ICIC Express Letters
Part B: Applications
Volume **12**, Number **7**, July **2021**

ICIC International ©2021 ISSN 2185-2766
pp. 635–643

WATER DEPTH ESTIMATION USING MEASUREMENT DISTORTION OF A TIME-OF-FLIGHT CAMERA AND ITS APPLICATION TO INTERACTIVE ART

Shunnosuke Kataoka[1], Noriaki Suetake[2] and Takanori Koga[3,*]

[1]Graduate School of Library, Information and Media Studies
University of Tsukuba
1-2 Kasuga, Tsukuba 305-8550, Japan
shunnosuke834@digitalnature.slis.tsukuba.ac.jp

[2]Graduate School of Science and Technology for Innovation
Yamaguchi University
1677-1 Yoshida, Yamaguchi 753-8512, Japan
nsuetake@yamaguchi-u.ac.jp

[3]Graduate School of Humanity-Oriented Science and Engineering
Kindai University
11-6 Kayanomori, IIzuka 820-8555, Japan
*Corresponding author: koga@fuk.kindai.ac.jp

Received November 2020; accepted February 2021

Abstract. *In this study, we propose a simultaneous multi-point estimation method for the estimation of the depths of water using the principle of distortion of measurement in a depth image taken by a time-of-flight camera. Generally, a float or an ultrasonic sensor is used to measure water depth, but these are not suitable for simultaneous measurement at multiple points. The proposed method realizes simultaneous estimation of the depths of water at multiple points in real time by image processing. Thus, there is no physical interference with floats and no need to install multiple ultrasonic sensors at each measurement target point. In this study, the effectiveness of the proposed method is demonstrated by the results of water depth estimation experiment. The experimental results show that the proposed method can estimate almost accurate water depths at several points simultaneously when the water depth falls within a certain range. Furthermore, we demonstrate an actual application of the proposed method, seeing how sound and images change when an interactive artwork is observed through water kept in the buckets on a tabletop as a medium.*

Keywords: Time-of-flight (ToF) camera, Depth image, Water depth estimation, Simultaneous multi-point estimation, Interactive art

1. **Introduction.** In recent years, many studies, based on image processing for surface shape estimation of transparent objects such as solids and fluids and shape estimation of objects in transparent fluids, have been actively conducted using various approaches [1, 2, 3, 4]. Tanaka *et al.* [1] proposed a method for restoring the shape of a transparent static object using the measurement distortion of a time-of-flight (ToF) camera [5]. Asano *et al.* [2] realized the shape estimation of underwater objects based on Beer-Lambert's law by using infrared images captured at two wavelengths. Qian *et al.* [3] proposed a method for estimating the surface shape of a fluid, which involved capturing images of geometric patterns observed in the fluid using multiple cameras and performing stereo matching. Ye *et al.* [4] showed that by adding an ultrasonic sensor to the ToF sensor, 3D shapes can be obtained even in images containing transparent objects. With technological

DOI: 10.24507/icicelb.12.07.635

advances in image processing for transparent objects, further development of human-computer interactions using the object as a medium is expected.

For example, in the field of interactive art, the method of using a transparent liquid such as water as a medium for the viewer to touch the work is attracting attention. Koga and Matsuo [6] proposed a work that performs adaptive image projection according to the changing shape of puddles on a tabletop caused by the viewers' touch. FairLift [7] is a work that applies an optical system to displaying an aerial image underwater and on water. In this study, the position of the projected image is adjusted after measuring the water depth using an ultrasonic sensor. An ultrasonic sensor is a device typically used for measuring the water level and the water depth, which is easy to operate and inexpensive. However, when water depths are simultaneously measured at multiple points, the sensors must be arranged at each measurement point, which complicates the configuration. Although using floats is one of the solutions, it cannot solve the problems of physical interference and appearance.

