Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-21 (2019) 291–296

Realistic Sonar Image Simulation Using Generative Adversarial Network

Minsung Sung* Jason Kim* Juhwan Kim* Son-Cheol Yu*

* *Department of Creative IT Engineering, Pohang University of Science and Technology, Pohang, Republic of Korea (e-mail: {ms.sung, js21kim, robot_juhwan, sncyu}@postech.ac.kr)*

Abstract: Sonar sensors are widely utilized underwater because they can observe long-ranged objects and are tolerant to measurement conditions, such as turbidity and light conditions. However, sonar images have low quality and hard to collect, so development and application of sonar-based algorithms are difficult. This paper proposes a method to generate realistic sonar images or to segment real sonar image, to better utilize the sonar sensors. A simple sonar image simulator was implemented using a ray-tracing method. The simulator could calculate semantic information of real sonar images including properties of highlight, background, and shadow regions. Then, a generative adversarial network translated the simulated images into more realistic images or real sonar images into simulated-like images. The proposed method can be used to augment or pre-process sonar images.

Keywords: Sonar simulator, Forward Scan Sonar, FSS, GAN, Sonar GAN.

1. INTRODUCTION

Sonar sensors are widely used to investigate underwater environments because they have wide scanning ranges and can observe objects in dark or highly turbid environments. Therefore, various sonar-based algorithms, such as underwater object detection (Kim and Yu 2017), localization (Kim et al. 2017), and mapping (Joe et al. 2018), have been developed.

One of the difficulties in developing sonar-based algorithms is collecting underwater sonar images. Developing sonar-based algorithms requires a large number of sonar images to verify and improve their approaches. For example, deep-learning-based underwater object detection methods require thousands of images of target objects to train neural networks. However, sonar sensors are not popular because of high price, and sonar images are rarely shared in public. Therefore, several underwater experiments should be performed to capture sonar images and to develop new sonar-based methods. Even if using an Unmanned Underwater Vehicle (UUV), these experiments take a lot of time and resources.

Sonar image simulators can be used to synthesize sonar images under various settings (Sung et al. 2019; Lee et al. 2019; Denos et al. 2017). However, simulating very realistic sonar images is difficult. First, the texture of an object depends on many factors, such as the material of the object and biofouling (Bozma and Kuc 1994). Modeling noise of sonar images is also complicated. There are various types of noise in sonar images, such as speckle noise and crosstalk noise. Moreover, each noise is hard to predict because it is affected by many different factors such as beamforming mechanism, backscattering, and multipath beam-reception (Hansen 2009). Therefore, sonar-based algorithms developed directly using the simulated sonar images many not work properly in the field.

This paper proposes a method to generate realistic sonar images or to translate real sonar images into simulated-like images. Using the sonar simulator based on a ray-tracing method, we could simulate simple sonar images of given objects. Because the noise and texture of images are hard to model, we just simulated images that were noise-free and simplified, but contained semantic information of real sonar images including properties of highlight and shadow regions. Then, we used the Generative Adversarial Network (GAN) to make the simulated sonar images more realistic. The GAN trained with pairs of the real sonar image and simulated image can translate the simulated image into realistic sonar image by adding noise and degradation effct. GAN can also perform an inverse translation; converting a real sonar image into a simulated-like image.

The proposed method can help development and utilization of various sonar-based algorithms. We can generate highly realistic sonar images of target objects and viewpoints with less cost and in a short time. The proposed method can also translate real sonar image into simulated-like images, which can be regarded as a denoising and segmenting the real sonar images. These images can be useful for extracting information such as edge and position of objects more reliably.

The rest of paper is organized as follows: In Section II, we addressed previous methods to implement sonar image simulators. Section III describes a method using GAN to translate simulated images into realistic sonar images and real sonar images into simulated-like images. Section IV explains experiments to obtain a sonar image dataset and to train the GAN. In section V, we analyzed the results of

10.1016/j.ifacol.2019.12.322

two types of the sonar image translation. The paper ends with the conclusion in section VI.

2. RELATED WORKS

Sonar image simulators have been developed to generate sonar images easily instead of conducting the experiments. One of the widely used methods is using a ray-tracing method (Bell and Linnett 1997; Gu et al. 2013; Kim et al. 2018b). Sonar images of given objects can be generated through the calculation of propagation and reflection of acoustic waves. However, due to computational complexity in modeling other characteristics such as reverberation and backscattering, these simulators assume propagation and reflection of the acoustic waves are ideal. As a result, the simulated results are different from real sonar images.

