

Key issues and technologies for AR/VR head-mounted displays

Byoung-ho Lee*, Chanhyung Yoo, Jinsoo Jeong, Byoung-hyo Lee, and Kiseung Bang
School of Electrical and Computer Engineering, Seoul National University
Gwanak-Gu Gwanakro 1, Seoul 08826, South Korea

ABSTRACT

We first discuss the key factors of augmented reality (AR) and virtual reality (VR) displays. Various requirements for immersive experiences are categorized as six factors that must be considered when designing the AR/VR head-mounted displays (HMDs). These factors have a strong correlation with other factors and should maintain a moderate balance between them. Based on recent researches we second introduce various technologies for AR/VR. By comparing the pros and cons of each method, we discuss the progression of AR/VR devices that can provide more affordable HMD devices for the public.

Keywords: Augmented reality, virtual reality, head mounted displays, near eye displays, 3D displays

1. INTRODUCTION

Recently, a demand for immersive visual experience has steadily been raised as high-quality images and videos are easily accessible due to the rapid development of display technologies. Head-mounted displays (HMDs), which are also referred to as near-eye displays (NEDs), are essential devices for implementing immersive augmented reality (AR) or virtual reality (VR). VR systems featuring a wide viewing angle offer vivid and realistic virtual experiences and is specialized in the entertainment field such as cinema and game. AR systems, which can simultaneously superimpose augmented virtual images on real objects, are expected to be useful in the real life such as industrial field, automobiles, education, and medical practice.

In this keynote paper, key factors and various technologies for AR/VR devices are outlined. AR/VR HMDs are wearable displays, unlike a flat-panel display, and bring design rules for interaction services. Accordingly, many factors should be significantly considered. Virtual or augmented images must be continuously updated according to eye or body movement, which requires high-performance and robust systems. Moreover, the device should be considered to be comfortably worn for long time. Consequently, for this, the design of AR/VR HMDs needs to consider a trade-off relation between various factors (optical qualities, form factor, cost and manufacturing condition). In the following chapter, we will explain the key issues of the AR/VR devices and discuss various techniques for AR/VR display systems.

2. KEY ISSUES FOR AR/VR DISPLAYS

2.1 Resolution

In AR/VR displays, the high-resolution of virtual images should be provided to make users feel immersive and not suffer from heterogeneous feeling from reality. The virtual images are typically given by floating the display plane such as organic light emitting diode (OLED) or liquid crystal display (LCD) panels using a lens, or directly projected to the user's retina using a projector with a converging lens. The lens is called eyepiece which magnifies the display part to widen the field of view (FOV). As the magnification increases, the FOV becomes wider by allowing the user to see a larger virtual image. However, the pixel of the display is also magnified, resulting in the degradation of the resolution. Therefore, the parameters of resolution and FOV have a trade-off relationship in AR/VR displays. The resolution can be analyzed from the spatial frequency of the perceived or magnified virtual image in a unit of cycles per degree (CPD). The cycles mean the number of black and white pairs. Figure 1 shows the relationship between the lateral resolution expressed in CPD and the lateral FOV. Even the results are acquired in an ideal case where any constraints such as vignetting and aberrations are not considered, it is a challenge to provide both wide FOV and high-resolution using

commercialized displays of full high definition (FHD) or ultra-high definition (UHD). For example, 30 CPD which corresponds to the 20/20 vision condition¹ meets FOV of 32° and 64° as the illustrated green line in Fig. 1. Therefore, it is required to optimize and design these parameters with the optical system to make users feel the immersive virtual image.

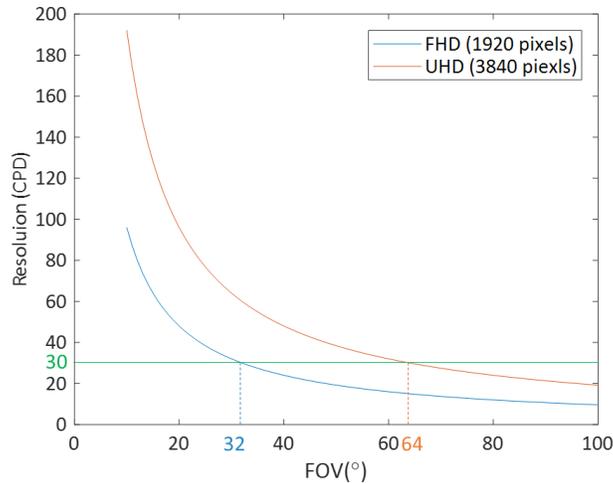


Figure 1. The relationship between the lateral resolution and lateral FOV. The green line corresponds to the 30 CPD which is the 20/20 vision condition.

