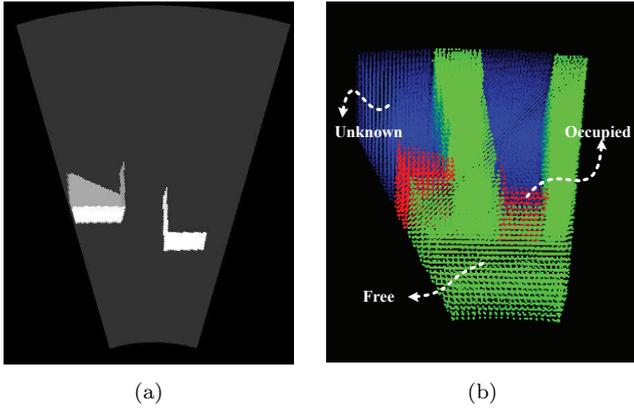


Fig. 4. Generation process of input point cloud data: (a) segmented acoustic image and (b) generated 3D point cloud data.

Here, green, red and blue respectively represent the free, occupied and unknown groups for one measurement.

After the 3D point cloud data are generated, the point clouds are down sampled by voxel grid filter (Rusu and Cousins (2011)) in order to prevent over updating of the probability in next occupancy mapping process—given that if there are  $k$  points existing in one voxel, this voxel is going to be updated by  $k$  times, which will not only increase computing burden, but also make the probability too high or too low.

### 3.3 3D Occupancy Mapping

The typical occupancy grid mapping algorithm is used to realize 3D reconstruction from point cloud data (Elfes (1989)). We use OctoMap library which produces Octree structure based 3D model in order to implement occupancy mapping (Hornung et al. (2013)).

*Bayesian inference* We define point cloud data from discrete time step  $t = 1$  to  $t = T$  as  $C_1, C_2, \dots, C_T$ . Here, the time step has the same meaning as the index of the image used. 3D environment map is denoted as  $\mathbf{m}$  and each voxel is denoted as  $\mathbf{m}_i$ . Under an assumption that each voxel is conditionally independent of one another given measurements, the posterior probability of the map information can be written as follows:

$$p(\mathbf{m}|C_{1:t}) = \prod_i p(\mathbf{m}_i|C_{1:t}). \quad (4)$$

By using Bayesian inference,  $p(\mathbf{m}_i|C_{1:t})$  can be calculated as:

$$p(\mathbf{m}_i|C_{1:t}) = \frac{p(\mathbf{m}_i|C_t)p(C_t)p(\mathbf{m}_i|C_{1:t-1})}{p(\mathbf{m}_i)p(C_t|C_1, \dots, C_{t-1})}. \quad (5)$$

To simplify calculation, we define  $L(x)$  using log odds as follows:

$$L(x) = \log \left[ \frac{P(x)}{1 - P(x)} \right]. \quad (6)$$

Therefore, Eq. (5) can be written as:

$$L(\mathbf{m}_i|C_{1:t}) = L(\mathbf{m}_i|C_{1:t-1}) + L(\mathbf{m}_i|C_t), \quad (7)$$

where  $L(\mathbf{m}_i|C_t)$  is called inverse sensor model. Each voxel has a probability of occupancy that after setting a threshold, voxels can be classified into occupied, free and unknown. Normally, occupied voxels are the dense 3D reconstruction result.

---

### Algorithm 2 Inverse sensor model

---

```

Input: 3D points
Output:  $L(\mathbf{m}_i|C_t)$ 
if It is an occupied 3D point then
    return  $l_{\text{occupied}}$ 
end if
if It is an free 3D point then
    return  $l_{\text{free}}$ 
end if
if It is an unknown 3D point then
    return  $l_{\text{unknown}}$ 
end if

```

---

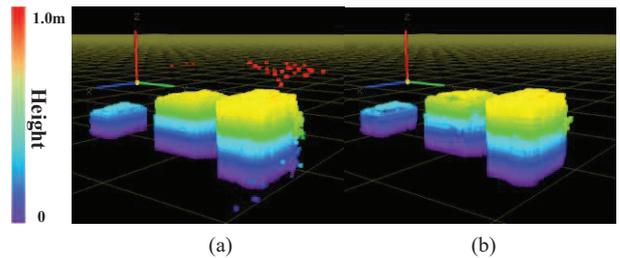


Fig. 5. Refinement process: (a) raw 3D reconstruction result and (b) 3D reconstruction result after filtering.

