Development of Volumetric Display Capable of Transmitting Information in Different Languages Using Language Identification

Taishin Murase, Ryuji Hirayama*,, Naoto Hoshikawa***, Hirotaka Nakayama****, Tomoyoshi Shimobaba, Tomoyoshi Ito, Atsushi Shiraki**

Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan.
* Research Fellow of the Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku Tokyo 102-0083, Japan.
** Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo 125-8585, Japan.
*** National Institute of Technology, Oyama College, 771 Nakakuki, Oyama, Tochigi 323-0806, Japan.
**** National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo 181-8588, Japan.
Keywords: Directional volumetric display, Multilingual signage, Language identification.

ABSTRACT

We developed a directional volumetric display that exhibits different images depending on the viewing direction. The display can be expected to be applied to multilingual signage that transmits information in different languages. In this study, we develop a display that exhibits images according to the language used by the observer.

1 INTRODUCTION

In previous studies, we developed directional volumetric displays that exhibit different images depending on the viewing direction [1][2]. Figure 1 and Fig. 2 show the results of those volumetric displays. Figure 1 is a display using a 3D crystal and Fig. 2 is a display using threads and a projector. One can see that different images are exhibited depending on the viewing direction on each display. However, the display using 3D crystal records a pattern directly on the 3D display, so only fixed still images can be exhibited. On the other hand, the display using threads and a projector can express moving images, but it can only exhibit images with a low resolution of 20x20 pixels. However, according to the result of Fig. 2, it is thought that it is sufficient to express characters and symbols.

A display that can exhibit different images depending on the viewing direction can be expected to be applied to multilingual signage that transmits information in different languages as shown in Fig. 3. A display that exhibits images according to the language used by the observer can be expected to have high information transmission capabilities. But to do so, it is necessary to identify the language used by the observer.

On the other hand, speech recognition technology that can identify the language from human voice has been developed in recent years [3][4]. By using this technology, it is possible to input voice data for a few seconds and identify the language of the voice data. Therefore, in this study, we aim to develop a thread volumetric display that can change the image exhibited on the display according to the language of the observer.

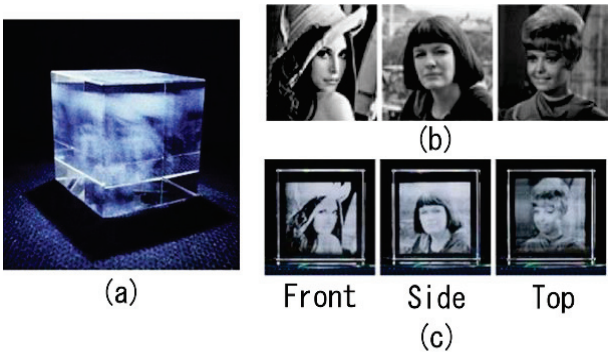

Fig. 1 3D crystal volumetric display.
(a) General view,
(b) Original images,
(c) Crystal seen each direction.

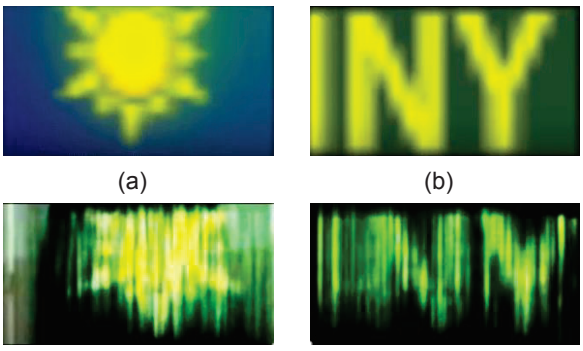

Fig. 2 Directional volumetric display.
(a) Original image viewing from the front,
(b) Original image viewing from the side,
(c) Projection result from the front,
(d) Projection result from the front.

2 METHOD

In this section, we describe the overview of volumetric display to be developed in subsection 2.1 and the language identification technology used in this study.

2.1 Volumetric Display Overview

The thread volumetric display that we have developed so far has succeeded in exhibiting images in any direction. However, the display developed in this study is considered to exhibit different images on the front and side.

