IEEE Open Journal of
Ultrasonics, Ferroelectrics, and Frequency Control

Received 31 January 2024; revised 12 June 2024; accepted 17 September 2024.
Date of publication 23 September 2024; date of current version 4 October 2024.

Digital Object Identifier 10.1109/OJUFFC.2024.3466090

3-D Object Reconstruction From Outdoor Ultrasonic Image and Variation Autoencoder

RYOTARO OHARA[1], YUTO YASUDA[1], RIKU HAMABE[1], SHUN SATO[1], ISHII TORU[2], SHINTARO IZUMI[2,3] (Member, IEEE), AND HIROSHI KAWAGUCHI[2] (Member, IEEE)

[1]Graduate School of System Informatics, Kobe University, Nada-ku, Kobe 657-8501, Japan
[2]Graduate School of Science, Technology and Innovation, Kobe University, Kobe 657-8501, Japan
[3]Osaka Heat Cool Inc., Minoh 562-0015, Japan

CORRESPONDING AUTHOR: R. OHARA (ohara.ryotaro@cs28.cs.kobe-u.ac.jp)

This work was supported by the Project Commissioned by the New Energy and Industrial Technology Development Organization (NEDO) under Grant P19005.

ABSTRACT We present a technique for three-dimensional (3D) object reconstruction utilizing an ultrasonic array sensor and a variational autoencoder (VAE) within a high-interference environment. In scenarios where optical instruments such as cameras and LiDAR are impractical, the utilization of air-coupled ultrasonic waves for 3D measurements has emerged as a viable alternative. Nevertheless, deploying this technology in high-interference settings, particularly outdoor environments, has presented significant challenges. To tackle this challenge, we have developed and established a methodology for the 3D reconstruction of stationary objects by combining the time-of-flight point cloud data acquired through beamforming with the deep learning model VAE. This study proceeds by elucidating the procedure for conducting beamforming and measuring distances using ultrasonic waves. Subsequently, we expound upon an experimental methodology that employs 3D object reconstruction and associated techniques. Finally, we present the results obtained from attaching an ultrasonic sensor to a utility pole and conducting ultrasonic measurements. Our investigation focuses on four distinct types of objects: boxes, motorbikes, humans, and reflectors. The measurement system was positioned 5 m above the ground on a utility pole situated alongside the road. The objects selected for measurement were situated in stationary positions within a 3 m^3 area, with a maximum distance of 10 m from the utility pole. The objective of this study is to assess the efficacy of ultrasonic measurements and object reconstruction techniques under these specified conditions. The results indicate a precision of 0.939, a recall of 0.868, and an F-value of 0.902, which are considered sufficient for the application of ultrasonic waves.

INDEX TERMS Beamforming, three-dimensional object reconstruction, ultrasound array, variation autoencoder.

I. INTRODUCTION

AIR-COUPLED ultrasonic measurement, which does not require direct contact with the object, finds application in a wide range of fields such as agriculture [1], robotics [2], [3], and self-position estimation [4]. Among these diverse applications, time-of-flight (ToF) ultrasound imaging [5], [6], [7], [8] stands out for its capability to measure the shape of an object in three dimensions. Although light detection and ranging (LiDAR) and millimeter-wave radars are commonly used for three-dimensional (3D) measurements [9], the effectiveness of LiDAR can be compromised by the optical properties of the object being measured, such as poor weather conditions and transparent objects [10], [11], unlike ultrasonic imaging which remains unaffected by these factors. Additionally, ultrasound imaging is less expensive than both LiDAR and millimeter-wave radar, providing a cost-effective alternative for 3D measurement applications.

Ultrasound imaging employs beamforming and ToF technologies. Beamforming enhances signals from various directions using the phase differences between transducers. ToF is a technology that estimates the distance to an object based on the travel time of sound waves emitted by a device, which are then reflected off the object and returned to the device. By integrating these techniques, it is possible to perform ToF calculations from multiple directions, thereby enhancing the accuracy and versatility of ultrasound imaging. These technologies find applications in diverse fields, including object

detection [12] and non-destructive testing [13]. Furthermore, research has explored 3D measurements using air-coupled ultrasound to investigate its potential in various scientific and industrial domains. However, most previous studies on 3D air-coupled ultrasound and beamforming imaging have been limited to indoor distances within 10 m, with only a few studies extending beyond this range [2], [5], [14], [15], [16].

Outdoor ultrasonic measurement presents challenges owing to the need to eliminate various sources of noise such as multipath and environmental disturbances. However, if ultrasonic 3D imaging can be successfully implemented outdoors, it could serve as an alternative in environments where optical cameras are impractical owing to privacy concerns, such as residential areas. This technology holds potential for diverse applications, including monitoring road conditions, human behavior, and traffic volume in residential areas during times of disaster.

We have developed a 3D reconstruction system that uses ultrasonic imaging technology. This system is designed to monitor road conditions, where approximations of the outer shapes of objects are sufficient without the need for details. The system achieves 3D reconstruction by integrating the ToF point cloud obtained through 25 kHz ultrasound beamforming and direct sequence spread spectrum (DSSS) [17] modulation with a deep learning model, specifically a variational autoencoder (VAE) [18]. We have developed a comprehensive computational pipeline for this system, encompassing both hardware development and 3D reconstruction processes, utilizing a VAE that integrates all necessary components from data acquisition to analysis. The system successfully reconstructs a $30 \times 30 \times 30$ voxel space from ultrasound signals sampled at 227 kHz, achieving a precision of 0.939, recall of 0.868, and F-value of 0.902, which highlights the effectiveness of our approach in 3D ultrasonic imaging.

The remainder of this paper is organized as follows: Section II describes the theoretical and computational pipelines that we have developed for ultrasound imaging. Section III details our measurement system, including the ultrasound array and recording hardware. Section IV presents our experimental results and the accuracy of the 3D reconstruction. In Section V, we compare our findings with those of previous studies. Finally, in Section VI, we present our conclusions.

II. PREVIOUS WORKS

This section reviews previous research on 3D reconstruction using air-coupled ultrasound and acoustic traffic monitoring.

The 3D measurements based on air-coupled ultrasound are realized using the ToF of ultrasound. In a previous study [15], two types of distances were targeted: the far field, which is more than 2 m away, and the near field, which is approximately 200 mm away in an acoustic chamber. High-precision 3D imaging was achieved by measuring an ultrasonic array of 64 transducers, which also served as microphones, arranged according to a Fermat spiral with an aperture of 200 mm. One of the major contributions of this study is the demonstration of the effectiveness of microphone placement based on the Fermat spiral in ultrasonic imaging. This study shows that microphones arranged in a Fermat's spiral pattern reduce the side lobes and grating lobes better than those arranged in a grid pattern, thus improving the quality of ultrasonic imaging. The evaluation experiments were conducted at far-field and near-field distances. In the far-field, all 28 corner reflectors at a distance of 2 m were accurately measured individually. In the near-field, precise measurements of a human finger were obtained at a distance of 20 mm.