Simulators to synthesize more realistic sonar images have been developed. Kim et al. (2018a) added speckle noise to sonar images generated by ideal ray-tracing method. Riordan et al. (2019) developed a more realistic simulator which the field operation is possible beyond a laboratory level by accounting for more models for sonar such as the movement of UUVs, terrains, and scattering. Despite the computational complexity, the simulator could generate a sonar image in real-time using General-Purpose computing on Graphics Processing Units. Cerqueira et al. (2017) also proposed a realistic and real-time simulator. They used rasterization instead of ray tracing, so they could calculate the acoustic frame in a real-time using precomputed data such as normal vectors and distances. They also addressed speckle noise and materials of the objects for more realistic representation. However, synthesizing a very realistic sonar image is still difficult, because real sonar images have various kinds of noises and modeling of some phenomena is challenging according to the surrounding environments (Etter 1995).

Recently, neural networks (NN) for style transfer have been developed. Some studies introduced these NNs to process their sonar images. Chen and Summers (2016) generated sonar imagery by applying the style transfer to a feature map extracted from the target object for automatic target recognition. Lee et al. (2018) simulated a realistic sonar image of divers by applying NN named StyleBankNet to the depth map of a scene. Trained using a large amount of data, NN can find a high dimensional image-to-image mapping which is complex to model. We also proposed a method to generate a realistic sonar image and to translate a real sonar image into a simplified image by applying NN-based style transfer to a simple sonar simulator.

3. PROPOSED METHOD

We propose an approach for the realistic sonar image simulation that consists of two steps: simple sonar image simulation and image translation using GAN. First, we implemented the sonar image simulator based on the ray-tracing method. The simulator simulates images including only semantic information such as the shape of highlight and shadow through simple calculations. Then, GAN translates these simple images into realistic sonar images by adding noises, or convert real sonar image into the simplified images through denoising and segmentation. In this section, we explained these two parts in more detail.

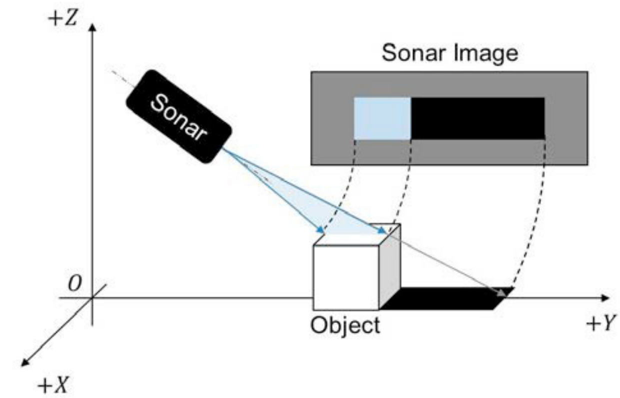


Fig. 1. Imaging mechanism of the sonar sensor.

3.1 Sonar Image Simulator

We implemented the ray-tracing-based sonar simulator by calculating propagation and reflection of acoustic waves according to the imaging principle of the sonar sensor as shown in Fig. 1. The sonar sensor shoots multiple acoustic waves at various azimuth angles to generate a sonar image. Then, the sonar sensor measures the time-of-flight (TOF) and the intensity of reflected waves. By multiplying the velocity of the acoustic wave and the TOF, the sonar sensor calculates the distance to the underwater objects. Then, the sonar image is constructed by mapping the measured intensity on the pixel of the corresponding azimuth angle and distance.

We modeled the imaging mechanism using vectors for convenient expressions. First, the reflection point of an acoustic wave is an intersection of the acoustic wave and the surface of the object. So, the position vector of the reflection point $\overrightarrow{p_\theta}$ can be expressed as:

$$\overrightarrow{p_\theta} = \frac{\overrightarrow{N} \cdot \overrightarrow{p_1}}{\overrightarrow{N} \cdot \overrightarrow{v_\theta}} \overrightarrow{v_\theta}. \tag{1}$$

where t is a constant, $\overrightarrow{v_\theta}$ is a unit direction vector of the acoustic wave transmitted in a direction of azimuth angle θ, $\overrightarrow{N}$ is the normal vector of the object surface, $\overrightarrow{p_1}$ is the position vector of one point of the object surface.

