3Dp1/3DSAp1 - 23L
(Late-News Paper)

Implemented of Images and Sounds Person Tracking System using Directional Volumetric Display

Mitsuru Baba[1], Ryuji Hirayama[2,3], Naoto Hoshikawa[4], Hirotaka Nakayama[5], Tomoyoshi Shimobaba[1], Tomoyoshi Ito[1], Atsushi Shiraki[1]

[1]Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan
[2] Research Fellow of the Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku Tokyo 102-0083, Japan
[3] Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo 125-8585, Japan
[4] National Institute of Technology, Oyama College, 771 Nakakuki, Oyama, Tochigi 323-0806, Japan
[5] National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo 181-8588, Japan
Keywords: Volumetric displays, Directional image, Person tracking, Digital signage.

ABSTRACT

In previous study, we developed the directional volumetric display which can display multiple images in different directions. In this study, we implemented a method of person tracking for the directional volumetric display to enable transmitting images and sounds following person using motion capture.

1 INTRODUCTION

In recent years, a display development which can display images not only in plane but also in space has a lot of attentions because of progress of technology such as Augmented Reality and Virtual Reality [1,2].

In previous study, we developed the directional volumetric display using strings, which can display multiple images in different directions by the method of displaying them in the same space [3,4]. Overview of the directional volumetric display is shown in Fig. 1. The directional volumetric display can display images in the specified direction in advance by the program, and has confidentiality that cannot obtain image information from other specified directions. For this reason, the directional volumetric display is expected to be used as an entertainment display such as a media art by realizing to transmit images and sounds to a position of person.

In this study, for the purpose of applying the directional volumetric display to a media application, we propose and implement a method of person tracking for the directional volumetric display to enable transmitting images and sounds following person using motion capture.

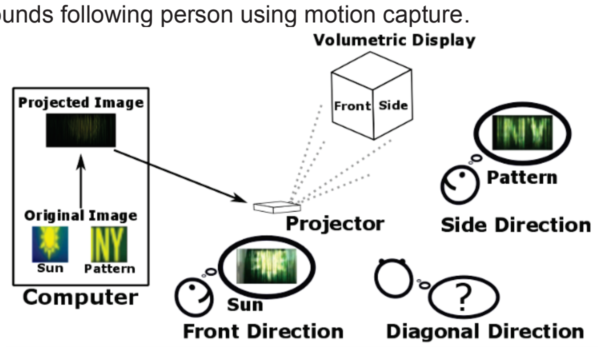


Fig. 1 Overview of the directional volumetric display.

2 The directional volumetric display

2.1 Create projection image

As the method of creating a projection image, consider a cube of size $\mathrm{P} \times \mathrm{Q} \times \mathrm{R}$ as shown in Fig. 2. $I_1(u_1, v_1)$ and $I_2(u_2, v_2)$ present a pixel values in the first and second original images, and the voxel values $\mathrm{V}(\mathrm{x}, \mathrm{y}, \mathrm{z})$ at coordinates $(\mathrm{x}, \mathrm{y}, \mathrm{z})$ is calculated by equation (1). Besides, the angle between $I_1(u_1, v_1)$ and $I_2(u_2, v_2)$ is θ, and the coordinates is expressed as $(u_1, v_1, w_1) = (\mathrm{x}, \mathrm{y}, \mathrm{z})$ and $(u_2, v_2, w_2) = (\mathrm{x} \cos\theta + z \sin\theta, \mathrm{y}, -\mathrm{x} \cos\theta\, \mathrm{z} \sin\theta)$. Therefore, the voxel values in any direction are calculated by equation (2). λ is a constant used to adjust total intensity. The directional volumetric display has three constraints, which determines the placement location on the condition that strings are placed in space without unevenness.

$$\mathbf{V(x,y,z)} = \boldsymbol{\lambda I_1(u_1, v_1) I_2(u_2, v_2)} \quad (1)$$

$$\mathbf{V(x,y,z)} = \boldsymbol{\lambda I_1(u_1, v_1) I_2(u_2 \cos\theta + w_2 \sin\theta, v_2)} \quad (2)$$

The directional volumetric display has three constraints. First, one ray from the projector must correspond to one string. Next, a string can be observed both from the front and side with no overlap between strings. Last, strings are arranged irregularly in order to create the directional volumetric display. In addition to three constraints, the projection image is created by adjusting the height of each ray in consideration of the depth of strings [4].