A previous study [14] focused on the hardware aspects of 3D measurements using air-coupled ultrasound. In this study, a microphone array, field-programmable gate array (FPGA), and graphics processing unit were employed to construct a highly compact ultrasonic measurement system that integrates all components from data acquisition to beamforming. One of the significant contributions of this study is the proposal of a cohesive hardware system that encompasses not only the ultrasonic array but also signal processing. Although acoustic sensing research often highlights the cost-effectiveness of microphones compared to LiDAR and millimeter-wave radar, only a limited number of studies have implemented them, especially within edge signal processing systems. Despite being an embedded system, it achieves 30 FPS with beamforming of 240 beams.

Previous research has focused on ultrasonic measurements at targeting distances exceeding 10 m. In a previous study, researchers successfully imaged objects positioned 16 mm away with a resolution of 25×25 pixels [16]. Furthermore, in our previous investigation [5], we successfully conducted position measurements of objects located 12.5 m away in an outdoor setting. However, these studies have not achieved detailed 3D measurements for discerning object shapes.

Prior research on outdoor acoustic monitoring has explored passive systems that utilize the sound emitted by vehicles during operation rather than the reflected sound. Acoustic traffic monitoring systems can be categorized into two types: rule-based and machine learning-based systems.

In a previous study [19], a microphone array arranged in a crisscross pattern was employed. The recorded audible sound was utilized to estimate parameters such as the number of passing vehicles, lane occupancy, average speed, and even vehicle types in multiple lanes. Machine learning-based acoustic traffic monitoring employs K-nearest neighbors and convolutional neural networks (CNN). In [20], audio was recorded and traffic was monitored using stereo microphones installed at the roadside. The power spectrogram and phase difference information of the audio were used as inputs to the CNN to estimate parameters such as the number of vehicles, speed, type, and occupancy.

III. ULTRASONIC IMAGING

In this section, the ToF calculation using air-coupled ultrasound and the theory of 3D reconstruction from the ToF

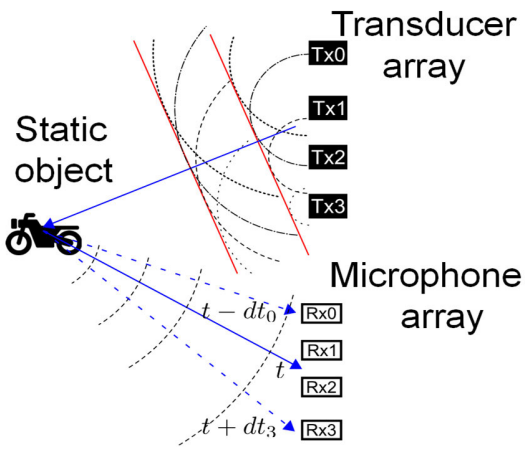

FIGURE 1. 3D measurement using beamforming and ToF.

point cloud are explained. First, the ToF is elucidated, followed by a description of the calculation pipeline for the 3D reconstruction. The ToF represents the duration it takes for the ultrasonic waves emitted from the transducer to bounce off the object's surface and reach the microphone.

Fig. 1 provides an overview of 3D measurements using air-coupled ultrasound. Initially, ultrasonic waves are emitted from speaker array toward the object. Phase-shifted ultrasound waves overlap and amplify, resulting in directional focus. The emitted ultrasonic waves are reflected off the surface of the object. The reflected ultrasonic waves are measured by a microphone array with a phase difference. By calculating the delays and sums based on phase differences, sound waves from an arbitrary direction can be enhanced. Next, the cross-correlation between the beamformed reflected wave and the transmitted waves is calculated. The closer the waveform shapes are, the greater the cross-correlation; the ToF is the time from the timing of transmission to the peak of the cross-correlation. In environments with high interference, such as outdoor settings, ultrasonic cross-correlation tends to deteriorate. However, in this study, we enhanced interference resilience by employing DSSS modulation for ultrasonic signals. Further details will be discussed in Section II.

Fig. 2 illustrates a block diagram of the proposed system utilized for ultrasonic measurements and 3D object reconstruction. A DSSS signal using an M-sequence as the transmission signal is emitted toward an object. In Stage (a) of the diagram, the signals reflected from the object surface are received by the microphone array and subjected to beamforming. In Stage (b), the beamformed reflected signal is normalized by the distance, and the cross-correlation of the transmitted signals is calculated, while removing the background to obtain a 3D ToF image. Consequently, the distance to an object can be measured. The ultrasonic data processed by this ultrasonic signal processing system serve as the input for the VAE machine learning in Stage (c).

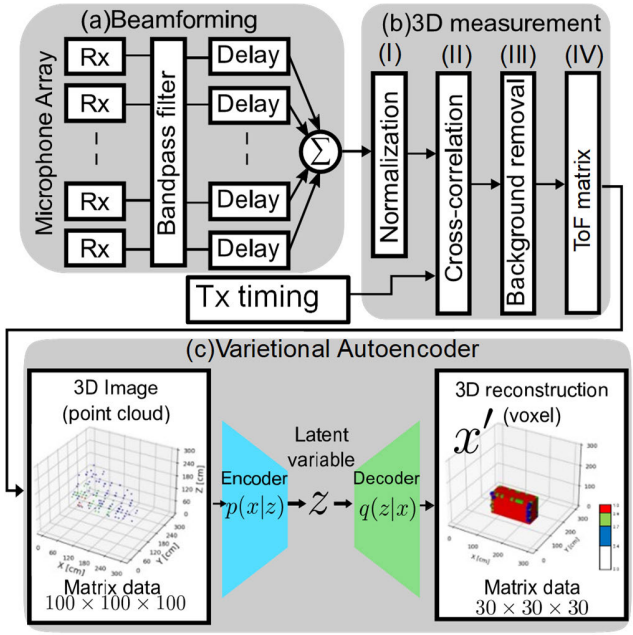

FIGURE 2. Block diagram of proposed ultrasonic imaging system.