This position vector is calculated in the local coordinate system of the sonar sensor. So, the distance from the sonar sensor to the object in direction of azimuth angle θ is calculated as:

$$r = \left|\overrightarrow{p_\theta}\right|. \tag{2}$$

The sonar sensor scans the range between r_{min} and r_{max} for the distance and θ_{min} and θ_{max} for the azimuth angle and maps the area into M by N image. Thus, the pixel coordinate of reflection point (i, j) is expressed as:

$$i = \frac{\left|\overrightarrow{p_\theta}\right| - r_{min}}{r_{max} - r_{min}} \cdot (M-1) + 1, \tag{3}$$

$$j = \frac{\theta - \theta_{min}}{\theta_{max} - \theta_{min}} \cdot (N-1) + 1. \tag{4}$$

The pixel intensity mapped to this coordinate is determined by the intensity of the echo. Considered transmis-

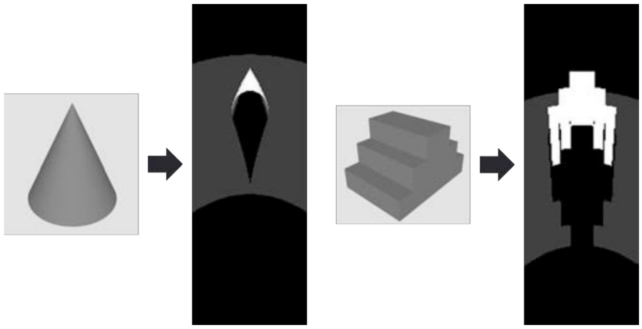


Fig. 2. Simulated sonar images of given 3D model.

sion loss, the incidence angle of the acoustic wave, and the acoustic impedance, the pixel intensity is calculated as (Catmull 1974, Kwak et al. 2015):

$$I = k\frac{z - z_0}{z + z_0} \cdot \frac{I_0}{|\overrightarrow{p_\theta}|} \cdot \cos^2 \alpha, \tag{5}$$

where k is a constant, z and z_0 are the acoustic impedance of the object and water, and I_0 is the initial intensity of acoustic wave, and α is the incidence angle.

As a result, the sonar simulator can predict the sonar image of a given model like Fig. 2. Because we only considered few sonar parameters, the simulated images are noise-free and a simplified versions of the real sonar images. Still, the simulator can well represent the semantic information of target objects, such as shape and location of highlight and shadow.

3.2 Generative Adversarial Network for Image Translation

We then proposed a method to translate the simulated images into realistic sonar images (sim-to-real) or real sonar images into simulated-like images (real-to-sim). Compared to the implemented simulator, the real sonar image contains additional noises and textures. Thus, to operate the simulator in a field, a method to make the simulated images more realistic or real sonar images noise-free is required.

We introduced a GAN-based style transfer for sim-to-real and real-to-sim sonar image translation. Sim-to-real and real-to-sim translation can be regarded as transferring the style of an image into the noisy and blurred form or noise-free and simplified form retaining contents such as shape and location of highlight and shadow of the input image. Deep learning can find high dimensional and non-linear image-to-image mapping relation, which is hard to model, through a large number of images. Among various deep learning-based method, GAN has recently recorded outstanding performance in generating some realistic images. Moreover, we can enable both sim-to-real and real-to-sim translation with a single network without implementing multiple image processing algorithms.

We adopted pix2pix model (Isola et al. 2016) which has proper structure for the proposed sonar-image-translation application. The generator of GAN is U-Net (Ronneberger et al. 2015) consisted of the 15 layers. U-Net has encoder-decoder architecture. In the encoder-decoder architecture, the feature map generated through the encoder captures the context and information of the input image. Thus, using this feature map, the generator can add or remove the noise preserving contents of the input image. Another feature of the generator is the skip connection. Through the skip connection, the decoder copies feature maps from earlier layers. So, the generator enables to localize the objects and features more precisely in the output image. As a result, the generator can translate sonar images accurately preserving semantic information such as highlights and shadows.

