Spatial Reality Enhancement Using Eye-sensing Light Field Display

Yuji Nakahata, Tomoya Yano, Kazuki Yokoyama, and Koji Aoyama

Yuji.Nakahata@sony.com
Sony Group Corporation, R&D Center, Atsugi, Kanagawa, Japan

Keywords: Light field display, Autostereoscopic display, Naked-eye 3D display, High-speed camera, Motion parallax image

ABSTRACT

Eye-sensing light field display (ELFD) was developed for the enhancing spatial reality. A comfortable and overwhelming sense of reality can be presented by consideration of human perceptual characteristics such as real-time expressions of motion parallax, accurate binocular parallax images, and clues for binocular fusion.

1 Introduction

In recent years, the advancement of the CG-image production technology and the development of the volumetric capturing technology has made a wide variety of 3D image data readily available. In addition, the data transmission technology has also advanced, and the infrastructures that can instantly send enormous amounts of 3D image data have been established. On the other hand, even if we have enough image data to give the sense of reality, there are no adequate displays to show it. When 3D-TVs were developed and commercialized, 3D displays based on only binocular parallax did not provide a sufficient sense of reality.

Autostereoscopic displays have been developed in various forms [1, 2, 3]. We also research the Ray-modeler [4] and the 360degree Cylindrical Transparent Display [5] displays, which tried to express a sense of reality. These devices were developed to display stereoscopic images effectively considering motion parallax or the effect of transparency. This article describes the eye-sensing light field display (ELFD), which we developed and exhibited at CES 2020 [6, 7] as shown in Figure 1. The ELFD has the unique optical and rendering technology based on eye sensing. In addition, we succeeded in imparting a very high sense of reality by providing functions based on the physiological elements of stereoscopic vision.

Figure 1. ELFD prototype showcased at CES 2020

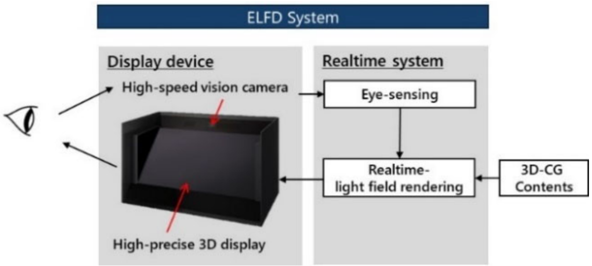


Figure 2. Overview of the ELFD system configuration

2 System Overview

The outline of the newly developed display system is described below. We attempted to realize a high-resolution 3D display by limiting the number of users to one. The user's eye position was recognized by a high-speed camera placed at the top of the display, and the image aligned with the position was rendered in real time. An LCD was used for the display, and the lenticular lens were laminated to the front of the LCD. The image to be output to the LCD was rearranged according to this lens in conjunction with the above-mentioned real-time rendering.

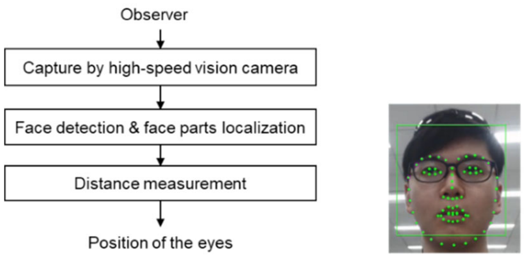
Besides, the display was installed diagonally to account for physiological factors. Both the pedestal-like structure and the triangular structure are placed along the display to improve stereoscopic visibility.

3 Approach

Our goal was to create a 3D display that has sense of reality and comfortableness. To achieve this, it is necessary to accurately reproduce motion parallax in real time, as well as to enhance the visibility of binocular disparity images. It is also important to provide users with clues to encourage the fusion of the binocular images. The following is a detailed description of our efforts to achieve these goals.

3.1 Real-time representation of motion parallax

Motion parallax as a stereoscopic representation was not reproduced on previously general 3D-TV. When the viewing position is moved on a display with binocular parallax and no motion parallax, a sense of discomfort occurs as if the image is following the viewer. This is caused by the discrepancy between the viewer's movement and the visible image. We generated a natural stereoscopic vision by creating motion parallax. This was done by sensing the user's eyes with a high-



(a) Block diagram (b) Example of face detection and face parts localization

Figure 3. The eye-sensing system

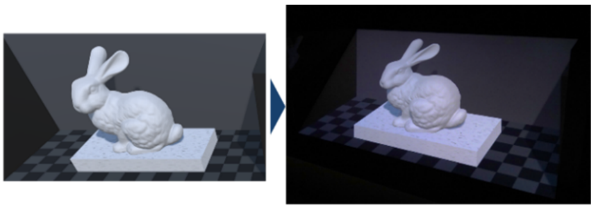
speed camera and rendering images according to the eye position in real time.