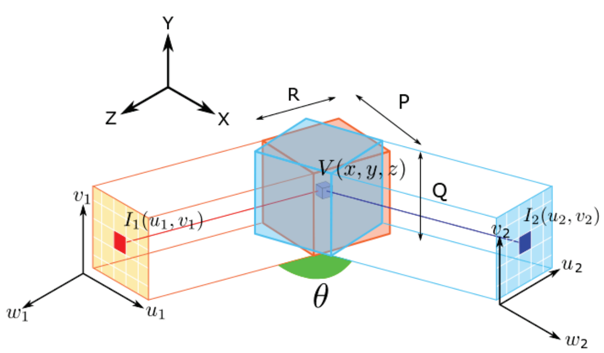


Fig. 2 Creation of projection image algorithm.

2.2 Person tracking

A person tracking algorithm is implemented using Kinect v2 (Microsoft Co., Japan). Specification of Kinect v2 is shown in Table 1, and flowchart of a person tracking algorithm is shown in Fig. 3. The person of tracking target who raises both hands is recognized by Kinect at first. Kinect recognizes a user to use the skeleton information acquired by the depth sensor and determine the person tracking to compare a current frame image that is the distance information from Kinect to the user with a previous frame image. The angle θ between Kinect and the tracking person is used to calculate the voxel values, display the images at the moving position, and transmit the sounds.

The location of Kinect is set at the center of the side direction $\theta = 90°$ of the directional volumetric display, so that the detection range of the person tracking is $60° \leq \theta \leq 130°$. Layout diagram for a person tracking is shown in Fig. 4.

Table 1 Specification of Kinect v2.

Feature	Specification
Color Camera	$\mathbf{1920 \times 1080}$
Depth Camera	$\mathbf{512 \times 424}$
Camera framerate	30 [fps]
Sensor	Time-of-Flight Camera
Camera Range	0.50 ~ 4.5 [m]
Angle of View (Horizontal / Vertical)	$\mathbf{70\ /\ 60}$ [°]

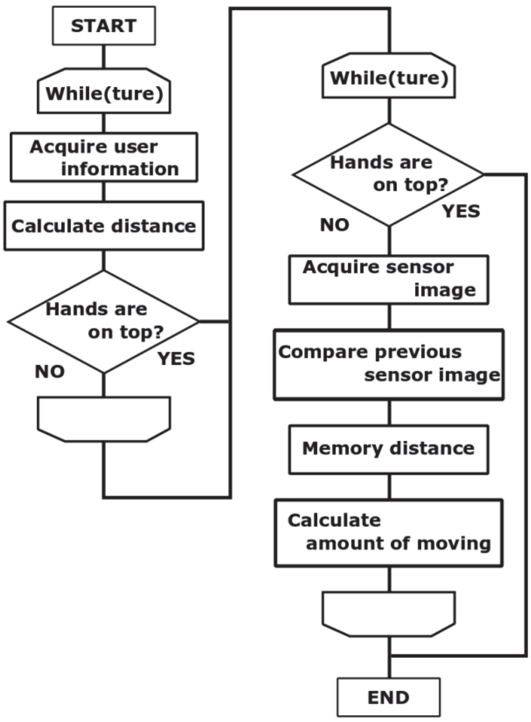


Fig. 3 Flowchart of a person tracking algorithm.

2.3 Sound transmission

Sound transmission is realized by controlling a parametric speaker and a turntable, which can transmit sound only in a single direction. A speaker and turntable are controlled by wireless communication, because the directional volumetric display is a large system and consideration of exhibiting. Two Arduino Uno microcontrollers, which is equipped with an XBee S2C module (Digi International K.K., Japan) and the Zigbee communication standard are used to control the turntable. A parametric speakers is used Clarielle JDS-U48 (Global Alliance Co., Ltd, Japan), which has a beam angle of $-10° \leq \theta \leq 10°$. Overview of sound transmission is shown in Fig. 5.