In this study, we propose a multi-point simultaneous measurement method of water depths using depth images, which overcomes the aforementioned problems, and apply it to an interactive artwork using water as a medium. For multi-point simultaneous measurement of water depths, a ToF camera is used to obtain the depth information of a transparent object and the resulting measurement distortion. This distortion is caused by the difference between the refractive indices of the transparent object and that of air. Because this measurement distortion is observed even when water is the measurement target, the proposed method models the information of the measurement distortion according to the water depth information obtained in advance, and performs the simultaneous multi-point estimation of the water depth using this model. In this study, the effectiveness of the proposed method is verified using a series of water depth estimation experiments. Besides, we describe an application of the proposed method to an interactive artwork that uses water in the buckets on a tabletop as a medium for changing sounds and images.

2. **Proposed Method.** The system configuration for the multi-point simultaneous measurement of water depth and the interactive artwork used in this study is shown in Figure 1(a). The depth information of the water in the buckets is obtained by the ToF camera (Microsoft Kinect V2), which is installed facing the floor surface, as shown in Figure 1(b). Considering the Kinect V2's sensing capability, the installation height was set to 1,000 mm from the floor. The acquired depth information is input to the workstation,

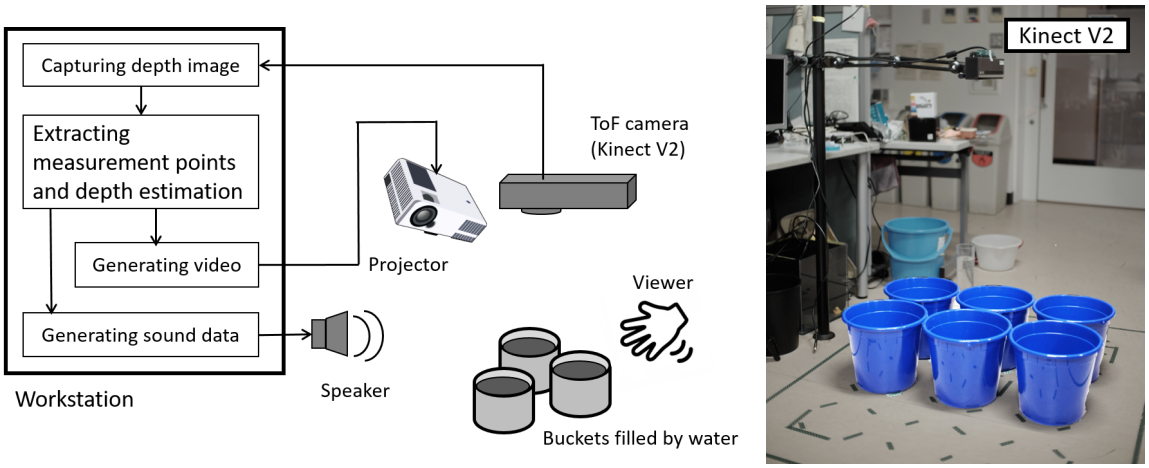


FIGURE 1. (a) System configuration; (b) state of the measurement experiment

which extracts the target points for measurement and estimates the water depth. Images and sounds generated based on the estimation results are output from the projector and speakers. In this section, a detailed explanation of the depth estimation is described.

2.1. **Measurement distortion of ToF camera.** The depth information sensed by the Kinect V2 ToF camera is obtained as a depth image with a depth value for each pixel. In the proposed method, simultaneous multi-point measurement of water depths is performed by accounting for the measurement distortion that occurs when the depth information of a transparent object is obtained using a ToF camera. Measurement distortion is a phenomenon in which when a depth image is obtained using a ToF camera, a distance longer than the actual distance to the object is measured in a region where the transparent object exists. This is caused by the slowing down of the speed of light traveling inside the transparent object due to the difference between the refractive indices of the transparent object and that of air [1].

