



# Object Surface Recognition Based on Standing Waves in Acoustic Signals

Makoto Kumon<sup>1\*</sup>, Rikuto Fukunaga<sup>2</sup>, Tomoya Manabe<sup>3</sup> and Kei Nakatsuma<sup>1</sup>

<sup>1</sup>Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Japan, <sup>2</sup>Graduate School of Science and Technology, Kumamoto University, Kumamoto, Japan, <sup>3</sup>Faculty of Engineering, Kumamoto University, Kumamoto, Japan

This paper proposes the use of the standing waves created by the interference between transmitted and reflected acoustic signals to recognize the size and the shape of a target object. This study shows that the profile of the distance spectrum generated by the interference encodes not only the distance to the target, but also the distance to the edges of the target surface. To recognize the extent of the surface, a high-resolution distance spectrum is proposed, and a method to estimate the points on the edges by incorporating observations from multiple measurement is introduced. Numerical simulations validated the approach and showed that the method worked even in the presence of noise. Experimental results are also shown to verify that the method works in a real environment.

## OPEN ACCESS

### Edited by:

Caleb Rascon,  
National Autonomous University of  
Mexico, Mexico

### Reviewed by:

Edgar Garduño Ángeles,  
National Autonomous University of  
Mexico, Mexico

Siqi Cai,  
National University of Singapore,  
Singapore

### \*Correspondence:

Makoto Kumon  
kumon@gpo.kumamoto-u.ac.jp

### Specialty section:

This article was submitted to  
Computational Intelligence in Robotics,  
a section of the journal  
Frontiers in Robotics and AI

Received: 10 February 2022

Accepted: 05 April 2022

Published: 25 April 2022

### Citation:

Kumon M, Fukunaga R, Manabe T and  
Nakatsuma K (2022) Object Surface  
Recognition Based on Standing  
Waves in Acoustic Signals.  
Front. Robot. AI 9:872964.  
doi: 10.3389/frobt.2022.872964

**Keywords:** robot audition, object recognition, echo location, standing wave, distance spectrum

## 1 INTRODUCTION

Robots are expected to be used in various environments, which requires them to have situational awareness and the autonomy to make proper decisions accordingly. Thus, environment recognition is one of the key functions for autonomous robots.

Visual sensors such as cameras and light detection and ranging devices have been intensively investigated for this purpose. Especially for three-dimensional recognition with visual sensors, cameras with a depth or range sensing function are developed; for example, a time-of-flight camera was incorporated for mapping the environment May et al. (2009), a structured-light approach was proposed for the range recognition Boyer and Kak (1987), and a light-section technique Wang and Wong (2003) was introduced for accurate range measurement of targets nearby. Another candidate is acoustic sensors e.g., ultrasonic range sensors for obstacle detection, and sound navigation and ranging for underwater guidance. While the measurement resolution of acoustic sensors is not as accurate as that of visual sensors, acoustic measurements can compensate visual methods because acoustic observation is robust to visual interference, such as changing lighting conditions, challenging medium quality due to small particles such as fog or dust, and occlusions. Specular surfaces or transparent objects (e.g. glass objects) cause “blind spots” in visual measurements, and acoustic methods can be employed to detect such objects.

Popular acoustic sensors use ultrasonic signals. Ultrasonic sensors are “active” in the sense that they emit the signal and receive a reflected signal. The features of the reflected signal, such as time-of-flight Santamaria and Arkin (1995); Kuc (2008), the spectrum Reijniers and Peremans (2007), and the envelope Egaña et al. (2008), are analyzed to extract information about the object. Such echolocation architecture can be also integrated into a robotic localization system (Steckel et al. (2012); Lim and Leonard (2000).

The detectable range of high-frequency signals with a short wavelength is limited because of attenuation during propagation; the common detection range of commercial ultrasonic sensors is up

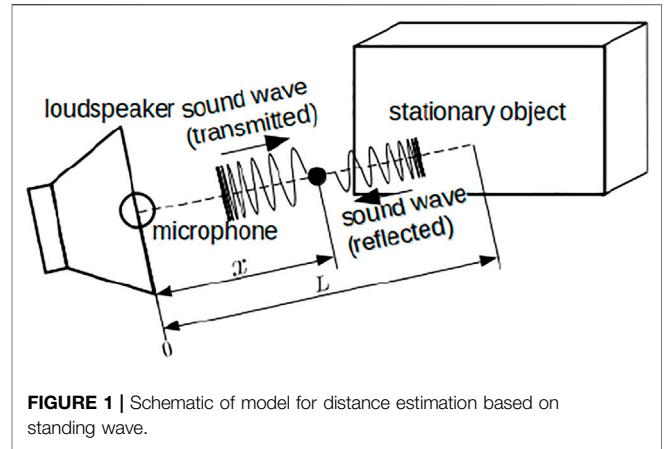
to several meters. In contrast, audible signals with a frequency of 20–20 kHz have longer wavelengths, and a longer penetration distance than ultrasonic signals, which may improve the detection range. An example of the use of such audible signals is visually impaired people using “clicking” tones to recognize their environments and objects Kolarik et al. (2014). Users trained with clicking have a detection range of more than 30 m Pelegrin-Garcia et al. (2018), and they can recognize not only the distance to targets but also their shapes Thaler et al. (2018).

Uebo et al. (2009) proposed to the use of audible signals to estimate the distance based on standing waves produced by interference between the transmitted and reflected signals; the power spectrum of the standing wave has a periodic structure in the frequency domain, and its period encodes the distance between the sensor and the object. Kishinami et al. (2020) proposed a method for recognizing multiple objects by using a distance estimation based on a standing wave incorporating the phase information with reliable frequency selection Takao et al. (2017).

Shape and orientation of the target object are useful features to recognize. For example, a method was proposed for recognizing the features of the target, such as flat or sharp corner from the profile of the reflected ultrasonic pulses Smith, (2001). Another approach is to estimate the shape of objects given a model of its shape (Santamaria and Arkin (1995)). Recently, BatVision (Christensen et al. (2020)) and CatChatter (Tracy and Kottege (2021)) proposed incorporation of statistical learning techniques to reconstruct a depth image from acoustic echoes to recognize complicated shapes.

Because a wide-beam audible signal can cover a large area with long penetration, this paper proposes extending the distance recognition based on the standing wave to recognize the shape and the orientation of the target of interest. Authors (Kumon et al. (2021)) previously reported that the profile of a distant spectrum (Uebo et al. (2009)) encodes not only the distance to the target but also the distance to the edges of the target surface; peaks in the distance spectrum profile correspond to reflections from points on the edges. Because the distance to the edge may slightly differ to that to the surface, it is important for the distance spectrum to have sufficient resolution to distinguish them. To recognize those distances, this paper proposes the use of a high-resolution distance spectrum to detect such peaks clearly. With this high-resolution distance spectrum, the surface recognition framework in the previous work (Kumon et al. (2021)) can improve the estimate. This paper also proposes the use of multiple observations at different locations to reconstruct the shape.

The rest of this paper is organized as follows. Distance measurements based on the standing wave in an acoustic signal, as proposed by Uebo (Uebo et al. (2009)), are briefly summarized in **Section 2**. The distance spectrum profile is examined in **Section 3**, with reference to our previous work (Kumon et al. (2021)), and a method to estimate the shape of the target object is proposed. Then, we describe how the proposed was validated by numerical simulations and experiments (**Section 4**). Conclusions are given in **Section 5**.



