

# Dense Depthmap Prediction from Ultrasonic Sensors

Koki Yasui Fumihiko Sakaue Jun Sato\*  
Yu Koyama Mitsuyasu Matsuura†

\* Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Japan  
E-mail: {yasui@cv, sakaue@, junsato@}nitech.ac.jp

† SOKEN INC., 500-20, Minamiyama, Komenoki-cho, Nisshin, Japan

**Abstract**—In this paper, we propose a method for estimating a dense depth map of a scene by combining information acquired from multiple ultrasonic sensors. In this method, the probabilities of the presence of an object at each discretized point in the scene are estimated based on the observed signals, and the depth map is estimated from this probability map. We also show that the depth map estimation can be stabilized by introducing prior knowledge such as the sparsity of the 3D scene as a regularization. Experiments on real and synthetic scenes confirm that the proposed method can estimate the depth map.

## I. INTRODUCTION

3D scene reconstruction using sensors such as cameras has been widely researched and developed due to its wide range of applications. In particular, various techniques for measuring 3D information using stereo cameras have been developed and are widely used due to the low cost of cameras. In recent years, methods such as LiDAR and ToF cameras are also widely used to measure distances by projecting invisible light such as infrared light onto an object and observing the reflected light with a camera. However, these methods using optical information work properly when proper observation is possible in the scene, but they cannot obtain accurate information when there is a medium in the scene that blocks the light, such as fog or rain.

On the other hand, distance sensors that use ultrasonic waves, i.e. ultrasonic sensors, are also widely used to acquire distance information. These sensors can directly measure the presence or absence of an object or the distance to an object by irradiating ultrasonic waves toward the object and measuring the reflected waves. It also has the advantage of being able to obtain accurate information in rainy and foggy environments and in the dark, which cameras are not good at. For this reason, ultrasonic sensors are widely used in situations where it is difficult to use a camera and where distance information to the object is important.

However, distance measurement techniques using ultrasonic sensors can only detect the approximate direction and distance to objects in the surroundings, and cannot measure the exact shape and size of the scene. Since the accurate scene information that cannot be obtained by ultrasonic sensors is very important for 3D information measurement, many systems use ultrasonic sensors and other types of sensors such as cameras to measure scene information by interpolating the obtained

information. However, sensors that use optical information, such as cameras, have the aforementioned problems related to fog and rain, and when they are used together, the final measurement results are greatly affected by the environment of the scene. Therefore, in this study, we propose a method for estimating depth maps, such as those obtained by ToF cameras and stereo cameras, using only the information obtained from ultrasonic sensors. If such a method can be realized, it will be possible to obtain scene information more stably even in scenes where optical sensors such as cameras cannot be used. We propose a method for estimating the distance between a camera and a scene, which can be used both indoors and outdoors.

## II. RELATED WORKS

Ultrasonic sensors are mainly used to determine whether there is an object in the vicinity of the sensor. However, since ultrasonic sensors have advantages such as being unaffected by changes in weather conditions such as fog and rain, and being able to directly acquire 3D information, research is also being conducted on how to use them for scene information analysis. One of them is a method for estimating a 3D scene by analyzing the information obtained from sensors using a neural network[6], [3]. In these methods, the shapes in the scene are represented by primitive 3D shapes such as cylinders and cubes, and detailed shape estimation of the entire scene is performed by detecting these shapes in the scene. In addition, with the recent development of neural networks[1], a method to obtain scene height information from ultrasonic sensor information has been proposed[5]. However, these methods use neural networks to analyze the patterns of information reflected from the primitive shapes to detect shapes similar to them, making it difficult to detect objects that do not match these conditions.

In this research, we focus on the amplitude of the sound wave information acquired by the ultrasonic sensors and aim to measure the scene information directly by analyzing it. In addition, we show how to calculate the probability that an object exists at each point in space using multiple ultrasonic sensors, and how to estimate the distance image of the space from the probability distribution.