2.2 Field of view (FOV)

FOV of AR/VR displays is the angular region where the virtual image can be displayed in the user's sight, which is usually given in the unit of degrees. FOV is one of the most important factors for immersive experiences. When an AR/VR display has a large enough FOV, the users may feel surrounded by the virtual environment. The desired FOV of AR/VR display can be decided as the FOV of the human eye. The FOV of a single eye is about 135 degrees in the horizontal axis, while the overlapped FOV of two eyes is 114 degrees, which can provide binocular depth perception². However, it is not an easy task to achieve wide FOV because of the trade-off between the FOV and form factor. In VR systems, the FOV is mainly limited by the eyepiece lens. To improve the FOV, the eyepiece lens should be closer to the user's eye or area of the lens should be wider. In other words, the lens should have a larger numerical aperture (NA) with acceptable imaging quality. However, it is even harder to achieve wide FOV in AR system than VR due to the higher standard of form factor and need for transparency. While many commercial VR devices in the market have wider FOV than 100 degrees, most AR devices in the market have 30~50 degrees. In AR systems, the see-through combiner is required to observe the light emitting from real objects to user's eye without distortions, which becomes the limiting factor of the FOV. The reason for limitation is depending on the types of the AR displays like birdbath, waveguide and Maxwellian view display. Detailed characteristics of each type and the limiting factor will be described in Section 3.

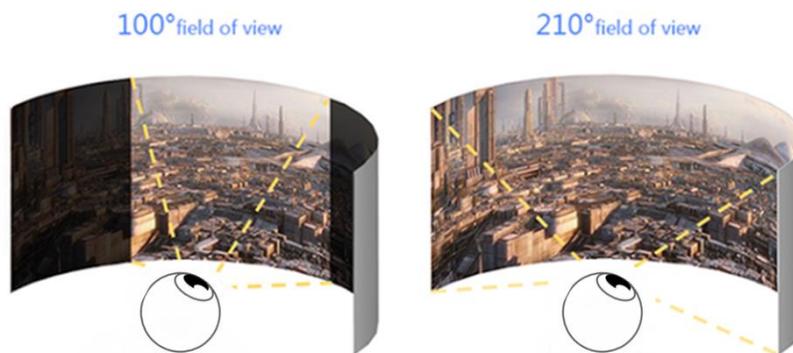


Figure 2. The concept diagram presenting a field of view (FOV).

2.3 Eyebox

The eyebox refers to the area where the user can watch the image through, as shown in Fig. 3. In general, the eyebox has a trade-off relationship with the FOV³. When the FOV is wider, the eyebox should be narrower, and when the eyebox is wider, the FOV should be narrower. Recently, researches on the eyebox extension with large FOV are actively studied. The eyebox extension is usually implemented through a pupil tracking technique, which can be divided into active and passive scanning method³⁻⁶. In the active scanning method, the whole bandwidth of the eyebox is steered by using MEMS, motorized stage, etc.^{3,4}. Even though they can utilize a whole bandwidth of the display, the weight and delay latency of the mechanical devices are critical problems. The passive method uses a specially designed optical component, such as the holographic optical element (HOE), to extend eyebox in several places and changes only the image according to the position of the eye^{5,6}. Therefore, there is an advantage of not needing a mechanical device, but it is difficult to represent continuous images between the pupil positions or it is difficult to utilize the entire bandwidth of the display.

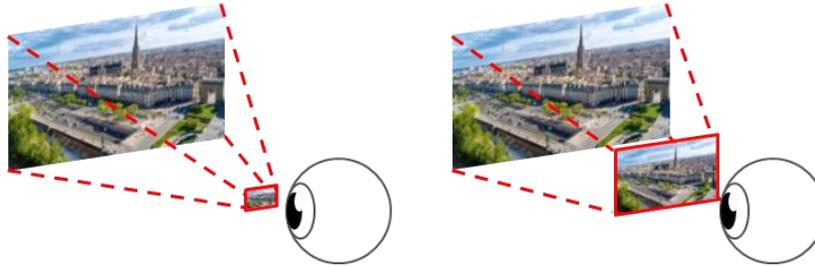


Figure 3. The concept diagram presenting an eyebox.