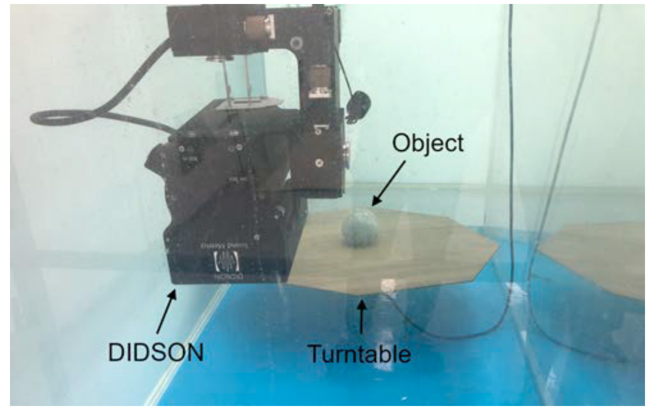


Fig. 3. Experimental setup of sonar image acquisition.

The discriminator of GAN is a convolutional neural network composed of four convolutional layers. The discriminator distinguishes whether a given image is a real image or an image synthesized by the generator. So, convolutional layers which are proper for classification were used to construct the discriminator. The discriminator observed input images by dividing them into small patches. Thus, the discriminator is not deceived easily if the generator does not represent the detail well, and the generator can produce more realistic images.

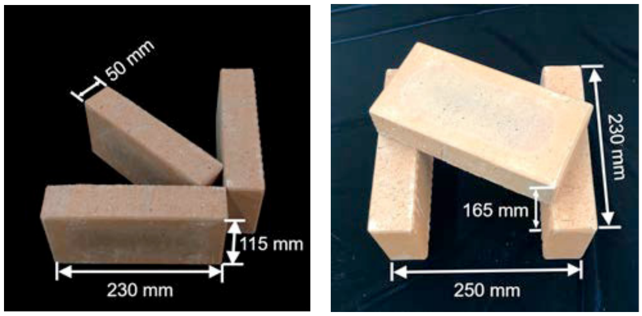
4. EXPERIMENT

Forward scan sonar (FSS) is a widely used sonar sensor because it has a high resolution compared to other sonar sensors. Thus, we applied the proposed method to the FSS.

4.1 Acquisition of Real Sonar Image

Pairs of images consisting of the real sonar image and the simulated sonar image are required to train the GAN. We first conducted indoor water tank experiments to acquire the real sonar image of the object. Training of NNs requires a large amount and various types of images. Addressing that even the same object appears differently in the sonar images according to the viewpoint, we designed a turntable to acquire various sonar images efficiently. The turntable can rotate the object on its wooden board at the desired angle about the z-axis of the object using a stepping motor.

Figure 3 illustrates the experimental setting for data acquisition. We installed the turntable on the floor of the water tank. Then, we put bricks in two types on the turntable. Finally, we captured the images of the objects rotating the turntable by one degree. DIDSON was used as the FSS (Belcher et al. 2002). Figure 4 shows the types of the objects we used, and Table 1 describes the



(a) Object type 1. (b) Object type 1.

Fig. 4. Types of objects used for the experiments.

Table 1. Experimental settings to capture images

Parameter	Value
Water tank size	1.35 m x 3 m x 1.7 m (width x length x height)
Sonar roll	0 °
Sonar pitch	-27 °
Sonar yaw	0 °
Object x position	1.9 m
Object y position	0 m
Object z position	-0.6 m

experimental conditions. We captured 360 sonar images of each type, so total 720 sonar images are acquired.

4.2 Training of Network

The simulated images corresponding to real sonar images are also required to train the GAN. Using the implemented simulator, we simulated the sonar images under the same sonar setting with the indoor experiments. First, parameters of sonar image simulator were set according to the specification of DIDSON like Table 2. Next, we modeled the bricks with computer-aided design (CAD). We located the CAD models on the same distance and viewpoint by multiplying translation and rotation matrix to the models. Finally, we simulated images corresponding to the real sonar images through (1)-(5).

Before the training of the GAN, we pre-processed the image pairs. Because the sonar sensor has limited vertical beamwidth, there is an area that no highlight appears in each image. Because this area contains only the noise, it disturbs the NN to be trained. So we manually cropped the images around the object. We also resized the cropped images. The original image size of DIDSON is 512 by 96. Since the decoder of generator extracts and pools the input image through seven layers, 96 is small for the input image. So we resize the input image size into 256 by 256.