For example, in the depth image obtained under the conditions shown in Figure 1, the area where water exists causes measurement distortion that has a depth value of 1,000 mm or more. Because the amount of distortion changes according to the water depth, this phenomenon can be used to detect the water area and estimate the water depth. Figure 2(a) shows a depth image of water at a depth of 20 mm in six buckets captured with a Kinect V2 ToF camera installed at a height of 1,000 mm from the floor. It can be seen from Figure 2(a) that the depth value measured is larger than the installation height of 1,000 mm from 130px to 350px, which denote the coordinates where the buckets were

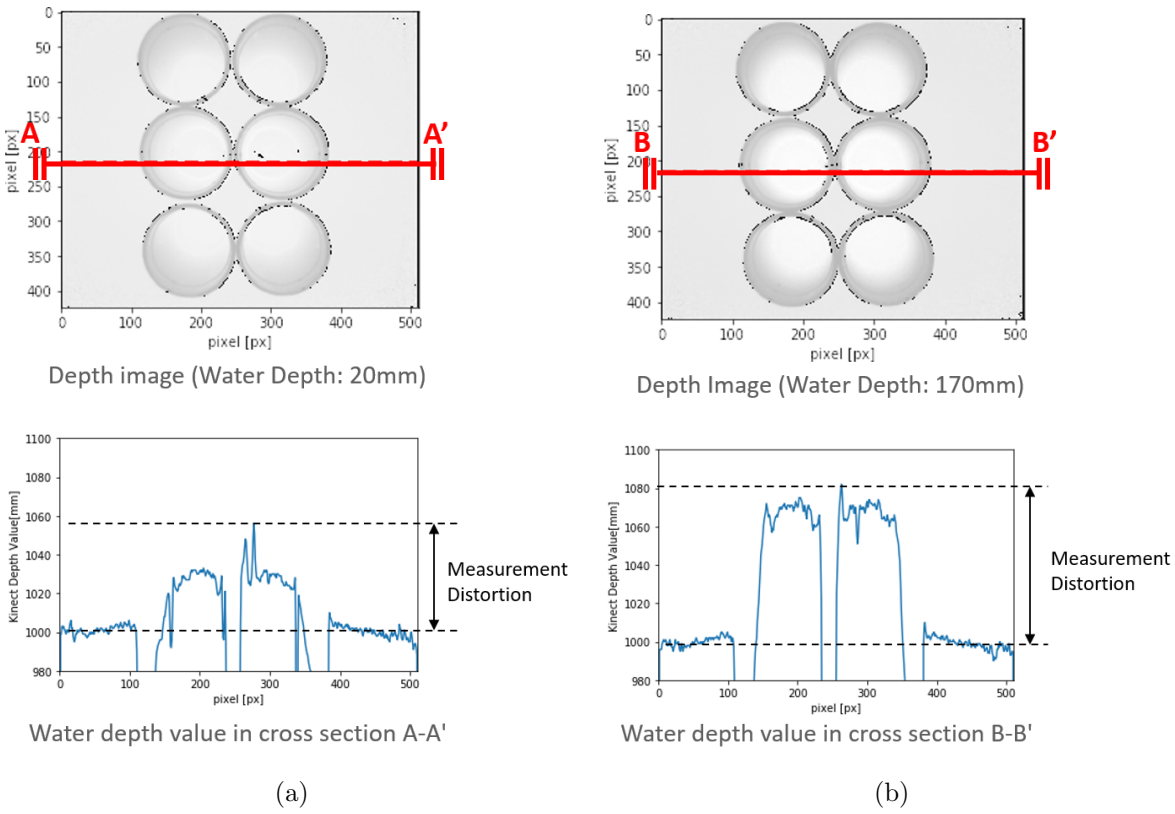


FIGURE 2. Illustration of measurement distortion. In (a) and (b), the upper parts depict the depth images and the lower graphs show the depth value in a cross-section of each depth image. The measurement distortion refers to the measured depth value exceeding the physically maximum depth value (1,000 mm in this case).

placed. Moreover, it can be confirmed that the measurement distortion is larger when the water depth is 170 mm than when it is 20 mm as shown in Figure 2(b).

In this study, the amount of measurement distortion obtained according to the water depth is modeled in advance, and this model is used to detect water areas in an arbitrary input depth image and to estimate their depths.