2.4 Form factor

The form factor should be considered in two categories: physical dimensions (volume and weight) and shape of the entire system. First, the compact form factor system with high-performance is a major challenge in AR/VR. The size and weight of the devices have gradually been decreased by employing state-of-the-art displays and optics. However, visual performance and form factors are interdependent. In order to provide moderate immersive experiences, the reduction in form factor is inevitably limited. Most of conventional AR/VR devices can be categorized into two groups: the relatively heavy HMD-type devices with high-performance and lightweight glasses-type devices with low-performance. The standalone HMDs especially are heavy and bulky due to a high-performance computing device, sensors and a battery in addition to optical devices. Meanwhile, glasses-type devices can pair with the smartphone or desktop, which can reduce the load of computing power and sensors. Second, the shape should be affordable and give a comfortable fit for the user's face or bone structure. Improperly designed devices may put pressure on the user's nose or sides of the head, causing dizziness and discomfort, when wearing for prolonged periods. Additionally, the features of the HMDs or glasses should be designed to be more visually attractive. The design factor is a more important concept in AR devices than VR which has been mainly used in the private or disconnected space from the outside. To gain more popularity to the public, it is necessary to provide various designs, such as normal spectacles, so that consumers can decide on a style according to personal preference.

2.5 Accommodation

In AR/VR displays, the stereo image can be easily provided due to the left-channel and right-channel separated displays, compared to typical 2D flat-panel displays. The stereoscopic vision is a very effective source to recognize the spatial position of three-dimensional (3D) objects and has been used mainly in 3D displays. In addition to binocular disparity, the physiological depth cues also include accommodation (focus) and motion parallax⁷. These depth cues always coincide when staring at objects in the real world, thus they provide a natural perception of depth. In the AR environment, the virtual image has to be aligned to the real object and augmented images rendered in consideration of binocular parallax are observed at the same depth as the real object. Nevertheless, the virtual image is located on a fixed focal plane determined by the display source and optics. This discrepancy between the focal distance of the real and virtual objects may hinder a sharp image. Furthermore, for AR images, the vergence distance by the binocular disparity and accommodation distance by the focal power may be mismatched, which is called vergence-accommodation conflict

(VAC), as shown in Fig. 4. The VAC problem limits comfort depth range and may provoke the various adverse effects such as visual fatigue, nausea and vertigo⁸⁻¹⁰. To mitigate this issue, focus-tunable technology has been proposed recently¹¹⁻¹³. By continuously or discretely adjusting the focal distance, discomfort or blurred vision can be alleviated compared to a single focal plane. In addition, researches have been reported to combine the holographic technique with AR/VR systems, which reconstructs the true wavefront as seeing real objects.

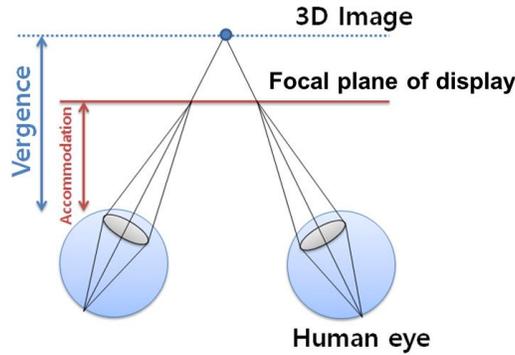


Figure 4. Schematic diagram of vergence-accommodation conflict (VAC).

2.6 Occlusion

Occlusion in AR displays refers to the technology to block the light from the outside world selectively. Basically, AR displays show virtual images with the real world scene, which means that the light from the display and light from the outside world are summed up in the user's sight. Therefore, the background image behind the virtual object is seen together, that is, the virtual object will be always seen transparent. Also, the image contrast will be degraded especially when the virtual object has a dark color. This effect may disturb a realistic experience. If we can block the light from the real scene at the position where the virtual object should occlude, in other words, if we can subtract the light, we can provide clearer and realistic virtual images.