We then employed data augmentation. To avoid overfitting, diversifying the input image is required. We used the translation, magnification, vertical and horizontal flips, and pixel intensity inversion. Picking out the similar images among 720 image pairs, we finally constructed a dataset composed of 412 pairs of images like Fig. 5.

Using this dataset, we trained the GAN in two models; sim-to-real and real-to-sim translation. For the sim-to-real translation, we used the simulated images as the input of the GAN and the real sonar images as the label of the GAN. For the real-to-sim translation model, we trained

Table 2. Parameters of sonar simulator for DIDSON

Parameter	Value
r_{min}	0.838 m
r_{max}	3.338 m
θ_{min}	-14.5 °
θ_{max}	14.5 °
M	512
N	96

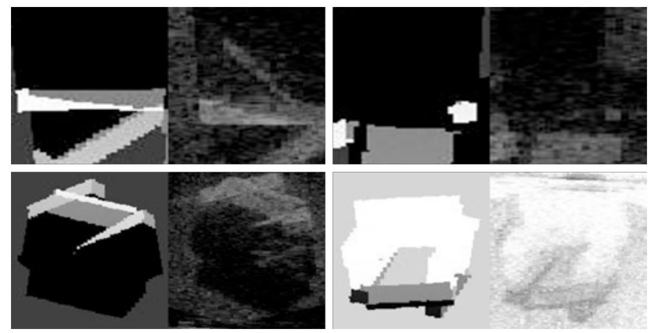


Fig. 5. Sample image pairs of training dataset.

the GAN with the real sonar images as input and the simulated images as the label.

5. RESULT

We trained each models for 850 epochs, and it took 7.6 hours using Graphics Processing Unit NVIDIA Titan X. As the training progresses, the generator and discriminator operate adversarially generating more and more realistic images. After completing the training of the GAN, we tested the GAN using 20 validation images. The validation dataset consisted of images of the same object as the training dataset but from the unknown viewpoint. In this section, we analyzed the results of the two trained models quantitatively.

5.1 Sim-to-Real Result

First, we tested the sim-to-real model. This model translates the simple image simulated by ray tracing into the realistic sonar image. We implemented another method for simulating realistic sonar image, which adds modeled speckle noise, to verify the performance of the proposed method. Figure 6 shows the results of the realistic sonar image simulation. The GAN predicted the realistic sonar image from the image generated by the simple sonar image simulator. The output preserved the important information such as highlight and shadow well. Compared to the noise-adding method, the output images of the proposed method better represented the blurred edges and overall distribution of the pixel intensities.

For quantitative analysis, we compared the similarity between the simulated image and the real sonar image. We used the two-dimensional (2D) discrete cross-correlation, expressed as:

$$r = \frac{\sum_{i,j}\left[I_1(i,j)-\overline{I_1}\right]\left[I_2(i,j)-\overline{I_2}\right]}{\sqrt{\sum_{i,j}\left[I_1(i,j)-\overline{I_1}\right]^2\sum_{i,j}\left[I_2(i,j)-\overline{I_2}\right]^2}}, \quad (6)$$

where $1 \le i \le M$, $1 \le j \le N$, $I(i,j)$ means the pixel intensity at a point (i,j), and $\overline{I}$ denotes the mean of pixel

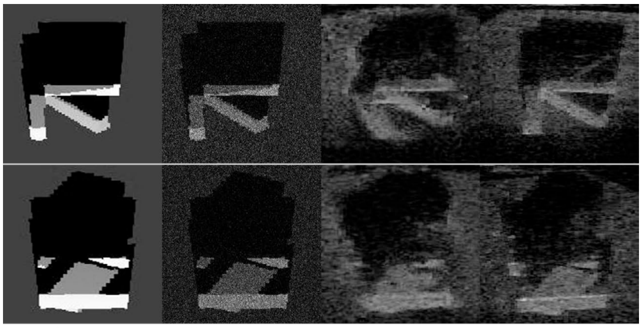


Fig. 6. Sim-to-real sonar image translation results.