**FIGURE 1** | Schematic of model for distance estimation based on standing wave.

## 2 USING A STANDING WAVE FOR DISTANCE MEASUREMENT

This section describes the method for estimating the distance to an object using acoustic signals as proposed by Uebo (Uebo et al. (2009)).

**Figure 1** shows a schematic of the distance estimation method based on a standing wave. The signal transmitted from a speaker at time  $t$  to a point  $x$  is denoted as  $v_{Tr}(t, x)$ . Further, we assume that the signal is a linear chirp given by

$$v_{Tr}(t, x) = Ae^{j\left(2\pi \int_0^{t-|x|/c} f(\tau)d\tau + \theta\right)}, \quad (1)$$

where  $f(\tau)$  is the instantaneous frequency defined as

$$f(\tau) = \frac{f_w}{T}\tau + f_1, \quad (2)$$

and  $A$ ,  $c$ , and  $\theta$  represent the amplitude, speed of sound and phase, respectively; and  $T$ ,  $f_1$ ,  $f_N$ , and  $f_w$  represent the duration, lowest frequency, highest frequency, and bandwidth  $f_w = f_N - f_1$ , respectively.

The signal reflected from a stationary object that locates at the distance  $L$  from the microphone is denoted as  $v_{Ref}(t, x)$ , and its assumed model is as follows:

$$v_{Ref}(t, x) = Aye^{j\phi}e^{j\left(2\pi \int_0^{t-(2L-|x|)/c} f(\tau)d\tau + \theta\right)}, \quad (3)$$

where  $\gamma$  and  $\phi$  are reflection parameters that are constant over the frequency range of interest. The mixed signal  $v_C(t, 0)$  at the origin ( $x = 0$ ), where a microphone is located, can be computed as  $v_C(t, 0) = v_{Tr}(t, 0) + v_{R}(t, 0)$ . The power of  $v_C(t, 0)$  becomes

$$|v_C(t, 0)|^2 = |A|^2 \left\{ 1 + \gamma^2 + 2\gamma \cos\left(2\pi \frac{f_w}{T} \frac{2L}{c} t - 2\pi \frac{f_w}{2T} \left(\frac{2L}{c}\right)^2 + 2\pi f_1 \frac{2L}{c} - \phi\right) \right\}. \quad (4)$$

The first and the second terms of the right-hand side show the power of the transmitted and received signals, and they are constant. The cosine term shows the effect of the interference

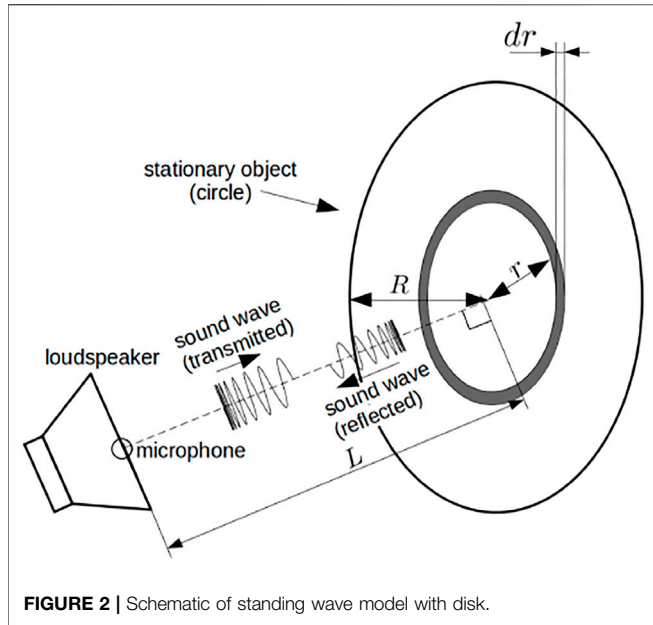


FIGURE 2 | Schematic of standing wave model with disk.

and it is independent of the initial phase  $\theta$ . Substituting (2) into (4) gives, the power of  $v_C$  in the frequency domain as

$$p(f, 0) = |A|^2 \left\{ 1 + \gamma^2 + 2\gamma \cos\left(\frac{4\pi L}{c} f + C_0\right) \right\}, \quad (5)$$

where  $C_0$  is a constant term for the interference. As (5) is a biased cosine function, the power spectrum of an unbiased  $p(f, 0)$  over the frequency range  $(f_1, f_N)$  forms an impulsive peak at the frequency of the cosine function. By rescaling this frequency with a factor  $\frac{c}{2L}$ , the power spectrum becomes the “distance spectrum” because it has a peak at  $L$  that represents the distance to the object.

### 3 OBJECT RECOGNITION USING DISTANCE SPECTRUM PROFILE

#### 3.1 Disk Model

The distance spectrum described in the previous section is based on a single reflected signal, which models the target object as a point. However, common objects in the real environment are difficult to approximate as points, and reflected signals from such objects can be affected by their spatial structure. To take the spatial effect into account, a disk reflection model is introduced (Figure 2) as previously reported (Kumon, M. et al. (2021)).

Consider a disk with a radius  $R$  located at a distance  $L$  as the target object, as shown in Figure 2. For simplicity, the observation point is located on the line that perpendicularly passes through the center of the disk. Note that the chirp signal during a short period can be approximated as a sine wave and that the phase of the transmitted signal does not affect the distance spectrum. For simplicity, the transmitted signal is modeled as a unit pure tone  $\sin \omega t$  in the following analysis. The reflected acoustic signal with an angular frequency

$\omega$  from a ring that has a radius  $r$  and width  $dr$  at the origin  $x = 0$  is denoted as  $v_{\text{Ref},r,\omega}(t, 0)dr$ . Under the assumption that the gain coefficient and the phase shift by the reflection are constant over the whole frequency range as in (Uebo, T. et al. (2009)), we modeled the reflected signal by applying Lambert’s law (Kuttruff (1995)) and inverse distance law as follows;

$$v_{\text{Ref},r,\omega}(t, 0)dr = \gamma \sin\left(\omega\left(t - \frac{\sqrt{L^2 + r^2}}{c}\right) + \phi\right) \times \frac{L}{\sqrt{L^2 + r^2}} \times \frac{1}{\sqrt{L^2 + r^2}} \times \frac{2\pi r}{\pi R^2} dr. \quad (6)$$

Lambert's law  
Inverse distance law    Area ratio

By integrating  $v_{\text{Ref},r,\omega}(t, 0)dr$  of (6) over the disk, the reflected signal  $v_{\text{Ref},\omega}(t)$  is computed as follows:

$$v_{\text{Ref},\omega}(t, 0) = \int_0^R v_{\text{Ref},r,\omega}(t, 0)dr = \frac{2\gamma L}{R^2} \{C_1 \sin(\omega t - \phi) + C_2 \cos(\omega t - \phi)\}, \quad (7)$$

where

$$C_1 = \int_0^R \frac{r}{L^2 + r^2} \cos\left(\frac{\omega}{c} \sqrt{L^2 + r^2}\right) dr = C_i\left(\frac{\omega}{c} \sqrt{L^2 + R^2}\right) - C_i\left(\frac{\omega}{c} L\right),$$

and

$$C_2 = S_i\left(\frac{\omega}{c} \sqrt{L^2 + R^2}\right) - S_i\left(\frac{\omega}{c} L\right).$$