2.2. **Construction of depth estimation model.** Figure 3 shows the procedure of model construction. First, to model the amount of measurement distortion observed according to the water depth in advance, the water depth in the buckets was changed to 0 mm, 50 mm, 100 mm, 150 mm, and 200 mm, and then 100 frames of depth images were acquired at each water depth. The images were captured while moving the bucket so that the bucket would appear in the entire depth image. After that, by synthesizing each depth image, the depth images were generated as shown in Figure 3(1).

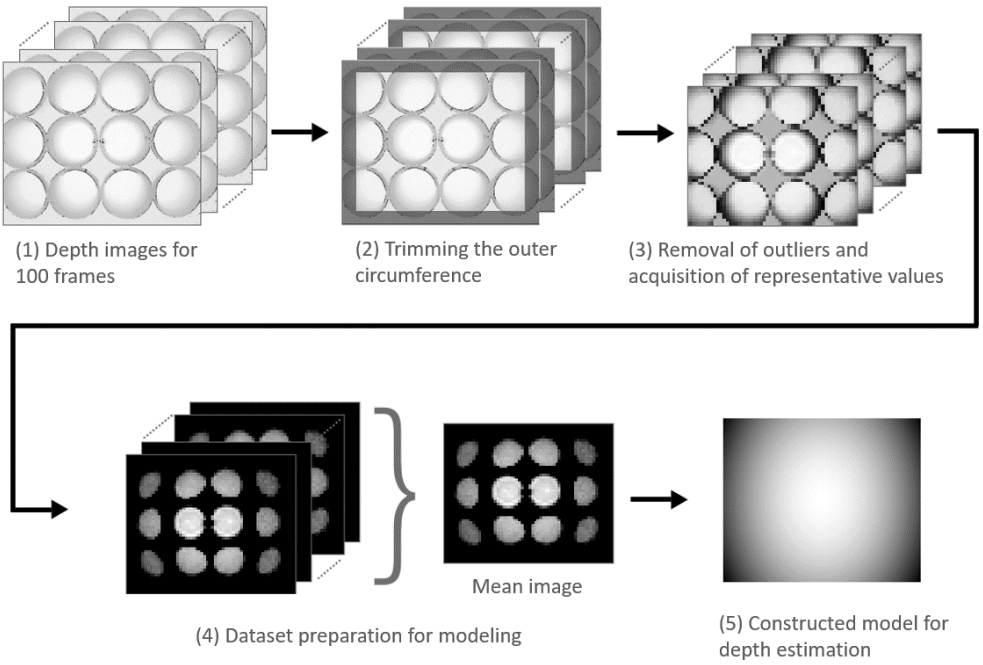


FIGURE 3. Procedure for creating a model for water depth estimation

The size of a depth image captured by Kinect V2 is 512×424 [pixels]. The 30 pixels around the image were cropped, as shown in Figure 3(2), so that they would not be included in the modeling data because the depth image tends to contain big random noise. In addition, because the captured depth image contains many outliers due to noise, these were removed using the following procedure. In the proposed method, the depth image was divided into 8×8 [pixels] of windows. Then, the top 15% pixels were trimmed in each window as outliers. From the depth values excluding outliers, the largest value was set to be the representative value of the window. By applying this process to all the windows in the depth image, 3,392 representative values were obtained, as shown in Figure 3(3). Subsequently, the values that exceed a threshold were extracted from the 3,392 representative values. The threshold value was set to 1,020 mm, which is the installation height of Kinect V2 from the floor of 1,000 mm plus an offset of 20 mm to account for the noise captured in the depth image. The mean image of the 100 representative images was used as modeling data, as shown in Figure 3(4). In this regard, the modeling data were acquired for each of the above-mentioned water depths.

Finally, a function-based approximation was performed for the modeling data obtained, and estimation models were generated for each water depth, as shown in Figure 3(5). In

this case, a quadratic function was used for the approximation. This is because it was observed the detected depth value tends to be larger toward the center of the depth image and smaller toward the edges. Specifically, the coefficients a and b, which minimize the following function, were calculated for each of the five actual depth cases:

$$f = \sum_{x=0}^{X} \sum_{y=0}^{Y} \left\{ d(x,y) - \left(-ax^2 - by^2 + c \right) \right\}^2, \tag{1}$$

where x and y are pixel coordinates, $d(x,y)$ is a sensor depth value, and c is the maximum value of depth in the modeling data.