First, transfer an angle information which is acquired by Kinect to Arduino which is connected with PC(Personal Computer) via serial communication, and then transfer it to another Arduino which is connected with turntable via wireless communication. After that, sound transmission is performed using the received information to rotate a stepping motor connected for the turntable to the person of tracking position. The rotation angle (θ_{diff}) of a turntable is obtained the difference between the angle ($\theta_{current}$) in the current frame image and the angle ($\theta_{previous}$) in the previous frame image by Equation (3), where K_p is proportional gain.

$$\boldsymbol{\theta_{diff} = K_p(\theta_{previous} - \theta_{current})} \quad (3)$$

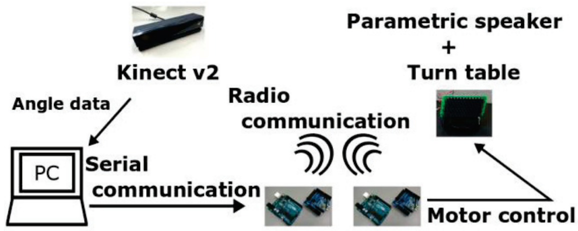


Fig. 5 Overview of sound transmission.

2.4 System configuration

The directional volumetric display used in this study is constructed as shown in Fig. 6. Specification of development system is shown in Table 2. The directional volumetric display is 95 cm in length, 95 cm in width, 187 cm in height, and the number of strings is 359. The projector that projects a projection image is located at a distance of 105 cm and a height of 70 cm from this display. The Kinect, parametric speaker, and turntable are located in the center of the display rotated $\theta = 90°$ counterclockwise from the front direction $\theta = 0°$. The directional volumetric display creates projection images based on two different images and displays different images in the front direction $\theta = 0°$ and the side direction $\theta = 90°$. In addition, the directional volumetric display follows images and sounds only in the side direction in synchronization with the movement of person tracking by Kinect.

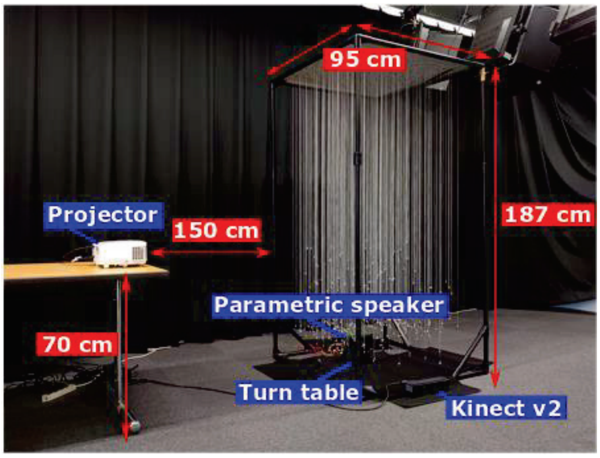


Fig. 6 Appearance of development system.

Table 2 Specification of development system.

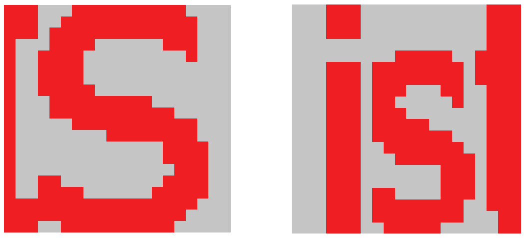
Feature	Specification
Video direction of front	$\boldsymbol{\theta = 0}$ [°]
Video direction of side	$\boldsymbol{60 \leq \theta \leq 130}$ [°]
$\mathbf{Height \times Width \times Depth}$	$\mathbf{187 \times 95 \times 95}$ [cm]
Number of threads	359
Range of person tracking	$\boldsymbol{60 \leq \theta \leq 130}$ [°]
Directional angle of sound	$\boldsymbol{-10 \leq \theta \leq 10}$ [°]

