Computational Near-Eye Displays: Engineering the Interface Between our Visual System and the Digital World

Gordon Wetzstein
Stanford University

Abstract

Immersive virtual and augmented reality systems (VR/AR) are entering the consumer market and have the potential to profoundly impact our society. Applications of these systems range from communication, entertainment, education, collaborative work, simulation and training to telesurgery, phobia treatment, and basic vision research. In every immersive experience, the primary interface between the user and the digital world is the near-eye display. Thus, developing near-eye display systems that provide a high-quality user experience is of the utmost importance. Many characteristics of near-eye displays that define the quality of an experience, such as resolution, refresh rate, contrast, and field of view, have been significantly improved in recent years. However, a significant source of visual discomfort prevails: the vergence-accommodation conflict (VAC). This visual conflict results from the fact that vergence cues, but not focus cues, are simulated in near-eye display systems. Indeed, natural focus cues are not supported by any existing near-eye display. Afforded by focus-tunable optics, we explore unprecedented display modes that tackle this issue in multiple ways with the goal of increasing visual comfort and providing more realistic visual experiences.

1. Introduction

In current VR/AR systems, a stereoscopic near-eye display is used to present two different images to the viewer's left and right eyes. Because each eye sees a different view of the virtual world, binocular disparity cues are created that generate a vivid sense of three-dimensionality. These disparity cues also drive the viewer's vergence state (the relative rotation of the eyeballs in their sockets) as they look around at objects with different depths in the virtual world. However, the accommodation, or focus state, of the viewer's eyes is optically fixed to one specific distance. This is because, despite the simulated disparity cues, the micro display inside a VR system is actually at a single, fixed optical distance. The specific distance is defined by the magnified image of the micro display, and the eyes are forced to focus at that distance and only that distance in order for the virtual world to appear sharp. Focusing at other distances (such as those actually simulated by the stereoscopic views) results in a blurred view.

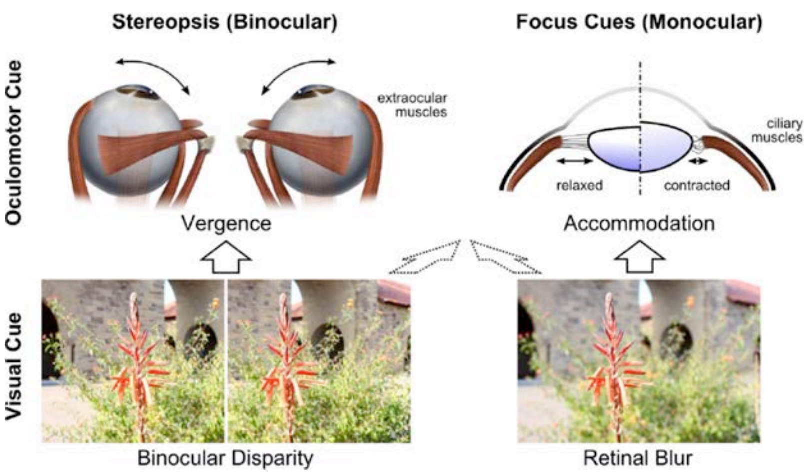

Figure 1. Overview of relevant depth cues. Vergence and accommodation are oculomotor cues whereas binocular disparity and retinal blur are visual cues. In normal viewing conditions, disparity drives vergence and blur drives accommodation. However, these cues are cross-coupled. Existing near-eye displays only support binocular cues, but not focus cues.

In the physical world, these two properties of the visual response – vergence and accommodation – work in harmony (see Figure 1). Thus, the neural systems that drive the vergence and accommodative states of the eye are neutrally coupled. VR/AR displays artificially decouple these cues due to their optical image formation. The resulting discrepancy between natural depth cues and those produced by existing VR/AR displays may lead to visual discomfort and fatigue, eyestrain, diplopic vision, headaches, nausea, compromised image quality, and may even lead to pathologies in the developing visual system of children. This discrepancy is referred to as the vergence-accommodation conflict (VAC).

The benefits of providing correct or nearly correct focus cues not only include increased visual comfort, but also improvements in 3D shape perception, stereoscopic correspondence matching, and discrimination of larger depth intervals. Thus, significant efforts have been made to engineer focus-supporting displays. However, all technologies that can potentially support focus cues suffer from undesirable tradeoffs in compromising image resolution, device form factor, brightness, contrast, or other important display characteristics. These tradeoffs pose substantial challenges for high-quality AR/VR visual imagery with practical, wearable displays.

2. Background

In recent years, a number of near-eye displays have been proposed that support focus cues. Generally, existing focus-supporting displays can be divided into several classes: adaptive focus, volumetric, light field, and holographic displays.