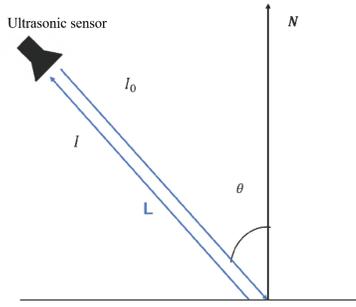


Fig. 1. Phong model for ultrasonic amplitude

### III. OBSERVATION MODEL

#### A. Phong Model

First, we consider the modeling of the signal observed by the ultrasonic sensor. The observed signal is a measurement of the reflected sound generated when the incident sound emitted by the ultrasonic sensor is reflected by an object in front of the sensor. Originally, the information of the measured sound wave includes various information such as phase. In this study, however, we will focus on the intensity of the sound wave. Since the intensity of the reflected sound is represented by the amplitude information of the sound wave, we will focus on the amplitude information of the observation signal to obtain the information of the target.

The amplitude of the observed signal depends on the normal direction of the object surface where the incident sound is reflected. In this paper, the amplitude of the observed signal is approximated by the Phong model [4], which is used to represent the reflection of light. The Phong model expresses the intensity of specularly reflected light. In this study, it is replaced by sound waves, and the amplitude  $I$  of the observed signal is expressed as follows:

$$I = I_0 k \cos^n \theta = I_0 k (\mathbf{L} \cdot \mathbf{N})^n \quad (1)$$

where  $\theta$  is the angle between the half-vector of the incident direction and the outgoing direction and the normal direction  $\mathbf{N}$ . When using an ultrasonic sensor, the incident and outgoing directions are equal, so the half vector is equal to the incident direction  $\mathbf{L}$ . Also,  $n$  is called the smoothness coefficient, and it represents the smoothness of the object's surface. Furthermore,  $I_0$  is the amplitude strength of the incident sound and  $k$  is the reflectance. These relationships are summarized in Fig.1.

#### B. Observation Model for Single Sensor

Next, we consider the signal observed by the ultrasonic sensor. If the speed of sound is 340m/s, the reflected sound wave is received by the sensor after  $2d/340$  seconds. Assuming that there is no interference between the sound waves, the amplitude of the received observation signal is the sum of the intensities of the sound waves reflected by objects on a sphere of radius  $d$  centered on the ultrasonic sensor. Using a point

$\mathbf{X}_d$  on the sphere and the normal to that point  $\mathbf{N}$ , we define the following function  $\delta$  as follows:

$$\delta(\mathbf{X}, \mathbf{N}) = \begin{cases} 1 & \text{when an object with normal } \mathbf{N} \text{ exists at point } \mathbf{X} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This function  $\delta$  expresses whether or not an object with normal  $\mathbf{N}$  exists at a point  $\mathbf{X}$ . Furthermore, if the Phong model can represent the amplitude, the amplitude of the observed signal  $I(d)$  corresponding to a certain distance  $d$  can be expressed as follows:

$$I(d) = \sum_{\mathbf{X}_d} \sum_{\mathbf{N}} \delta(\mathbf{X}_d, \mathbf{N}) I_0 k (\mathbf{L} \cdot \mathbf{N})^n \quad (3)$$

where  $I(d)$  is the amplitude of the observed signal corresponding to a certain distance  $d$ , and  $\mathbf{X}_d$  is the 3D point whose distance from the ultrasonic sensor is  $d$ .

According to this model, 3D shape reconstruction from ultrasonic sensor information is equivalent to estimating  $\delta$  from  $I(d)$ . However, it is difficult to stably estimate a function expressed in a discrete form as described above. Therefore, we consider approximating this function using the probability density function. The probability that an object exists at point  $\mathbf{X}$  and that the normal direction of the object's surface is  $\mathbf{N}$  is expressed as  $P(\mathbf{X}, \mathbf{N})$ . Using this, we can rewrite the equation (3) as follows:

$$I(d) = \sum_{\mathbf{X}_d} \sum_{\mathbf{N}} P(\mathbf{X}, \mathbf{N}) I_0 k (\mathbf{L} \cdot \mathbf{N})^n \quad (4)$$

