視線方向に応じた焦点可変画像による立体提示手法

Active Focus Control with Eye Gaze Detection for Depth Cognition

大谷慶一郎/東京大学, 吉田匠/東京大学大学院, 新居英明/東京大学大学院, 川上直樹/東京大学大学院, 舘暲/東京大学大学院
Keiichiro Otani, Takumi Yoshida, Hideaki Nii, Naoki Kawakami, and Susumu Tachi The University of Tokyo
t60518@mail.ecc.u-tokyo.ac.jp, takumi_yoshida@ipc.i.u-tokyo.ac.jp, hideaki_nii@ipc.i.u-tokyo.ac.jp,
kawakami@star.t.u-tokyo.ac.jp, tachi@star.t.u-tokyo.ac.jp

Abstract: This paper describes the display system which has an internal active focus control system with the camera for an eye gaze detection to enhance depth information.. In order to enhance the depth information of two-dimensional images, blur effect is often used in creating the CG image. However, in this situation, even if the observer gazes at unfocused objects, he cannot see the objects clearly. Hence, such the effect is static effect and different from the focal accommodation of eyes. We propose a prototype system for actively controlling the focus of an image on the basis of the point of gaze of the observer in order to simulate the effect of focal accommodation; our system can be mounted to the existing displays. By using this system, we conducted comparative experiments using an all-in-focus image and a controlled-focus image with regard to the gaze of the observer. The results indicated that we can improve the three-dimensional appearance of two-dimensional images.

Keywords: three-dimensional appearance, active focus control, eye gaze detection, depth-of-field blur

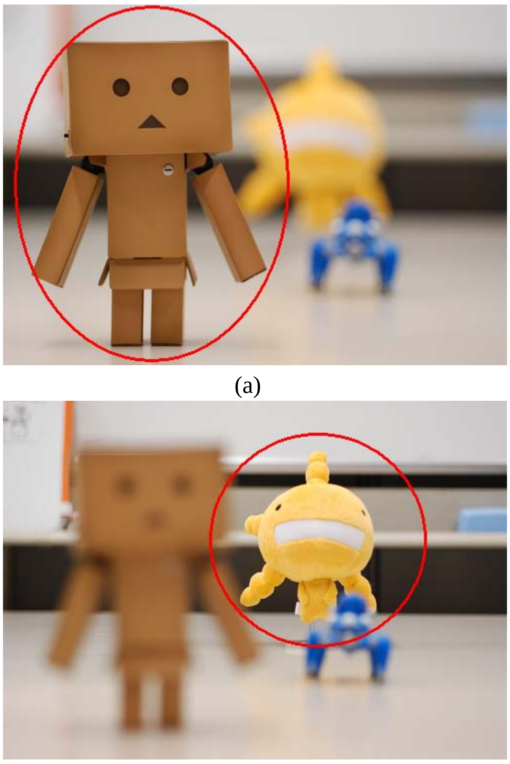
1. Introduction

It is possible to improve the sense of reality by enhancing the depth information of two-dimensional images. By using techniques to present two-dimensional images with high sense of reality, it is possible to design equipments for attractions in theme parks or transmit the visual information of a camera of a remote vehicle to its operator with high sense of reality. Thus, a technique for enhancing the depth information of two-dimensional images is required in various fields.

In the real world, various depth cues influence our depth cognition. The cues are classified into three categories [John01]—binocular depth cues such as binocular disparity and motion disparity, monocular depth cues such as convergence, and pictorial cues such as shade and linear perspective. There are many methods that use these depth cues to enhance the depth information of two-dimensional images. In this study, we focus on one of the pictorial cues—lens blur.