Table 3. Similarity values of the simulated images

	Simulator		**Simulator +Speckle noise**		**Simulator +Sim-to-real translation**	
	with GT	with others	with GT	with others	with GT	with others
Cross-correlation	0.4251	0.0737	0.3244	0.0555	**0.5855**	0.1140
SSIM	0.2364	0.2063	0.1750	0.1427	**0.3105**	0.2216

intensity of image, and structural similarity (SSIM) (Wang et al. 2004) to calculate the similarity between the two images. The 2D cross-correlation showed the similarity between pixel units of the two images, and the SSIM showed the similarity based on the contextual or semantic information.

Table 3 shows the calculated similarity between the images. We calculated the similarity of one simulated image with the corresponding ground-truth (GT) real sonar image and with the images of different objects and viewpoints. The images generated by adding speckle noise looked similar to the real sonar image in human eyes as shown in Fig. 6, but the similarity based on the pixel and structure, which the computer used to recognize the image, was not high. On the other hand, the outputs of the proposed sim-to-real method showed a high similarity with the ground truth. Moreover, the similarity with the images of different objects and viewpoint did not increase. Therefore, we verified that the proposed method generates the realistic sonar images of the target object, and the proposed method can be used to develop the algorithms used for field operations.

5.2 Real-to-Sim Result

We also tested the real-to-sim model. This model translates the real sonar images into the simulated-like images. Figure 7 shows that the proposed method predicts the simulated-like image similar to the ground truth. The proposed method removed the noise from the input image. Moreover, like segmentation, the GAN preserves the important contents such as shape, size, and location of highlights and shadows, of the input image. We also analyzed the real-to-sim translation results quantitatively by measuring peak signal to noise ratio (PSNR). We assumed that the image generated by the simulator as the ground truth. Then, we calculated the PSNR of the sonar images before and after applying the real-to-sim translation to verify whether the proposed method can remove the noise effectively.

As a result, the real-to-sim translation was able to improve the PSNR by about 6 dB from 32.88 dB to 38.89 dB. It shows that the proposed method generates a clear and simplified image from the real sonar image.

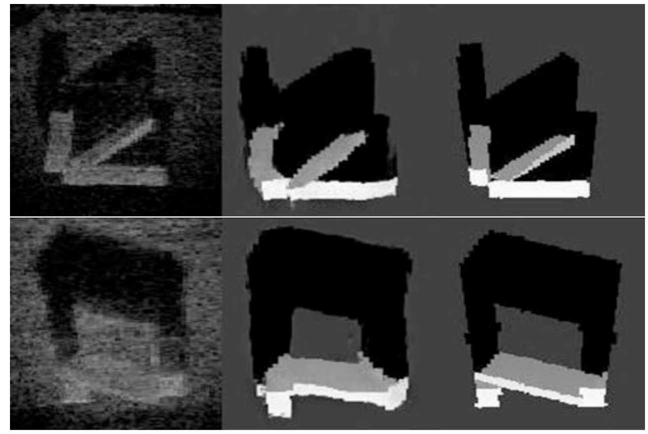


Fig. 7. Sim-to-real sonar image translation results.

6. CONCLUSION

In this paper, we proposed the sim-to-real and real-to-sim translation method for the sonar image. The proposed method can generate realistic sonar image by simulating the semantic information and applying the sim-to-real translation. The proposed method can also generate the noise-free and segmented image by applying the real-to-sim translation to real sonar image captured in the field. The proposed method can be used as data augmentation or pre-processing such as denoising and segmentation of sonar images.

For future works, we plan to improve the proposed method to process more general sonar images. In this paper, the proposed method successfully generated the realistic or simplified sonar images of target objects even if the input images were taken at a new viewpoint and not included in the training dataset. However, the results when unknown objects were given were not very good. To tackle this problem, We first plan to construct larger dataset composed of more various shapes of objects. We also plan to apply more diverse data-augmentation techniques. Then, the GAN may be robust to the semantic information of the input images and focus on the noises to add or remove. Next, we also plan to quantify the robustness of the proposed method by analyzing the geometry of the sonar and modeling the parameters for the sonar images. We expect the proposed method can be a framework for constructing dataset needed to develop underwater sonar algorithms.

ACKNOWLEDGEMENTS

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICT Consilience Creative program(IITP-2019-2011-1-00783) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2017R1A5A1014883). This research was a part of the project titled 'Gyeongbuk Sea Grant', funded by the Ministry of Oceans and Fisheries, Korea.

