PROCEEDINGS OF THE INTERNATIONAL DISPLAY WORKSHOPS, VOL. 27, 2020

3D2-2
Invited

Augmented Reality Displays Based on 3D Gaze Sensing

Sei Ikeda[1]

ikeda.sei.jp@ieee.org
[1]Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka, 560-8531 Japan
Keywords: 3D Gaze Sensing, Depth of Focus, AR X-Ray Vision, In-Vehicle AR Displays.

ABSTRACT

This paper presents several display systems for augmented reality (AR) developed by the author's group. All of them are based on sensing the three-dimensional position of user's gaze (3D gaze), and they are AR user interfaces that provide some benefit to the user without the user's intention to operate the systems. In particular, we focused on the depth of a user's gaze for (1) presenting a more realistic representation of depth of focus, and (2) gaze-driven AR x-ray vision to make real objects semi-transparent and show the hidden parts. Additionally, an application example for an in-vehicle AR display is shown as a situation where such technology will be necessary or important in the near future.

1 INTRODUCTION

Augmented reality (AR) is a technology that transforms ordinary, everyday spaces into information presentation domains, such as web browsers and smartphones. Just as a vast array of new services have been created for the many office workers and students who casually glance at their phones on the train to work, AR has the ability to greatly expand such opportunities. Although AR displays have the advantage of superimposing semi-transparent images, it is difficult for the system to distinguish between real and virtual images, which causes various problems as described below. One of the solutions to this problem is a user interface based on 3D eye sensing.

There are two types of gaze-based virtual/augmented reality (VR/AR) user interfaces. One is an interface in which the user intentionally inputs some information with the line of sight instead of using the mouse or keyboard. The gaze interface is a high priority input mode, along with voice input and other input modes, when both hands are not free to move. However, there are several problems of incorrect inputs when the direction the user wants to look is different from the required input [1]. For example, the interface that moves icons on the screen by eye movement does not allow the user to look elsewhere while moving, which results in an operation error.

The other type of interface does not require the user to make any intentional input motions but provides a function by reading the user's intentions from the natural eye movements. This kind of user interface has the advantage that what the user wants to do does not conflict with the motions for system operation so that other problems are unlikely to arise from the introduction of the interface.

This paper introduces several AR displays developed by the author's group that improve visibility and reality by sensing the depth of gaze as the latter type of interface. Specifically, this paper reviews a study on the representation of depth of field in AR displays in Section 2 and AR X-ray vision in Section 3. We also present examples of applications for in-vehicle AR displays as a necessary or important part of these technologies in the near future.

2 DEPTH OF FOCUS IN AR DISPLAYS

Most commercially available AR displays are unable to present virtual images at multiple focal distances. The incorrect focus blur between the real object image and the CG image can occur, especially in the augmented reality display. The authors' groups have proposed several solutions to this issue.

My colleagues and I proposed a method to reduce the focal blur of the virtual image caused by the difference in depth between the presented virtual image and the real scene [2]. Assuming that the blur is generated by a certain point spread function (PSF), we optimize the presented image so that the presented image becomes the target image by blurring with the PSF. Specifically, we calculated the PSF from the gaze depth and pupil size by eye tracking and generated the presented image by deconvolution of the target image. In this method, only virtual objects with a certain depth could be represented.

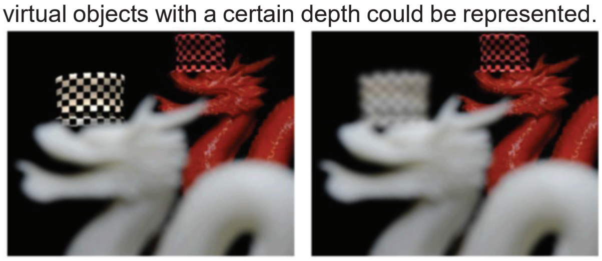

Fig. 1 AR based on accommodation sensing.

Fig. 1 shows a display system that uses an autorefractometer to determine the focal length of each lens of the eye and accordingly reproduces the focal blur of virtual objects with arbitrary shapes using real-time ray tracing techniques [3]. We showed that applying the accommodation sensing to AR would make it impossible to distinguish between virtual and real objects.

We are now developing a video see-through varifocal binocular magnifier whose focus changes according to the user's vergence angle [4]. Compared with light field-based microscopes [5], this system has a much simpler structure and can be mounted on the tip of an endoscope.

There are many display systems that use light field displays to reproduce the depth of focus instead of using eye tracking. In recent years, various optical see-through displays using spatial light modulator or variable focus lens have been proposed [6][7]. The above three systems developed by the authors' group have the distinction of being structurally simpler than such systems.

