Ultimate augmented reality displays with passive optics: fundamentals and limitations

Barmak Heshmat*[a], Leihao Wei[b], Moqian Tian[a]
[a]Meta Augmented reality, 2855 Campus drive, San Mateo, CA, USA 94403; [b]University of California at Los Angeles, 420 Westwood Plaza Los Angeles, CA, USA 90095.

ABSTRACT

We first discuss the ultimate specifications of an augmented reality display that would saturate the human perception. Thereafter our study identifies fundamental limitations and trade-offs enforced by laws of optics for any augmented reality display that uses passive optical elements such as visors, waveguides, and meta-surfaces to deliver the image to the eye. The limitations are categorized into 7 rules that optics designers must consider when they are designing augmented reality glasses. These rules are directly drawn from Fermat's principle, perturbation theory, linear optics reciprocity, and human visual perception principles. Based on psychophysical theories we further work toward defining and quantizing levels of depth that would saturate the human depth perception. Our results indicate that passive optics acts as a passive system with less than unity pulse response function that would always reduce the performance of the original light source. Additionally, our investigations reveal the dynamics between allocation of depth levels and number of depth levels for ultimate lighfield experiences.

Keywords: Augmented reality, virtual reality, head mounted displays, lightfield, field of view, accommodation and vergence, depth perception, waveguide, visor

1. INTRODUCTION

While mixed reality lightfield head-mounted displays with eyeglass-like appearance and full 154° per eye field-of-view (FoV) is considered as the ultimate specifications of future head mounted displays, in practice, the advancement toward such form factor and capabilities has been delayed by numerous engineering challenges. In the last decade, transparent slab waveguides have paved the way for smaller formfactors [1], and active optical components and computational methods have enabled low-resolution lightfield capabilities [2]. While there have been many studies that have proposed innovative solutions to these challenges, there have been very few studies on what are the actual fundamental trade-offs and limitations enforced by physics of light, geometry of human eye, and human visual perception limits.

This study discusses and compares the fundamental limitations and trade-offs that exist between different parameters of an arbitrary augmented reality display based on passive optical elements and pinpoints several intrinsic limitations and presents guidelines for designing optimal augmented reality experiences. In summary we pin point seven rules that should be considered when designing an augmented reality headset with passive optics and we present a theoretical framework for designing ultimate light-field experiences based on sparse monocular depth allocation.

2. FUNDAMENTAL LIMITATIONS SET BY PASSIVE OPTICS

Both waveguides and visors are passive optical elements that convey or form the image that is generated by an external electronic source (usually either a projector, a micro display, or an LCoS display). This has been and most likely will be the dominant approach toward designing augmented reality displays in the coming decade for two major reasons: first, the advancements in lithography techniques and electronics have provided lower cost, higher resolution, and yet smaller form factors for such image engines, which makes them appealing for wearable headsets; and second, transparent electronics and other active optical components (e.g., tunable lenses, switchable Bragg gratings, etc.) are still not fully mass-producible for display purposes and they typically suffer from diffraction, speed, or efficiency issues.

Optical Data Science II, edited by Bahram Jalali, Ken-ichi Kitayama, Proc. of SPIE Vol. 10937, 1093707 · © 2019 SPIE · CCC code: 0277-786X/19/$18 · doi: 10.1117/12.2505803

Although appealing for mass-production, using an external electronic image engine with a set of passive optics to deliver the image to the eye comes with several fundamental drawbacks:

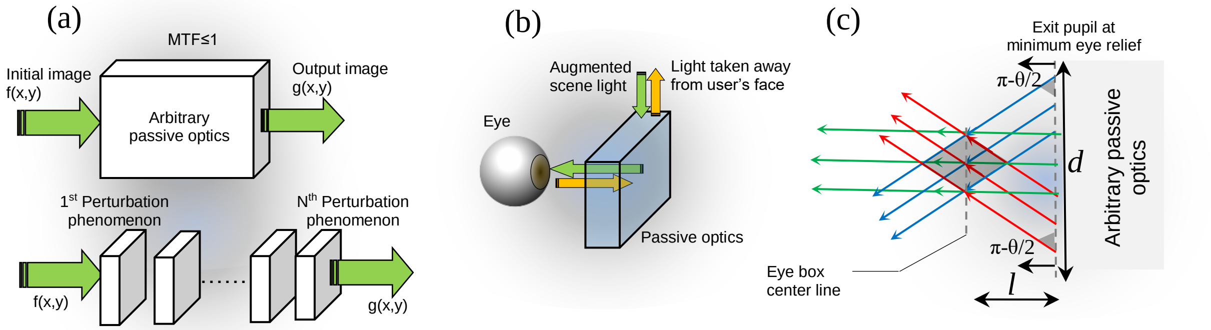

Figure 1. (a) Top diagram: the image accuracy will never be better than the original image source. Bottom diagram: as the number of perturbation phenomena increases, on average the color accuracy drops. (b) Due to reciprocity, a more efficient optics appears darker from the outside world. (c) Schematics of light exiting an arbitrary passive optics of an AR headset. *d* is the exit pupil extent, *l* is the minimum eye relief, *θ* is the FoV.