1.1. Related study

In order to enhance the depth information of two-dimensional images, blur effects is often used. For example, the method of rendering with regard to the depth-of-field is an established technique in CG image. Using this method, the creator attracts the attention of the audience to an object by emphasizing it by focusing on it and unfocusing the other areas. Potmesil and Chakravarty [Potmesil81] first proposed the simulation of an optical lens to simulate depth-of-field blur. Barsky [Barsky04] proposed another method for vision realistic rendering, which accurately simulates the image we perceive in the real world. The common point among these methods is that the focus point is constant and the position is dependent on the creator of the CG movie. On the other hand, when we see the real world, the objects near the point of gaze appear clear, and the objects in the other areas appear blurred, as shown in figure 1. This phenomenon is attributed to focal accommodation. Therefore, the conventional methods do not accurately simulate the effects of focal accommodation. Hillaire et al. [Hillaire08] improved the three-dimensional appearance of images by using the observer's focus point as input. The research shows that focusing actively on the point of gaze of the observer improves sensations and perception of the virtual environment. But the near objects from the observer which require the exact focal accommodation are unconsidered.



(a)

(b)

Figure 1: In the real world, the object we gaze at appears clear and the other objects appear blurry. When we gaze at the object in the front, the vision resembles that in (a). When we gaze at the object in the back, the vision resembles that in (b).

1.2. Purpose of this study

The purpose of this study is to develop the display which has internal active focus control system with the camera for eye gaze detection to enhance depth information.

By conducting comparative experiments using a prototype system, we studied whether the depth cognition of the two-dimensional images are improved by simulating the effect of focal accommodation. From the experiments, we conclude that our method may have great possibilities.

In Section 1, we present the background of this study, related studies, and purpose of our study. In Section 2, we present our proposed active focus control system. In Section 3, we describe in detail the techniques used in the system. In Section 4, we describe the comparative experiments conducted using the system. In Section 5, we provide the results; in Section 6, we discuss the results. In Section 7, we summarize this study and provide a conclusion.

2. The eye gaze detection and blur generation system

We propose the display which has internal active focus control system with the camera for eye gaze detection to enhance depth information. The system actively focuses the CG object on the point of gaze of the observer. By focusing actively on the point of gaze, the system accurately simulates the phenomenon wherein objects situated at the same distance as the depth of focal plane appear clear and the other objects appear blurry. Figure 2 shows the structure of the system. We describe the process flow.

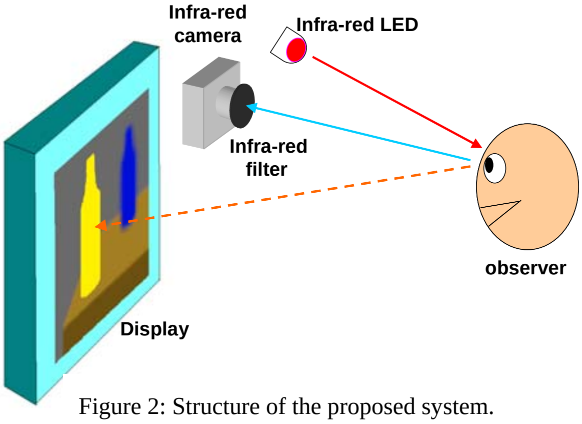


Figure 2: Structure of the proposed system.

Firstly, the eye gaze detection component actively measures the eye motion of the observer. The eye is illuminated by an infrared LED. The image of the eye is captured by an infrared camera through an infrared pass filter. Second, the image is transmitted to a computer, which calculates the point of gaze by using the pupil-center-corneal reflection method. After the point of gaze is computed, the depth of focal plane is calculated. Then, the blur is calculated and the blur per pixel is drawn on the CG image.

3. Implementation

In this section, we explain the eye gaze detection algorithm and blur generation algorithm.

3.1. Eye gaze detection

We describe an algorithm for detecting a user's eye gaze. There are various techniques to detect eye gaze, such as the EOG method, optical lever method, search coil method, and pupil-center-corneal reflection method [Yusuke04]. Each method has both merits and demerits. We adopt the pupil-center-corneal reflection method because it does not task the observer and it is easy to implement. The pupil-center-corneal reflection method involves the computation of the point of gaze of the observer by using the pupil position and the first Purkinje image. The first Purkinje image is formed from light reflected on the corneal; it corresponds to light borrowed from infrared LED. The algorithm for this technique is described in detail as follows.

