Comfort in Head-Mounted Displays

Christopher Stanwyck

Christopher_Stanwyck@student.uml.edu

Abstract

As virtual and augmented reality environments pick up steam head-mounted displays are proliferating as the primary visual device. This technology creates the illusion of a truly 3D environment, but at the potential cost of user discomfort. This paper explores the causes and potential mitigation for user visual discomfort.

Introduction

In Virtual Reality (VR) and Augmented Reality (AR) systems, 3D environments viewed on head-mounted near eye displays. These images are displayed in such a way that each eye of the user sees a different image to create a stereoscopic cue. This stimulates the vergence effect and the appearance of an object at a distance further than the actual screen. The ability to accommodate and focus on objects of varying distances is also an important cue for depth perception. Since accommodation and vergence responses are nearly linear, they are tightly coupled in the human visual system (HVS). However, there are no cues in mainstream HMDs that create the accommodation cues. This results in the vergence-accommodation (VA) conflict which is the primary driver of visual discomfort for users wearing HMDs since the user must force an accommodation to an unexpected distance which conflicts with the expected distance based on the vergence effect. There are a number of research areas in HMDs that aim to create or remove the need for the accommodation cues so that the VA conflict can be resolved and user comfort can be increased. As the effect of the VA conflict can be seen in as little as 25 minutes and has a building and lasting effect [Hua], solving this problem is crucial for the mainstream adoption of HMDs.

Accommodation Invariant Displays

One approach to solving the VA conflict is to create a display that allows the eyes to accommodate to any distance and still see an in-focus image. This can be done using the principles of a pinhole camera. If the display is created in such a way as to create an infinite depth of field, the eye will be able to accommodate to any distance and still see an in-focus image.

One problem with this technique is that restricting the image aperture to create an infinite depth of field results in an extremely narrow margin for aligning the user's eye to the image. It also restricts the field of view since the user would not able to look around or else the pupil would not be aligned with the optics. However, there are other techniques which can create an accommodation invariant display.

In "Accommodation-invariant Computational Near-eye Displays", the authors use another technique for creating an accommodation invariant display. Rather than use a pinhole camera effect, they use a focal sweep via a focus-tunable lens. A user will perceptually integrate the varying focal depth into one coherent image and focus on the image that matches the accommodation state driven by binocular disparity. In this way, no matter what the eyes accommodate to, there will not be a significant blur. This concept, however, suffers the same limitations as the pinhole camera technique. The computations involved create a resulting image which suffers from a loss of resolution compared to an image with a single focal depth.

[Reference AI paper]

Varifocal Displays

Another approach for solving the vergence-accommodation conflict is to use a focus-tunable lens to match the focus of the image to the expected accommodation state. In “Accommodation and Comfort in Head-Mounted Displays”, the authors show objectively that this method, along with depth-of-field rendering, produces near accurate accommodation responses (as compared to the provided stimulus). The focus tunable lens is calibrated for each user and is controlled electronically to match the simulated depth of a simulated object in the scene.

To make such a technique work in a real world scenario, the focus must be allowed to dynamically adjust per object that the user focuses on. This requires eye-tracking and gaze contingent rendering approaches along with one or more depth maps for the displayed scene. This technique may be limited by the latency of the eye tracking software and the fact that retinal blur must be rendered and is not accurate (only near-accurate).

Focal Surface Displays

Matsuda et al. propose another display type for handling the VAC, a focal surface display. This technique uses a spatial light modulator (SLM) to program a spatially varying focal length so that the focus can change across the image. The authors create planes or “surfaces” for each scene which are rendered at a specific focal depth to reduce computational complexity. Their results show that with just 3 surfaces the focal cues for a scene can be rendered very accurately. They found that just a small focus disparity is within a comfortable margin for viewing and triggering correct accommodation.