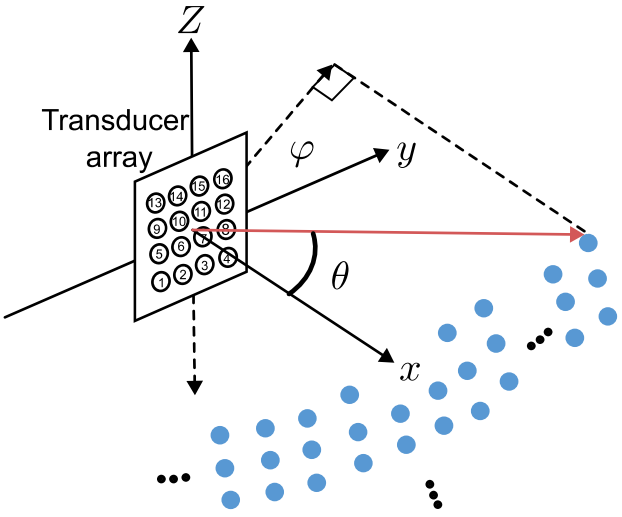

FIGURE 3. Beamforming of a rectangular array.

A. BEAMFORMING

This subsection delineates the beamforming process depicted in Fig. 2(a). As depicted in Fig. 3, when the array is situated on the Y–Z plane, the main beam direction in 3D is denoted by the zenith angle θ around the X-axis and the azimuth angle φ relative to the Y–Z plane. In the nth element, the displacement in the y-axis direction from the origin is dy_n, and the displacement in the z-axis direction is dz_n. In beamforming, the delay time dt_n of the n-th element in the main beam direction (θ, φ) is given by (1), where c represents the propagation velocity of ultrasonic waves in air.

$$dt_n = (dy_n sin\,(\theta)\,cos(\varphi) + dz_n sin\,(\theta)\,sin(\varphi))/c. \quad (1)$$

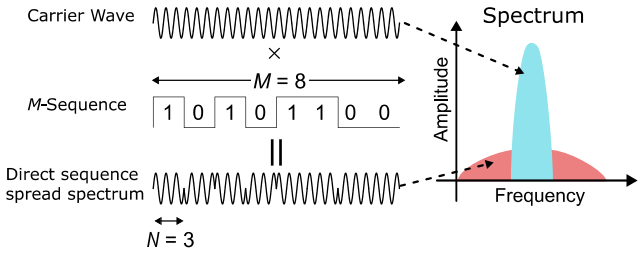

FIGURE 4. Direct-sequence spread spectrum.

Here, c represents the propagation velocity of ultrasonic waves in air.

B. 3D MEASUREMENT

The 3D measurement illustrated in Fig. 2(b) comprises four substages: (I) normalization, (II) cross-correlation, (III) background removal, and (IV) ToF matrix. (I) Normalization mitigates the effect of distance attenuation. Propagation attenuation [21] is represented using a loudspeaker that uniformly radiates a spherical wave. If the power at the receiver is P_r, the power at the transmitter is P_t. The distance from the transmitter to receiver is R. The power at the receiver is expressed as follows:

$$P_r = \frac{P_t}{4\pi R^2} \tag{2}$$

Based on this equation, the received signal can be normalized by R^2, signifying distance attenuation.

This normalization alleviates the effects of the distance attenuation on the intensity of the reflection from the object surface, thereby enhancing the detection accuracy of distant objects for the DSSD cross-correlation in Substage (II).

As depicted in Fig. 4, the cross-correlation between the received and transmitted signals is computed in (II). In this study, the transmitted signal was modulated by DSSS to enhance interference and noise immunity in the cross-correlation. Illustrated on the right side of Fig. 4, DSSS is a modulation technique where the transmitted data are multiplied by a spread signal with a higher frequency to broaden the bandwidth over a wider frequency band than the original signal, ensuring secure communication. Although this method consumes a larger bandwidth, it can improve the ability to reject interfering and noise waves. The XOR operation is performed between the M-Sequence and the transmitted data. The carrier wave remains non-inverted if each bit of the XOR result is 1 and is inverted if each bit is 0. In this instance, a data bit of one was modulated with an 8-bit M-sequence, with each M-sequence corresponding to a carrier wave of three wavelengths.

When an object is situated on a road, the signal reflected from the road often manifests as a strong interference wave. Additionally, the grating lobe of the ultrasonic array transmitter may introduce significant interference from directions other than that of the main beam. Consequently, to eliminate signals originating from sources other than the object itself, the received signal must undergo background removal (III), especially in scenarios where no object is present. Fig. 5 illustrates the distribution of signal strength when measuring reflected signals in front of the origin point. The signal strength of the background data exhibited non-uniformity. Therefore, it is assumed that the signal magnitude follows a normal distribution, and the magnitude of the signal to be subtracted is thus determined accordingly.

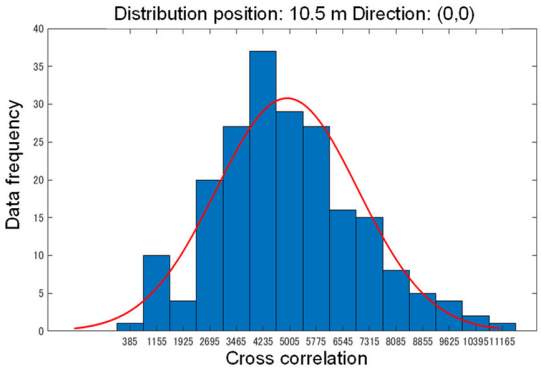

FIGURE 5. Histogram of correlated values.

The ToF data, from which background information has been removed, constitutes a point cloud expressed in a polar coordinate system denoted by (r, θ, ϕ). In this study, the input to the VAE is a 3D matrix. To input the ToF point cloud into the VAE, the point cloud is partitioned into a grid and transformed into a (100, 100, 100) Cartesian matrix represented by (x, y, z) with dimensions of (100, 100, 100).

C. VARIATIONAL AUTO ENCODER

The 3D VAE encoder-decoder model utilized to obtain the 3D coordinates of a stationary target object is depicted in Fig. 6. A VAE functions as a type of autoencoder that compresses input voxels into a 500-dimensional latent variable space using an encoder, which are subsequently reconstructed as voxels in three dimensions by a decoder.

Supervised learning is employed to eliminate noise, such as interference waves, and to perform object detection by imaging a target object using ultrasonic data. The ultrasonic wavelengths range from ten to several millimeters. The input data dimensions are (100, 100, and 100) by segmenting the data into 30 mm intervals, resulting in a cube measuring 3000 mm per side. Each voxel represents a cube with 30 mm sides. These 3D data undergo convolution using a convolutional layer and a max-pooling layer to extract features in 3D space [22]. Subsequently, they are imaged in 3D space using a reverse convolutional layer.

For the training data utilized for supervision, the object is represented as (30, 30, 30) by dividing a cube of 3000 mm on one side into sections of 100 mm each. The positions where the target object is located are assigned a value of 1.0, while the positions where the target object does not exist are filled with values of 0.1 in a voxel for use as training data.