First, we obtain the observer's pupil position. The infrared camera captures the image of the eye through the infrared filter. We convert the raw image into a binary image. Subsequently, the contour of the pupil is detected by searching the most likely contour length. Then, the contour is approximated to an ellipse by the least square method. The center of the ellipse is the observer's pupil position. Second, we obtain the observer's first Purkinje image position. Similarly, we convert the raw image, which is captured when obtaining the pupil position, into a binary image. Now, we detect not only the first Purkinje image but also the second Purkinje image. However, we use only the former to compute the point of gaze. Hence, we set the ROI (region of interest) around the pupil which do not include the second Purkinje image, and calculate the barycenter of ROI. The barycenter is the observer's first Purkinje image position. Figure 3 shows the pupil and the first Purkinje image obtained by the above process.

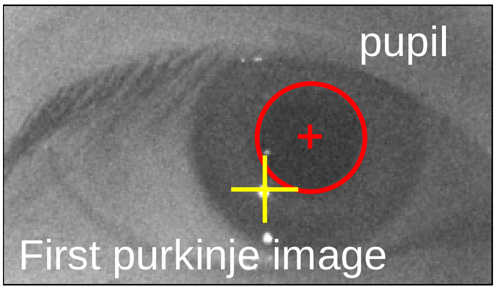


Figure 3: Pupil and first Purkinje image.

The point of gaze on the display is calculated by using the pupil position and the first Purkinje image position. By using a conversion matrix F, we represent the relationship between the three points, as shown in equation 1.

$\begin{bmatrix} u & v \end{bmatrix}^t$: The point of gaze position on the display

$\begin{bmatrix} x_1 & y_1 \end{bmatrix}^t$: The pupil position

$\begin{bmatrix} x_2 & y_2 \end{bmatrix}^t$: The first Purkinje image position

$$\begin{bmatrix}u\\v\end{bmatrix} = F\begin{bmatrix}x_1\\y_1\\x_2\\y_2\end{bmatrix} \quad \cdots(1)$$

Here, F represents the 2 × 4 conversion matrix. F has eight parameters; hence, we must calculate them after calibration of the system. The observer gazes in turns at the nine fixed grid points, and at the time the system detects the pupil position and the first Purkinje image position. Then the grid point positions, the pupil positions, and the first Purkinje image positions are substituted into equation 1. The least squares solutions of the simultaneous equation are the eight parameters of F. The calibration is required in each time the system gets started.

3.2. Scale of blur circle

We describe the lens blur effect in our eye. When an object is situated at a distance different from the distance of the focal plane, blurring appears in the image on the retina. First, we derive the formula for the scale of the blur. We consider the following assumptions.

1: As shown in figure 4, the crystal lens is a single converging lens and the retina is a single plane. They are parallel to one another [Rokita96].

2: The blur scale depends on the distances of the object from the crystal lens.

3: The radius of the crystal lens is equivalent to that of the pupil.

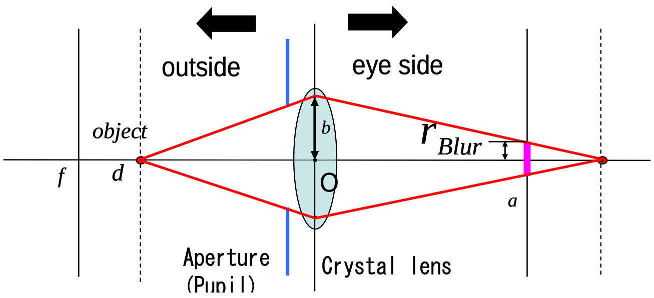


Figure 4: Eye model

According to the first assumption, the shape of the blur appearing on the retina is a circle. According to the second and third assumptions, geometrically, the radius of the blur circle r_{Blur} is expressed by the following formula. Here, f denotes the depth of focal plane; d, the distance of the object from the eye; a, the distance from the crystal lens to the retina; and b, the radius of the crystal lens.