Also, assuming that the probability of  $\mathbf{X}$  and the probability of  $\mathbf{N}$  are independent, this equation can be rewritten as follows:

$$I(d) = \sum_{\mathbf{X}_d} \sum_{\mathbf{N}} P(\mathbf{X}) P(\mathbf{N}) I_0 k (\mathbf{L} \cdot \mathbf{N})^n \quad (5)$$

where  $P(\mathbf{X})$  is the probability that an object exists at point  $\mathbf{X}$  and  $P(\mathbf{N})$  is the probability that the normal direction of the surface of an object is  $\mathbf{N}$ . If we assume that the normal direction is uniformly distributed regardless of the 3D point, then  $P(\mathbf{N})$  is constant regardless of  $\mathbf{X}$ . In this case, the equation (5) can be rewritten as follows:

$$I(d) = a \sum_{\mathbf{X}_d} P(\mathbf{X}) \quad (6)$$

where,  $a = \sum_{\mathbf{N}} P(\mathbf{N}) I_0 k (\mathbf{L} \cdot \mathbf{N})^n$ . Furthermore, the amplitude of the sound wave reflected from all points on the sphere is defined as follows:

$$I_{all} = \sum_{\mathbf{X}_d} a \quad (7)$$

If the probability that an object exists at all points on a sphere with distance  $d$  is uniform, then  $P(\mathbf{X}_d)$  can be determined as follows:

$$P(\mathbf{X}_d) = \frac{I(d)}{I_{all}} \quad (8)$$

From equation (8), we can calculate the probability that an object exists at point  $\mathbf{X}_d$ .

### C. Model for Multiple Sensors

Next, we consider a method for estimating the probability of the existence of an object using multiple observation signals obtained when multiple ultrasonic sensors are used. For simplicity, we assume that there is only one object in the scene and that the distance to the object is measured by ultrasonic sensors placed at different locations. In this case, the information obtained from a single ultrasonic sensor is the probability that the object exists somewhere on the surface of a sphere with a certain radius centered on the ultrasonic sensor, and it is not possible to determine where on the sphere the sound wave was reflected from based on this information alone. When two ultrasonic sensors are used, it is possible to limit the location to the circumference of the intersection of the surfaces of the two hemispheres centered on each ultrasonic sensor, but it is still not possible to obtain a unique location. When three ultrasonic sensors are used, the intersection of the surfaces of the three hemispheres becomes a point, and the object can be determined to exist at that one point. Therefore, at least three ultrasonic sensors are required to estimate the position of an object accurately. Therefore, in this paper, we use multiple ultrasonic sensors to achieve dense shape estimation of the scene.

We assume that the probability of existence of a point  $\mathbf{X}_d$  observed by ultrasonic sensor  $k$  is  $P_k(\mathbf{X}_d)$ . Assuming that the information obtained from each ultrasonic sensor is independent, the existence probability  $P_m(\mathbf{X}_d)$  at point  $\mathbf{X}_d$  can be defined as follows:

$$P_m(\mathbf{X}_d) = \prod_{k=1}^K P_k(\mathbf{X}_d) \quad (9)$$

where  $K$  is the number of ultrasonic sensors. This allows us to estimate the probability of existence at each point using the information obtained from all ultrasonic sensors.

## IV. DEPTH PREDICTION

### A. Probability Model

Next, we assume that the target space is sampled by a set of voxels, and consider a method to calculate distance information from the existence probability at each point. Here, we assume that multiple ultrasonic sensors are placed on a certain plane, and consider a method for estimating the distance image from that plane.

The probability of the presence of an object in each voxel can be obtained from the equation (9). Assuming that sound waves do not pass through objects, there is only one voxel that exists on the straight line connecting a point and a sensor. Therefore, if we determine the position of the object on this line, we can obtain the dense 3D shape of the scene. If we assume that an object exists independently on each line, we can determine that the object exists in the voxel with the maximum probability of existence on each line. Therefore, the depth from the plane where the multiple ultrasonic sensors are installed to the object can be obtained as follows:

$$\hat{d}(x, y) = \operatorname{argmax}_d P_m(\mathbf{X}_d(x, y)) \quad (10)$$

where,  $\hat{d}(x, y)$  is the depth of a straight line from a point  $(x, y)$  on the plane perpendicular to the plane. Also,  $\mathbf{X}_d(x, y)$  denotes a point on the same line whose depth is  $z$ . By performing these operations for all points on the plane, we can estimate the distance image based on the information obtained from multiple ultrasonic sensors.