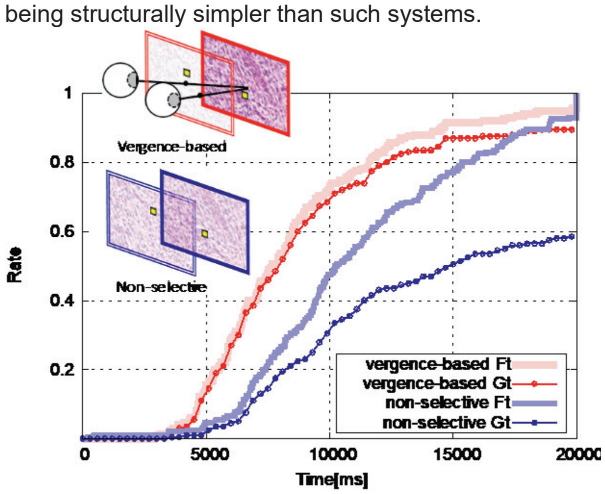

Fig. 2 Vergence-based AR x-ray vision.

3 AR X-RAY VISION

AR x-ray vision, also known as ghosting, is a technology to visualize a real scene (occludee) occluded by a real object (occluder). Users can recognize both texture and geometrical relationship between the occluder and occludee at once. In a broader interpretation, the problem is how to ensure the visibility of both surfaces in a scene where the real and virtual surfaces exist at different depths. There is a well-known problem of losing depth cues by simply making the occluder semi-transparent as a naive solution [8].

There are two main approaches to solve this problem. One approach is to partially change the transmittance of each layer (occluder or occludee) to increase the visibility of both layers. There are methods for determining pixel-by-pixel opacity using image edge and saliency maps [9], super pixels [10], and curvature of a three-dimensional model [11]. One of the other sophisticated methods optimizes the pixel-wise opacity of images based on a model of the human primary visual cortex [12]. The problem with this approach is that the main focus is on controlling the visibility of either the occluder or the occludee layer, and there are limitations to increasing the visibility of both layers.

Another approach is to vary the transmissivity of each layer over time according to the layer the user wants to look at. The most ideal way to display the layers that the user wants to see is to use the Brain-Machine Interface (BMI). Blum et al. proposed a gaze-driven augmented reality system which uses electromyographic (EMG) and eye-tracking technology to visualize the hidden internal organ around the point of gaze [13]. It was called a BMI but actually a gesture interface in which the user intentionally opens his or her eyelids to induce EMG peaks.

My colleagues and I proposed a method to reduce the mixing ratio of non-gazing objects by estimating the gaze depth by sensing the angle of congestion [14]. Fig. 2 shows the result of the comparison between gaze-based AR and the constant 50% transparency. Ft and Gt mean the cumulative probability density of respondents and correct answers, respectively. The visibility of the texture was improved by up to about 1.5 times compared to the method with a constant mixing ratio (alpha value) of 50%. On the other hand, there were a certain number of participants who were unable to gaze at the layers with the low alpha value.

In our new study, the authors' group developed a method that mixes both of the above approaches [15]. As shown in Fig. 3, the layer the user is looking at is rendered with the 100% alpha value. The non-gazing layer is rendered as a mesh whose most areas have low alpha. If the gazed-at surface is a distant layer, the far layer is partially hidden by the near meshed layer. If the gazing surface is the near layer, the far layer is rendered after rendering the near layer without using depth buffer. In this way, the user perceives the image of the far layer exists at the back of the near layer, regardless of the z-order of rendering, because the far layer produces binocular disparity consistent with the position of the far layer. Although the near layer is occluded by the meshed layer in each eye view, users perceive the near layer does not occluded because there are only a few areas that are simultaneously occluded in both eye views.

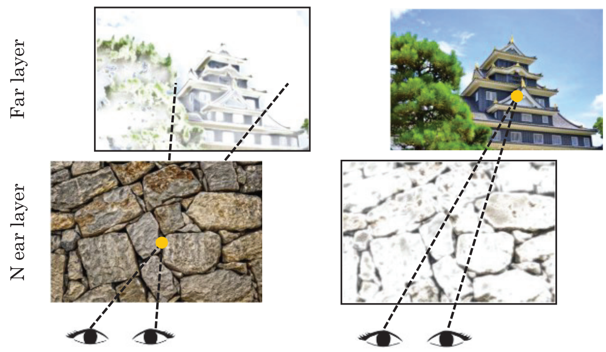

Fig. 3 Mesh-like representation for AR x-ray vision.

4 APPLICATIONS FOR IN-VEHICLE AR DISPLAYS

The above sections described that sensing gaze depth can improve the capability of AR display systems. This chapter shows an example of applications of the techniques described in Chapter 2 and 3. Two major problems must be solved for the applications.

Commercially available personal AR displays, such as Microsoft HoloLens, Magic Leap, and HTC Vive, have become much smaller than they were 10 or 20 years ago, and are now affordable for individuals to own. However, they have not become as popular devices as smartphones, in part because they do not provide power for long periods of time or achieve the lightness and thinness of ordinary glasses. These problems are still challenging.

The lack of accuracy of gaze depth estimation is a more fundamental problem. The range of depth that can be computed geometrically and optically accurately by eye sensing is, at most, within a few meters. AR displays based on gaze depth may be useful for the tasks at hands. However, the applications of such displays are limited.