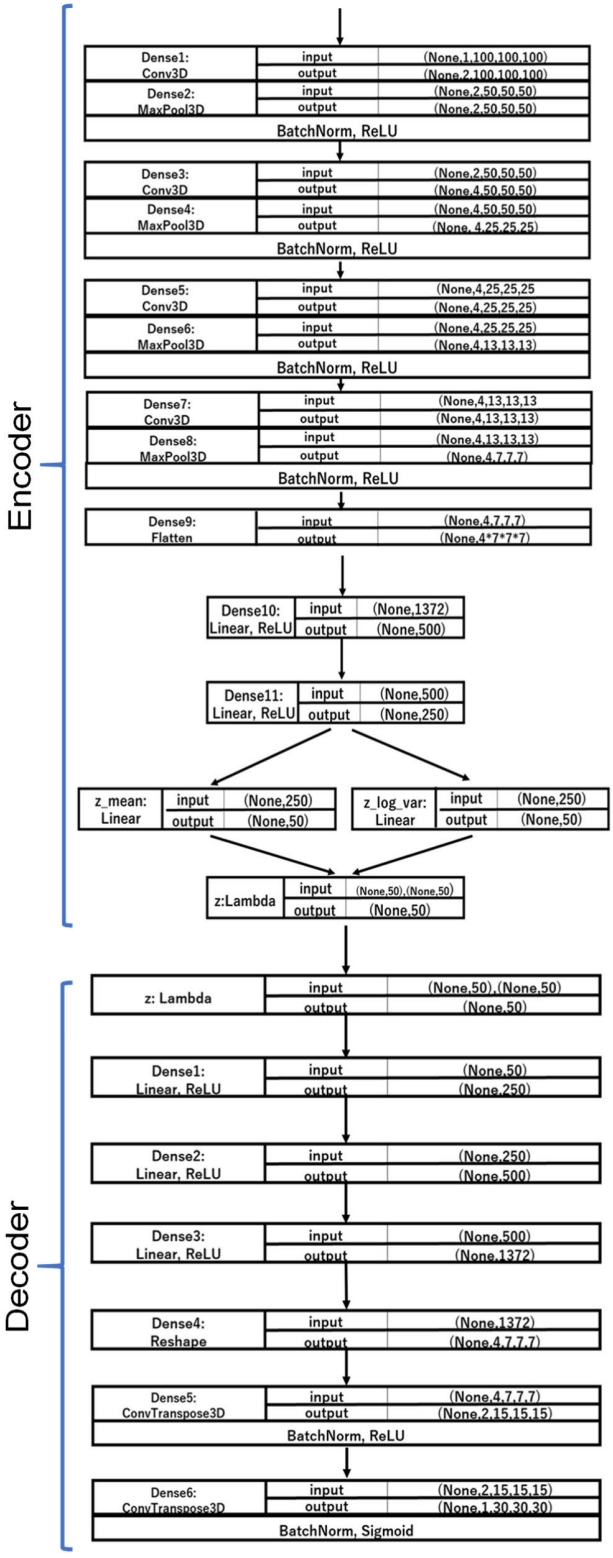

FIGURE 6. **Network architecture of 3D reconstructed VAE.**

D. 3D EVALUATION

In the 3D VAE machine learning framework, although the input has dimensions of (100 × 100 × 100), the output is represented as (30 × 30 × 30), where 1.0 signifies the ideal value for a specified object. We assessed the proximity of this region to the area where the object exists in the training data using *Precision*, *Recall*, and F-value (where the F-value is was recognized as the harmonic mean of the precision and recall). The precision, recall, and F-value are denoted in (3), (4), and (5), respectively, with explanations of the true positive (*TP*), false negative (*FN*), false positive (*FP*), and true negative (*TN*) presented in Table 1 to clarify their use in the equations.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + TN} \tag{4}$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

TABLE 1. **Confusion matrix for binary classification.**

		Predict: Positive	Predict: Negative
Reference	Positive	TP: True Positive	FN: False Negative
	Negative	FP: False Positive	TN: True Negative

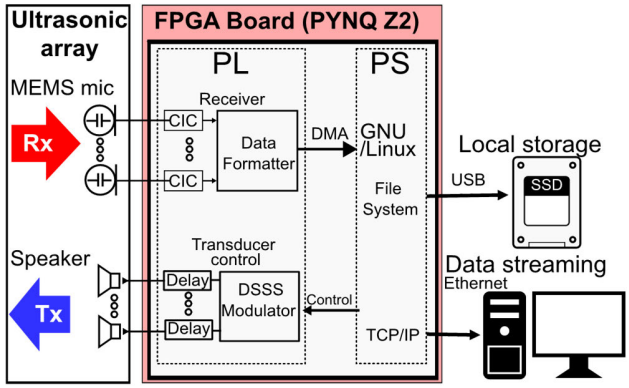

FIGURE 7. **Ultrasonic measurement system structure.**

IV. EXPERIMENTAL SETUP

In this section, we outline the measurement system. Subsection A details the configuration of the proposed system, whereas Subsection B covers the experimental conditions.

A. MEASUREMENT SYSTEM

Both hardware and software components were developed to implement the proposed measurement system. Fig. 7 illustrates the structure of the system, which consists of three main components. The first component is the ultrasonic array, which integrates microphones and speakers. The second component is a FPGA, which is responsible not only for processing signals but also for the DSSS modulation

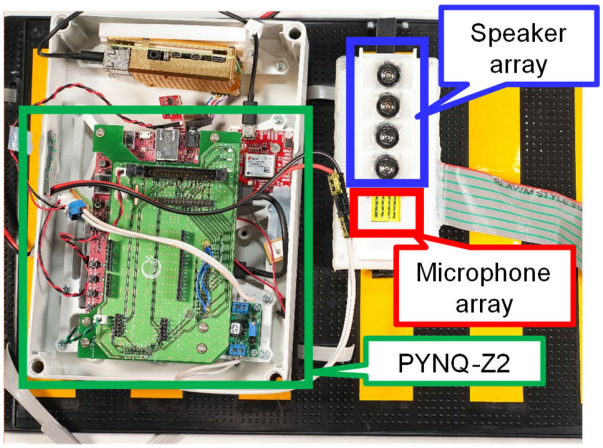

FIGURE 8. **Ultrasonic array and FPGA.**

of ultrasonic waves, and it includes its own local storage. The third component is a recording server tasked with receiving and storing data.

Our measurement system utilized a PYNQ-Z2 [23]. The PYNQ-Z2 is equipped with a Xilinx ZYNQ system on a chip (SoC), which includes an internal ARM central processing unit (CPU) and a processing system comprised of peripheral circuits alongside programmable logic serving as an FPGA, facilitating CPU-to-FPGA communication. Additionally, the PYNQ-Z2 supports Linux, allowing for the utilization of TCP/IP and Linux file systems. We leveraged these features to achieve the streaming and logging of measurement data.