REFERENCES

Belcher, E., Hanot, W., and Burch, J. (2002). Dual-frequency identification sonar (didson). In *Proceedings of the 2002 Interntional Symposium on Underwater Technology (Cat. No. 02EX556)*, 187–192. IEEE.

Bell, J.M. and Linnett, L. (1997). Simulation and analysis of synthetic sidescan sonar images. *IEE Proceedings-radar, sonar and navigation*, 144(4), 219–226.

Bozma, O. and Kuc, R. (1994). A physical model-based analysis of heterogeneous environments using sonar-endura method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5), 497–506.

Catmull, E. (1974). A subdivision algorithm for computer display of curved surfaces. Technical report, UTAH UNIV SALT LAKE CITY SCHOOL OF COMPUTING.

Cerqueira, R., Trocoli, T., Neves, G., Joyeux, S., Albiez, J., and Oliveira, L. (2017). A novel gpu-based sonar simulator for real-time applications. *Computers & Graphics*, 68, 66–76.

Chen, J.L. and Summers, J.E. (2016). Deep neural networks for learning classification features and generative models from synthetic aperture sonar big data. In *Proceedings of Meetings on Acoustics 172ASA*, volume 29, 032001. ASA.

Denos, K., Ravaut, M., Fagette, A., and Lim, H.S. (2017). Deep learning applied to underwater mine warfare. In *OCEANS 2017-Aberdeen*, 1–7. IEEE.

Etter, P.C. (1995). *Underwater acoustic modeling: principles, techniques and applications*. CRC Press.

Gu, J.H., Joe, H.G., and Yu, S.C. (2013). Development of image sonar simulator for underwater object recognition. In *2013 OCEANS-San Diego*, 1–6. IEEE.

Hansen, R.E. (2009). Introduction to sonar. *Course Material to INF-GEO4310, University of Oslo,(Oct. 7, 2009)*.

Isola, P., Zhu, J.Y., Zhou, T., and Efros, A.A. (2016). Image-to-image translation with conditional adversarial networks. *arxiv*.

Joe, H., Cho, H., and Yu, S.C. (2018). Implementation of point cloud algorithm on three different installations of mechanically scanning imaging sonar for seabed mapping. In *TENCON 2018-2018 IEEE Region 10 Conference*, 1097–1101. IEEE.

Kim, B., Cho, H., Song, S., and Yu, S.C. (2017). Imaging sonar based navigation method for backtracking of auv. In *OCEANS 2017-Anchorage*, 1–5. IEEE.

Kim, B., Kim, J., Lee, M., Sung, M., and Yu, S.C. (2018a). Active planning of auvs for 3d reconstruction of underwater object using imaging sonar. In *2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV)*, 1–6. IEEE.

Kim, B. and Yu, S.C. (2017). Imaging sonar based real-time underwater object detection utilizing adaboost method. In *2017 IEEE Underwater Technology (UT)*, 1–5. IEEE.

Kim, J., Sung, M., and Yu, S.C. (2018b). Development of simulator for autonomous underwater vehicles utilizing underwater acoustic and optical sensing emulators. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, 416–419. IEEE.

Kwak, S., Ji, Y., Yamashita, A., and Asama, H. (2015). Development of acoustic camera-imaging simulator based on novel model. In *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, 1719–1724. IEEE.

Lee, M., Kim, J., and Yu, S.C. (2019). Robust 3d shape classification method using simulated multi view sonar images and convolutional nueral network. In *OCEANS 2019-Marseille*, 1–5. IEEE.

Lee, S., Park, B., and Kim, A. (2018). Deep learning from shallow dives: Sonar image generation and training for underwater object detection. *arXiv preprint arXiv:1810.07990*.

Riordan, J., Flannery, F., Toal, D., Rossi, M., and Dooly, G. (2019). Interdisciplinary methodology to extend technology readiness levels in sonar simulation from laboratory validation to hydrography demonstrator. *Journal of Marine Science and Engineering*, 7(5), 159.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Sung, M., Cho, H., Kim, J., and Yu, S.C. (2019). Sonar image translation using generative adversarial network for underwater object recognition. In *2019 IEEE Underwater Technology (UT)*, 1–6. IEEE.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.