The term  $C_i(\cdot)$  and  $S_i(\cdot)$  are the cosine integral and the sine integral, respectively (Weisstein, (2002)):

$$C_i(x) = \int_0^x \frac{1 - \cos t}{t} dt, \quad S_i(x) = \int_0^x \frac{\sin t}{t} dt.$$

The power of the measured signal  $\sin \omega t + v_{\text{Ref},\omega}(t, 0)$  at the origin can be written as

$$p(\omega, 0) = \left[ 1 + \frac{2\gamma L}{R^2} \{C_1 \cos \phi + C_2 \sin \phi\} \right]^2 + \left[ \frac{2\gamma L}{R^2} \{-C_1 \sin \phi + C_2 \cos \phi\} \right]^2 = 1 + \alpha^2 (C_1^2 + C_2^2) + 2\alpha (C_1 \cos \phi + C_2 \sin \phi),$$

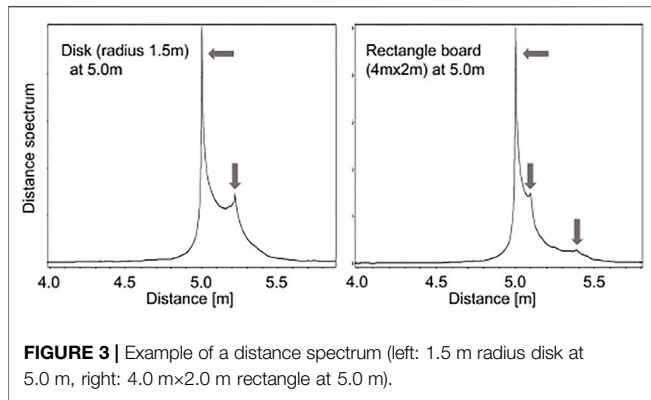
where  $\alpha = \frac{2\gamma L}{R^2}$ .

The cosine integral  $C_i(x)$  and the sine integral  $S_i(x)$  can be approximated for sufficiently large  $x$ ,  $|x| \gg 1$  (Airey (1937)) as

$$C_i(x) \approx \frac{\sin x}{x} - \frac{\cos x}{x^2}, \quad S_i(x) \approx \frac{\pi}{2} - \frac{\cos x}{x} - \frac{\sin x}{x^2}. \quad (8)$$

Let  $a = \frac{\sqrt{L^2 + R^2}}{c}$  and  $b = \frac{L}{c}$  ( $a > b$ ). Using (8), the following approximation holds (see the Appendix for details):

$$\begin{cases} C_1^2 + C_2^2 \approx 0 \\ C_1 \cos \phi + C_2 \sin \phi \approx \frac{\sin(a\omega - \phi)}{a\omega} - \frac{\sin(b\omega - \phi)}{b\omega}. \end{cases} \quad (9)$$



Finally, the power  $p(\omega, 0)$  can be approximated as

$$p(\omega, 0) \approx 1 + 2\alpha \left( \frac{\sin(a\omega - \phi)}{a\omega} - \frac{\sin(b\omega - \phi)}{b\omega} \right). \quad (10)$$

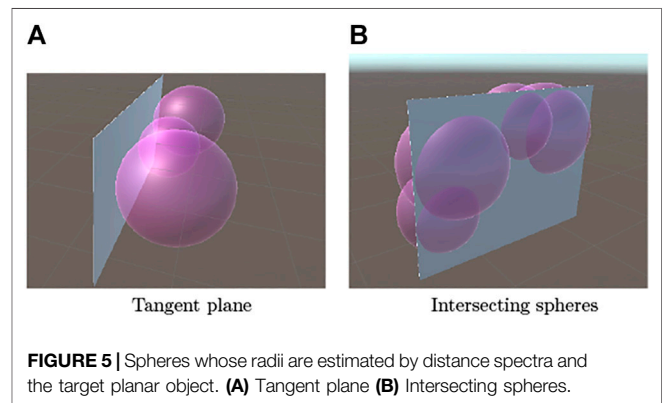
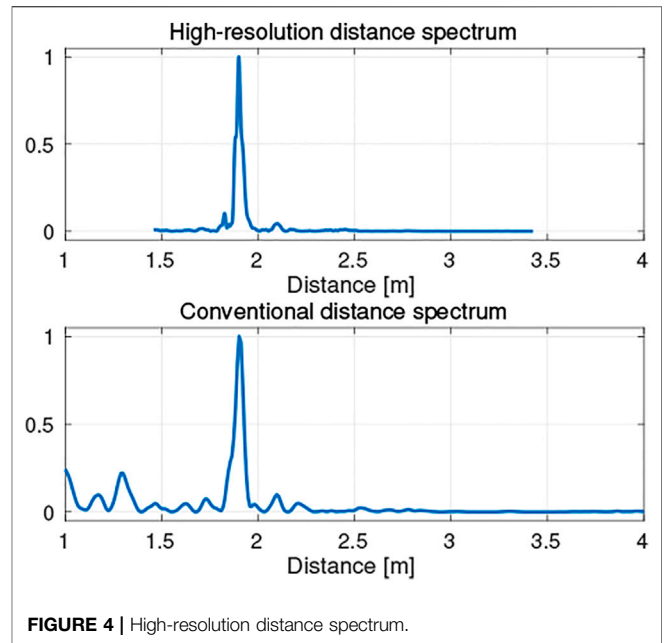
The approximation (10) shows that the power spectrum  $p(\omega, 0)$  can be approximated by a constant and two sinc functions with coefficients  $a$  and  $b$ . Recalling that  $\omega$  is given over a bounded frequency range, the “power spectrum” of  $p(\omega, 0)$  for  $\omega$  forms the distance spectrum that has two peaks at  $a$  and  $b$  that correspond to the edge ( $a$ ) and distance ( $b$ ), respectively, of the disk.

To verify this approximation model for the distance spectrum, a simple numerical acoustic simulation was conducted. In the simulation, a chirp signal was emitted from  $L = 5.0$  m away from a disk of radius  $R = 1.5$  m that had a reflection gain  $\gamma = 0.05$  and a phase shift  $\phi = \pi$  rad. The distance spectrum shown in the left figure of **Figure 3** has two peaks at 5.0 and 5.22 m that correspond to the distance to the disk and the distance from its center to its edge, respectively ( $\sqrt{5.0^2 + 1.5^2} \approx 5.22$ ). It is worth noting that this approach also works for a rectangle; a numerical result with a 4.0m x 2.0 m rectangle board at 5.0 m shows three peaks corresponding to its distance and two edges (5.0 m,  $\sqrt{5.0^2 + (\frac{2.0}{2})^2} \approx 5.10$  m and  $\sqrt{5.0^2 + (\frac{4.0}{2})^2} \approx 5.39$  m).

### 3.2 High-Resolution Distance Spectrum

Because the distance spectrum can be computed using a Fourier transform of the power spectrum  $p(f, 0)$ , a fast Fourier transform (FFT) (Cooley and Tukey, (1965)) is an effective method for this computation. The FFT-based conventional distance spectrum has a uniform resolution, and Uebo (Uebo et al. (2009)) showed that the highest resolution for the distance spectrum is given by  $\frac{c}{2f_w}$ .