Two-dimensional adaptive focus displays do not produce correct focus cues – the virtual image of a single display plane is presented to each eye, just as in conventional near-eye displays. However, the system is capable of dynamically adjusting the distance of the observed image, either by physically actuating the screen (Sugihara, 1998) or using focus-tunable optics (programmable liquid lenses). Because this technology only enables the distance of the entire virtual image to be adjusted at once, the issue with these displays is that the correct focal distance at which to place the display will depend on where in the simulated 3D scene the user is looking. Peli (1999) lists several references that proposed the idea of gaze-contingent focus, but the author is not aware of anyone having built a practical gaze-contingent, focus-tunable display prototype. The challenge for this technology is to engineer a robust gaze and vergence tracking system in a head-mounted display with custom optics. A software-only alternative to gaze-contingent focus is gaze-contingent blur rendering (Mauderer 2014), however because the distance to the display is still fixed in this technique, it does not affect the VAC. Konrad et al. (2016) recently evaluated several focus-tunable display modes in near-eye displays and also proposed the idea of monovision as a practical alternative to gaze-contingent focus, where each eye is optically accommodated at a different depth.

Three-dimensional volumetric and multi-plane displays represent the most common approach to focus-supporting near-eye displays. Instead of using 2D display primitives at some fixed or adaptive distance to the eye, volumetric displays either mechanically or optically scan out the 3D space of possible light emitting display primitives in front of each eye (Schowengerdt and Seibel 2006). Multi-plane displays approximate this volume using a few virtual planes that are generated by beam splitters (Dolgoff 1997, Akeley 2004) or time-mulitplexed focus-tunable optics (Rolland 2000, von Waldkirch 2004, Liu 2008, Love 2009, Llull 2015). Whereas implementations with beam splitters compromise the form factor of a near-eye display, temporal multiplexing introduces perceived flicker and requires display refresh rates beyond those offered by current-generation microdisplays.

Four-dimensional light field and holographic displays aim to synthesize the full 4D light field in front of each eye. Conceptually, this approach allows for parallax over the entire eyebox to be accurately reproduced, including monocular occlusions, specular highlights, and other effects that cannot be reproduced by volumetric displays. However, current-generation light field displays provide limited resolution (Lanman 2013, Hua 2014, Huang 2015) whereas holographic displays suffer from speckle and have extreme requirements on pixel sizes that are not afforded by near-eye displays also providing a large field of view.

3. Emerging Computational Near-eye Display Systems

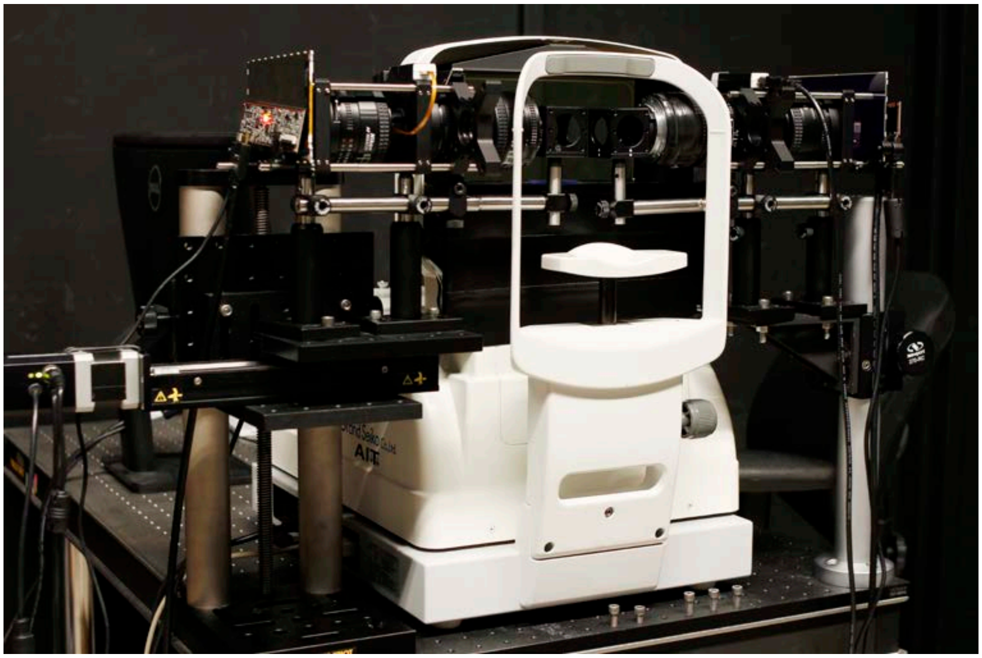

Figure 2. Prototype focus-tunable stereoscopic display. This setup allows for a range of different focus-tunable and monovision display modes to be tested with user studies. An autorefractor is integrated in the setup to measure where a user accommodates for a displayed stimulus. The outcome of these studies will inform the design of future near-eye displays.

In this work, we ask whether it is possible to provide natural focus cues and to mitigate visual discomfort using focus-tunable optics, i.e. programmable liquid lenses. For this purpose, we demonstrate a prototype focus-tunable near-eye display system (Figure 2) that allows us to evaluate several advanced display modes via user studies. Detailed results and insights of these studies will be presented in the accompanying Frontiers of Engineering Symposium talk.

The following display modes will be discussed:

Conventional Near-eye Displays are simple magnifiers that enlarge the image of a microdisplay and create a virtual image at some fixed distance to the viewer.