Figure 4 shows a schematic diagram of the display to be developed. In the example of Fig. 4, information is initially exhibited in Japanese from the front of the display and in English from the side. Microphones are installed on the front and side of the display, respectively. By inputting voice data in a predetermined language from the microphone, information corresponding to the language of the voice data is exhibited. In the example of Fig. 4, Korean voice data is input from the front microphone, and information exhibited on the front display is switched from Japanese to Korean. At this time, the input voice data is language-identified, and the image exhibited from the projector is changed according to the language. The input voice data corresponds to any word in each language. As described above, a display that changes the exhibited images according to the input language by using language identification is developed.

2.2 Language Identification

In this study, Cloud Speech-to-Text provided by Google is used for the language identification [5]. Cloud Speech-to-Text is a powerful speech recognition API that uses the latest deep learning neural network algorithms. One of the functions of Cloud Speech-to-Text is a language identification function that uses voice data as input and automatically identifies the language. In this study, this function is used to identify the language used by the observer.

3 RESULTS

In this section, we describe the overview of the actual thread volumetric display used for study in subsection 3.1, Cloud Speech-to-Text parameters in subsection 3.2, and the study results in subsection 3.3.

3.1 Volumetric Display Used for Study

Figure 5 shows the thread volumetric display used in this study. The thread volumetric display that can exhibit images of 20 x 20 pixels was created by arranging 20 x 20 threads in an 80 cm square. However, 359 threads could be actually placed due to the placement constraint. In addition, microphones for voice input were installed on the front and side of the display.

3.2 Cloud Speech-to-Text

Cloud Speech-to-Text has several parameters that should be set for language identification. This section describes two important parameters in this study. The first is setting the language to be identified. For language identification, it is necessary to set the language to be identified in advance. This time, it was set to be identified in three types: Japanese, Korean, and English. The second is the number of seconds of input voice data. The longer the time, the higher the accuracy, but it was set to 2 seconds because it was judged that the identification accuracy was sufficient even in a short time.

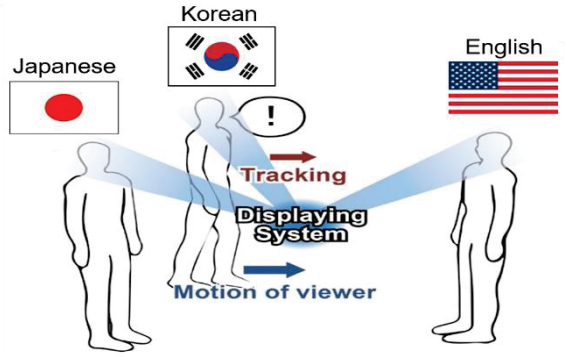

Fig. 3Directional volumetric display.

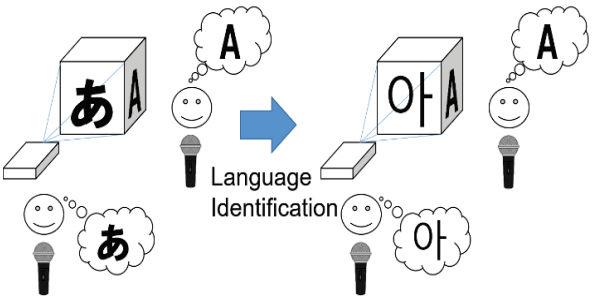

Fig. 4 Volumetric display overview.

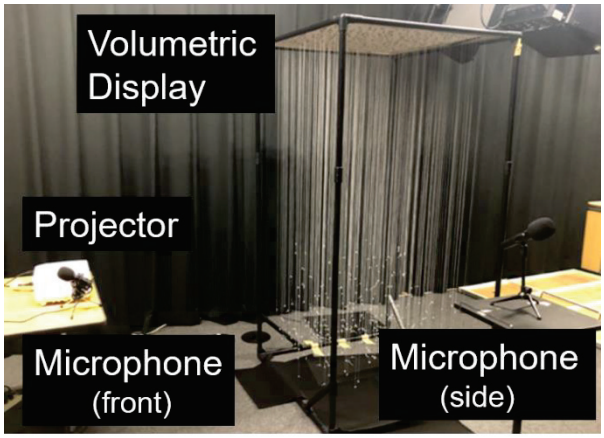

Fig. 5 Volumetric display used in study.