As shown in **Figure 3**, the peaks corresponding to the edges of the target object are closely located around the most significant peak. This study developed a method to use a discrete Fourier transform (DFT) (Smith, (2002)) with dense query points around the first peak to compute the distance spectrum. **Figure 4** shows examples of distance spectra, and the proposed result indicates distinct peaks while the conventional one contains false peaks. Note that the query points can be selected arbitrarily with the DFT, and that the obtained distance spectrum can have higher resolution than that obtained using a FFT. However, since the DFT approach requires more computation, in the present study



the conventional FFT-based distance spectrum was cascaded to find a region of interest indicated by a major peak, and to then select the finer query points around the major peak so as to simultaneously obtain high resolution, and improve the computational efficiency.

### 3.3 Object Recognition Using Multiple Observations

A single observation gives a distance spectrum that provides the distance to the object and its edge(s). This subsection explains how multiple observations can be used to determine the shape and orientation of the object’s surface. Note that the proposed method is designed for the offline processing after all observations.

#### 3.3.1 Surface Detection

Recall that the first peak in the distance spectrum contains information about the shortest distance to the target surface.

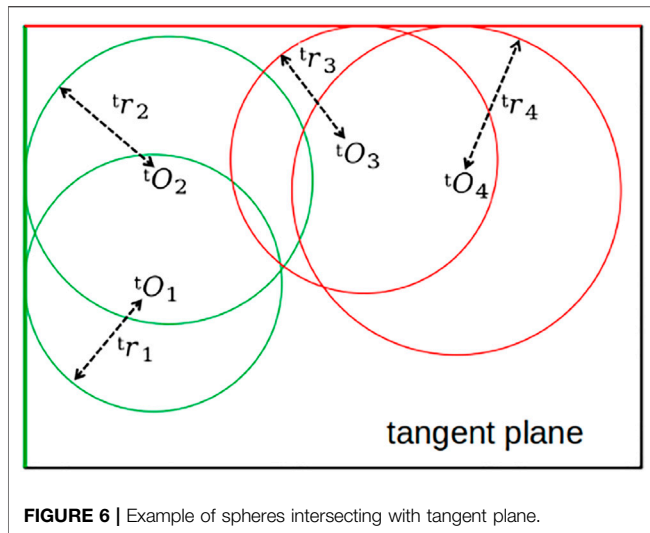


FIGURE 6 | Example of spheres intersecting with tangent plane.

Hence, the surface of the target object is on the common tangent plane of spheres whose radii are given by the most significant peaks in the distance spectra (Figure 5A).

This common tangent plane is computed as follows. The center of the sphere  $i$  and its radius are denoted as  $O_i$  and  $r_i$ , respectively, and the unit normal vector for the plane is denoted as  $n$ . The distance between the plane and the origin is denoted by  $d$ . Then, the sphere and the plane can be modeled as

$$\|O_i - x\| = r_i, \quad (11)$$

$$n^T x + d = 0. \quad (12)$$

Because the contact point  $x_i$  satisfies both (11) and (12),  $\pm r_i + n^T O_i + d = 0$  holds. For  $N$  observations, the equation can be combined in the following form:

$$\begin{bmatrix} O_1^T & 1 \\ O_2^T & 1 \\ \cdots & \cdots \\ O_N^T & 1 \end{bmatrix} \begin{bmatrix} n \\ d \end{bmatrix} + \begin{bmatrix} \pm r_1 \\ \pm r_2 \\ \vdots \\ \pm r_N \end{bmatrix} = 0, \quad \|n\| = 1. \quad (13)$$

The relation (13) may not have a unique solution if the observed radius  $r_i$  is affected by measurement noise. To overcome this, the random sample consensus (RANSAC) (Fischler and Bollers, (1981)) approach was incorporated, in which randomly sampled subsets of  $N$  observations are used to form a sub-problem of (13), and the best-fit solution of the sub-problems is selected as the solution of (13). The sign of  $r_i$  is also determined in this process. The fitness to evaluate the solution for RANSAC is computed based on the residual of (13).

### 3.3.2 Edge Recognition

The second and subsequent peaks in the distance spectrum give the radii of spheres whose centers exist at the observation locations. These spheres intersect the common tangent plane obtained above as in Figure 5B. According to the disk model proposed in Section 3.1, these spheres contact the edge of the surface (Figure 6).

In this model, the shape of the surface is estimated by finding the contact points of spheres intersecting with the estimated tangent plane. This is an ill-posed problem because the contact points cannot be determined uniquely, but it may be acceptable to assume that multiple circles share the tangent line if that line corresponds to the actual edge of the target. Based on this idea, the developed method uses the frequency of the tangent lines contacting the circles as the likelihood of the surface's edges. An algorithm inspired by Hough transform (Hough, (1962)) was used to solve this problem, as described below. The center of the circle  $i$  on the tangent plane is denoted as  ${}^tO_i$  with a radius of  ${}^t r_i$ . The unit normal vector for the tangent line on the tangent plane is represented by  ${}^t n(\xi)$  where  $\xi \in [0, 2\pi)$  rad indicates the angle between the line and the reference axis. The distance between the line and the origin is denoted as  ${}^t d$  that hold, and then the contact point is  ${}^t d = {}^t O_i^T n + {}^t r_i$ .

We introduce a function to evaluate the likelihood of  $\xi$  and  $d$  as

$$l(\xi, d) = \sum_{i=1}^M \exp(-\beta(d - {}^t d(\xi; {}^t O_i, {}^t r_i))^2), \quad (14)$$

where  $M$  is the number of detected circles and  $\beta$  is a positive constant. Peaks of  $l(\xi, d)$  that are greater than a given threshold  $l_{th}$  are selected, and the corresponding parameters  $\{(\xi, d) | l(\xi, d) \geq l_{th}, \nabla l(\xi, d) = 0\}$  are considered as the lines that form the shape of the target object on the tangent plane.

To determine the valid segments of the tangent lines, the contact points are computed. Let  $\xi_j$  and  $d_j$  be the estimated parameters for the line  $j$ . Then, the contact point between circle  $i$  and line  $j$ , denoted by  ${}^t p_{i,j}$ , is  ${}^t p_{i,j} = {}^t O_i^T n^\perp(\xi_j) n^\perp(\xi_j) + d_j {}^t n(\xi_j)$ , where  ${}^t n^\perp$  is a vector perpendicular to  ${}^t n$  on the tangent plane. Such contact points are considered as points on the edges of the target that allow us to recognize its shape.

## 4 VALIDATION

### 4.1 Numerical Simulation

Acoustic numerical simulations were conducted to verify the developed method for recognizing the shape of a target. The considered target object was a board measuring 1.2 m  $\times$  0.9 m, as shown in Figure 7. A chirp signal in the 20 kHz frequency band was emitted from a transmitter, and the signal was sampled at 48 kHz. The emitter and the receiver were located at the same position, which was 2.0 m away from the board. Seven observations at the points shown in the figure were simulated. The surface located on the plane could be represented as  $(0, 1, 0) x - 2 = 0$  in the coordinate frame of the figure. For efficient computation of the DFT to obtain a high-resolution distance spectrum, the non-equispaced FFT technique (Dutt and Rokhlin (1993)) was incorporated.

The obtained distance spectra are shown in Figure 8. The figure shows that the primary peaks that correspond to the distance to the board were extracted. Some of the peaks of the edges were also detected.

