Vision-based Interaction with Aerial 3D Display

Hayato Watanabe, Hiroyo Ishikawa, Satoshi Aoki, Seiji Suzuki, Hideo Saito Dept. Information and Computer Science Keio University Yokohama, Japan

Paper No.4-4

{hayato, hiroyo, aoshi, suzuki, saito}@hvrl.ics.keio.ac.jp

Abstract—This paper presents vision-based interaction for the aerial 3D display that can generate illuminating dots sequence in a volumetric screen by laser scanning device. The 3D contents shown in the volumetric screen can be interactively changed and controlled according to the user's action, which is captured with cameras. As an example of the interactive system, we developed "Jan-Ken Game (rock paper scissors)" system.

Keywords; 3D display, behavior recognition, computer vision

I. INTRODUCTION

Nowadays, the technology of the 3D display which shows 3D images attracts attention. To display 3D visual information, some physiological factors, such as the binocular disparity, the degree of convergence, the distance of the focus, and the monocular motion disparity, need to be recognized by human. In order to implement these factors, 3D display is being developed by using various approaches [1].

Most of the 3D displays reported until now draw pseudo-3D images on 2D planes [2]. The 3D display which uses the flat panel display is one of these pseudo-3D displays. Some of 3D flat panel display needs 3D glasses, and others do not. By using 3D glasses, the different images can be shown to two eyes, which make the viewers feel 3D perception. Even without such 3D glasses, the viewers can see different images for two eyes by using the barrier parallax or reticular lens, which is attached onto the surface of the flat panel display. Computer-generated holography is other type of 3D display. The holographic display [3] and the display which uses the integral imaging [4] method are this type of 3D display. The advantage of these 3D displays is that we can see high resolution stereoscopic images, and the disadvantage is the 3D viewing appearance is limited and we may get a feeling of exhaustion.

The volumetric display is other type of the 3D display [5]. This display draws images in the space so it has the advantage that we can implement physiological factors easily. Light dots or plane images are drawn in space while the driveline moves and rotates the screen. We can get the stereoscopic effects because of the persistence of vision. We can see the stereoscopic images in the 3D display which projects the images with the screen rotating at the high speed [6].

Because these pseudo-3D displays just show a number of different 2D images that depend on viewing direction, 3D viewing appearance is actually limited. To overcome such limitations, we have developed the Aerial 3D Display that

Satoru Shimada, Tatsumi Kimura Masayuki Kakehata, Yuji Tsukada Photonics Research Institute AIST Tsukuba, Japan satoru-shimada@aist.go.jp Hidei Kimura

Aerial Systems Inc. Burton Inc. Kawasaki, Japan kimura@burton-jp.com

generates real illuminating dots at arbitrary positions using laser-plasma phenomenon, so that viewers can see 3D images in the aerial true 3D space [7, 8]. Using the similar technology, we have also developed a Desktop 3D Display System as shown in Figure 1. In those display systems, the object can be observed from omni directional angle - 360° view in both horizontal and vertical directions.

For the Aerial 3D Display, we need to develop an interactive display controlling functionality for practical uses such as entertainment, advertisement, 3D object design, etc. To achieve the interactive functionality, we need to recognize human behavior that affects the contents displayed in the Aerial 3D Display. We employ a depth camera [9] for capturing the human behavior. The camera can capture the silhouette and the shape of the human hand in real-time without any influence by the illumination and the background of the environment, so that the interaction via the human behavior can be robustly achieved.

We developed "Jan-Ken Game (rock paper scissors)" system using the depth camera and the Aerial 3D Display. This system can automatically synchronize with the human interaction by recognizing the behavior of the human. It can also win the game always by deciding rock, paper or scissors based on the recognized shape of the human.



Figure 1. Desktop 3D Display. The green dot can be diplayed at arbitrary position in the 3D screen. An example of "Jan-Ken Game (rock paper scissors)" scene is also shown in the right image.

This work has been supported by "Foundation of Technology Supporting the Creation of Digital Media Contents" project (CREST, JST), Japan.

II. AERIAL 3D DISPLAY

A. The drawing mechanism

The Aerial 3D Display is composed of a high power infrared pulse laser and a 3D scanner as shown in Figure 2. The laser frequency of the mid-air type Aerial 3D Display is 1 kHz while that of the desktop type Aerial 3D Display is 50 kHz. In the result, the mid-air type Aerial 3D Display can draw one thousand dots in one second, while the desktop type can draw fifty thousand dots in one second. When the laser is focused, the luminous body is generated. Though emission phenomenon's time is very short, we can see a dot in 0.1 - 0.2 second because of the persistence of vision. The 3D scanner controls the direction of the laser and the distance of the focus. The Aerial 3D Display draws the 3D contents by vector scanning.