However, the implementation of occlusion in AR displays is very difficult. Due to its difficulty, the occlusion is one of the most slowly developed technologies in AR displays. One main reason that makes it difficult is that the occlusion mask should be imaged at a far distance. Therefore, a lens should be located between the mask and the user's eye. But since the lens will distort the real scene from outside, other relay lens systems should be added after the occlusion mask as seen in Fig. 5. This relay lens system occupies too much volume and limits the FOV of the real scene as well as the virtual image. Recently, a similar system using lens arrays rather than the big lenses has been suggested¹⁴. Since the focal length of the lens array can be much shorter than the big lenses, the size of the whole system could be reduced much. But this system also could not avoid the problem that the real scene may not be perfectly recovered but distorted in practice. Another big obstacle to overcome is the screen door effect in a real scene. Since the occlusion mask should be actively controlled in practice, pixelated structures like liquid crystal array are usually used. But in this case, the real scene is also blocked by the black matrices of the pixelated structures, which makes black mesh-shape noise in the real scene. This effect is difficult to avoid as long as the pixelated structures are used as the active occlusion mask. To overcome these two main problems, a new type of occlusion system needs to be suggested.

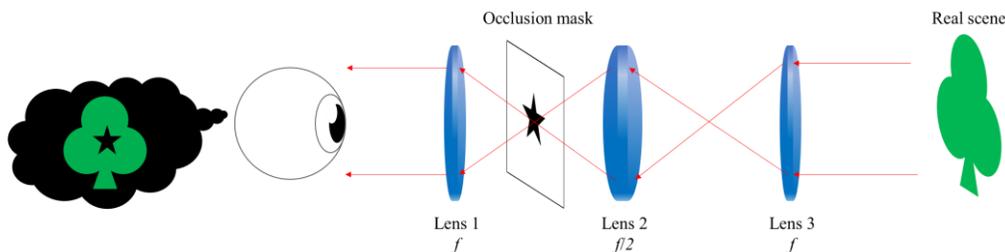


Figure 5. Basic configuration of occlusion system in AR displays.

3. VARIOUS AND ADVANCED TECHNOLOGIES

3.1 Birdbath-based techniques

Birdbath type display is the simplest type of AR displays. The system is described in Fig. 6. The light from the display is reflected by the concave mirror so that the display is imaged to a far distance. The half-mirror reflects the light toward the user's eye and also the light from the outside passes through the half-mirror toward the eye. Therefore, the half-mirror operates as an image combiner and the user can observe the virtual image with the real world scene. This kind of system was used in Google Glass. It has strength in its simple system and stable eyepoint, but the FOV is limited by the geometrical reason and the size of the whole system is not compact. Since the half-mirror should be placed between the concave mirror and the user's eye, the user can observe the virtual image only within a small FOV. Also, since the half-mirror should be placed in front of the eye at 45 degrees, it is hard to make the system thin. Figure 6 shows the result of the birdbath display with dual focal planes. In addition to the basic birdbath system, Savart plate is inserted for dual focal planes. Since the Savart plate has two different optical path lengths depending on the polarization state, the display is imaged to two different planes. The result shows that it has a FOV of 8.9 degrees.

In addition, there is another type of AR display that uses curved half-mirror instead of the curved mirror and half-mirror. The curved half-mirror both images the display at far distance and operates as an image combiner. Curved half-mirror type display has been widely studied and applied to many commercialized products such as Meta 2. Due to the absence of additional half-mirror after the curved mirror, the FOV can be improved compared to the birdbath type display. However, it still has large form factor because off-axis optic is still needed.