The method proposed in Section 3.1 was applied to estimate the plane containing the board from the first peaks

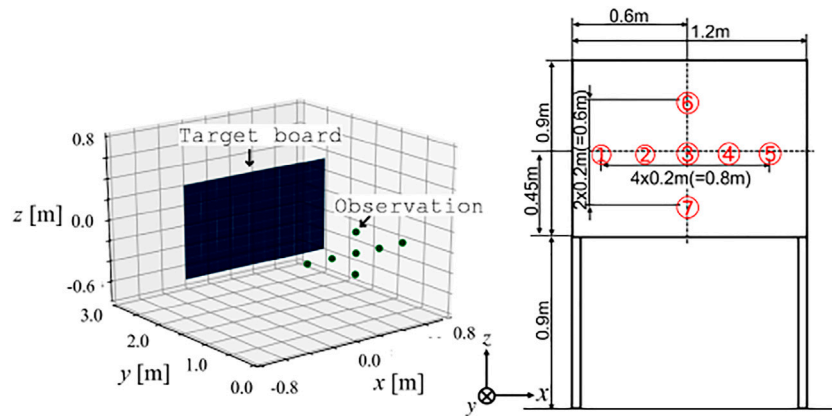
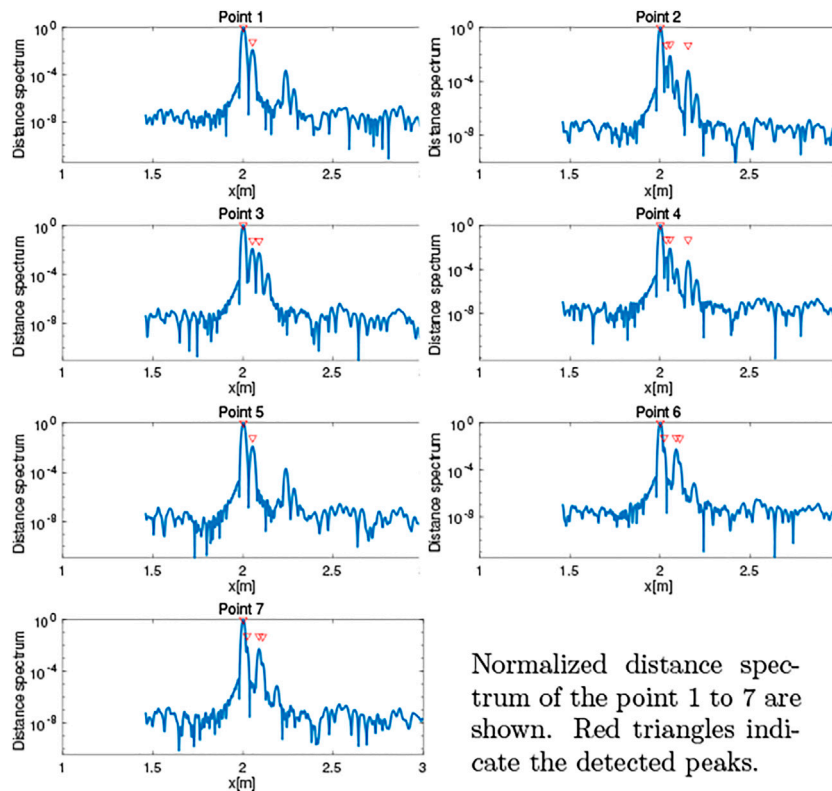


FIGURE 7 | Target board and observation points for validation.

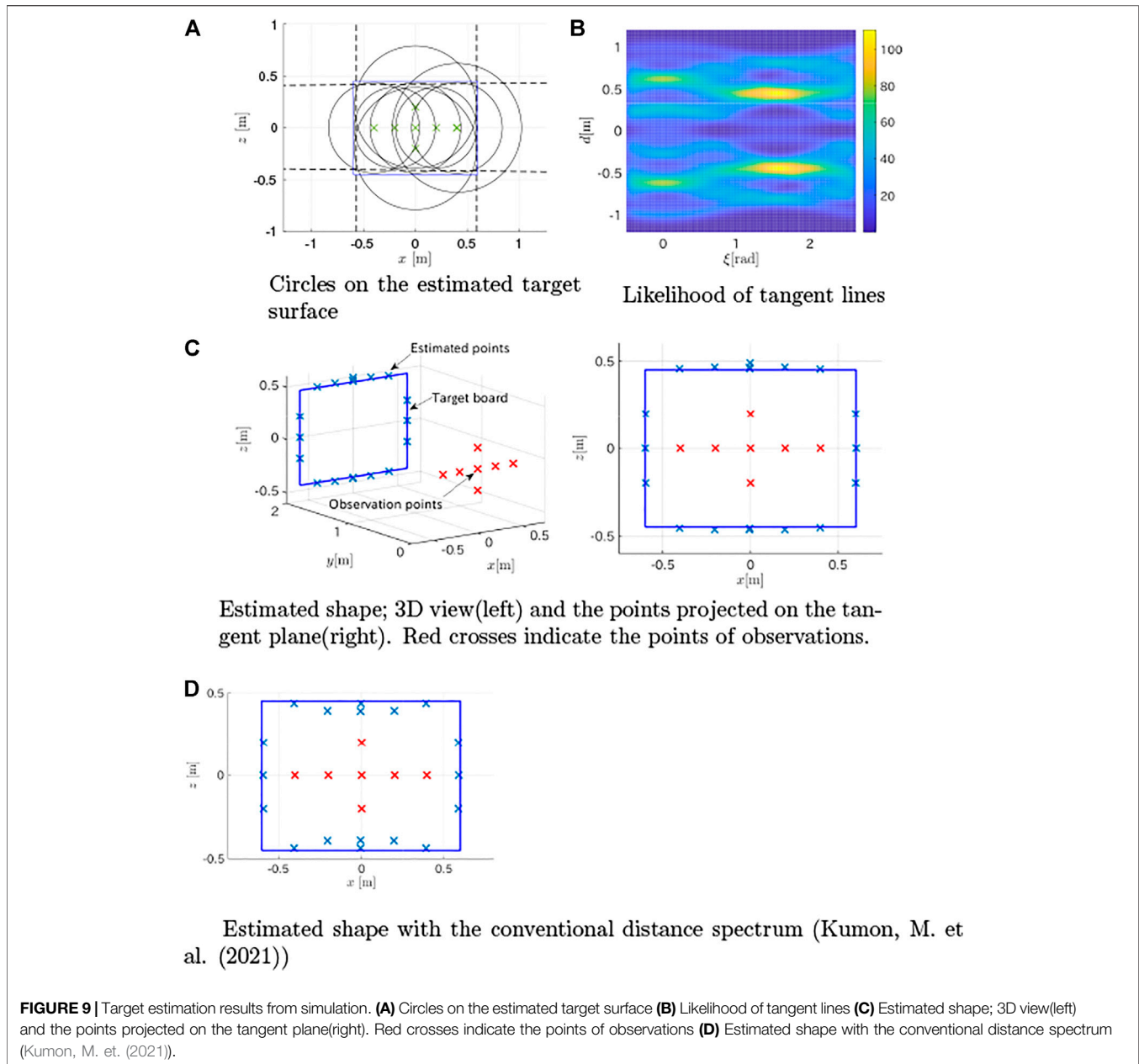


Normalized distance spectrum of the point 1 to 7 are shown. Red triangles indicate the detected peaks.