3 RESULTS

The projection image displays “THE SIDE IS THE FRONT.” in the front direction ($\theta = 0°$) and “Diese Seite ist die Seite.” in the side direction ($60° \leq \theta \leq 130°$). In the side direction, the same German sound as the projection image was output from the speaker. Original images in the front and side directions are shown in Fig. 7. The person of tracking target moved from $\theta = 90°$ to $\theta = 60°$ starting from $\theta = 90°$, and then moved from $\theta = 60°$ to $\theta = 120°$ through $\theta = 90°$.

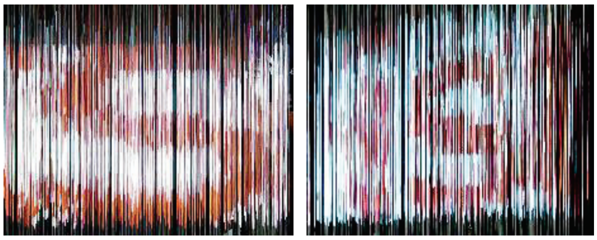
The results before the person tracking is shown Fig. 8. The frame rate was approximately 6.4 frames per second. The results indicated that English character strings in the front direction (a) and German character strings in the side direction (b) were confirmed, but the original images used could not be confirmed from any other direction. Therefore, the development system was able to display images with directivity to the position of the front direction ($\theta = 0°$) and the side direction ($\theta = 90°$). The results during the person tracking is shown Fig. 9. The results indicated that German character strings at the tracking position (a) and tracking position (b) could be confirmed, and the original image in the side direction used could not be confirmed from any other side directions. Hence, the development system was able to display images with directivity to the person of tracking position in the side direction. For this reason, the development system can follow images with directivity in the front direction ($\theta = 0°$), and the side direction ($60° \leq \theta \leq 130°$) which is the person of tracking position.

Sound waveform of side direction ($\theta = 90°$) is shown in Fig. 10. The vertical axis is shown sound volume, and the horizontal axis is shown time. The result indicated that the volume was large when the person of tracking target was in the side direction ($\theta = 90°$), and the volume was small when the person of tracking target was in $\theta = 60°$ and $\theta = 120°$. In addition, the result indicated that the volume has changed due to passing through $\theta = 90°$, which is the sound measurement position, when the person of tracking target moved from $\theta = 60°$ to $\theta = 120°$. Accordingly, the development system can follow sound with directivity to the person of tracking target.



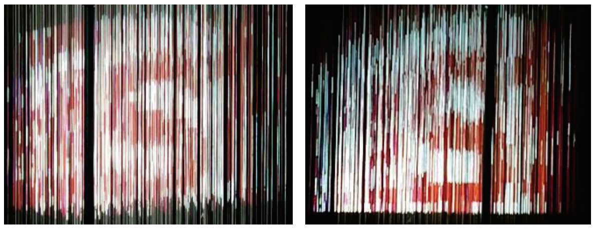
(a) **Front image.** **(b) Side image.**

Fig. 7 Original input images.



(a) Front image($\boldsymbol{\theta = 0°}$). (b) Side image($\boldsymbol{\theta = 90°}$).

Fig. 8 The results before the person tracking.



(a) Side image ($\boldsymbol{\theta = 60°}$). (b) Side image ($\boldsymbol{\theta = 120°}$).

Fig. 9 The results during the person tracking.

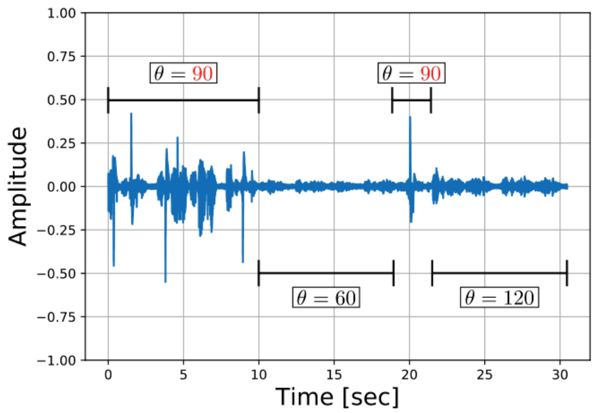


Fig. 10 Sound waveform of side direction ($\boldsymbol{\theta = 90°}$).