The FPGA consists of two types of circuits: a speaker control circuit and a receiving circuit. The speaker control circuit generates an ultrasonic signal projected onto the object to be measured and drives the speaker. It operates by generating an ultrasonic signal projected onto the target object and controlling the speaker. The speaker functions based on DSSS modulation with parameters such as the M-sequence and frequency configured from the processing system. In the experiment, modulation was carried out with a carrier wave of 25 kHz, an M-sequence length of 8 bits, and 8 waves per bit. The receiving circuit demodulates the output of the 16-channel microphone array as a pulse code modulation (PCM) signal and transfers it to the PS. The ultrasonic microphone produces a 4.55 MHz Pulse-density modulation (PDM) output, which is demodulated into a 227 kHz PCM signal using a cascaded integrator comb (CIC) filter. Additional information is added to the 16-channel PCM signal and transferred to the PS via DMA.

Fig. 8 displays the appearance of the measurement system. Enclosed within a waterproof case designed for utility pole installation, the FPGA and Raspberry Pi serve as the data-streaming target, storage, and server, respectively. The system operates via Wi-Fi, with a Raspberry Pi acting as the gateway. The speakers and microphone array are housed within a dedicated 3D-printed enclosure and connected to the PYNQ-Z2 via flat cables.

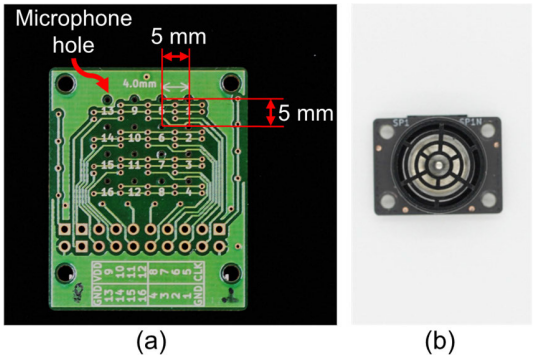

FIGURE 9. **Ultrasonic array device (a) 4 × 4 microphone array (b) speaker.**

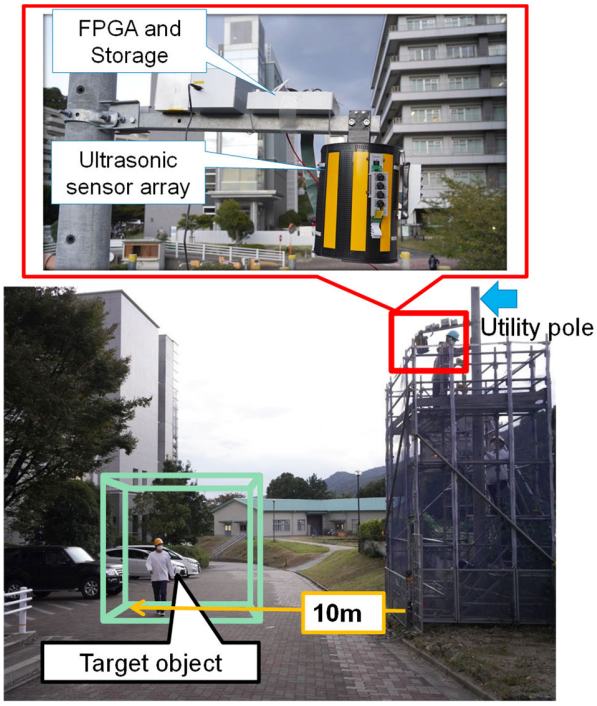

FIGURE 10. **Measurement environment and measurement system installation.**

Fig. 9 (a) showcases the microphone array, whereas (b) exhibits the transducers. (a) The microphone array comprises 4 × 4 MEMS microphones arranged on a grid, spaced 5 mm apart, which is less than half the wavelength of the carrier wave. All microphones share a common clock source and produce a PDM signal with a maximum frequency of 4.55 MHz. (b) The speakers transmit modulated waves. Employing an H-bridge configuration, the circuit can drive microphones with an amplitude that is twice the supply voltage. Four speakers are integrated into the speaker array.

B. MEASUREMENT ENVIRONMENT

Fig. 10 illustrates the installation of the ultrasonic array during measurement. The study aims to monitor the road environment, with the measurement system installed on a utility

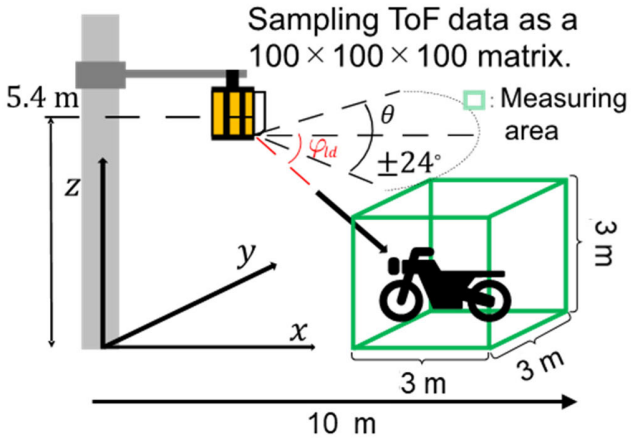

FIGURE 11. Experiment environment. θ is azimuth angle, φ is elevation angle).

pole by the roadside. The measurement area encompasses a 3000 mm cubic space positioned 10 m from the pole, where the objects of interest are located and measured. Mounted on a fixed arm 5.4 m above the pole, the ultrasonic array is designed to encircle the pole in a belt-like structure.

V. EXPERIMENT RESULTS

Experimental results are presented in this section. Subsection A outlines the training data, detailing the data sampling method and the classes of objects intended for 3D reconstruction. Section B discusses the results of the 3D reconstruction.

A. SIGNAL STRENGTH DISTRIBUTION AND TRAINING DATA

The experimental environment for acquiring ultrasonic data in 3D space is depicted in Fig. 11. Following the placement of the sensor on the utility pole, beamforming, as described in Section II, was employed to gather the data. During the experiment, the pole height was 5.4 m, and the radiation direction φ_{ld} was 53°. The target measurement range extended 10 m from the sensor in the direction of distance. The elevation angle ranged from −24° to 24°, and similarly, the azimuth angle spanned from −24° to 24°. The area under consideration measured 3000 × 3000 × 3000 mm^3, delineated by a green cube in the figure.