### B. Smoothness Constraint

In the method of depth estimation described above, the distance image is estimated by assuming that all the points on the line exist independently. However, they are not independent by nature but are continuous with each other. Therefore, in this section, we assume that the shape change of the 3D scene to be estimated is smooth, and investigate how to estimate the 3D shape stably using this assumption.

First, we describe the regularization based on the smoothness constraint. In general, it is known that in our 3D scenes, the depth of the image does not fluctuate rapidly, and the fluctuation is smooth. Therefore, we can assume that the spatial changes in the estimated 3D scene will occur smoothly as well. Therefore, if the spatial smoothness constraint is satisfied, the following evaluation equation becomes small.

$$E_z = \sum_{x,y} ( \|d(x, y) - d(x + 1, y)\|^2 + \|d(x, y) - d(x, y + 1)\|^2 ) \quad (11)$$

where  $D(X, Y)$  is the depth at point  $(X, Y)$ .

Therefore, the distance to the object is such that it satisfies the regularization term expressed as equation (11) and maximizes the probability of the object's presence. This can be thought of as maximizing the following evaluation equation.

$$E' = \prod_{x,y} P_m(\mathbf{X}_d(x, y)) - w'_r E_z \quad (12)$$

where  $w'_r$  is the weight for regularization. This equation can be replaced by the problem of minimizing the following evaluation equation by calculating the logarithm of the first term.

$$E = \sum_{x,y} \log P_m(\mathbf{X}_d(x, y)) + w_r E_z \quad (13)$$

By minimizing such an evaluation equation, we can estimate the distance image taking into account the smoothness of the space.

### C. Distribution Approximation by Gaussian Distribution

It is possible to obtain an appropriate shape by minimizing the equation (13), but when  $P_m$  can take any probability distribution, a large amount of calculation is required for its minimization. However, when  $P_m$  can take any probability distribution, the minimization of  $P_m$  requires a large amount of calculation. Therefore, in this study, the probability distribution on each line is approximated by a normal distribution to simplify the calculation. In this case, the equation (13) can be rewritten as follows:

$$E = \sum_{x,y} \left\| \frac{d(x, y) - \mu(x, y)}{\sigma(x, y)} \right\|^2 + w_r E_z \quad (14)$$

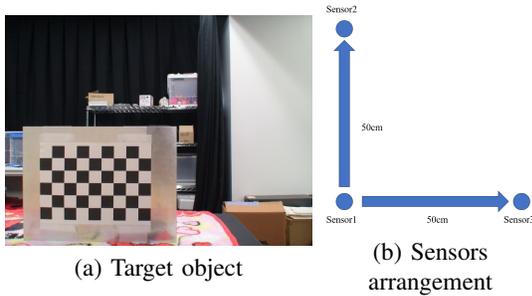


Fig. 2. Arrangement of the target and the sensors.

where  $\mu(x, y)$  and  $\sigma(x, y)$  are the mean and standard deviation at point  $(x, y)$ , respectively, and are obtained from  $P(\mathbf{X}(x, y))$ . Since the newly defined evaluation equation can be minimized by the general least-squares method, the optimal depth can be obtained with a limited number of calculations. This makes it possible to efficiently estimate the depth information including the regularization term based on the smoothness constraint.