$$r_{Blur} = a \cdot b\left|\frac{1}{f} - \frac{1}{d}\right| \quad \cdots(2)$$

3.3. Drawing the blur on the display

Based on equation (2), we draw the blur per pixel on the display. In this study, we apply blur to an image by averaging the color of the pixels. The pixels used for the averaging are the pixels which have their centers inside the blur circle. In case the blur appears around the pixel in $(i \quad j)$, as shown in figure 5, the color of the pixel is the average of the color of the shaded blue pixels.

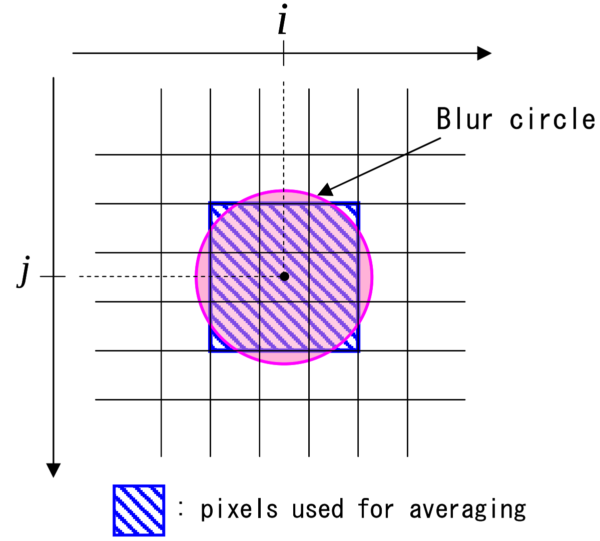


Figure 5: Drawing the blurred pixels in the display. The pixels of which center is inside the blur circle are used for averaging.

3.4. Prototype system

In this section, we describe the prototype system that we have implemented. Figure 6 shows the structure of the prototype system we have implemented.

The system consists of the following three components: an eye gaze detection component, a computer to create two-dimensional images with blurring at distances, except that equivalent to the depth of focal plane, and a display to show the image. In the eye gaze detection component, an infrared camera captures images of the observer’s eye through the infrared filter over a half mirror. The camera should be located near the eye to capture the clear image of the eye. However if the camera is in front of the eye, it blocks the observer’s view. Hence we locate the camera at a point which is optically near to the eye by using the half mirror. Then the system detects the point of gaze by image data processing by using the pupil-center-corneal reflection method. A two-dimensional image blurred with the focus point on the basis of the point of gaze is created on the GPU. The display presents the two-dimensional image. The process mentioned above is performed in real time.

Figure 7 shows the actual system, and figure 8 shows the image of the eye gaze detection component.

We used a personal computer with an Intel PentiumR 4 CPU (2.4 GHz, 512 MB RAM) and a SAPPHIRE X1950 PRO graphic board. The infrared camera is the Dragonfly Express from Point Grey Research. The size of the display is 17 inches and the distance from the observer to the display is 18.5 inches. The resolution of the display is 1280 × 1024 pixels and the resolution of the image is also 1280 × 1024 pixels. The eye gaze detection rate is approximately 25 frames per second.