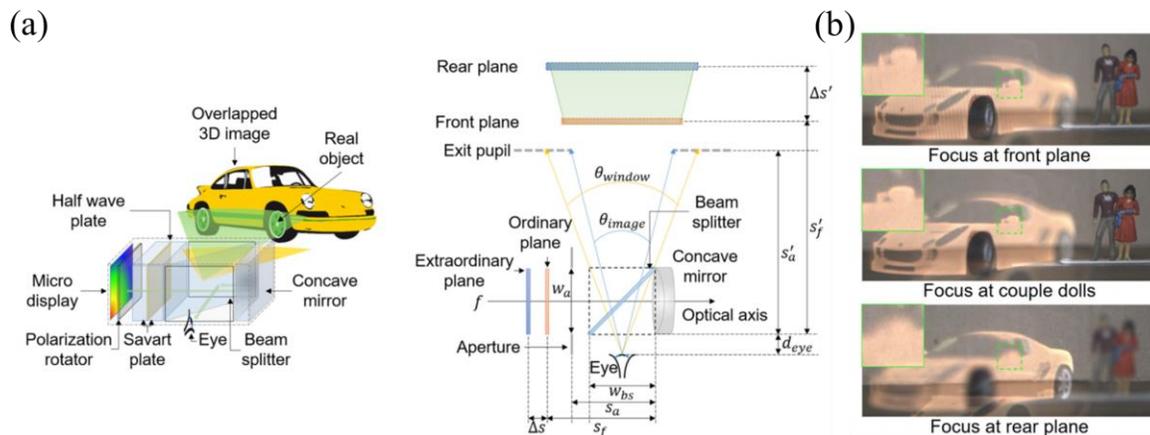


Figure 6. (a) Configuration of the birdbath type AR display with dual focal planes using Savart plate and (b) display results.

3.2 Guide-based techniques

Lightguide and waveguide techniques are promising technologies for wearable devices since optical combiner can be manufactured in very thin volumes in comparison with the above-mentioned birdbath type. Basically, guide techniques use total internal reflection (TIR) to propagate the input beam to the observer's eye. Consequently, this is possible to obtain sufficient spacing between the display and combiner, causing the reduction of heavy components in front of the eyes. In addition, the guide-based HMDs provide a large eyepoint that can allow eye movement or rotation by using the exit-pupil expanding (EPE) technique¹⁵. These guide techniques can be distinguished into two concepts: the mirror and diffraction optics. First, the mirror-type guide-based display uses small embedded reflectors in the guide substrate, which was adopted by Lumus, Google, and Epson¹⁶. The second concept uses diffractive optical elements (DOEs) such as surface-relief gratings (SRGs) or volume holographic gratings (VHGs) in order to deflect the input or output beam to a specific angle. These DOEs can be fabricated in micro-scale and only diffract the predefined beam with a particular angle or wavelength, which is suitable for applications in the AR systems. Despite these various advantages, the guide-based techniques present some intrinsic problems such as color degradation called "rainbow effect" and small FOV. The DOEs must satisfy the phase-matching condition which results in a sharp drop of diffraction efficiency at a wide incident

angle, as shown in Fig. 7(b). In order to resolve the small FOV problem, the DOE materials with a high refractive index have to be developed. Also, various multiplexing techniques have been adopted^{17, 18}. Spatial multiplexing by simply stacking multiple guide channels is limited since this method increases the form factor. VHGs can enlarge narrow FOV by recording using angular-multiplexing, but this method also has a limitation without improving the refractive index. Recently, several researches have been proposed to overcome these limitations by using meta-gratings or polarization-dependent gratings^{19, 20}.

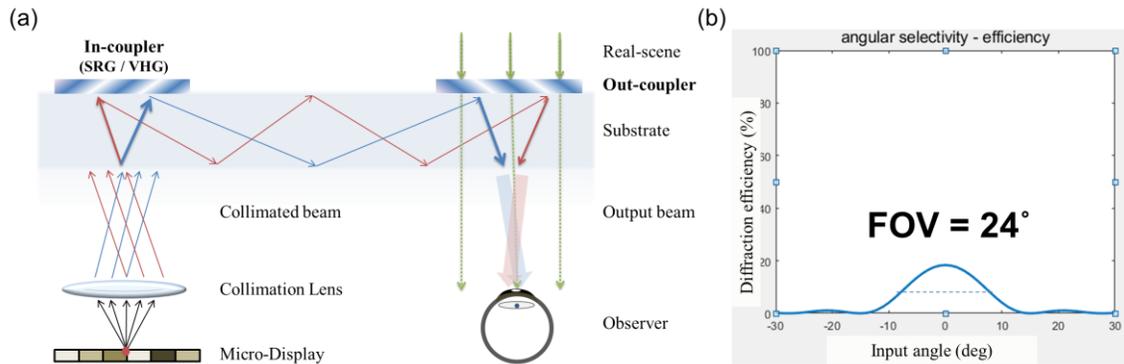


Figure 7. (a) Schematic diagram of the waveguide-type HMD. (b) Simulation graph of the angular-selectivity in VHGs which means the variation of the diffraction efficiency depending on the input angle. The thickness of VHG is 4 μm and the modulation of refractive index is 0.03.