### V. EXPERIMENTAL RESULTS

#### A. Environment

We experimented to confirm the effectiveness of the distance image estimation from ultrasonic sensor information proposed in this paper. First, we describe the environment used in this experiment. In this experiment, three ultrasonic sensors were placed on the same plane, and a rectangular plate with a length of 30 cm and a width of 40 cm was placed 90 cm in front of the ultrasonic sensors as shown in Figure 2(a). The three ultrasonic sensors were installed as shown in Figure 2(b). The three ultrasonic sensors were set up as shown in Fig. 2(b), and we adjusted the orientation of the sensors to increase the common range of the sound waves emitted by the three sensors. In order to suppress the reflection of sound from the object to be restored, we placed a thick cloth that absorbs sound waves and reduces the reflection of sound on the table.

#### B. Depthmap Estimated Results

In this environment, we show the results of distance image estimation from the information of multiple ultrasonic sensors. First, the observation signals obtained from the three ultrasonic sensors used in this experiment are shown in Figure 3. The RGB image obtained from the camera placed at the same location as the ultrasonic sensor 1 is shown in Figure 2(a). The result of estimating the distance image from the observation signal shown in Figure 3 is shown in Figure 4. From this result, we can see that the object is estimated as a distance closer to the lower left than to the other regions. This is consistent with the fact that the object to be estimated is located in the lower-left corner of the correct scene. The distance to the object is estimated to be about 80cm, which means that the distance is estimated with an error of about 10cm.

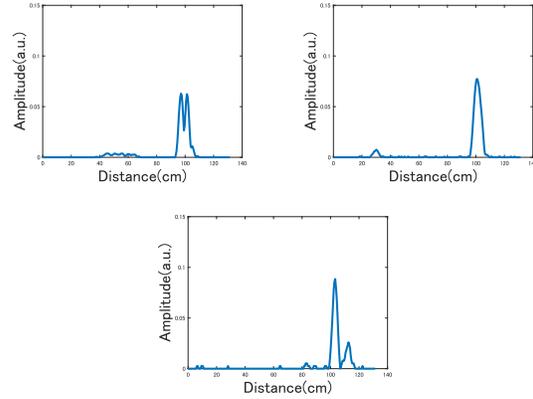


Fig. 3. Observation signals from three ultrasonic sensors.

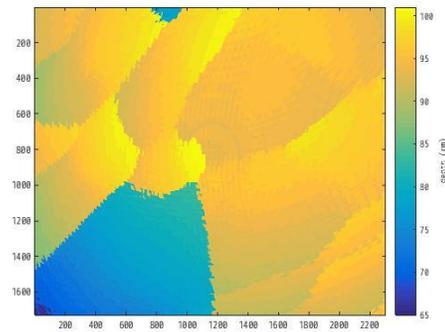


Fig. 4. 3D scene in real environment experiment

However, while the shape of the object is rectangular, the estimated distance image does not show the rectangular shape of the object. This indicates that it is difficult to estimate the detailed shape of the object using only the information obtained from the three ultrasonic sensors, especially for the edge regions of the object. The reason for this is that the direction of the reflected wave and the normal direction of the object are shallow in these edge regions, and the reflected wave is measured weakly. This problem can be improved by increasing the number of ultrasonic sensors used in the experiment and increasing the amount of information obtained from the 3D scene to be estimated.

#### C. Evaluation

Next, we evaluated the accuracy of the estimation results of a dense 3D scene in a real environment experiment. In order to evaluate the accuracy, we placed the target object in front of the three ultrasonic sensors on the same plane, and measured it while moving this position to check whether appropriate distance measurement was possible. The object was a rectangular plate, and we moved the object from 60 cm in front of the ultrasonic sensor to 90 cm in front at 10 cm intervals. The observation signals acquired by the three