2.3. **Estimation of water depth.** Using the constructed model, the water depth is estimated for the area that exceeds the threshold (i.e., the water area) in the input depth image. Let us consider $f_n(x,y)$ to be the n-th estimation model. Here, n is 0, 1, 2, 3, or 4, which corresponds to a water depth of $w_0 = 0$ mm, $w_1 = 50$ mm, $w_2 = 100$ mm, $w_3 = 150$ mm, and $w_4 = 200$ mm. If $d_r(x,y)$ denotes a sensor depth value at (x,y), two models $f_n^*(x,y)$ and $f_{n+1}^*(x,y)$ are found, which satisfy $f_n^*(x,y) < d_r(x,y) < f_{n+1}^*(x,y)$. Using these n-th and $(n+1)$-th models, corresponding depth values w_n and w_{n+1} are given. The final estimate value $\hat{w}(x,y)$ is calculated as follows:

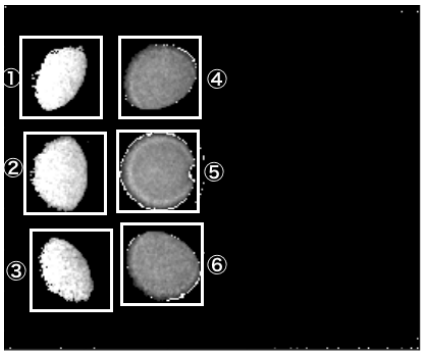
$$\hat{w}(x,y) = \frac{d_r(x,y) - f_n^*(x,y)}{f_{n+1}^*(x,y) - f_n^*(x,y)} \times (w_{n+1} - w_n) + w_n. \tag{2}$$

This procedure is applied to all the points to be estimated where water exists.

3. **Experiment and Evaluation.** To verify the accuracy of the proposed method for estimating the water depth, specific results of model building and estimation under the conditions described in Subsections 2.2 and 2.3 with the equipment installed under the conditions described in Figure 1 and Section 2 are shown below. The results of water depth estimation are shown in Figures 4-7. The shades of the pixels in the figures indicate the estimated water depth values. The brighter the pixel value, the deeper the estimated water depth. To observe the effect of change in the position in the field of view (FOV) and the actual water depth on the estimation accuracy, we conducted the experiments shown in the following subsections.

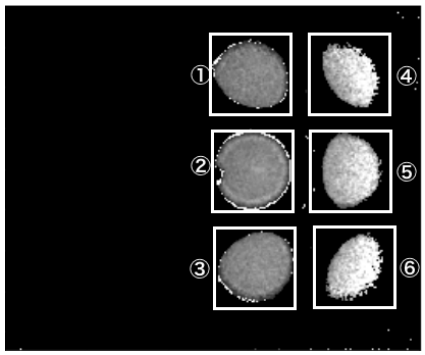
In the evaluation, the areas of the buckets were specified for the depth image as shown in Figures 4-7. The mean and standard deviation of the estimated water depth were calculated for the pixels whose depth exceeded 0 in the specified areas. This calculation was performed for all the areas of the buckets. Specifically, the evaluation was conducted for the following three situations.

3.1. **Depth estimation at the side ends of the field of view.** Figures 4(a) and 4(b) show the estimation results when the buckets were placed on the left and right sides of the FOV, respectively. When the buckets were placed on the left side of the FOV, the estimation results demonstrated that the means of the estimated water depths of the buckets No. 1 to No. 3 placed at the end were significantly different from the actual water depth of 200 mm. By contrast, the means of the estimated water depths in the buckets No. 4 to No. 6 placed on the center side were found to be closer to the actual water depth. A similar tendency was observed in the estimation results when the buckets were placed on the right side of the FOV. From these results, it was confirmed that the estimation accuracy of the proposed method is significantly reduced at the side ends of the FOV.