*Inverse sensor model* In order to apply acoustic camera images to occupancy mapping, a novel inverse sensor model is designed based on principles of acoustic camera. The inverse sensor model can be represented as:

$$L(\mathbf{m}_i|C_t) = \begin{cases} l_{\text{occupied}} \\ l_{\text{free}} \\ l_{\text{unknown}} \end{cases} \quad (8)$$

and calculated by Algorithm 2.

In inverse sensor model, we take acoustic shadow into consideration and several assumptions below are made for the inverse sensor model. Here, we define  $|\frac{l_{\text{free}}}{l_{\text{occupied}}}|$  as *ratio of empty*.

- Free space is more likely to be free that *ratio of empty*  $> 1$ .
- Acoustic shadows are difficult to be distinguished from background, they are thought to be mixed with background.
- Since shadows are not detected from images, areas that shadows are supposed to be existed is processed as unknown.

*Ratio of empty*  $> 1$  is important that if there is no reflection signal in one radial direction, we can be sure that there is nothing in this direction. On the other hand, occupied measurement may conclude plenty of “false measurement” due to the point cloud generation method. Assume that  $l_{\text{free}} = -2.2$  with an occupancy probability of 0.1,  $l_{\text{occupied}} = 0.41$  with an occupancy probability of 0.6 and the threshold of occupancy probability is 0.5 which is zero in log odds. If we acquired a measurement that this voxel is free, we need more than five occupied measurements to change the state of this voxel from free to occupied. This is also the reason why we take shadow into consideration. If one shadow area is simply processed as

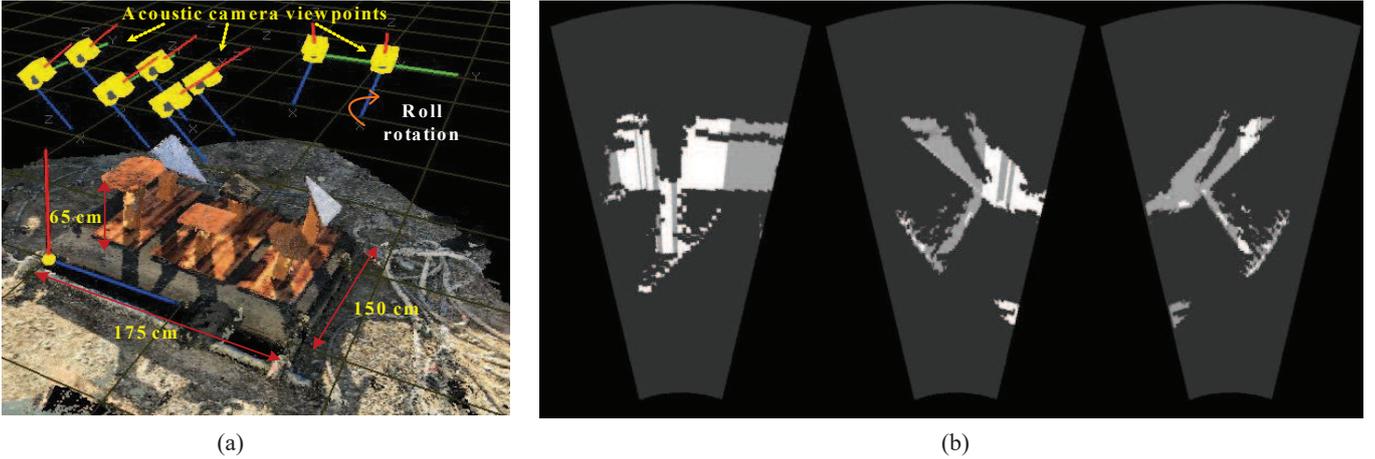


Fig. 6. Simulation experiment with virtual images: (a) is the simulation environment and (b) shows examples of the generated virtual images.