First, the intensity of the light that reaches both eyes was calculated from the 3D Object placed in the display device. It was projected onto the autostereoscopic display surface installed as a source for that light. Next, the pixel of the light displayed to either the left or right eye was calculated and rendered in real time. Figure 3 (a) shows the block diagram of the system we developed to achieve this real time display method. The eye-sensing system uses a high-speed vision camera and highly accurate face parts localization technology. The face parts recognition technology was trained and optimized using the deep neural network technology. An example of the result is shown in Figure 3 (b).

Furthermore, the image on the display surface is a kind of *trompe l'oeil* illusion that is called Anamorphosis [8]. Figure4 (a) is an image taken from a position other than the viewer's position, and it can be seen that the image is distorted as compared with Figure4 (b) when taken from the viewer's position. This dynamic stereo anamorphosis technology, which dynamically updates the illusion according to the user's position, makes the user feel as if the object is "actually there".

3.2 Visibility of binocular parallax images

Image display by eye sensing considerably contributes to not only motion parallax but also accurate delivery of binocular parallax images to the user. The visibility of a binocular parallax image is an important factor for viewing a stereoscopic image. In the presentation of binocular parallax, the effect of crosstalk, in which the right and left images are mixed, is considerably strong. Minimizing this effect is an important key to realizing a display with a high sense of reality.

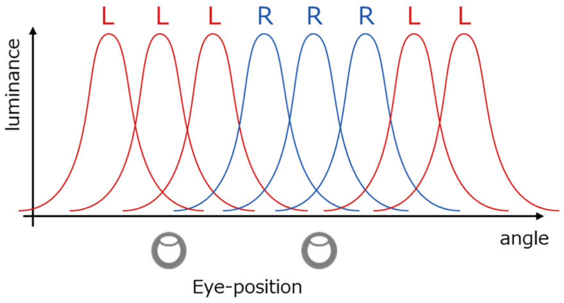


(a) screen shot of generated image (b) view from observer's position

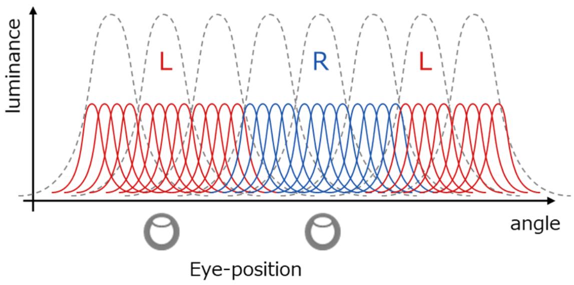
Figure 4. Effect of Dynamic Stereo Anamorphosis

We constructed an optical system based on the assumption that the eye position is clearly known by eye-sensing. Since the position of the user's eyes is known, it is possible to use a method of delivering correct viewpoint images for the right eye and the left eye respectively. This makes it possible to obtain a significantly higher resolution than the general method of presenting a large number of images for each viewing area. In addition, by providing sufficient pseudo viewpoints in the optical design and displaying images for the two viewpoints on them, it was possible to obtain a large margin for the positions of the left and right eyes. Thus, we succeeded in suppressing the generation of crosstalk in not only the state in which the user is stationary but also when the user moves and the positions of the left and right eyes change.

Figure 5 (a) shows a case where the light division corresponding to six viewpoints is generally assigned to the left and right eyes. In this case, the image cannot be updated unless the position of the user's eyes significantly moves. Since the image of the opposite eye leaks in during that time, the crosstalk would be observed. Furthermore, even when there is no eye movement, because the angle distribution per viewpoint is wide, the crosstalk is easy to occur. Figure 5 (b) shows the design of ELFD, and large number of pseudo viewpoints can be observed. This makes it possible to update the image even when the position of the viewer's eyes changes slightly, and the crosstalk can be suppressed. In addition, because the angle distribution per viewpoint is narrow and the number of viewpoints is sufficient, it is evident that it also greatly contributes to the reduction of crosstalk when there is no change of the eye position.



(a) Conventional method



(b) This method

Figure 5. Design of the viewpoint images

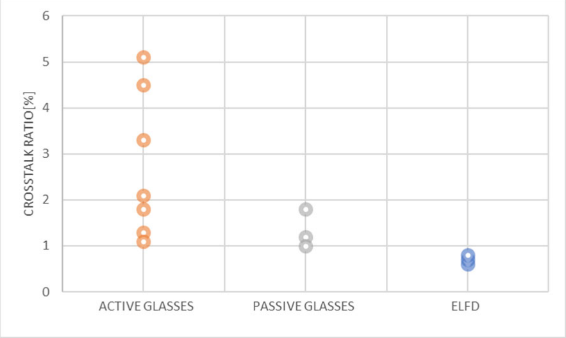


Figure 6. Crosstalk ratio

The measurement results of actual crosstalk are shown in Figure 6. Crosstalk is calculated based on Equation 1 below. Because the crosstalk performance is determined by two viewpoints in the end, it was compared with that of a 3D display using glasses which also reproduces two viewpoints [9].

$$X = \frac{L_{bw} - L_{bb}}{L_{ww} - L_{bb}} \qquad (1)$$

Here, X is the crosstalk value, L_{bw} is the brightness observed when black is displayed for the observing eye and white is displayed for the other eye; L_{ww} is the brightness observed when white is displayed for both eyes, and L_{bb} is the brightness observed when black is displayed for both eyes.