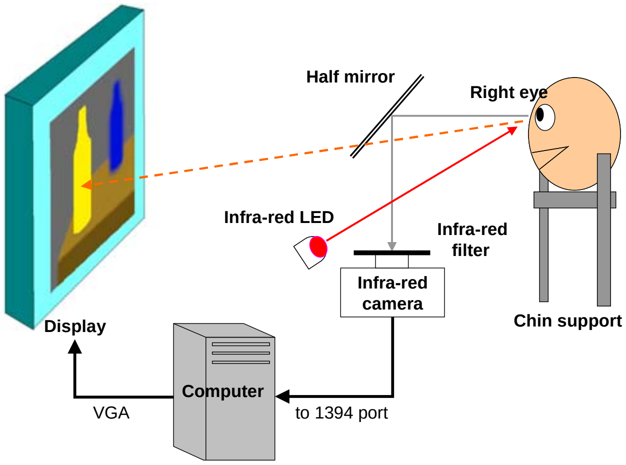


Figure 6: Structure of the implemented system.

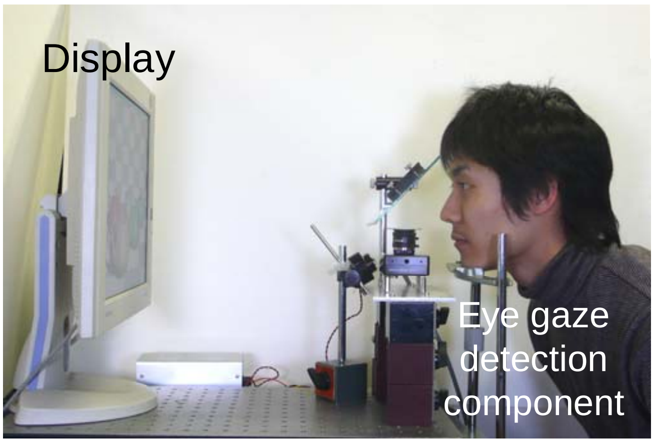


Figure 7: The experimental setup

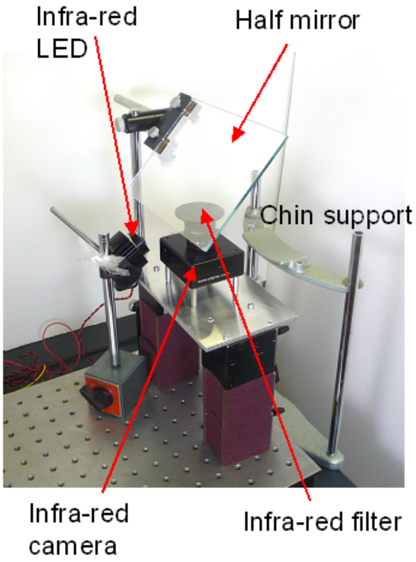


Figure 8: Image of the eye gaze detection component.

4. Experiment

We conducted comparative experiments to quantitatively evaluate the effects of presenting the image by using the prototype system. The images we compared are an all-in-focus image and focus-controlled image. Figure 7 shows the pictorial content of the images we presented.

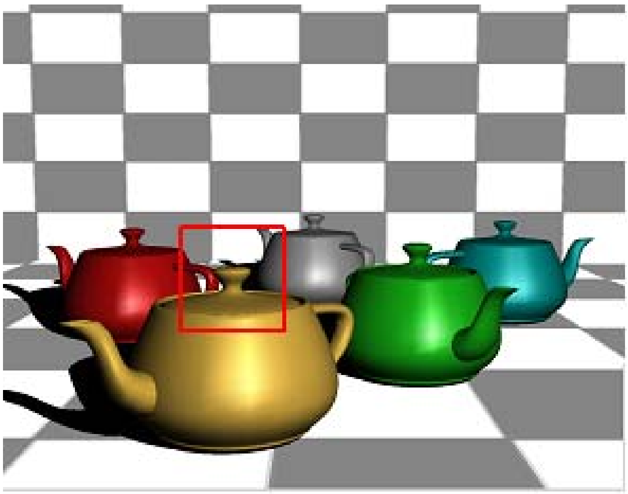


Figure 9: Image we presented in the experiment. The five teapots are positioned within the same scale. The wall and floor are drawn and represented by a checked pattern for observers to perceive the blur. The closeups image of the area inside the red square are shown in figure 10 and figure 11.