No.	Mean [mm]	SD [mm]
1	414.5	91.8
2	364.0	78.6
3	369.8	83.3
4	201.1	41.0
5	215.0	44.7
6	203.7	40.7

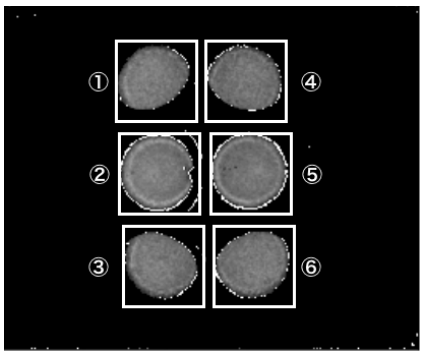
(a)



No.	Mean [mm]	SD [mm]
1	192.6	38.4
2	202.2	40.3
3	201.7	41.2
4	321.6	75.6
5	295.6	66.9
6	348.5	86.5

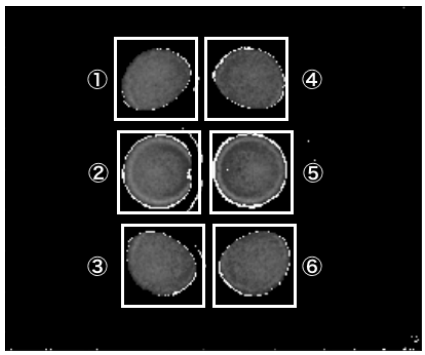
(b)

FIGURE 4. Depth estimation results in the left and right halves of the field of view (Actual depth: 200 mm)



No.	Mean [mm]	SD [mm]
1	182.1	31.6
2	188.2	34.7
3	178.0	32.0
4	171.1	28.7
5	176.9	29.7
6	178.6	30.4

(a)

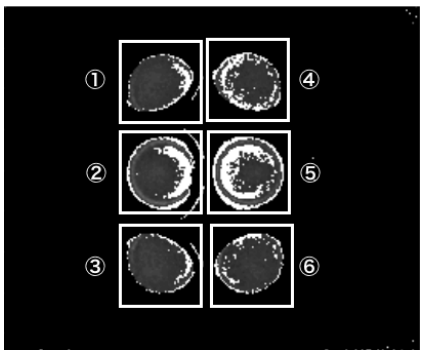


No.	Mean [mm]	SD [mm]
1	143.0	22.8
2	146.6	28.1
3	139.1	21.7
4	129.2	18.5
5	134.4	26.3
6	135.9	20.3

(b)

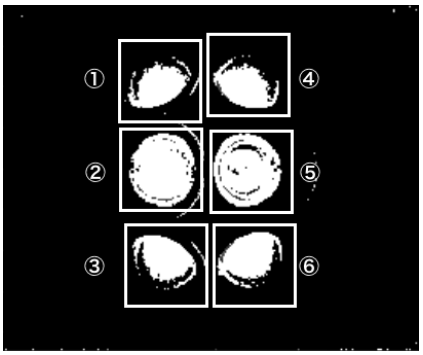
FIGURE 5. Depth estimation results in the center of field of view: (a) actual depth: 170 mm; (b) actual depth: 130 mm

3.2. **Water depth estimation in the center of the field of view.** Following the estimation at the side ends of the FOV, the water depth estimation results when the buckets were placed in the center of the FOV were evaluated as follows. The evaluation areas at water depths of 170, 130, 80, and 20 mm are shown in Figures 5(a), 5(b), 6(a), and 6(b), respectively. It was confirmed that at water depths of 170, 130, and 80 mm, the estimated mean values were within an error of ±35 mm from the actual water depth. However, it was also observed that the estimation accuracy was significantly lower when



No.	Mean [mm]	SD [mm]
1	93.4	10.0
2	98.7	19.4
3	94.0	10.0
4	85.2	7.8
5	93.2	17.5
6	88.6	9.3