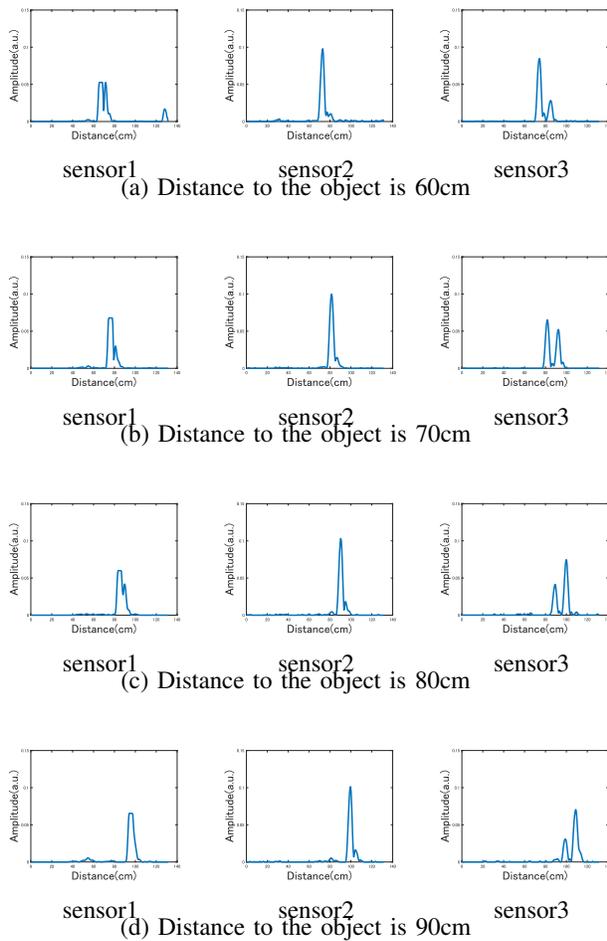


Fig. 5. Observation signal for each distance

ultrasonic sensors in each scene are shown in Figure 5. The observed images and estimated distance images for each scene are shown in Figure 6. Furthermore, the average distance error in each scene is shown in Table I. From these results, we can see that the position of the object is generally estimated, but the distance to the object is shifted forward from the original distance. This is probably due to the strong reflection from the desk on which the object is placed in the real environment experiment, which caused the estimated distance to be shifted forward. Note that, although the wall behind the object cannot be measured properly, this is due to the limitation of the sensor. The sensor utilized in this experiment is mainly designed to measure the distance to the object placed at a short distance, so it is not able to measure the distance to the far wall properly. This problem can be easily solved by using sensors that can measure a long distance.

*D. Evaluation in Synthesized Environment*

Next, we show the results of evaluating the proposed method by simulation experiments. In this experiment, we especially

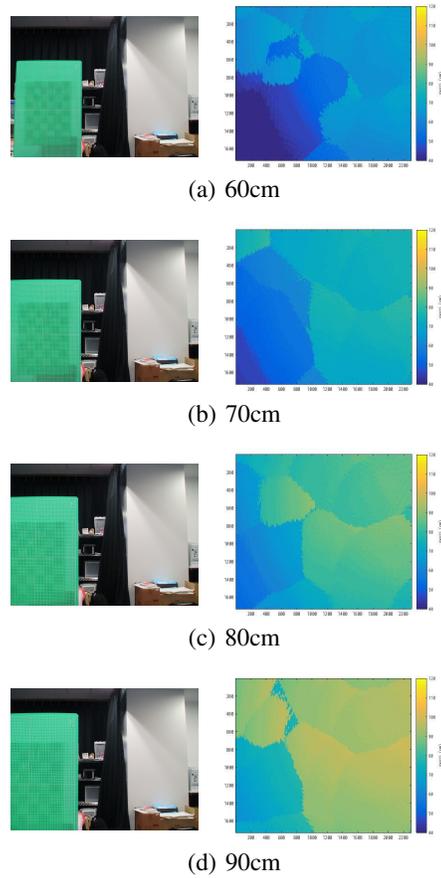


Fig. 6. Observed and recovered distance images for each distance

TABLE I  
AVERAGE DISTANCE ERROR FROM THE CORRECT 3D SCENE IN REAL-WORLD EXPERIMENTS

distance	Average error(cm)
60cm	18.0
70cm	11.7
80cm	22.3
90cm	18.4

evaluated the effect of the regularization term. To create the simulation data, we used the NYU Depth Dataset V2[2], which contains RGB images and distance images captured by Microsoft Kinect. From this data, we created observation signals acquired by ultrasonic sensors and used them as input to reconstruct the shape of the scene. We confirmed the effect of regularization by varying the regularization weights during the reconstruction process.