There are some applications that may benefit many ordinary users by avoiding these problems. One of these situations is the display of advertisements on the in-vehicle display. In case of in-vehicle displays, problems such as downsizing and power saving can be avoided, and the cost per person is unlikely to be a problem since the same display is used by many people in a vehicle. It is very important for advertisers to check the effectiveness of semi-projected advertisements presented on an in-vehicle display, and it is an important issue whether users have seen the AR advertisement or not. For translucent advertisements, where the advertisements are placed in a fixed position relative to the real world or window glass, the depth of the fixation can be determined by observing it for a certain time, as shown in Figure 4, because the user is moving in parallel with the car [16]. Combining with our method [3] and [15], the visibility can be improved.

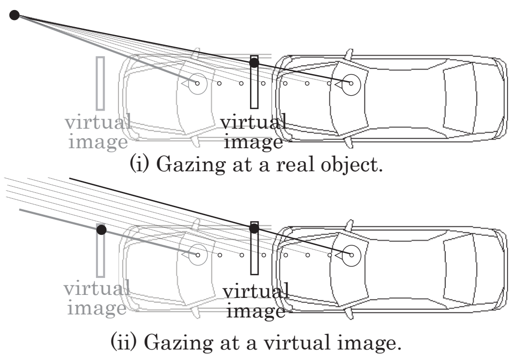

Fig. 4 Identification of passenger's fixation.

5 CONCLUSIONS

This paper showed that 3D gaze position sensing can improve the visibility of AR displays. The detection of gaze depth is a bottleneck to apply these techniques to various applications. One application is AR advertisement on in-vehicle AR displays according to passengers' gaze depth.

REFERENCES

[1] R. Jacob and S. Stellmach, "What You Look at is What You Get: Gaze-Based User Interfaces," Interactions, vol. 23, no. 5, pp. 62–65 (2016).

[2] K. Oshima, K. Moser, J. Rompapas D. II, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato: "SharpView: Improved clarity of defocussed content on optical see-through head-mounted displays", Proc. 3DUI2016, pp. 173-181 (2016).

[3] D. C. Rompapas, A. Rovira, A. Plopski, C. Sandor, T Taketomi, G. Yamamoto, H. Kato, and S. Ikeda, "EyeAR: Refocusable augmented reality content through eye measurements," Multimodal Technologies and Interaction, Vol. 1, No. 4, Article 22 (2017).

[4] T. Sawai, S. Ikeda, S. Yoshimoto, H. Ishizuka, Y. Kuroda, and O. Oshiro, "Video see-through varifocal binocular magnifier," Proc. SCI'20, pp.751-752 (2020) in Japanese.

[5] O. Bimber and D. C. Schedl: "Light-Field Microscopy: A Review," Journal of Neurology, vol. 4, no. 1, pp. 1-6 (2019)

[6] K. Rathinavel, G. Wetzstein and H. Fuchs: "Varifocal Occlusion-Capable Optical See-through Augmented Reality Display based on Focus-tunable Optics," IEEE Trans. on Visualization and Computer Graphics, vol. 25, no. 11, pp. 3125-3134 (2019)

[7] P. Chakravarthula, D. Dunn, K. Akşit and H. Fuchs: "FocusAR: Auto-focus Augmented Reality Eyeglasses for both Real World and Virtual Imagery," IEEE Trans. on Visualization and Computer Graphics, vol. 24, no. 11, pp. 2906-2916 (2018)

[8] V. Buchmann, T. Nilsen, and M. Billinghurst, "Interaction with Partially Transparent Hands and Objects," Proc. Sixth Australasian Conference on User Interface, pp. 17–20, (2005).

[9] C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila, "An augmented reality x-ray system based on visual saliency," Proc. ISMAR, pp. 27–36 (2010).

[10] S. Zollmann, D. Kalkofen, E. Mendez, and G. Reitmayr, "Image-based ghostings for single layer occlusions in augmented reality," Proc. ISMAR, pp. 19–26 (2010).

[11] D. Kalkofen, E. Veas, S. Zollmann, M. Steinberger, and D. Schmalstieg, "Adaptive ghosted views for Augmented Reality," Proc. ISMAR, pp. 1–9 (2013).

[12] T. Fukiage, T. Oishi, and K. Ikeuchi, "Visibility-based blending for real-time applications," Proc. ISMAR2014, pp. 63–72 (2014).

[13] T. Blum, R. Stauder, E. Euler, and N. Navab, "Superman-like X-ray vision: Towards brain-computer interfaces for medical augmented reality," Proc. ISMAR, pp. 271–272 (2012).

[14] Y. Kitajima, S. Ikeda, and K. Sato: "Vergence-based AR x-ray vision", Proc. ISMAR, pp. 188-189 (2015).

[15] S. Ikeda, F. Shibata, Y. Kikuchi, and K. Hara: Japan Patent Application No. 2019-110139 (2019).

[16] D. Mardanbegi, T. Langlotz, and H. Gellersen: "Resolving target ambiguity in 3D gaze interaction through VOR depth estimation," Proc. CHI (2019).