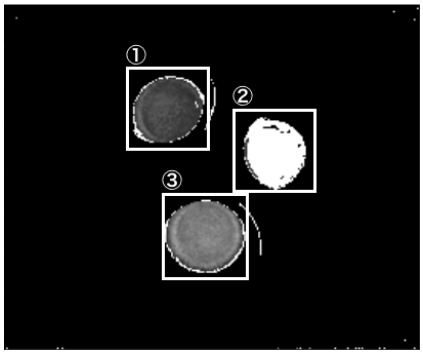
(a)



No.	Mean [mm]	SD [mm]
1	78.6	2.8
2	80.5	4.2
3	78.5	1.8
4	101.3	21.8
5	84.7	20.7
6	108.4	28.5

(b)

FIGURE 6. Depth estimation results in the center of field of view: (a) actual depth: 80 mm; (b) actual depth: 20 mm



No.	Actual depth [mm]	Mean [mm]	SD [mm]
1	110	115.2	20.1
2	30	101.7	22.9
3	180	190.3	33.3

FIGURE 7. Results of simultaneous estimation of different water depths

the water depth was low, such as 20 mm. The reason for this is believed to be that the amount of measurement distortion obtained from the depth image becomes less due to the lower water depth.

3.3. **Simultaneous estimation of different water depths.** Figure 7 shows the estimation results for three buckets with different water depths in the input frame. It can be seen that the mean of the estimated values in each area increases in proportion to the water depth even when the buckets with different water depths are placed. Moreover, it was observed that the estimation was almost accurate in the areas where the buckets with depths of 180 mm and 110 mm were placed. However, in the area where the bucket with a water depth of 30 mm was placed, accurate estimation of the water depth was possible at very few points. It can be assumed that many points did not exceed the threshold value

when selecting the measurement target points because the water depth was too shallow at those points, as we observed during the estimation at a water depth of 20 mm.

The results of the experiment established that the estimation can be performed effectively when the water depth is in the range of 80 mm and 200 mm in the center of FOV. The standard deviation tended to increase with the increase of the actual water depth. On the other hand, the accuracy decreased when the water depth was less than 80 mm or in the area at the side ends of the FOV.

4. **Application to Interactive Art.** In this study, we performed the prototyping of an interactive art system that changes sound and video using water as a medium, as shown in Figure 1(a). The coordinates of the buckets in the x and y directions, the estimation result of the water depth, and the state of the fluctuation of the water surface are transmitted to the projection image generation unit and the acoustic data generation unit, shown in Figure 1(a) using Open Sound Control (OSC). The projected image generator generates an image by processing the projected content, which varies according to the depth of the area where the water is detected, as shown in Figure 8. In this system, the ToF camera and the projector are calibrated in advance to project the generated images to the corresponding water areas. Meanwhile, the acoustic data generation unit uses the visual programming environment, Pure Data (Pd), to associate the data obtained from the varying water conditions with various effects such as pitch, volume, effect, and noise, and generates an interactive sound.

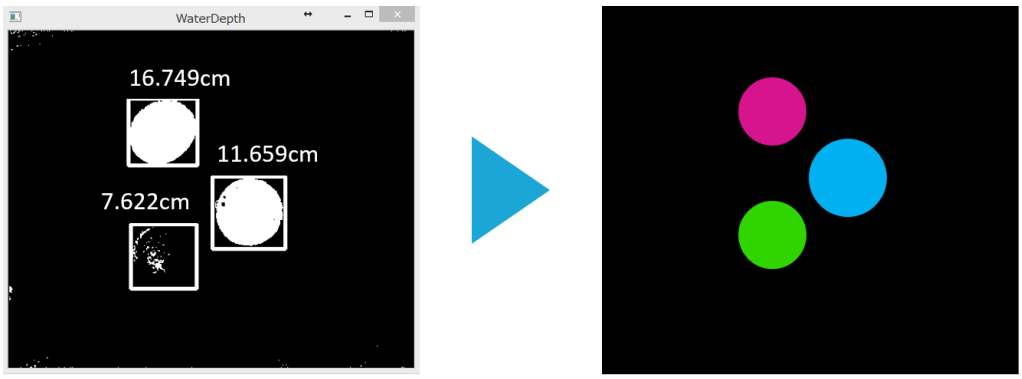
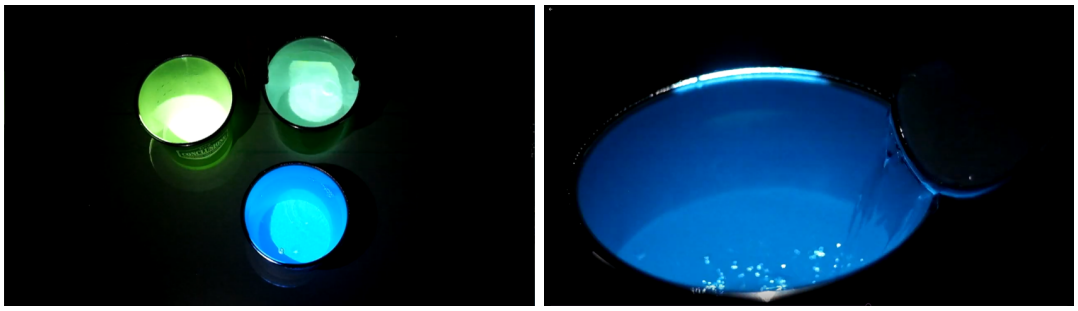