The 3D display with the active-type glasses showed crosstalk values ranging from 1.1% to 5.2%, and the 3D display with the passive-type glasses showed crosstalk values of 1.0%–1.8%. The data of the active-type glasses contains various types of LCDs such as 120Hz LCD and 240Hz LCD. Therefore, there is a variation in the crosstalk data values. In contrast, the ELFD had crosstalk values of 0.6%–0.8%. Thus, crosstalk performance equal to or higher than that of a glass-based 3D display can be achieved with an autostereoscopic 3D display without glasses.

3.3 Clues for Binocular Fusion

In order to provide users with a comfortable and realistic image experience, it is also important to consider physiological factors and to provide clues for binocular fusion. We worked to realize a comfortable and highly realistic display by incorporating the following two elements.

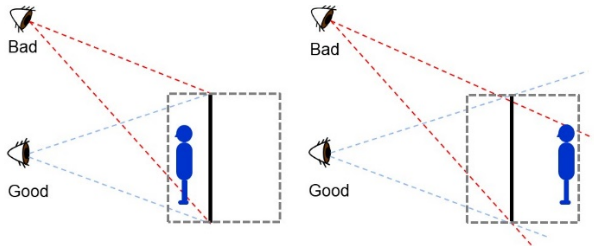
The first element is to arrange the display diagonally. Figure 7(a) shows a normal display layout, and Figure7(b) shows a diagonal arrangement we proposed. In the left figures, the object is displayed in front of the screen, and in the right figures, the object is displayed behind the screen. There are three reasons why diagonal placement makes binocular fusion easier.

First, the amount of parallax can be suppressed in the diagonal arrangement configuration. Since most of the observer's attention is focused on the image at the center of the screen, by positioning the screen diagonally, it is possible to suppress the amount of parallax actually seen by the observer. The suppression of the amount of parallax leads to the reduction of crosstalk and discomfort when viewing images with binocular parallax.

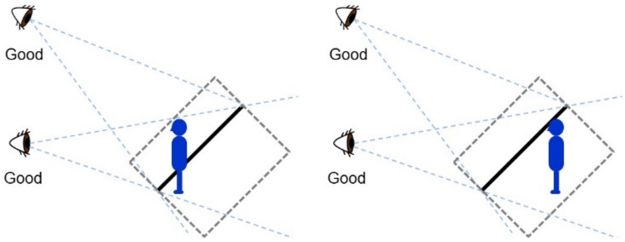
The second reason is the reduction of the frame effect. In the case of Figure7(a), when the observer sees the image from above, the image cut off occurs easily. This cut off makes it difficult for the observer to fuse the binocular image due to the frame effect, and thus the sense of existence is lost. On the other hand, when it is arranged diagonally as shown in Figure 7(b), the entire image can be seen even when the viewing position moves upward. Therefore, it is possible to maintain a state in which fused binocular image is easily seen by the observer.

The last one is continuous binocular fusion clues. As can be seen from Figure 7(a), in a normal vertical arrangement, when the object is fully displayed in front of or behind the screen, there is no part in binocular images without any parallax in the displayed image. Accordingly, the observer only sees the parallax image. On the other hand, in Figure 7(b), there always exists a part of binocular images having no parallax. This makes the observer's binocular fusion easier, because no parallax part becomes starting point for seeing the fused binocular image. The image with continuous parallax starting from no parallax part serves the continuous binocular fusion clues to the observer. When there is no parallax part in the displayed image, there is no starting point for seeing the fused binocular image and this makes binocular fusion harder for the observer.

The second element is the arrangement of structures that provide clues to fuse binocular images. As shown in Figure 8, we placed the pedestal-shaped structure A aligned with the edge of the display and the triangular structures B along the diagonally arranged display. The pedestal-shaped structure was arranged parallel to the position of the floor surface of the image. As a result, the floor position of the image seems to be connected to the real object, and the image seems to be localized in the real space. The floor surface of the image is



(a) Normal display configuration



(b) Proposed display configuration

Figure 7. Display configuration improving clues for binocular fusion

also an important criterion for depth perception. This depth information makes the binocular fusion easy for the observer when the depth information of the floor surface can be recognized accurately. The presence of the triangular structures around the display, i.e., at a position visible to the user, allows the user to recognize the depth position of the image presented by binocular parallax as compared to an object in real space. These structures trigger the user to fuse the images and provide the user with ease of binocular fusion. The diagonally placed display as the first element allows the triangle structures arrangement.