Fig. 12 illustrates images depicting ultrasonic data and training data for a circular reflector, box, motorcycle, and person. These images were captured from the position of the sensor installed on the utility pole. As previously mentioned, the ultrasonic data formed a cube measuring 3000 mm on each side. Consequently, each voxel was subdivided every 30 mm to create dimensions of (100 × 100 × 100). The obtained ultrasonic data were normalized from 0 to 1.0 for display purposes: blue voxels indicate a strength between 0.6 and 0.7; green voxels represent values between 0.7 and 0.9; and red voxels signify values greater than 0.9. The training data were utilized to approximate a 3D object based on the size and shape of each object. The box, motorcycle, and human are approximated as cuboids, whereas the reflector is approximated as a sphere.

TABLE 2. Approximated size of the target object.

	(a) Box	(b) Motorcycle	(c) Human	(d) Reflector
Length [mm] (Diameter)	700	1600	300	100
Width [mm]	300	600	300	

TABLE 3. Reconstruction result (3D stationary target).

	(a) Reflector	(b) Box	(c) Motorcycle	(d) Human	Average
Precision	0.935	0.957	0.959	0.911	0.939
Recall	0.835	0.957	0.950	0.825	0.868
F value	0.880	0.957	0.954	0.866	0.902

Table 2 displays the sizes of the approximated solids up to 1,600 mm. Considering the challenge of imaging objects in 3D space within a few centimeters, we divided the object into 100 mm pieces and represented them as (30, 30, 30). A value of 1.0 was assigned to the position where the object existed, and voxel plots were generated in blue for visualization. A total of 1628 ultrasonic data points were measured, with 1304 used for training data and 324 for test data.

B. EVALUATION RESULTS OF 3D STATIONARY OBJECTS

Fig. 13 showcases the input and output results achieved using the proposed VAE model for 3D stationary objects: (a) box, (b) motorcycle, (c) person, and (d) reflector at the center. The different plot colors denote different values: values between 0.4 and 0.7 are represented in blue, values between 0.7 and 0.9 are represented in green, and values greater than 0.9 are represented in red after data normalization. A comparison of (a) and (b) illustrates the difference in size between the box and motorcycle. Furthermore, comparing (a) and (b) with (c) and (d) reveals variations in size and shape, as identified and classified.

Table 3 presents the results of the precision, recall, and F-values obtained when the threshold value was set to 0.5. For evaluation, a value above 0.5 was regarded as 1. Values below 0.5 were taken as 0. The overall accuracies for precision, recall, and F-value were 0.939, 0.868, and 0.902, respectively.

This paper targets the reconstruction of small objects ranging from a few hundred millimeters to 1,600 mm. Reconstruction would likely be even more feasible for larger measurement targets such as cars and trucks. As in Substage (III) of our method's pipeline, the background signals are were removed to intensify the signal reflected from the target object. Consequently, for larger objects such as cars

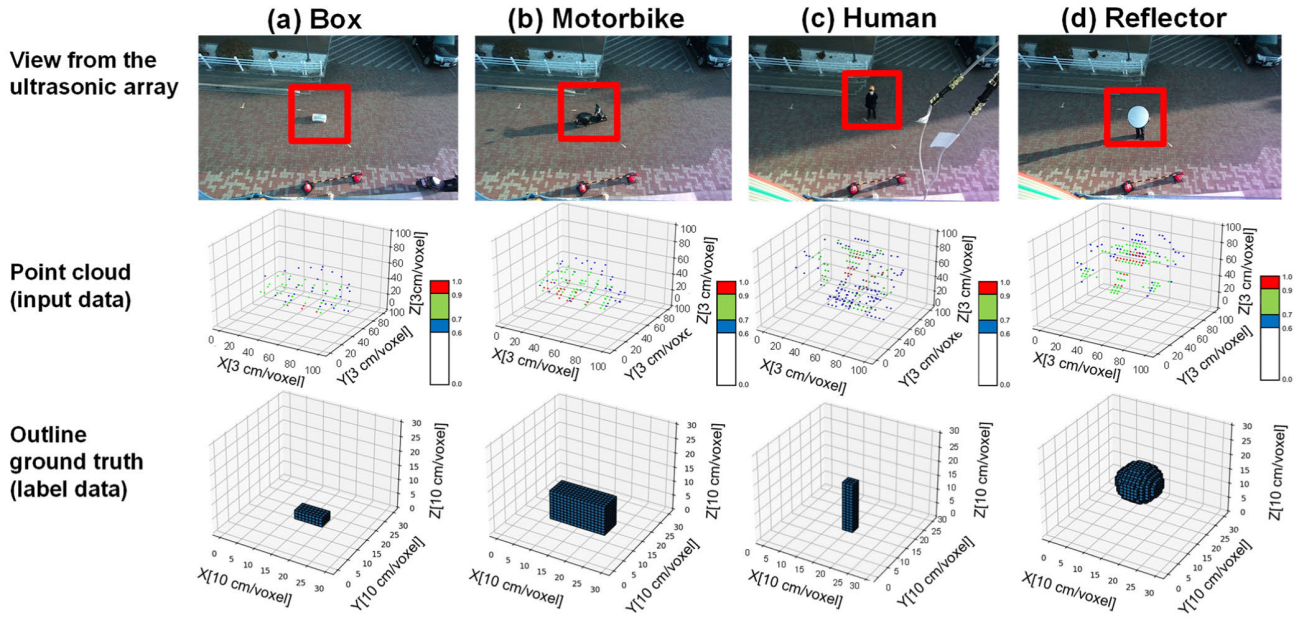

FIGURE 12. **3D static target dataset: appearances from ultrasonic array, point clouds, and ground truth.**