3.3 Maxwellian-view display (retinal projection display)

Maxwellian-view displays can produce a clear image on the retina regardless of vision or focal length of eye lens. As shown in Fig. 8, using only the rays that converge into the pupil, the retinal image can be formed independently of the refractive power of the eye²¹. The Maxwellian view has been commonly used for low vision people, but there is also a lot of research into AR displays that need to focus on real scenes and virtual images at the same time. By collecting the light rays in a single point reproduced on the display by a lens to realize the Maxwellian view, the FOV of the HMDs can be dramatically increased. However, due to the trade-off between the eyepiece and the FOV mentioned above, the eyepiece becomes very small so that the image is not visible even if the eye pupil is slightly out of the original position.

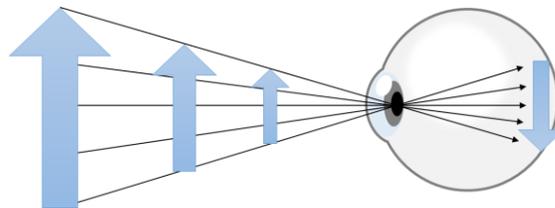


Figure 8. The concept diagram of the Maxwellian view. Since only the rays that converge into the eye pupil are utilized, the retinal image can be seen clear regardless of the focused plane or refractive power of the eye lens.

3.4 Holographic display

Holographic displays are considered as a next-generation technique for AR/VR HMDs due to its expression for virtual scenes using wavefront modulation. 3D holographic scenes rather than 2D images can be generated from the wavefront considering the wave optical characteristics including the interference, and the diffraction of the coherent light. The VAC problem in AR/VR HMDs is resolved through the holographic technique. Figures 9(a) and 9(b) show schematic diagrams of the holographic HMD. The reflective spatial light modulator (SLM) dynamically modulates the incident beam to generate the 3D holograms. Figure 9(a) shows a conventional optical system of the holographic HMDs. The modulated wavefront is propagated through the optical relay system with spatial filtering and perceived by the user

through the eyepiece lens. The type of AR/VR HMDs is set by designing the eyepiece and image combiner having see-through characteristics. Figure 9(b) is an example of the holographic AR HMDs which adopts the HOE as the eyepiece as well as the image combiner³. The HOE has a see-through characteristic because it only diffracts the incident beam whose k -vector matches the period of HOE's grating. Figure 9(c) shows that holographic display can provide focus cues. Although the holographic displays have some limitations such as the resolution, FOV, eyebox, form factor, etc., it is a viable candidate for future AR/VR HMDs in that it can fully display 3D scenes.

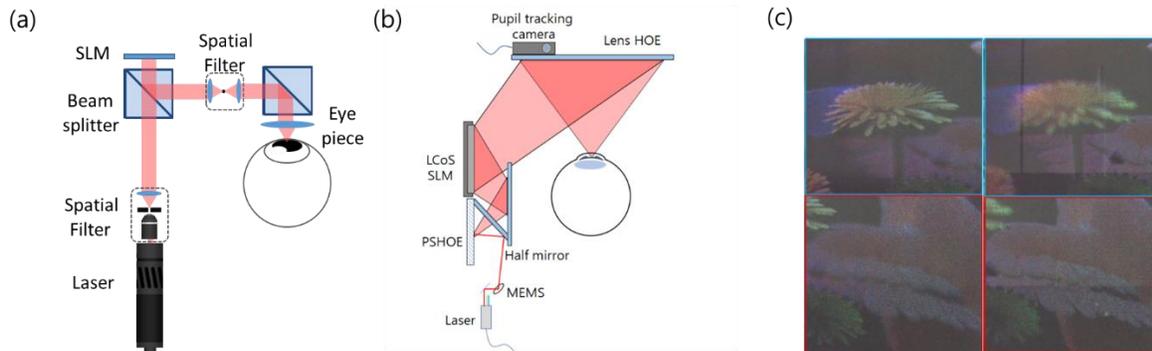


Figure 9. (a) Schematic diagram of the holographic HMD. (b) An example of the holographic AR HMD where lens HOE is utilized for eyepiece and image combiner with see-through characteristic. (c) The experimental results of focus cue from the setup of (b). The figures of (b) and (c) are extracted from [3].