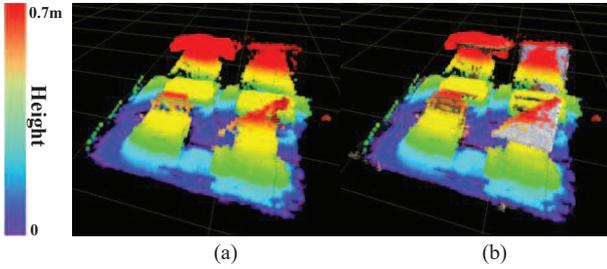


Fig. 7. Simulation result: (a) is the dense 3D reconstruction result of complex environment and (b) renders 3D reconstruction result and model used as ground truth together.

free, we need plenty of measurements to change the state of these voxels which will make our reconstruction sometimes unsuccessful. If  $ratio\ of\ empty \rightarrow \infty$  occupancy mapping can be seen as a paradigm of space carving (Klingensmith et al. (2014))(Aykin and Negahdaripour (2017)).

### 3.4 Refinement

After generating 3D occupancy map, noise exists which will degenerate our reconstruction result. Therefore, we exploit radius outlier filter (Rusu and Cousins (2011)), which is one of the simplest refinement method. Figure 5 shows our raw result and the result after filtering.

## 4. EXPERIMENT

To prove the feasibility of our 3D reconstruction framework, simulation experiments were conducted with virtual acoustic images generated from acoustic camera imaging simulator we developed (Kwak et al. (2015)). First, we tried the 3D reconstruction of some simple objects as Fig. 5 which performed successfully. Then, we attempted our framework on complex environment. We also applied real experiment data to our framework to verify its effectiveness.

### 4.1 Simulation

Figure 6b shows the complex simulation environment we designed which is generated by structure from motion (SfM). The size of the simulation environment is  $1750\ W \times 1500\ D \times 650\ H\ mm^3$  including six artificial objects fixed on boards. Here, the simulated acoustic camera viewpoints (i.e., 6-DOF poses) are represented in yellow boxes as Fig. 6a, where blue, green and red axes represent  $x$ ,  $y$  and  $z$  axes, respectively.

Camera motion may affect the result of 3D reconstruction (Huang and Kaess (2015))(Aykin and Negahdaripour (2017)). Elevation angle has an uncertainty of 14 deg that we have to narrow the occupied points using intersection. Arcs with uncertainty from different observations are better to intersect rather than coincide or parallel. In other words, translation of  $z$ -axis, rotations of pitch (i.e.,  $y$ -axis rotation) and roll (i.e.,  $x$ -axis rotation) are relatively effective. This is not a necessary condition, but if one observation merely increases the occupied possibility of voxels supposed to be free, we need more observations to decrease the possibility of these voxels to get a correct final result. For single motion, roll rotation can be the most effective way for leading to intersection of arcs with uncertainty.

Camera motion is generated as:

- Moving the acoustic camera around the environment to be inspected and stop at an arbitrary appropriate position.
- Making roll rotation (i.e.,  $x$ -axis rotation), storing images and poses.
- Moving the acoustic camera again to another arbitrary appropriate position.

Some corresponding simulated images are illustrated in Fig. 6b. Note that we treated these images as real image data from multiple acoustic camera viewpoints in our simulation experiment. 3D reconstruction process was performed using simulation data mentioned above. As shown in Fig. 7, we reconstruct the environment successfully. In Fig. 7b, dense 3D reconstruction result and model we used