This technique is able to render high resolution images with near correct focal cues. Like other implementations, it has its tradeoffs. Notable issues include chromatic aberrations. This is due in part to the wave-length dependence on focal length using the SLM. Some aberrations can be corrected for. The authors could resolve this with time multiplexing but the desire is to have all colors displayed simultaneously. Lastly, this technique currently suffers from a restricted field of view which is a limitation of the current optics and refresh-rates.

Holographic displays

Near-eye Fresnel holographic displays are another potential solution to producing focal cues which can solve the VAC. In “Holographic Near-Eye Displays for Virtual and Augmented Reality”, Maimone et al. propose several prototypes which demonstrate parts of the capabilities required for a holographic display. This display makes use of a SLM to steer the light through phase modulation to provide a holographic image. In theory, this method would allow for a truly per pixel depth cue to be rendered since it creates a “true” 3D image. In practice, this is still computationally complex so the authors rely on eye-tracking and foveated rendering so that only part of the display requires high resolution computation. The other parts can use a low resolution computation since the periphery in the HVS is not as detailed. Their algorithm also divides the hologram in to sub-holograms due to limitation in spatial frequencies of the SLM since a high spatial frequency will increase the degree of deflection past what the SLM can support. The SLM used also had other drawbacks in resolution and refresh rate.

The authors also claim that this method can be used for vision correction so that additional corrective lenses are not required. The size and sleekness of HMDs is important for mainstream use so this is an important feature. They did not show a fully integrated device however due to technical reasons and the fact that the optical stack is still quite large. New advances in optics or computational methods will be required to further shrink the solution.

Analysis

Many new display technologies for near-eye VR and AR displays rely on computational methodologies to correct for optical deficiencies. Many systems are limited by FOV, resolution, refresh rates, combined size of optics, contrast etc. The additional focal cues needed to resolve the VA conflict add additional burdens on optics and computational cost. Depth of field rendering and gaze contingent rendering have been shown to be valuable both for reducing computational complexity as well as providing a more realistic model of the HVS and possibly reducing discomfort due to the VAC. True accommodation cues must be generated with accurate blur rather than a rendered blue to have a significant effect on VAC.

Another common core component has been the use of an SLM for control of light waves. This has been used to create changes in focus as well as create holography. Current SLM technology has imposed limitations on several different techniques. This is mainly due to the limited resolution, limited refresh rates, and limited FOV. It is expected that as the technology in displays continues to improve that the limitations of these methods will slowly dissipate.

In addition, several techniques make use of planes to encode depth information or scenes with varying rendered depths. Some form of time or spatial multiplexing is used in these cases to create more computationally efficient images. This shows another trade-off between image quality, VAC resolution, and frame rate.

Conclusions

Each approach has its strengths and weaknesses, and each approach seeks to create an optimal balance between current technological limitations and design goals. Improvement in optical technologies as well as computer hardware will add de-facto improvements to each of the aforementioned technologies. Future areas of research for addressing VAC will benefit from improved computational algorithms which appear to be able to correct for any remaining optical deficiencies. As optical and processing hardware continue to improve, technologies which have long been waiting for emergence or re-emergence (holograms and VR itself) will be re-visited and used to create truly 3D scenes with correct focal cues.

Bibliography

Konrad, Robert and Padmanaban, Nitish and Molner, Keenan and Cooper, Emily A. and Wetzstein, Gordon. Accommodation-invariant Computational Near-eye Displays. *ACM Transactions on Graphics 36, 4 (2017).*

Matsuda, Nathan and Fix, Alexander and Lanman, Douglas. Focal Surface Displays. *ACM Transactions on Graphics 36, 4 (2017).*

Koulieris, George-Alex and Bui, Bee and Banks, Martin S. and Drettakis, George. Accommodation and Comfort in Head-mounted Displays. *ACM Transactions on Graphics 36, 4 (2017).*

Hua, Hong. Enabling Focus Cues in Head-Mounted Displays. *Proceedings of the IEEE 105, 5 (2017).*