3.5 Stereoscopic display with focus cue

The above-mentioned VAC problem is well known for AR/VR displays and should be addressed to this issue. To resolve the VAC problem, stereoscopic displays with adjustable focus have been reported¹¹. Most of them can be described as either multifocal or varifocal method. First, the multifocal displays provide multiple focal planes of virtual images with multiplexing techniques. In this case, the zone of comfort (ZOC) can be expanded by stacked focal planes, as shown in Fig. 10(b). Although this method should require a tunable-focusing optical element, most of these optical elements such as a focus-tunable lens have a small aperture size, causing a narrow FOV. Furthermore, it requires a fast switchable display synchronized with a tunable-focusing optics, which makes optical system bulky. In order to apply to AR systems, focus-tunable optic modules need to be compact. Second, the varifocal display can adjust the focal plane at proper distance matched a vergence distance by using adjustable-focus optics and gaze-tracking technique. Both solutions can be reduced adverse effects by VAC and provide unblurred images. In guide-based technique, the Magic Leap One with two focal planes was already released in 2017, which was the first commercial product with multiple focal stacks. This HMD device is implemented by using two guide channels corresponding to each focal plane. A similar study, which provides dual focal planes with a single guide channel using polarization multiplexing, has also been reported²². The next-generation AR/VR HMDs will have to offer multiple or adjustable focal plane.

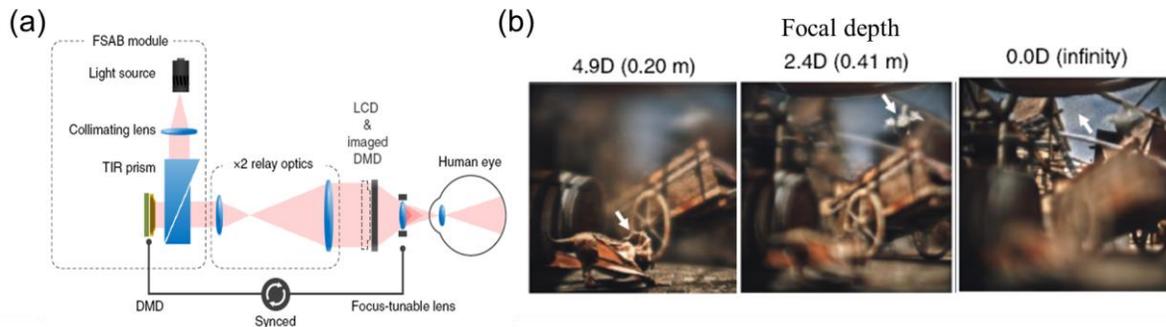


Figure 10. (a) Schematic diagram of a benchtop-type tomographic near-eye display. (b) Experimental results of tomographic near-eye display according to the focal depth. Figures are extracted from [13].

4. CONCLUSION

This keynote paper presents major issues and technologies for AR/VR display. We briefly overviewed the key factors and related issues in the first chapter. Then we focus on the optical properties and components of AR/VR devices. Many peoples are accustomed to high-resolution 2D images, which is a major obstacle to the commercialization of AR/VR devices. Ultra-resolution microdisplays and many solutions for a wide FOV and eyebox have been vigorously reported and realized. But miniaturization is still challenging due to excessive computing load and power consumption for high-standard quality. Therefore, AR/VR systems must be designed to take into account the trade-off relation of various key factors and issues. In the second chapter, various methods and advanced technologies for AR/VR HMDs are summarized. By adopting novel optical devices and methods, AR/VR devices can become been compact and lightweight, breaking through the inherent limitations. AR/VR technologies have a substantial potential and the proliferation of AR/VR devices is expected to create innovations in daily life.

ACKNOWLEDGEMENT

This work was supported by Institute for Information Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00787, Development of vision assistant HMD and contents for the legally blind and low visions).