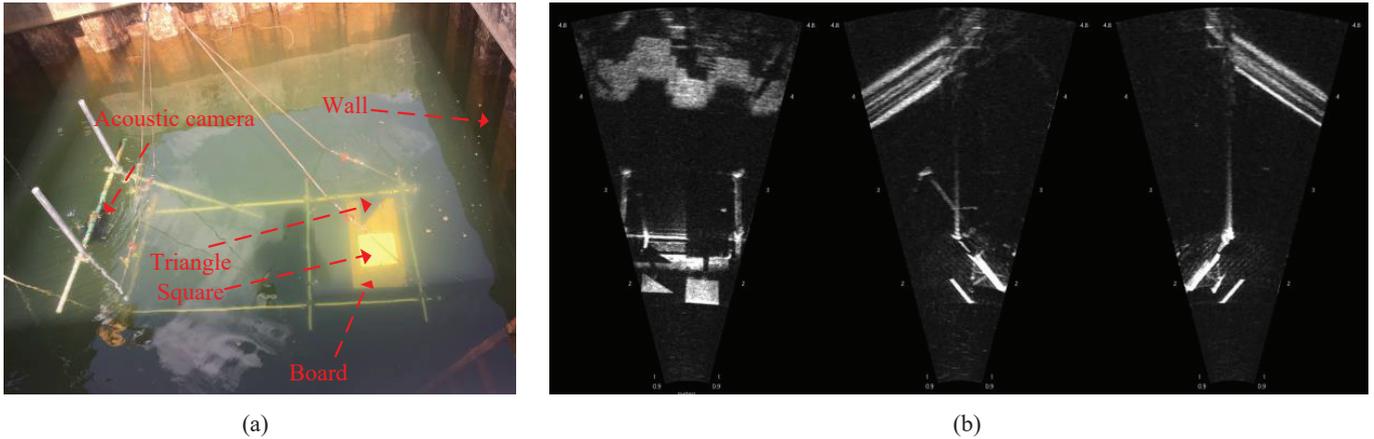


Fig. 8. Real experiment using ARIS EXPLORER 3000: (a) shows the experiment environment. Acoustic camera ARIS EXPLORER 3000 is mounted on a pan tilt which is fixed to steel bar and (b) shows examples of the captured real images.

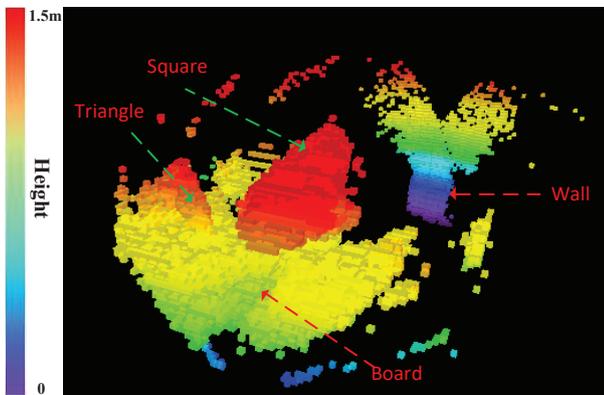


Fig. 9. Dense 3D reconstruction result with real acoustic images captured by roll rotation in one position.

as ground truth are rendered simultaneously. From this result, we confirmed the effectiveness of our framework.

As for computing cost, each acoustic image is processed in about 2 s using an Intel Core i7 vPro. 8,960,000 points are generated from one image and after applying voxel grid filter, about 1,000,000 points left for updating occupancy map. The resolution of voxels is 20 mm. For one roll rotation, image is taken every 5 deg so that 72 images are used as input data from one position. However, it is possible to take less images from one position. The computing time shows that this algorithm can generate map online.

#### 4.2 Real Data

We also applied some real experiment data to our framework using ARIS EXPLORER 3000 which is a state-of-the-art acoustic camera. In order to verify the validity of our method, first of all, we took 27 acoustic images in a real underwater environment as shown in Fig. 8. Triangle and square objects were fixed on a wood board and were submerged by turbid water. Acoustic camera was mounted on a pan-tilt camera mounting module which can make roll and pitch rotation. In this experiment, position of acoustic camera was fixed and pitch angle was kept at about 30 deg.

To inspect these objects, roll rotation, from  $-60$  deg to  $60$  deg was carried out for one time. It is worth mentioning that images are binarized by conventional Otsu's method (Otsu (1979)). The dense 3D reconstruction result is shown in Fig. 9. We can recognize objects and environment successfully. More viewpoints are necessary for a better result which is one of our future work.