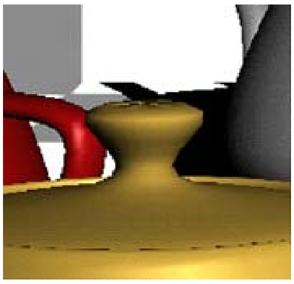
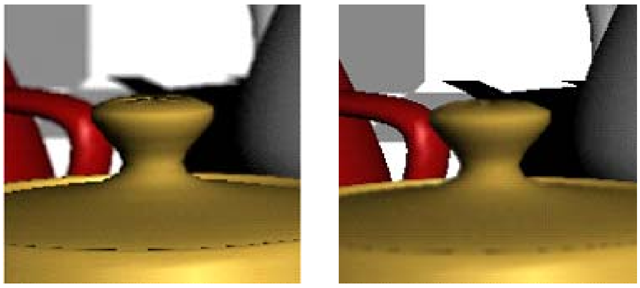


Figure 10: Closeup image of the all-in-focus image. Every pixel is drawn clearly.



(a) (b)

Figure 11: The closeup image of focus-controlled image. In the image (a), the point of gaze of the observer is the cover handle of the yellow teapot. In the image (b), the point of gaze of the observer is the wall.

4.1. Experimental procedure

At the beginning of the experiment, the eye gaze detection system was calibrated. Then, the two types of images were presented in turns to the subjects. First, the system showed the all-in-focus images to the subject for approximately 20 s. Then, the system showed the focus-controlled images for approximately 20 s. Subsequently, the subject was given three questionnaires and asked to score points ranging from 1 to 5 for each image by comparing the images. The questionnaires were based on “depth cognition,” “stress,” and “reality.” The questionnaire on depth cognition required the evaluation of the amount of depth information. The questionnaire on stress required the evaluation of the amount of stress experienced when observing the image. The questionnaire on reality required the evaluation of the extent of reality conveyed by the image. Further, the subjects were asked to comment or remark freely. Seven men and one woman participated in the experiment.

5. Results

The results of depth cognition, observation stress, and reality are shown in figure 12, figure 13, and figure 14, respectively.

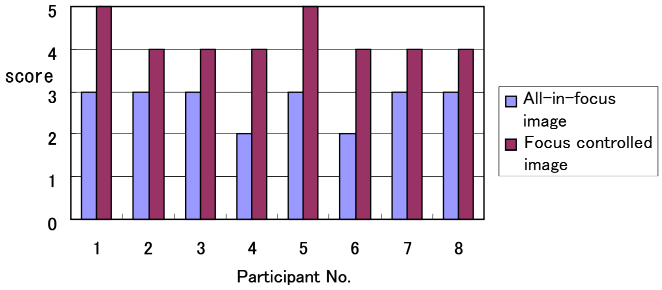


Figure 12: Score of the depth cognition per subject.

As shown in figure 12, all the subjects answered that the focus-controlled images could enhance more depth information as compared to the all-in-focus images. The subjects commented that it was easy to perceive the depth information of focus-controlled images.

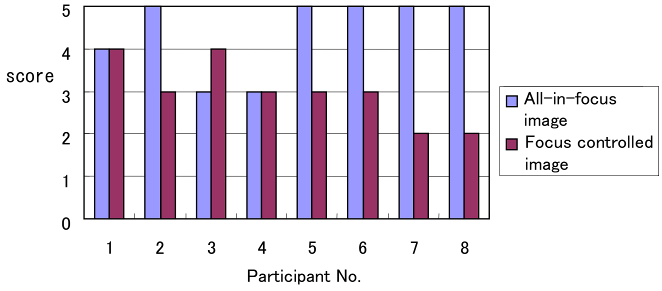


Figure 13: Score of the stress per subject.