Figure 2. The drawing mechanism. 3D scanner controlls the direction of the laser and the distance of focus. We can display dots at the arbitrary position.

B. 3D content data

A 3D content data is a point sequence with 3D positions. We adopt for vector scanning so we need to create "one-stroke drawing" content data. To create such one-stroke drawing 3D contents for the Aerial 3D Display, we calculate 3D positions of a point sequence mathematically by programming [10, 11].

A 3D content data is sent to 3D scanner through DA-Converter and then we can see a 3D content by the Aerial 3D Display.

C. Example of 3D contents

We show the example of 3D contents in the desktop type Aerial 3D Display. This is the content of shape of "Tornado" which was created by rolling the spiral orbit. Unlike the ordinary pseudo-3D, this image can be seen from omni directional angle. Accordingly, it can be also seen from vertical direction. The image as shown in Figure 3 (b) is captured from the upper side. The center of the tornado can be observed.



(a) Different views within horizontal direction for "Tornado."



(b) "Torado" observed from upper viewpoint.





(c) "Sphere"



(e) The letters of "3D" looked from the front side (upper left), the upper side (upper right) and the lateral (lower left and lower right).

Figure 3. Examples of 3D contents shown with the desktop type 3D display.

Also, the image as shown in Figure 3 (c) is the contents of the sphere. If we add the brightness data to this content, the image as shown in Figure 3 (d) can be created. This is the content of the globe showing the Eurasia continent and the Pacific Ocean. The fact that South America can be seen through proves that the image is displayed in the aerial true 3D space not using the plane image method.

Lastly, we show the content of letters. This is the content that two letters of "3" and "D" are orthogonally displayed as shown in Figure 3 (e). Looking from the front side, the letters of "3D" can be visually recognized. By looking from the lateral or the upper side, we can also recognize that two letters are displayed on different planes in 3D space.

These images of contents prove that the Aerial 3D Display is the genuine 3D display unlike the other types.

III. Algorithm

In this chapter, we describe the method of the hand shape recognition of Jan-Ken Game and the interactive display controlling functionality of the Aerial 3D Display. Display content is changed according to the player's hand shape.



Figure 4. Depth images (top), silouette images (center) and vectors by principal component analysis (bottom).

A. Hand shape recognition

The hand shape is recognized by counting the finger number from the silhouette image as shown in Figure 4. We obtain the silhouette image by using the depth camera. The depth camera can scan distance between the target and the camera. It gives depth data as a point cloud with 3D positions to a user. This voxel data can be presented as a depth image. We extract the object at the arbitrary distance by setting thresholds to this depth image. In this paper, we assume that the hand is near by camera. We eliminate the far object. So we can extract the hand without any influence by the illumination and the background of the environment, for example, the situation that someone walks across behind player's hand.

To count the finger number, we apply for principal component analysis to the silhouette image. This method transforms a data into a small number of characteristic variables. We can get the arm vector and the arm vector's vertical vector from the result of principal component analysis. We draw several parallel lines against the arm vector's vertical vector and count the number of intersection between the lines and the silhouettes. We can assume that the line which has the greatest number of intersection intersects with the fingers. To recognize the hand shape, we use two lines which have the greatest number of intersection. The reason of using two lines is to decrease false detection when a line crosses Rock shape's arthros. If we use only one line, the Rock shape may be recognized as Scissors. Hand shape is recognized by intersection. In case the number of intersection is two, we assume that the hand shape is "Rock", and four is "Scissors" while over six is "Paper".

B. Interactive display controlling functionality

Display content is changed according to the recognized hand shape. We made three types of display contents, namely Rock, Scissors and Paper in advance from 3D polygon data as shown in Figure 5. To change display contents smoothly, we developed the controlling functionality. It saves each content data's controlling position and set the frame number. We can change the content by choosing the frame number. In Jan-Ken Game, the frame number is automatically changed according to player's hand shape. The displayed content always wins the game against the players as shown in Table 1.



Figure 5. 3D contents of hand shapes. Rock (left), Scissors (center), Paper (right).