FIGURE 8. Screenshots of real-time water depth estimation software and an illustration of the projected image that changes according to the states of water (i.e., positions and depths)

After building the system, we confirmed that the viewer was able to enjoy the changes in the video and sound in real time by changing the water condition, as shown in Figure 9.

5. **Conclusions.** In this study, we proposed a multi-point simultaneous estimation method of water depth using the measurement distortion of the depth image obtained by a ToF camera and evaluated its estimation accuracy. A series of experiments were performed to establish that the proposed method can simultaneously estimate accurate water depths at multiple points, when the depth falls within a certain range of values. Moreover, as an application of the proposed method, we performed art prototyping that allowed viewers to interactively enjoy real-time variations in video and sound by changing the state of water.

In future, we will improve the estimation accuracy and depth range by resolving the errors that are observed at the side ends of the FOV and when the water depth is shallow.



(a) (b)

FIGURE 9. Interactive art named *Aqua Synthesizer* created by employing the proposed method. (a) The projected image and the generated sound change according to the depth of the water in the buckets placed on the table. When the viewer moves the buckets, the projected image follows, and the sound changes. (b) By increasing the water depth in the buckets, it is possible to change the projected image and the sound in real time.

Acknowledgment. This work was supported by JSPS KAKENHI 19K12091.

REFERENCES

[1] K. Tanaka, Y. Mukaigawa, H. Kubo, Y. Matsushita and Y. Yagi, Recovering transparent shape from time-of-flight distortion, *Proc. of CVPR*, pp.4387-4395, 2016.

[2] Y. Asano, Y. Zheng, K. Nishino and I. Sato, Shape from water: Bispectral light absorption for depth recovery, in *Computer Vision – ECCV 2016 Lecture Notes in Computer Science, vol.9910*, B. Leibe, J. Matas, N. Sebe and M. Welling (eds.), Springer, Cham, 2016.

[3] Y. Qian, M. Gong and Y.-H. Yang, Stereo-based 3D reconstruction of dynamic fluid surfaces by global optimization, *Proc. of CVPR*, pp.6650-6659, 2017.

[4] M. Ye, Y. Zhang, R. Yang and D. Manocha, 3D reconstruction in the presence of glasses by acoustic and stereo fusion, *Proc. of CVPR*, pp.4885-4893, 2015.

[5] F. Remondino and D. Stoppa (eds.), *TOF Range-Imaging Cameras*, Springer-Verlag, Berlin Heidelberg, 2013.

[6] T. Koga and T. Matsuo, Table-top projector-camera system with transparent water detection for interactive art, *Proc. of IMECS*, 2016.

[7] Y. Matsuura and N. Koizumi, FairLift: Interaction with mid-air images on water surface, *ACM SIGGRAPH 2018 Emerging Technologies*, Article No. 6, 2018.