## 5. CONCLUSION

In this paper, we applied a probabilistic method to realize 3D reconstruction from acoustic images with unknown elevation angles. A new inverse sensor model is designed to apply acoustic image data to occupancy mapping theory. Simulation is implemented and real experiment data is also used to confirm the feasibility of our algorithm. Our proposed framework can generate any shape of objects without relying on features.

Future work related to this paper will involve valuing our result quantitatively and mounting acoustic camera on an underwater robot to carry out real experiment with more viewpoints. Because it is difficult to acquire real value of robot pose, localization method may be necessary to get a precise result that a method to realize simultaneous localization and mapping (SLAM) will be implemented. Furthermore, after 3D reconstruction, 3D recognition is also very important for automatic inspection of the underwater environment. Finally, further proof is also necessary to find out whether our algorithm works on real sea bed with non-artificial objects.

## ACKNOWLEDGEMENTS

The authors would like to thank S. Fuchiyama, A. Ueyama, K. Okada and their colleagues at KYOKUTO Inc. for full access to their facilities for the real experiment. We would also like to thank Y. Yamamura, F. Maeda, S. Imanaga at TOYO Corp. who helped in this research project over a day for the equipment set up and the acquisition of the sonar data.

## REFERENCES

- Aykin, M.D. and Negahdaripour, S. (2017). Three-dimensional target reconstruction from multiple 2-d forward-scan sonar views by space carving. *IEEE Journal of Oceanic Engineering*, 42(3), 574–589.
- Belcher, E., Hanot, W., and Burch, J. (2002). Dual-frequency identification sonar(didson). *Proceedings of the 2002 IEEE International Symposium on Underwater Technology (UT)*, 187–192.
- Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6), 46–57.
- Fairfield, N., Kantor, G.A., and Wettergreen, D. (2005). Three dimensional evidence grids for slam in complex underwater environments. *Proceedings of the 14th International Symposium of Unmanned Untethered Submersible Technology (UUST)*.
- Guerneve, T. and Petillot, Y. (2015). Underwater 3d reconstruction using blueview imaging sonar. *Proceedings of the 2015 MTS/IEEE conference on OCEANS-Genova*, 1–7.
- Hornung, A., Wurm, K.M., Bennewitz, M., Stachniss, C., and Burgard, W. (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 34(3), 189–206.
- Huang, T.A. and Kaess, M. (2015). Towards acoustic structure from motion for imaging sonar. *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 758–765.
- Klingensmith, M., Hermann, M., and Srinivasa, S. (2014). Noisy data via voxel depth carving. *Proceedings of the 2014 International Symposium on Experimental Robotics (IESR)*.
- Kwak, S., Ji, Y., Yamashita, A., and Asama, H. (2015). Development of acoustic camera-imaging simulator based on novel model. *Proceedings of the 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, 1719–1724.
- Kwon, S., Park, J., and Kim, J. (2017). 3d reconstruction of underwater objects using a wide-beam imaging sonar. *Proceedings of the 2017 IEEE International Symposium on Underwater Technology (UT)*, 1–4.
- Mai, N.T., Woo, H., Ji, Y., Tamura, Y., Yamashita, A., and Asama, H. (2017a). 3-d reconstruction of line features using multi-view acoustic images in underwater environment. *Proceedings of the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligence Systems (MFI)*, 312–317.
- Mai, N.T., Woo, H., Ji, Y., Tamura, Y., Yamashita, A., and Asama, H. (2017b). 3-d 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 (IFAC2017)*, 1066–1072.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62–66.
- Rusu, R.B. and Cousins, S. (2011). 3d is here: Point cloud library (pcl). *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 1–4.
- Sun, N., Shim, T., and Cao, M. (2008). 3d reconstruction of seafloor from sonar images based on the multi-sensor method. *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligence Systems (MFI)*, 573–577.
- Teixeira, P.V., Kaess, M., Hover, F.S., and Leonard, J.J. (2016). Underwater inspection using sonar-based volumetric submaps. *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4288–4295.