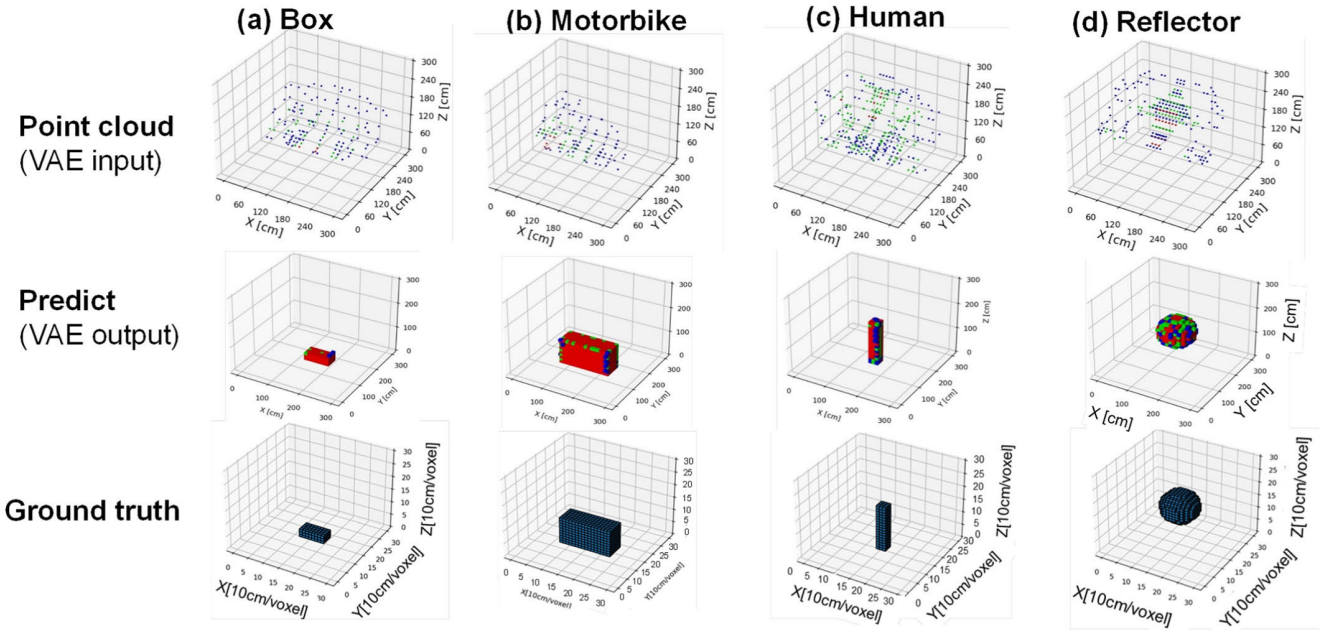

FIGURE 13. **3D reconstruction results and ground truth for each class.**

and trucks, the difference from the background would become more significant.

VI. CONCLUSION

Here, we present a 3D object reconstruction technique tailored for high-interference environments utilizing ultrasonic array sensors and a VAE. Leveraging ultrasonic data and supervised VAE, we accurately image the position, size, and shape of 3D stationary objects. Across the four object types examined, the average results demonstrate a precision of 0.939, recall of 0.868, and F-value of 0.902. It is worth noting that the experiment focused solely on stationary objects.

For a comprehensive assessment of road conditions, accounting for the Doppler effect becomes imperative, particularly for measuring moving vehicles and other mobile objects. Moreover, the measurement system employed in this

study isn't limited to ultrasonic waves; it can also capture audible sounds. By incorporating audible sounds emitted by vehicles and pedestrians, akin to previous studies, we can achieve a more nuanced understanding of road conditions.

ACKNOWLEDGMENT

An earlier version of this paper was presented at the 2022 IEEE International Ultrasonics Symposium [DOI: 10.1109/IUS54386.2022.9958685].

REFERENCES

[1] J. S. Dvorak, M. L. Stone, and K. P. Self, "Object detection for agricultural and construction environments using an ultrasonic sensor," *J. Agric. Saf. Health*, vol. 22, no. 2, pp. 19–107, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:24771869

[2] J. Steckel, A. Boen, and H. Peremans, "Broadband 3-D sonar system using a sparse array for indoor navigation," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 161–171, Feb. 2013, doi: 10.1109/TRO.2012.2221313.

[3] V. A. Zhmud, N. O. Kondratiev, K. A. Kuznetsov, V. G. Trubin, and L. V. Dimitrov, "Application of ultrasonic sensor for measuring distances in robotics," *J. Phys., Conf. Ser.*, vol. 1015, May 2018, Art. no. 032189, doi: 10.1088/1742-6596/1015/3/032189.

[4] I. Toru, Y. Yasuda, S. Sato, S. Izumi, and H. Kawaguchi, "Millimeter-precision ultrasonic DSSS positioning technique with geometric triangle constraint," *IEEE Sensors J.*, vol. 22, no. 16, pp. 16202–16211, Aug. 2022, doi: 10.1109/JSEN.2022.3188007.

[5] Y. Yoshikawa, Y. Yasuda, I. Toru, S. Izumi, and H. Kawaguchi, "12.5-mdistance measurement in high-interference environment using ultrasonic array sensors," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2021, pp. 1–6, doi: 10.1109/I2MTC50364.2021.9459822.

[6] H. Furuhashi, Y. Kuzuya, Y. Uchida, and M. Shimizu, "Three-dimensional imaging sensor system using an ultrasonic array sensor and a camera," in *Proc. IEEE Sensors*, Nov. 2010, pp. 713–718, doi: 10.1109/ICSENS.2010.5690053.

[7] S. Harput and A. Bozkurt, "Ultrasonic phased array device for acoustic imaging in air," *IEEE Sensors J.*, vol. 8, no. 11, pp. 1755–1762, Nov. 2008, doi: 10.1109/JSEN.2008.2004574.

[8] P. N. Keating, T. Sawatari, and G. Zilinskas, "Signal processing in acoustic imaging," *Proc. IEEE*, vol. 67, no. 4, pp. 496–510, 1979, doi: 10.1109/PROC.1979.11279.

[9] R. Whyte, L. Streeter, M. J. Cree, and A. A. Dorrington, "Application of LiDAR techniques to time-of-flight range imaging," *Appl. Opt.*, vol. 54, no. 33, p. 9654, Nov. 2015, doi: 10.1364/ao.54.009654.

[10] S.-W. Yang and C.-C. Wang, "On solving mirror reflection in LiDAR sensing," *IEEE/ASME Trans. Mechatronics*, vol. 16, no. 2, pp. 255–265, Apr. 2011, doi: 10.1109/TMECH.2010.2040113.

[11] A. Ziebinski, R. Cupek, H. Erdogan, and S. Waechter, "A survey of ADAS technologies for the future perspective of sensor fusion," in *Proc. Int. Conf. Comput. Collective Intell.*, 2016, pp. 135–146, doi: 10.1007/978-3-319-45246-3_13.

[12] L. Jiang, T. Cai, Q. Ma, F. Xu, and S. Wang, "Active object detection in sonar images," *IEEE Access*, vol. 8, pp. 102540–102553, 2020, doi: 10.1109/ACCESS.2020.2999341.

[13] Y. Yari et al., "Automated measurement of ovary development in Atlantic salmon using deep learning," *Ultrasound Med. Biol.*, vol. 50, no. 3, pp. 364–373, Mar. 2024, doi: 10.1016/j.ultrasmedbio.2023.11.008.

[14] G. Allevato, M. Rutsch, J. Hinrichs, M. Pesavento, and M. Kupnik, "Embedded air-coupled ultrasonic 3D sonar system with GPU acceleration," in *Proc. IEEE Sensors*, Oct. 2020, pp. 1–4, doi: 10.1109/SENSORS47125.2020.9278601.

[15] G. Allevato et al., "Air-coupled ultrasonic spiral phased array for high-precision beamforming and imaging," *IEEE Open J. Ultrason., Ferroelectr., Freq. Control*, vol. 2, pp. 40–54, 2022, doi: 10.1109/OJUFFC.2022.3142710.