4 DISCUSSION

The directional volumetric display is possible to transmit images and sounds with directivity in the front direction $\theta = 0°$ and the side direction $60° \leq \theta \leq 130°$.

There are points for improving in this system. One is improvement of processing efficiency by parallel computation using GPU(Graphics Processing Units) because the calculation speed of the projection image is slow and the frame rate is low due to the CPU(Central Processing Units) calculation. In addition, the proposed image quality improvement algorithm [5] is applied because of the improvement of processing efficiency. The other is to arrange multiple speakers on the ceiling in the directional volumetric display, and select the speaker transmitting sound according to a person of tracking position. Furthermore, we consider tracking more than one person using an omnidirectional camera. Overview of the directional volumetric display using omnidirectional camera is shown in Fig. 11.

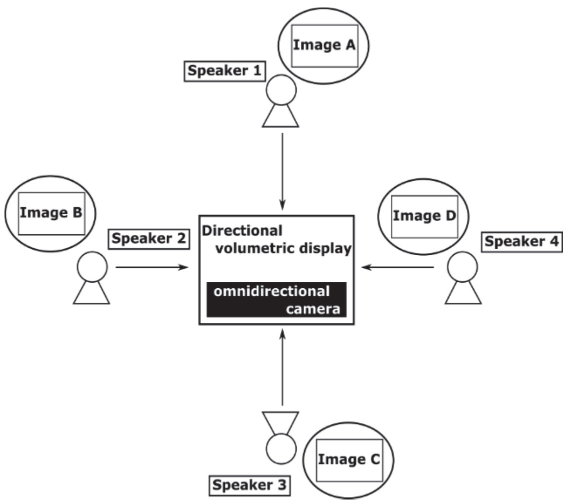


Fig. 11 Overview of the directional volumetric display using omnidirectional camera and multiple speakers.

5 CONCLUSIONS

In this study, for the purpose of applying the directional volumetric display to media application, we proposed and implemented a method of person tracking for the directional volumetric display to enable transmitting images and sounds following person using motion capture. As a result, the development system can transmit images and sounds only to the person of tracking target, and is possible to use as a media application. Further studies are required to implement real-time processing of voxel values calculation by GPU, and transmit images and sounds from multiple directions.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 18K11599.

REFERENCES

[1] Tao Zhan, Yun-Han Lee, and Shin-Tson Wu, "High-resolution additive light field near-eye display by switchable Pancharatnam–Berry phase lenses", OPTICS EXPRESS, Vol. 26, Issue 4, pp.4863-4872 (2018).

[2] Hong Hua, and Bahram Javidi, "A 3D integral imaging optical see-through head mounted display", OPTICS EXPRESS, Vol. 22, Issue 11, pp.13484-13491 (2014).

[3] H.Nakayama, A,Shiraki, R.Hirayama, H.Matsuda, T.Shimobaba and T.Ito, "Three-dimensional volume containing multiple two-demensional information patterns", Scientific Reports, 3, Article number 1931, pp.1-5 (2013).

[4] A.Shiraki, M.Ikeda, H.Nakayama, R.Hirayama, T.Kakue, T.Shimobaba, T.Ito, "Efficient method for fabricating a directional volumetric display using strings displaying multiple images", Applied Optics, Vol. 57, Issue 1, pp. A33-A38 (2018).

[5] Hirayama R, Nakayama H, Shiraki A, Kakue T, Shimobaba T, Ito T, T.Shimobaba, T.Ito, "Image quality improvement for a 3D structure exhibiting multiple 2D patterns and its implementation.", Opt Express, Vol. 24, Issue 7, pp. 7319-7327 (2016).