TABLE I.	CORRESPONDENCE RELATION BETWEEN THE PLAYER'S
	HAND SHAPE AND DISPLAY CONTENTS

Player's hand shape	Display contents
Rock	Paper
Paper	Scissors
Scissors	Rock

IV. EXPERIMENTS

We played Jan-Ken Game with the Desktop 3D display. We connected the depth camera and DA-converter to the laptop computer. Depth camera captured the player's hand. We set depth threshold of extraction from 0.0m to 0.5m so background information was eliminated. In this experiment, the player changed the hand shape randomly in one minute. We captured the movie of playing the game by camera. The Desktop 3D Display always won against the player as shown in Figure 6.

Case of failure happened when the player's hand was shaped inappropriately. In the case of Figure 7 (a) and (b), as the distance between a finger and another one was too near, we could not recognize the hand shape correctly. We recognized the hand shape by counting the number of intersections between the silhouettes and the vectors found with principal component analysis. The fingers touching together were recognized as one finger. Therefore, the hand shape of Figure 7 (a) was recognized as Scissors, and that of Figure 7 (b) was recognized as Rock. In the case of Figure 7 (c), the direction of the finger was horizontal so we could not count the finger number. Therefore, that hand shape was also recognized as Rock. We need to search solutions of these problems.

V. DISCUSSIONS AND CONCLUSION

In this paper, as vision-based interaction, we presented Jan-Ken Game which changes display contents according to the player's hand shape. By using the depth camera, we could recognize the player's hand shape without any influence by the illumination and the background of the environment.

We recognized a shape of the hand from a silhouette image so we could not recognize an inappropriate hand shape as shown in Figure 7. One solution of these problems is to recognize hand shape by using not only a silhouette image but also a depth image. Utilizing the depth data of the depth image and matching the 3D hand shape data with Rock, Paper and Scissors, we may be able to solve this problem. Other solution to get rid of the problems of Figure 7 (c) would be to use more than two depth cameras because any direction of hand shape can be available by using a number of cameras.

By applying for the method in this paper, we can develop any interactive software other than Jan-Ken Game. In the future, we would like to display the captured object in real-time by using depth camera. In addition, we would like to seek the potentiality of the Aerial 3D Display.

REFERENCES

- J.-Y. Son, B.Javidi, and K.-D. Kwack, Methods for Displaying Three-Dimensional Images, Proc. IEEE, Vol. 94, 502-523, 2006.
- [2] Shunsuke Yoshida, Roberto Lopez-Gulliver, Sumio Yano, Naomi Inoue, A Study to Realize a Box-shaped 3D Display: A Calibration Method to Align Lens Array and Display, 3DTV-CON'08, pp. 169-172, Istanbul, Turkey, 2008.
- [3] Gabor, A New Microscopic Principle, Nature, Vol 161, 777-778, 1948
- [4] M.G.Lippmann, Epreuves reversibles donnant la sensation du relief, J. de Phys, Vol 7, 821-825, 1908

- [5] Favalora E.G., Volumetric 3D Displays and Application Infrastructure, Computer, 37-44, 2006
- [6] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, Paul Debevec, Rendering for an Interactive 360 Light Field Display, ACM SIGGRAPH 2007 Emerging Technologies, 2007.
- [7] Hidei Kimura, Taro Uchiyama, Satoru Shimada, True 3D Display Using Laser Plasma in the Air, ACM SIGGRAPH 2006 Emerging Technologies, 2006.
- [8] Hideo Saito, et al., Laser-plasma scanning 3D display for putting digital contents in free space, Proc. SPIE Vol. 6803, pp. 680309, 2008.
- [9] Thierry Oggier, et al., An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger), Proc. SPIE, Vol. 5249, 534, 2004.
- [10] Hiroyo Ishikawa, Hideo Saito, Point Cloud Representation of 3D Shape for Laser-Plasma Scanning 3D DisplayThe 34th Annual Conference of the IEEE Industrial Electronics Society, pp.1913-1918, Orlando, Norvember, 2008.
- [11] Hiroyo Ishikawa, Hideo Saito, Billboard representation for Laser-Plasma Scanning 3D Display, Proceedings of Asiagraph 2009, Vol. 3, No.9, pp.41-46, Oct., 2009.



Figure 6. Images captured at 10 seconds (upper left), 15 seconds (center), 20 seconds (right), 30 seconds (lower left), 45 seconds (center) and 50 seconds (right) in the movie.



Figure 7. Case of failure. In the case of (a) and (b), fingers touched together. In the case of (c), the direction of fingers was horizontal.