FIGURE 8 | Distance spectrum obtained in simulation.

in the distance spectra, and the estimated plane was  $(0.001, 0.9999, 0.000)x - 2.002 = 0$ . Then, the observations were fused using the method in Section 3.3 based on the estimated plane. The likelihood of tangent lines intersecting circles (Figure 9A) obtained by 14 is shown in Figure 9B. Then, the tangent lines with large likelihood values were estimated as the dotted lines shown in Figure 9A, and contact points were computed (crosses in Figure 9C). The figure shows that the method

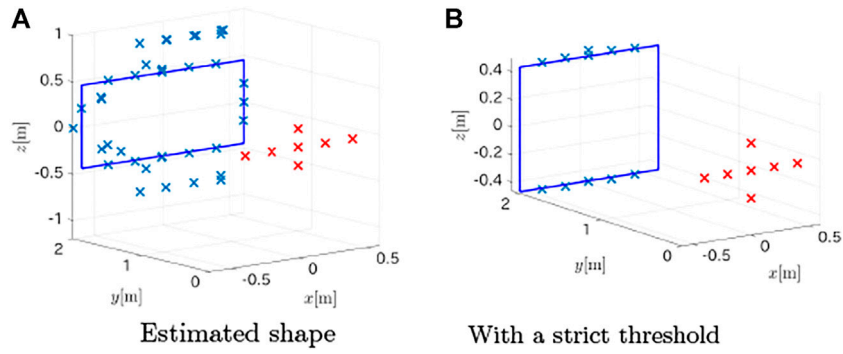
closely estimated the points of the edges. The distances between the 19 detected points and the nearest edges were computed as the estimation error. The mean absolute error was 0.0194 m with a standard deviation of 0.0297 m. This was better than the result (mean absolute error of 0.0417 m with a standard deviation of 0.0301 m, and shown in Figure 9D) obtained in our previous work (Kumon, M. et al. (2021)), which used the conventional FFT-based distance spectra.



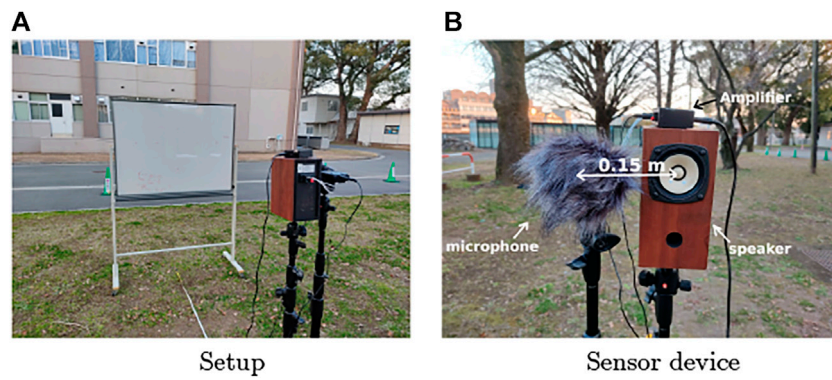
To evaluate the robustness of the method, it was tested under noisy measurement conditions. The signals were disturbed by uniform random noise whose magnitude was scaled to create a signal-to-noise ratio (SNR) of 60dB. The proposed method could accurately estimate the plane that contained the target as  $(0.001, 0.9999, 0.000)x - 2.002 = 0$ , but the estimated points for the target shape increased to 38, and some of the estimates were false positives, as shown in **Figure 10A**. The mean absolute error for these points was 0.1482 m with a standard deviation of 0.1441 m. By tuning the threshold to select the peaks from the distance spectra, the false detection rate could be reduced as shown in **Figure 10B**. The top and the bottom segments were estimated with a mean absolute error of 0.0290 m and a standard deviation of 0.037 m, but the method failed to detect the left and the right edges.

## 4.2 Experiment

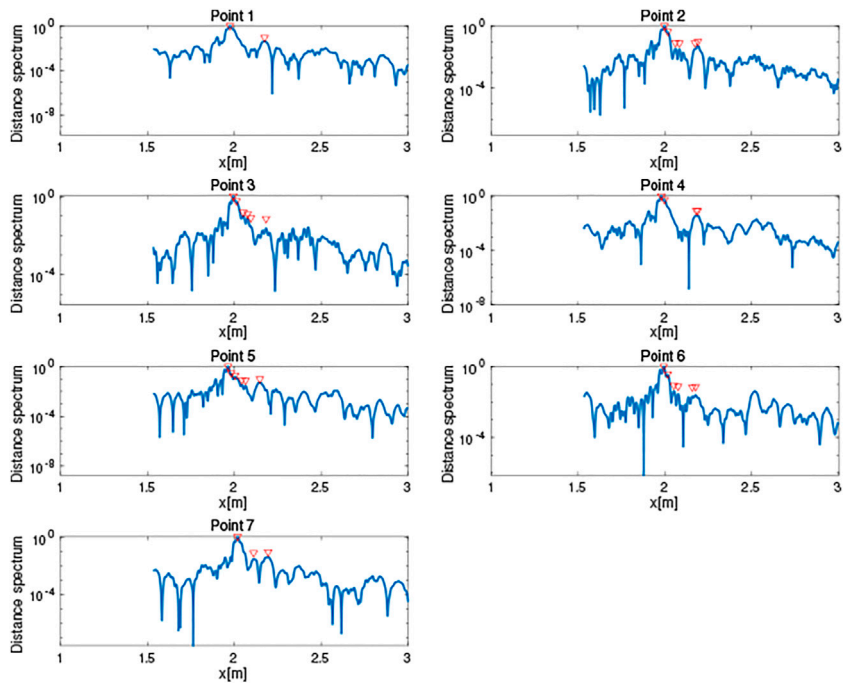
Next, we conducted experiments to show the validity of the proposed method. The target object was a board that had the same dimensions as in the numerical simulations depicted in **Figure 11**. A loudspeaker emitted a chirp signal, and the acoustic signal was recorded by a mobile audio recorder (Zoom, H4n Pro) that was located next to the speaker at a fixed distance of 0.15 m (**Figure 11B**). Although we placed the sensor device at multiple locations for the measurements instead of using an array of microphones, the principle still works for this case because the observation positions were precisely controlled. A synchronous averaging of 100 measurements was performed taking the background noise at the experiment site into account.



**FIGURE 10 |** Estimated points of the target in the presence of noise (SNR = 60dB). **(A)** Estimated shape **(B)** With a strict threshold.