[16] S. Kumar and H. Furuhashi, "Long-range measurement system using ultrasonic range sensor with high-power transmitter array in air," *Ultrasonics*, vol. 74, pp. 186–195, Feb. 2017, doi: 10.1016/j.ultras.2016.10.012.

[17] W. Kim, "Signal-to-noise ratio and bandwidth for pseudo-random codes in an ultrasonic imaging system," *Ultrason. Imag.*, vol. 6, no. 3, pp. 313–323, Jul. 1984, doi: 10.1016/0161-7346(84)90016-6.

[18] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," Dec. 2013, *arXiv:1312.6114*.

[19] Y. Na, Y. Guo, Q. Fu, and Y. Yan, "An acoustic traffic monitoring system: Design and implementation," in *Proc. IEEE 12th Int. Conf. Ubiquitous Intell. Comput., IEEE 12th Int. Conf. Auton. Trusted Comput., IEEE 15th Int. Conf. Scalable Comput. Commun. Associated Workshops (UIC-ATC-ScalCom)*, Aug. 2015, pp. 119–126, doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.41.

[20] T. Takahashi et al., "Acoustic traffic monitoring based on deep neural network trained by stereo-recorded sound and sensor data," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2023, pp. 935–939, doi: 10.23919/eusipco58844.2023.10290031.

[21] M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed., New York, NY, USA: McGraw-Hill, 2002.

[22] J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum, "Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling," 2016, *arXiv:1610.07584*.

[23] *PYNQ-Z2*. Accessed: Jan. 25, 2024. [Online]. Available: https://www.pynq.io

RYOTARO OHARA received the Foundation degree from Tokyo Metropolitan College of Industrial Technology in 2016, the B.Eng. degree in environmental and life science from Toyohashi University in 2018, and the Master of Science degree in technology and innovation from Kobe University, Hyogo, Japan, in 2021. His current research focuses on 3D environment monitoring using ultrasonic technology.

YUTO YASUDA received the B.Eng. degree in computer science and systems engineering and the M.Eng. degree in system informatics from Kobe University, Kobe, Japan, in 2021 and 2023, respectively. His research was development of a bather monitoring system in a bathroom using ultrasonic array sensors and machine learning.

RIKU HAMABE received the B.Eng. degree in computer science and systems engineering and the M.Eng. degree in system informatics from Kobe University, Kobe, Japan, in 2021 and 2023, respectively. His research was ultrasonic measurements and simulation of ray tracing in a bathroom environment.

SHUN SATO received the B.Eng. degree in computer science and systems engineering from Kobe University, Kobe, Japan, in 2022, where he is currently pursuing the master's degree. His current research interests include development of a bather monitoring system in a bathroom using ultrasonic array sensors and machine learning.

ISHII TORU received the B.Eng. degree in electronic engineering from Kyoto University, Kyoto, Japan, in 1986, and the M.B.A. and Ph.D. degrees in science, technology, and innovation from Kobe University, Kobe, Japan, in 2013 and 2022, respectively. He began his career with Minolta, Osaka, Japan, in 1986, where he developed digital imaging devices. In 1999, he transitioned to Murata Manufacturing, Kyoto, as the Development Manager of automotive radars, communication modules, and sensors. Since 2018, he has been involved in research on ultrasound sensing with Kobe University.

SHINTARO IZUMI (Member, IEEE) received the B.Eng. and M.Eng. degrees in computer science and systems engineering from Kobe University, Hyogo, Japan, in 2007 and 2008, respectively, and the Ph.D. degree in engineering from Kobe University in 2011. From 2009 to 2011, he was a JSPS Research Fellow with Kobe University, where he was an Assistant Professor with the Organization of Advanced Science and Technology from 2011 to 2018 and an Associate Professor with the Institute of Scientific and Industrial Research, Osaka University, from 2018 to 2019. Since 2019, he has held the position of an Associate Professor with the Graduate School of System Informatics, Kobe University. His current research interests include biomedical signal processing, communication protocols, low-power VLSI design, and sensor networks. He has served as a Technical Committee Member for IEEE Biomedical and Life Science Circuits and Systems, a Student Activity Committee Member for IEEE Kansai Section, and a Program Committee Member for IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips). Previously, he chaired the IEEE Kansai Section Young Professionals Affinity Group and received the 2010 IEEE SSCS Japan Chapter Young Researchers Award.

HIROSHI KAWAGUCHI (Member, IEEE) received the B.Eng. and M.Eng. degrees in electronic engineering from Chiba University, Chiba, Japan, in 1991 and 1993, respectively, and the Ph.D. degree in electronic engineering from The University of Tokyo, Tokyo, Japan, in 2006. He began his career with Konami Corporation, Kobe, Japan, in 1993, where he developed arcade entertainment systems. He transitioned to The Institute of Industrial Science, The University of Tokyo, as a Technical Associate in 1996 and was later appointed as a Research Associate in 2003. In 2005, he joined The Graduate School of Engineering, Kobe University, Kobe, as a Research Associate. He was a Visiting Researcher with the Politecnico di Milano from 2015 to 2016. Since 2016, he has been a Full Professor with The Graduate School of Science, Technology, and Innovation, Kobe University. Additionally, he is also a Collaborative Researcher with The Institute of Industrial Science, The University of Tokyo. His current research interests encompass a wide range of topics, including low-voltage operating circuits, soft error characterization and mitigation, ubiquitous sensor networks, organic semiconductor circuits, data converters, healthcare devices, and neuro computer architecture. He has received several awards, including the IEEE ISSCC 2004 Takuo Sugano Outstanding Paper Award, ACM/IEEE ASP-DAC 2013 University Design Contest Best Design Award, and IEEE ICECS 2016 Best Paper Award. He has actively contributed to various technical committees and conferences, serving as a member for the Design and Implementation of Signal Processing Systems (DISPS) Technical Committee for IEEE Signal Processing Society and Technical Program Committees for conferences, such as IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE Workshop on Signal Processing Systems (SiPS), IEEE Custom Integrated Circuits Conference (CICC), and IEEE Symposium on Low-Power and High-Speed Chips (COOL Chips). He has also been involved in organizing committees for conferences like IEEE Asian Solid-State Circuits Conference (A-SSCC) and ACM/IEEE Asia and South Pacific Design Automation Conference (ASP-DAC). Furthermore, he serves as an Associate Editor for various journals, including *Journal of Signal Processing Systems* (Springer), *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *IEICE Transactions on Electronics*, and *IPSJ Transactions on System LSI Design Methodology* (TSLDM). He is a member of the ACM and the IEICE.