The 3D scene shown in Figure 7 was used as the experimental data. Three ultrasonic sensors were placed on this distance image at intervals of 50 cm as shown in Figure 8, as in the real environment experiment, and we calculated the observation signals acquired by each ultrasonic sensor. We assumed that the camera is placed at the same position as the ultrasonic sensor 1, and the correct 3D scene is the distance image at

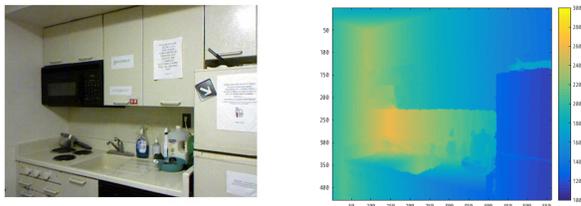
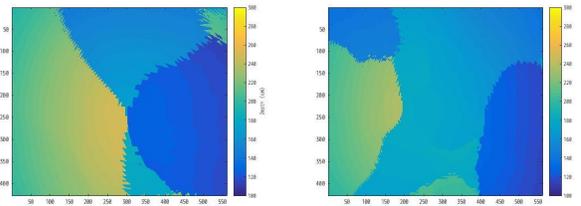


Fig. 7. Target scene



(a) Results by three sensors (b) Results by five sensors

Fig. 10. Predicted depth map



(a) Arrangement of three sensors (b) Arrangement of five sensors

Fig. 8. Arrangement of ultrasonic sensors

TABLE II  
MEAN ABSOLUTE ERROR

# of sensors	MAE (cm)
3	27.5
5	22.8

that camera position. The observation signal was calculated assuming that the sound emitted from the ultrasonic sensor does not attenuate and that the sound hits all pixels in the scene. The amplitude of the reflected wave was assumed to be equal to the magnitude of the incident wave. This dataset contains distance images of distant objects. However, in this experiment, we assumed that the measurable range of the sensor was 300cm, and excluded such data, and conducted the experiment with 178 data consisting of objects within 300cm. The temperature in the room was assumed to be 14 degrees Celsius, and we assumed the speed of sound to be 340 meters per second. In order to investigate the change in estimation accuracy depending on the number of sensors, we calculated the observed signals for the case where five ultrasonic sensors were used in the same way. The observation signals acquired by each sensor are shown in Figure 9.

E. Evaluation of the number of ultrasonic sensors

The results of the distance image estimated by three ultrasonic sensors and the distance image estimated by five sensors are shown in Figure 10. The results show that the

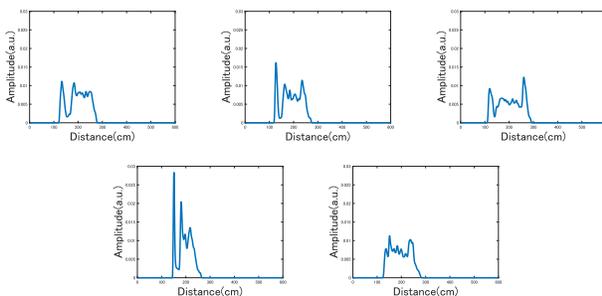


Fig. 9. Observed signals

object on the right side of the scene can be estimated as a short distance object in both cases. In particular, when five sensors are used, the proposed method can estimate the distance more accurately. This result indicates that the more ultrasonic sensors are used, the more accurately the method can estimate the object’s shape in the 3D scene. The average distance error from the correct 3D scene for all scene is shown in Table II. This result also shows that our proposed method can estimate the correct 3D scene using five ultrasonic sensors. This indicates that by increasing the number of sensors, it is possible to obtain more information about the scene and improve the accuracy of the distance image estimation.