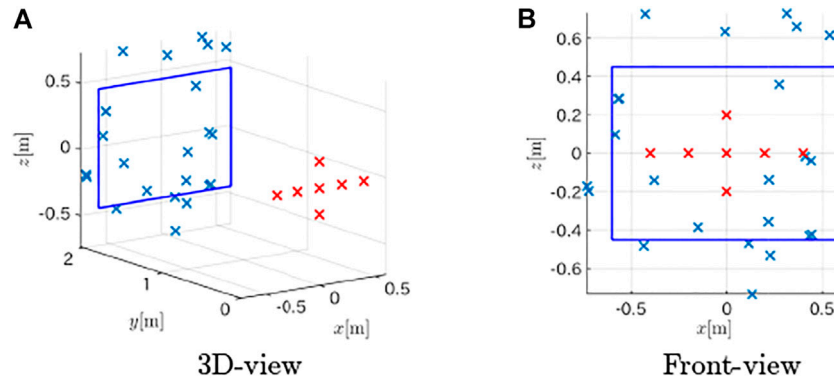


**FIGURE 11 |** Photos of experimental setup and device. **(A)** Setup **(B)** Sensor device.

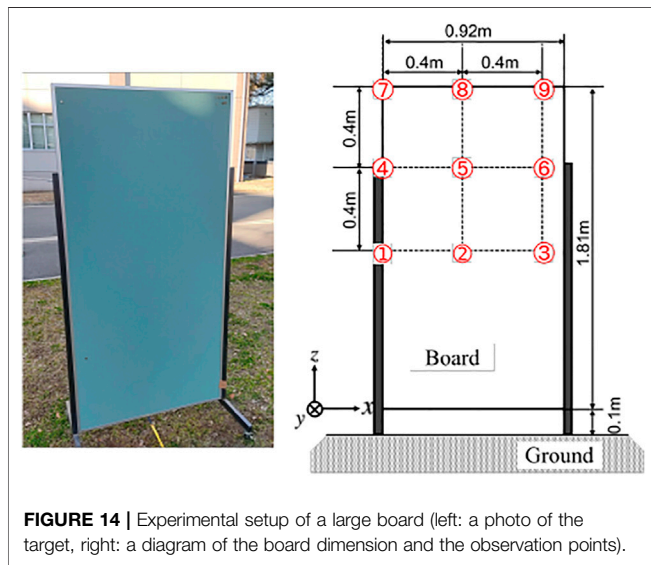


**FIGURE 12 |** Distance spectrum obtained in experiment.





**FIGURE 13** | Estimated points on the target from the experiment. **(A)** 3D-view **(B)** Front-view.



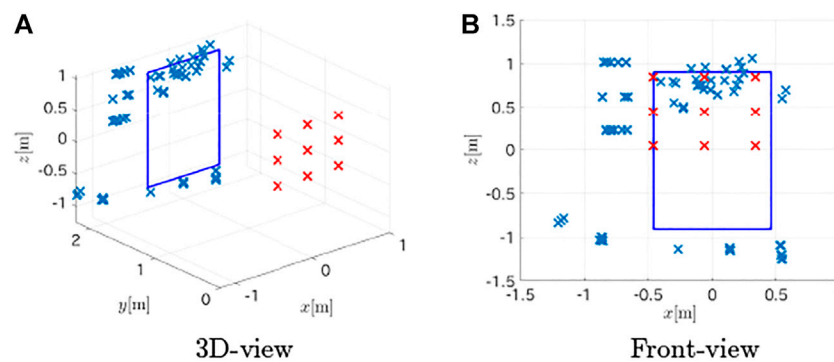
**FIGURE 14** | Experimental setup of a large board (left: a photo of the target, right: a diagram of the board dimension and the observation points).

The obtained distance spectra are shown in **Figure 12**, and the estimated plane was  $(-0.0399, 0.9973, 0.0619)x - 1.989 = 0$ ; that is, the cosine similarity of the normal vector was 0.9973 and the  $d$

error was 0.011 m. As shown in **Figure 13**, 23 estimated edge points were detected and the mean deviation from the nearest edge and its standard deviation were 0.1410 and 0.0849 m, respectively. The surface plane was estimated accurately, and the shape of the board was roughly obtained, although the estimated points were scattered around the target. The limited performance for the shape estimation might be because of noise in the environment, installation errors of the device, and acoustic distortion by the microphone and speaker.

Next, a larger board of 1.81 m × 0.92 m (**Figure 14**) was also used as another example to evaluate the proposed method. As shown in **Figure 14**, the bottom edge of the board was close to the ground which made the recognition challenging because the boundary of the target was difficult to detect. The observation pattern was designed to deviate from the center of the board to test in a more practical situation than in **Figure 11**.

The same observation process as in the previous experiment was conducted to record acoustic observations, and the proposed method was used to estimate the target. The estimated plane containing the target board was  $(0.0521, 0.9948, 0.0871)x - 2.060 = 0$ ; the cosine similarity of the normal vector was 0.9948 and the  $d$  error was 0.06 m; the performance to detect the target plane was at the same level as the previous experiment. Then, the method extracted 89 points as the points of the target as shown in



**FIGURE 15** | Estimated points on the large board target. **(A)** 3D-view **(B)** Front-view.

**Figure 15;** 48 of those points corresponded to the top and the bottom edges while the rest 41 points located at the left of the board were false detections.

Because the method could detect the bottom segment that was distant from the center of the observation pattern, it can be concluded that the method has the potential to recognize the target with the observation pattern that partially covers the target. Of course, the large search space to detect distant boundaries may cause false detection as in the result, and further studies are necessary to determine the optimal choice of parameters.

## 5 CONCLUSION

This study shows that the profile of the distance spectrum encodes surface range information from the target, and a high-resolution distance spectrum can recognize the distance to the surface and that to the edges accurately. Cues from the high-resolution distance spectra can be used to estimate the shape of the target object surface. Numerical simulation validated the proposed approach, and the robustness of the method under noisy conditions was also verified. The method was also tested by experiments, which showed that the approach is feasible, although there remains room for performance improvement.

Future work is to account for outliers that occur under noisy measurement conditions, or for observations when the receivers are not pointing to the targets. Techniques to monitor the quality of the observations may be effective for such inappropriate readings. And the method needs to be tested with more targets in various environments. Another challenge is to extend the method for a more complicated environment, such

## REFERENCES

- Airey, J. R. (1937). The Converging Factor in Asymptotic Series and the Calculation of Bessel, Laguerre and Other Functions. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* 24, 522–553. doi:10.1080/14786443708565133
- Boyer, K. L., and Kak, A. C. (1987). Color-encoded Structured Light for Rapid Active Ranging. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-9, 14–28. doi:10.1109/tpami.1987.4767869
- Christensen, J. H., Hornauer, S., and Stella, X. Y. (2020). “Batvision: Learning to See 3d Spatial Layout with Two Ears,” in 2020 IEEE International Conference on Robotics and Automation (ICRA) (Piscataway, NJ, USA: IEEE), 1581–1587. doi:10.1109/icra40945.2020.9196934
- Cooley, J. W., and Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Math. Comp.* 19, 297–301. doi:10.1090/s0025-5718-1965-0178586-1
- Dutt, A., and Rokhlin, V. (1993). Fast Fourier Transforms for Nonequispaced Data. *SIAM J. Sci. Comput.* 14, 1368–1393. doi:10.1137/0914081
- Egaña, A., Seco, F., and Ceres, R. (2008). Processing of Ultrasonic Echo Envelopes for Object Location with Nearby Receivers. *IEEE Trans. Instrumentation Meas.* 57, 2751–2755.
- Fischler, M. A., and Bolles, R. C. (1981). Random Sample Consensus. *Commun. ACM* 24, 381–395. doi:10.1145/358669.358692
- [Dataset] Hough, P. V. C. (1962). *Method and Means for Recognizing Complex Patterns*. US Patent US3069654A.
- Jong Hwan Lim, J. H., and Leonard, J. J. (2000). Mobile Robot Relocation from Echolocation Constraints. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 1035–1041. doi:10.1109/34.877524

as an indoor room that has significant reflections from several walls because it may become difficult to distinguish appropriate peaks of distance spectra. As multiple observations are used to estimate the target, a microphone array can be used to reduce the installation error among the microphones, to focus the reflections from the target of interest, and to reduce the noise.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The data examined in this study was taken from the authors’ previous work; this is necessary to clarify the improvement by the proposed method. Requests to access these datasets should be directed to kumon@gpo.kumamoto-u.ac.jp.