As shown in figure 13, one subject answered that stress was increased by focus-controlled image; however, the other five subjects observed the all-in-focus images with less stress than that that required for focus-controlled images. Many of them said that they felt stress when observing focus-controlled images because the images flickered during blurring.

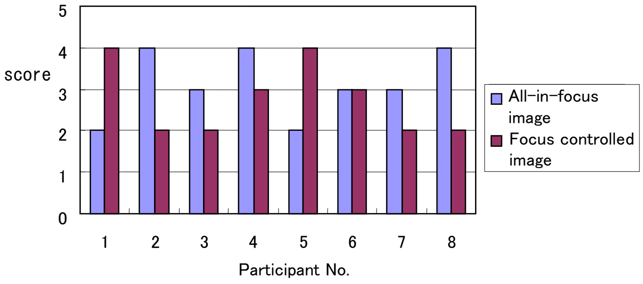


Figure 14: The score of the reality per subjects.

As shown in figure 14, two subjects answered that the reality is higher in observing focus-controlled images as compared to that in observing all-in-focus images. However, the other five subjects answered to the contrary. The reality tended to be decreased by using this system as a whole.

6. Discussion

The results clearly indicate that our proposed system enhances the depth information of a two-dimensional image. However, the stress and reality experienced with this system shoed the tendency to decrease. We discuss the reasons for this by considering the opinions and comments of the subjects.

A cause is that the system does not simulate the time for focal accommodation. When we shift our gaze from one object to another in the real world, it takes time to accommodate the focus of the eyes. However, this system could not simulate this phenomenon; hence, when the observer shifted his/her gaze from one object to another object in the images presented by the system , the focal plane was changed in no delay. We consider this as the cause of the flicker of the image during blurring, which resulted in the stress. In order to decrease the stress due to flicker, we improved the system to allow for more time for changing the focal plane smoothly. The improved system decreased the stress due to flicker.

Another cause is the low accuracy of eye gaze detection. The accuracy is significantly different for each subject because it depends on the contrast of the pupil or the length of the contour of the pupil. When the accuracy is low, the point of gaze calculated by the system and the point at which the subject gazes may be different, and the point at which object the observer gazes cannot be drawn clearly. In order to improve the eye gaze detection, we must use a camera with higher resolution or change the least square method for approximating the contour of the pupil.

As discussed above, improving the stress and reality of the system by improving the accuracy of the eye gaze detection is the topic of our further study.

7. Conclusions

We have studied the display which has internal active focus control system with the camera for eye gaze detection to enhance depth information. By using the methods we described in sections 3, which detect the eye gaze and generate the blur per pixel, we implemented the prototype system and mounted the system on an existing display. We conducted comparative experiments by using this system in order to qualitatively evaluate the effect of presenting the focus-controlled image. As a result, the system succeeds in enhancing the depth information. Our system presents images with less stress and high sense of reality by improving the accuracy of the eye gaze detection. In the future, it will become easy to perceive the depth information by combination our method and the other methods using other depth cues such as binocular disparity or convergence.

References

Barsky,B.A. 2004. Vision-realistic rendering: Simulation of the scanned foveal image from wavefront data of human subjects. Proceedings of Symposium on Applied Perception in Graphics and Visualization, 73–81.

Hillarie,S., Lecuyer,A., Cozot,R., and Casiez,G. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments.

John Krantz. 2001. Experiencing Sensation and Perception, Prentice Hall. New Jersey.

Potmesil,M. and Chakravarty,I. 1981. A lens and aperture camera model for synthetic image generation. In Proceedings of ACM SIGGRAPH, vol. 19, 298–306.

Rokita,P. 1996. Generating depth-of-field effects in virtual reality applications. In Prodeedings of IEEE Computer Graphics and Applications, vol.16, 18–21.

Yusuke,S., Hironobu,F., and Yutaka,H. 2004. Measurement of 3D Eye Movements Using Image Processing.