(a) (b) (c)

Fig. 6 Original images.

3.3 Implementation Results

Figure 6 shows the original images for each language used for exhibit on the volumetric display. (a), (b) and (c)

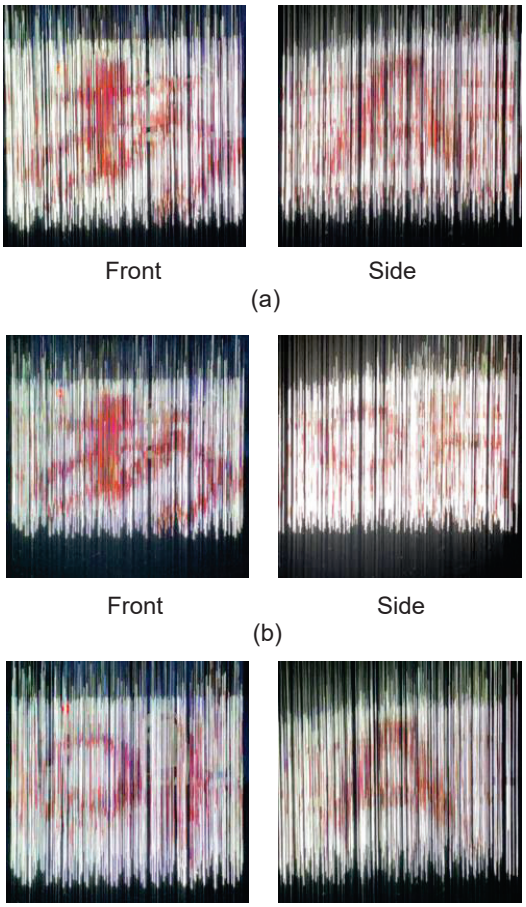

Fig. 7 Implementation results.

in Fig. 6 are Japanese, English and Korean, respectively.

Figure 7 shows the results exhibited from the thread volumetric display. Figure 7 (a) shows that the original Japanese image is exhibited on the front and the original English image on the side. From this situation, when any Korean was input to the side microphone, the result shown in Fig. 7 (b) was obtained. One can see that the original Japanese image is exhibited from the front and the original Korean image in the side. Furthermore, when inputting any Korean from the front microphone and any English from the side microphone, the result of Fig. 7 (c) was obtained. One can see that the original Korean image is exhibited from the front and the original English image in the side.

4 CONCLUSIONS

The purpose of this study was to develop a thread volumetric display that can change the image exhibited on the display according to the language used by the observer. By using language identification, we have succeeded in developing a display that fulfills that purpose.

In this study, we developed a thread volumetric display that supports only the front and side directions. As future work, we aim to develop a volumetric display that can track a person in real time and input voice data from any directions. In addition, in order to improve information transmission ability, we aim to develop a display that can exhibit higher resolution images.

ACKNOWLEDGEMENT

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

REFERENCES

[1] H. Nakayama, A. Shiraki, R. Hirayama, N. Matsuda, T. Shimobaba, T. Ito, “Three-dimensional volume containing multiple two-dimensional information patterns,” Scientific Reports, vol.3, article number 1931, pp.1–5, Jun. 2013.

[2] 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, No. 1, pp. A33-A38, (2018.01). doi.org/10.1364/AO.57.000A33.

[3] D. Snyder, D. Garcia-Romero, A. McCree, G. Sell, D. Povey, S. Khudanpur, “Spoken language recognition using x-vectors”, Proc. Odyssey: The Speaker and Language Recognition Workshop, pp. 105-111, 2018.

[4] L. Wan, P. Sridhar, Y. Yu, Q. Wang, I. L. Moreno, “Tuplemax loss for language identification”, in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 5976-5980.

[5] “Cloud Speech-to-Text – Speech Recognition | Cloud Speech-to-Text | Google Cloud “, Google, 19 Aug 2019.