REFERENCES

- [1] Kagadis G. C., Kloukinas C., Moore K., Philbin J., Papadimitroulas P., Alexakos C., Nagy P. G., Visvikis D., and Hendee W. R., "Cloud computing in medical imaging," *Med. Phys.* 40, 070901 (2013).
- [2] Stragburger, H., Rentschler, I., and Jüttner, M., "Peripheral vision and pattern recognition: a review," *Journal of Vision*, 11(5), 1-82 (2011).
- [3] Jang, C., Bang, K., Li, G., and Lee, B., "Holographic near-eye display with expanded eye-box," *ACM Trans. Graph.* 37(6), 195 (2018).
- [4] Kim, J., Jeong, Y., Stengel, M., Akşit, K., Albert, R., Boudaoud, B., Greer, T., Kim, J., Lopes, W., Majercik, Z., Shirley, P., Spjut, J., McGuire, M., and Luebke, D., "Foveated AR: dynamically-foveated augmented reality display," *ACM Trans. Graph.* 38(4), 1–15 (2019).
- [5] Park, J.-H., and Kim, S.-B., "Optical see-through holographic near-eye-display with eyebox steering and depth of field control," *Opt. Express* 26(21), 27076–27088 (2018).
- [6] Jeong, J., Lee, J., Yoo, C., Moon, S., Lee, B., and Lee, B. "Holographically customized optical combiner for eye-box extended near-eye display," *Opt. Express* 27, 38006-38018 (2019)
- [7] Teittinen M., "Depth cues in the human visual system," *The encyclopedia of virtual environments*, 1993, <http://www.hitl.washington.edu/projects/knowledge_base/virtual-worlds/EVE/III.A.1.c.DepthCues.html> (1993)
- [8] Yano S., Emoto M., Mitsuhashi T., and Thwaites H., "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays* 23(4), 191–201 (2002).
- [9] Hoffman D. M., Girshick A. R., Akeley K., and Banks M. S., "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.* 8(3), 33 (2008).
- [10] Shibata T., Kim J., Hoffman D. M., and Banks M. S., "The zone of comfort: predicting visual discomfort with stereo displays," *J. Vis.* 11(8), 11 (2011).
- [11] Kramida G. and Varshney A., "Resolving the vergence-accommodation conflict in head-mounted displays," *IEEE Trans. Vis. Comput. Graph.* 22(7), 1912–1931 (2016).
- [12] Matsuda N., Fix A., and Lanman D., "Focal surface displays," *ACM Trans. Graph.* 36(4), 86 (2017).
- [13] Lee S., Jo Y., Yoo D., Cho J., Lee D., and Lee B., "Tomographic near-eye displays," *Nat. Commun.* 10, 2497 (2019).
- [14] Yamaguchi, Y., and Takaki, Y., "See-through integral imaging display with background occlusion capability," *Appl. Opt.* 55(3), A144-A149 (2016).
- [15] Å yräs P., Saarikko P., and Levola T., "Exit pupil expander with a large field of view based on diffractive optics," *J. Soc. Inf. Disp.* 17, 659 (2009). 123

- [16] Sarayeddine K. and Mirza K., "Key challenges to affordable see through wearable displays: the missing link for mobile AR mass deployment," Proc. SPIE 8720, 87200D (2013).
- [17] Yu C., Peng Y., Zhao Q., Li H., and Liu X., "Highly efficient waveguide display with space-variant volume holographic gratings," Appl. Opt. 56(34), 9390–9397 (2017).
- [18] Han J., Liu J., Yao X., and Wang Y., "Portable waveguide display system with a large field of view by integrating freeform elements and volume holograms," Opt. Express 23(3), 3534–3549 (2015).
- [19] Shi Z., Chen W. T., and Capasso F., "Wide field-of-view waveguide displays enabled by polarization-dependent metagratings," in Digital Optics for Immersive Displays, 10676 (International Society for Optics and Photonics), 1067615 (2018).
- [20] Zhiqin H., Daniel L. M., and David R. S., "Polarization-selective waveguide holography in the visible spectrum," Opt. Express 27(24), 35631-35645 (2019).
- [21] Takaki, Y. and Fujimoto, N., "Flexible retinal image formation by holographic Maxwellian-view display," Opt. Express 26(18), 22985–22999 (2018).
- [22] Yoo C., Bang K., Jang C., Kim D., Lee C.-K., Sung G., Lee H.-S., and Lee B., "Dual-focus waveguide see-through near-eye display with polarization dependent lenses," Opt. Lett. 44(8), 1920-1923, (2019).