Adaptive Depth of Field Rendering is a software-only approach that renders the fixated object sharply while blurring other objects according to their relative distance. When combined with eye tracking, this mode is known as gaze-contingent retinal blur (Mauderer 2014). Due to the fact that the human accommodation system may be driven by the accommodation-dependent blur gradient, this display mode does not reproduce a physically correct stimulus.

Adaptive Focus Display is a software/hardware approach that either changes the focal length of the lenses or the distance between the micro display and the lenses (Konrad 2015). When combined with eye tracking, this mode is known as gaze-contingent focus. In this mode, the magnified virtual image observed by the viewer can be dynamically placed at arbitrary distances, for example at the distance where the viewer is verged (requires vergence tracking) or at the depth corresponding to their gaze direction (requires gaze tracking). No eye tracking is necessary to evaluate this mode when the viewer is asked to fixate on a specific object, for example one that moves.

Monovision refers to a common treatment for presbyopia, a condition that often occurs with age in which people lose the ability to focus their eyes on nearby objects. To improve visual clarity, monovision places two lenses with different prescription values in front of each eye such that one eye dominates for distance vision and the other for near vision. Monvision was recently proposed and evaluated for emmetropic viewers in VR/AR applications (Konrad 2015).

4. Discussion

The primary insights of our studies are that both the focus-tunable mode and the monovision mode demonstrate improvements over the conventional display, but both require optical changes to existing VR/AR displays. A software-only solution (i.e. depth of field rendering) proved ineffective. The focus-tunable mode provided the best gain over conventional VR/AR displays. We implemented this display mode with focus-tunable optics, but it could also be implemented by actuating (physically moving) the microdisplay in the VR/AR headset.

How does this study inform next-generation VR/AR displays?

Based on our study, we conclude that the adaptive focus display mode seems to be the most promising direction for future display designs. Dynamically changing the accommodation plane depending on the user’s gaze direction would improve visual comfort and realism in immersive VR/AR applications in a significant way.

However, refractive errors, including myopia and hyperopia, have to be corrected adequately with the near-eye display, so the prescription of the user must be known or measured. For presbyopic users, dynamically changing the accommodation plane would almost certainly always create a worse experience than the conventional display mode. Since presbyopes cannot actually accommodate, it is crucial for the display to present a sharp image within the user’s accommodation range, rather than attempting to drive their accommodation somewhere where they cannot accommodate. This would result in compromised image sharpness.

Therefore, a personalized experience that adapts to the user, whether emmetropic, myopic, hyperopic, or presbyopic, is crucial to deliver the best possible experience.

References

1. T. Sugihara, T. Miyasato “A lightweight 3-D HMD with Accommodative Compensation”, SID Digest 29(1), 1998
2. E. Peli “Optometric and perceptual Issues with Head-mounted Displays”, Visual Instrumentation: Optical Design & Engineering Principles, McGraw-Hill, 1999
3. M. Mauderer, S. Conte, M. Nacenta, D. Vishwanath “Depth Perception with Gaze-contingent Depth of Field”, ACM SIGCHI 2014
4. R. Konrad, E. Cooper, G. Wetzstein "Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus", ACM SIGCHI, 2016
5. B. Schowengerdt, E. Seibel “Trued 3-D Scanned Voxel Display using Single or Multiple Light Sources”, Journal of the Society of Information Displays 14(2), 2006
6. E. Dolgoff “Real-depth Imaging: A New 3D Imaging Technology with Inexpensive Direct-view (No Glasses) Video and other Applications”, Proc. SPIE 3012, 1997
7. K. Akeley, S. Watt, A. Girshick, M. Banks “A Stereo Display Prototype with Mutliple Focal Distances”, ACM Transactions on Graphics (SIGGRAPH) 23(3), 2004
8. J. Rolland, J. Krueger, A. Goon “Multifocal Planes Head-mounted Display”, Applied Optics 39(19), 2000
9. M. von Waldkirch, P. Lukowicz, G. Troester “Multiple Imaging Technique for Extending Depth of Focus in Retinal Displays”, Optics Express 12(25), 2004
10. S. Liu, D. Cheng, H. Hua “An Optical See-through Head Mounted Display with Addressable Focal Planes”, Proc. ISMAR 2008
11. G. Love, D. Hoffman, D. Hands, J. Gao, J. Kirby, M. Banks “High-speed Switchable Lens Enables the Development of a Volumetric Stereoscopic Display”, Optics Express 17(18), 2009
12. P. Llull, N Bedard, W. Wu, I. Tosic, K. Berkner, N. Balram “Design and Optimization of A Near-eye Multifocus Display System for Augmented Reality”, OSA Imaging and Applied Optics, 2015
13. D. Lanman, D. Luebke “Near-eye Light Field Displays”, ACM SIGGRAPH Asia 32(6), 2013
14. H. Hua, B. Javidi “A 3D Integral Imaging Optical See-through Head Mounted Display”, Optics Express 22(11), 2014
15. F.C. Huang, K. Chen, G. Wetzstein "The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-Eye Light Field Display with Focus Cues", ACM Transactions on Graphics (SIGGRAPH), 2015