F. Evaluation with Smoothness Regularization

Next, to evaluate the regularization based on the spatial smoothness constraint, we examined the change in the accuracy of 3D scene estimation with and without the regularization based on the spatial smoothness constraint. The results for the case with three ultrasonic sensors and the case with five ultrasonic sensors are shown in Figure 11. The results for the case with three ultrasonic sensors and the case with five ultrasonic sensors are shown in Fig. 11. The results show that the overall estimation is smoother than the case without regularization. It can also be confirmed that the lower right part of the image is closer to the sensor as in the case without regularization. It can be seen that the regularization term based on the spatial smoothness constraint is a more natural representation of the 3D scene. The average distance error between the correct 3D scene with/without regularization is shown in Table III. The results show that the average distance error without spatial regularization is smaller than the average distance error with spatial regularization, confirming that the estimation accuracy is better with spatial regularization. Therefore, it is confirmed that the regularization of the shape is effective in the proposed method.

VI. CONCLUSIONS

In this paper, we proposed a method for estimating the distance image using only the information obtained from

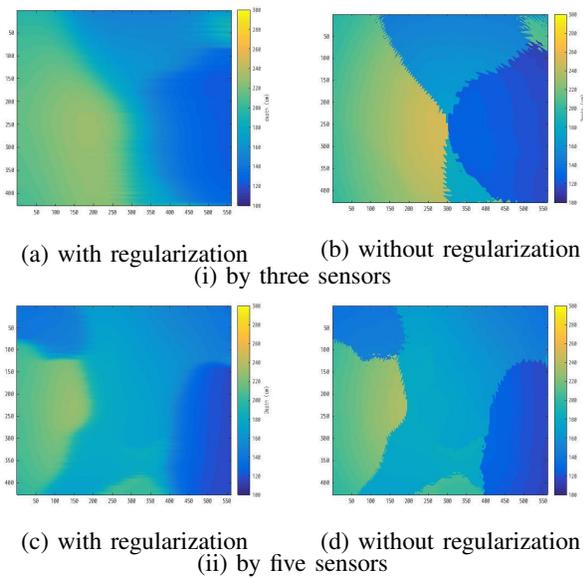


Fig. 11. Effect of regularization

TABLE III  
MEAN ABSOLUTE ERRORS

	MAE (cm)
w/o regularization (three)	27.5
with regularization (three)	27.1
w/o regularization (five)	22.8
with regularization (five)	22.6

multiple ultrasonic sensors. For this purpose, we modeled the observation signals acquired by ultrasonic sensors based on the Phong model. We showed that the amplitude information of the modeled observation signals can be expressed based on probability. The amplitude information of the modeled signals can be expressed based on probability. Using the amplitude information expressed based on probability, we showed how to derive the probability that an object exists at a certain point in 3D space when multiple ultrasonic sensors are used. Then, a method for estimating a dense 3D scene using the derived probability of the existence of an object is presented. We also introduced smoothness constraints as prior knowledge of the 3D scene to be estimated, and showed an optimization method that includes the smoothness constraints as regularization terms. Finally, a method for estimating a dense 3D scene based on the existence probability of an object is implemented in real space and shown to be feasible in practice. Furthermore, we conducted an evaluation experiment to determine the validity of the regularization term based on the smoothness constraint, and evaluated the regularization term based on the smoothness constraint.

In the future, we would like to develop a method for estimating the 3D scene by adding a regularization term related to the normal direction, because the observation signals acquired by ultrasonic sensors are easily affected by the normal direction of the object surface in the 3D scene.

REFERENCES

- [1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [2] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.
- [3] K. Ohtani, M. Baba, and T. Konishi. Position and posture measurements and shape recognition of columnar objects using an ultrasonic sensor array and neural networks. *Systems and Computers in Japa*, 33(11):27–38, 2002.
- [4] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.
- [5] Maximilian Popperli, Raghavendra Gulagundi, Senthil Yogamani, and Stefan Milz. Capsule neural network based height classification using low-cost automotive ultrasonic sensors. In *Proc. IEEE Intelligent Vehicles Symposium(IV2019)*, 2019.
- [6] N. Yoshizawa, H. Tada, and T. Yabuta. Object recognition by neural network using thickness data from acoustic image. *System and Computers in Japan*, 24(8):95–106, 1994.