## AUTHOR CONTRIBUTIONS

The method was developed by MK and the coauthors at Kumamoto University. The validations by numerical simulations and experiments were conducted by RF and TM, and the analysis was led by MK and KN. All the authors contributed to writing the manuscript.

## FUNDING

This work was supported by the Japan Society for the Promotion of Science, KAKENHI Grant Numbers 19H00750 and 17K00365.

- Kishinami, H., Itoyama, K., Nishida, K., and Nakadai, K. (2020). “Two-Dimensional Environment Recognition Based on Standing Wave of Auditory Signal with Weighted Likelihood,” in Proceedings of the 38th Conference of Robotics Society of Japan, October 9, 2020. Robotics Society of Japan, 1D3–04. (in Japanese).
- Kolarik, A. J., Cirstea, S., Pardhan, S., and Moore, B. C. J. (2014). A Summary of Research Investigating Echolocation Abilities of Blind and Sighted Humans. *Hearing Res.* 310, 60–68. doi:10.1016/j.heares.2014.01.010
- Kuc, R. (2008). Generating B-Scans of the Environment with a Conventional Sonar. *IEEE Sensors J.* 8, 151–160. doi:10.1109/jsen.2007.908242
- Kumon, M., Fukunaga, R., and Nakatsuma, K. (2021). “Planar Object Recognition Based on Distance Spectrum,” in JSAI Technical Report. vol. SIG-Challenge-58-10, November 26, 2021. Japanese Society for Artificial Intelligence, 53–59. (in Japanese).
- Kuttruff, H. (1995). A Simple Iteration Scheme for the Computation of Decay Constants in Enclosures with Diffusely Reflecting Boundaries. *The J. Acoust. Soc. America* 98, 288–293. doi:10.1121/1.413727
- May, S., Dröschel, D., Fuchs, S., Holz, D., and Nüchter, A. (2009). “Robust 3d-Mapping with Time-Of-Flight Cameras,” in 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (St. Louis, MO, USA: IEEE), 1673–1678. doi:10.1109/iroso.2009.5354684
- Pelegrín-García, D., De Sena, E., van Waterschoot, T., Rychtáriková, M., and Glorieux, C. (2018). Localization of a Virtual Wall by Means of Active Echolocation by Untrained Sighted Persons. *Appl. Acoust.* 139, 82–92. doi:10.1016/j.apacoust.2018.04.018
- Reijniers, J., and Peremans, H. (2007). Biomimetic Sonar System Performing Spectrum-Based Localization. *IEEE Trans. Robot.* 23, 1151–1159. doi:10.1109/tro.2007.907487

- Santamaria, J. C., and Arkin, R. C. (1995). "Model-Based Echolocation of Environmental Objects," 1994 IEEE/RSJ International Conference on Intelligent Robots and Systems, September 12–16, 1994. Munich, Germany: IEEE, 64–72.
- Smith, P. P. (2001). *Active Sensors for Local Planning in Mobile Robotics, Vol. 26 of World Scientific Series in Robotics and Intelligent Systems*. Singapore: World Scientific.
- Smith, S. W. (2002). *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. California: California Technical Publishing, chap. 10.
- Steckel, J., Boen, A., and Peremans, H. (2012). Broadband 3-D Sonar System Using a Sparse Array for Indoor Navigation. *IEEE Trans. Robotics* 29, 161–171.
- Takao, M., Hoshiba, K., and Nakadai, K. (2017). "Assessment of Distance Estimation Using Audible Sound Based on Spectral Selection," in JSAI Technical Report. vol. SIG-Challenge-049-5, November 25, 2017. Yokohama, Japan: Japanese Society for Artificial Intelligence, 29–34. (in Japanese).
- Thaler, L., De Vos, R., Kish, D., Antoniou, M., Baker, C., and Hornikx, M. (2018). Human Echolocators Adjust Loudness and Number of Clicks for Detection of Reflectors at Various Azimuth Angles. *Proc. R. Soc. B.* 285, 20172735. doi:10.1098/rspb.2017.2735
- Tracy, E., and Kottege, N. (2021). CatChatter: Acoustic Perception for Mobile Robots. *IEEE Robot. Autom. Lett.* 6, 7209–7216. doi:10.1109/lra.2021.3094492
- Uebo, T., Nakasako, N., Ohmata, N., and Mori, A. (2009). Distance Measurement Based on Standing Wave for Band-Limited Audible Sound with Random Phase. *Acoust. Sci. Tech.* 30, 18–24. doi:10.1250/ast.30.18
- Wang, W., and Wong, P. L. (2003). The Dynamic Measurement of Surface Topography Using a Light-Section Technique. *TriboTest* 9, 305–316. doi:10.1002/tt.3020090403
- [Dataset] Weisstein, W. E. (2002). *Sine Integral. MathWorld – Wolfram Web Resource*. Available at: <https://mathworld.wolfram.com/SineIntegral.html> (Accessed February 25, 2021).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kumon, Fukunaga, Manabe and Nakatsuma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

The approximated cosine and sine integrals are denoted as  $\overline{C}_i(x)$  and  $\overline{S}_i(x)$ , that is,

$$\overline{C}_i(x) = \frac{\sin x}{x} - \frac{\cos x}{x^2}, \quad \overline{S}_i(x) = \frac{\pi}{2} - \frac{\cos x}{x} - \frac{\sin x}{x^2}. \quad (15)$$

Then, the following holds.

$$\begin{aligned} \overline{C}_i^2(x) + \overline{S}_i^2(x) &= \frac{\pi^2}{4} + \frac{1}{x^2} + \frac{1}{x^4} - \pi \left( \frac{\cos x}{x} + \frac{\sin x}{x^2} \right), \\ \overline{C}_i(x)\overline{C}_i(y) + \overline{S}_i(x)\overline{S}_i(y) &= \frac{\pi^2}{4} - \frac{\pi}{2} \left( \frac{\cos y}{y} + \frac{\sin y}{y^2} + \frac{\cos x}{x} + \frac{\sin x}{x^2} \right). \end{aligned}$$

Furthermore,

$$C_1^2 + C_2^2 \approx \frac{1}{(a\omega)^2} + \frac{1}{(b\omega)^2} + \frac{1}{(a\omega)^4} + \frac{1}{(b\omega)^4} - \frac{2}{ab\omega^2} \sqrt{1 + \frac{a^2 + b^2}{(ab\omega)^2} + \frac{1}{(ab)^2\omega^4}} \times \sin((b-a)\omega + \phi_0),$$

where

$$\tan \phi_0 = \frac{1 + \frac{1}{ab\omega^2}}{\frac{1}{\omega} \left( \frac{1}{a} - \frac{1}{b} \right)} = \frac{ab\omega^2 + 1}{(b-a)\omega},$$

and

$$\begin{aligned} C_1 \cos \phi + C_2 \sin \phi &\approx \frac{\sin(a\omega - \phi)}{a\omega} - \frac{\sin(b\omega - \phi)}{b\omega} - \frac{\cos(a\omega - \phi)}{(a\omega)^2} \\ &\quad + \frac{\cos(b\omega - \phi)}{(b\omega)^2}. \end{aligned}$$

For large  $\omega$ ,  $\frac{1}{\omega^2} \ll 1$  and the above formulas give (9).