4 Discussion and Conclusions

Figure 9 shows an example of the viewing experience from various view positions. We could confirm that motion parallax could be reproduced by changing the viewpoint to the 3D object in the display according to the viewing position of the user.

We believe that an important element of a highly realistic display is that the user "does not feel the screen surface." We took measures to provide natural motion parallax and binocular parallax images with less crosstalk and the clues for binocular fusion, respectively, so that the screen would not feel like a screen. It is hard to express on the paper, we believe that users will be able to immerse themselves in a visual experience that as if what they were seeing is "actually there", without knowing where the screen is.

As mentioned above, herein, we focused on the development of an eye-sensing based 3D display technology and the improvement in the reality expression of spatial reproduction. As a result, we succeeded in enhancing some important factors for generating the stereoscopic effect. This ELFD made it possible for anyone to comfortably enjoy a space with a high sense of reality "as if it was there" and provide a system that gives an unprecedented video experience.

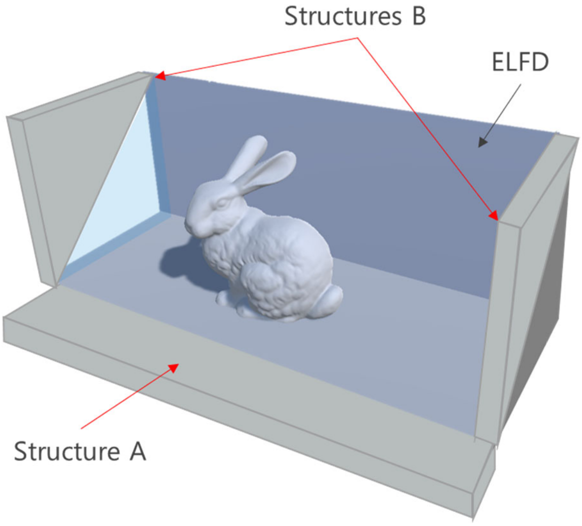
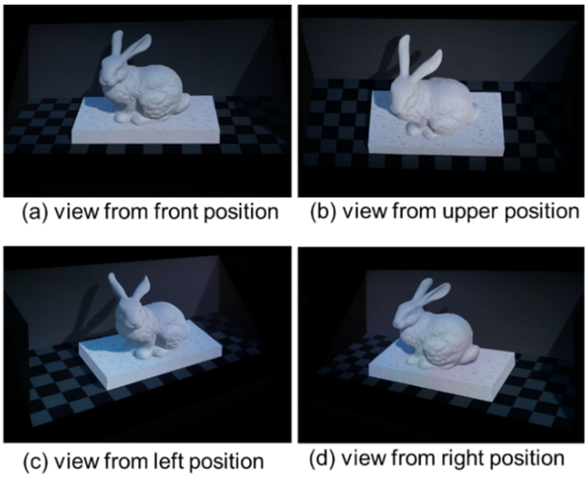


Figure 8. Structures for Clues of binocular fusion


(a) view from front position (b) view from upper position

(c) view from left position (d) view from right position

Figure 9. Viewing experience from various view positions

5 References

[1] Y.Takaki, "Multi-view 3-D display employing a flat-panel display with slanted pixel arrangement," *Journal of Society for Information Display*; 476-482 (2010).

[2] T. Koito *et al.*, "Newly Developed Light Field Display with Ultra-Wide Viewing Angle and High Resolution," *SID Digest*, 263-266 (2018).

[3] M. Makiguchi *et al.*, "Smooth Motion Parallax Glassless 3D Screen System Using Linear Blending of Viewing Zones and Spatially Imaged Iris Plane," *SID Digest,* 903–906 (2017).

[4] K. Ito *et al.*, "360-degree Autostereoscopic Display," *ACM SIGGRAPH 2010 Emerging Technologies*, *No. 1,* (2010).

[5] T. Nakamura *et al.*, "Cylindrical Transparent Display with Hologram Screen," *International Display Workshops,* 1178-1180 (2019).

[6] "Sony Highlights its Evolution as a "Creative Entertainment Company with a Solid Foundation of Technology" at CES 2020", *CES* (2020). https://www.sony.net/SonyInfo/News/Press/202001/20-001E/

[7] K.Aoyama, "Eye-sensing Light Field Display for Spatial Reality Reproduction", *SID Digest,* 669-672 (2021)

[8] Wikipedia. https://en.wikipedia.org/wiki/Anamorphosis

[9] Y. Takaki, "Three-Dimensional Display: Present and Future," *SID Display Week* (2009).