2.1 Rule 1: Sub-unity MTF effect

The image quality delivered through passive optics to the eye *cannot be better* than the original quality of the image produced by the electronic engine. This is a fundamental degradation that happens naturally due to absorption, spectral characteristics of components, inaccuracy in surfaces, unwanted reflections, and many other deficiencies in passive optical components. This is the same reason why a camera lens can never capture an image that is better than the original scene in terms of spatial or color accuracy. Therefore, based on this principle, it is impossible to get better pixel accuracy, color accuracy, or brightness than the original image that was produced by engine using passive optics such as waveguides or visors. In practice, the image quality *can only get worse*, as there are no perfect modulation transfer function (MTF) optical components (Fig. 1 (a) top).

2.2 Rule 2: Accumulative degradation from perturbation

By increasing the number of perturbation events (refraction, reflection, diffraction) in the path of light from the engine to the eye, *on average, the color inaccuracy (color nonuniformity or cross talk) can only increase* compared to engine image (original light source that feeds the passive optics). This is independent of optical design and is the direct result of wave behavior of light. Since perturbation events (e.g., reflecting from a surface, refracting through a periodic structure, or transmitting through multiple refractive index changes) induce diffraction or scattering due to natural geometrical imperfection of surfaces at atomic level, one can expect to always lose some photons into scattering and diffraction once the light passes through a perturbating structure. This fundamental drawback might not be a notable concern for visor-based designs with few reflections, but it can be critical for multilayered diffractive waveguides, as each layer or nanostructure along the way of light will scatter some of the light to the wrong direction and thus cause color inaccuracy (Fig. 1 (a) bottom).

2.3 Rule 3: Reciprocity vs transparency trade-off

Based on the Helmholtz reciprocity principle for passive optical elements, *the increase in efficiency of the optics on delivering the light from the image source to the eye in an arbitrary fixed geometry is monotonically tied to reducing the user's eye visibility from the outside world*. This is a substantial observation which is a direct result of reciprocity; the passive optical elements that bring the light to the eye would have to also work in the opposite direction for rays of light (Fig. 1 (b)), and therefore, would have to take the reflection from the surface of the user's eye and face back to the image engine. This intrinsically will make the eyes less visible from outside, which means the visors or waveguides would seem more dull or dark to an outside viewer. Obviously, some designs are less efficient than others while being darker to outside world.

Based on these three fundamental trade-offs and drawbacks it is fair to say that an AR headset based on passive visors or passive waveguides can only have comparable, but not better image accuracy than conventional flat monitors or projectors with the same optical specifications. This is a fundamental limit on optical quality of the image and does not indicate that AR headsets cannot have an edge over conventional monitors in terms of applicability, portability, or 3D perception. Further, based on the third trade-off, to have an AR headset that has more transparent appearance from outside world, at some point one must sacrifice the efficiency of the optics and pump higher intensity of light into the system while keeping the optical efficiency of the components low.

Additionally, it is obvious that each geometry for passive optics will enforce its own set of limitations and trade-offs between form factor, FoV, and eye box. Here we assume that the optics delivers a wavefront that mimics a lightfield generated from a natural scene in front of the eye (Fig. 1 (c)). In such geometry, which covers majority of waveguides and visors, there are the following relations and trade-offs between the given parameters.

2.4 Rule 4: Larger FoV increases the exit pupil size based on Fermat's principle

If l is the minimum eye relief (minimum distance allowed from surface of eye lens plane positioned at the middle of eye box to the closest optical surface) and θ is the horizontal or vertical FoV, then the exit pupil extension d of the passive optics at minimum eye relief increases with FoV by at least $2l\tan(\theta/2)$. Based on Fermat's principle of light the optical surface cannot be smaller than the exit pupil extension at minimum eye relief.

2.5 Rule 5: Larger FoV reduces eye box depth based on conservation of etendue

If etendue of the light bundle for each point in the scene is kept constant, the eye box depth (the distance ranging from the exit pupil plane to where the entire image is visible) *can only get smaller by increase in FoV*. Also, the decrease of the eye box depth with increase in FoV is on the order of $-d/(4\sin^2(\theta/2))$.

2.6 Rule 6: larger eye box reduces brightness

For a constant optical power exiting from the exit pupil, the increase in eye box vertical or horizontal extent will always reduce the brightness of the perceive image. This reduction in brightness is linear with the increase in one axis and quadratic with increase in both horizontal and vertical axes.

2.7 Rule 7: Additive augmentation reduces image dynamic range or contrast

The additive nature of augmented reality with passive optics means that the light intensity from visual content is added to the light intensity from outside environment. This reduces the dynamic range of the image (typically by orders of magnitude) as below:

$$D_{total} = \frac{I_{ARMax}+\alpha I_{SCMax}}{I_{ARMin}+\alpha I_{SCMin}} \tag{1}$$

$$D_{AR} = \frac{I_{ARMax}}{I_{ARMin}+\alpha I_{ave}} \ll \frac{I_{ARMax}}{I_{ARMin}} = D_{El} \tag{2}$$

In equation (1) D_{total} is the total dynamic range of the image seen through the glasses which includes both the scene and the augmented image. Here I_{AR} and I_{SC} is the augmented image and environment image intensity as function of x and y and the *Max* and *Min* subscripts indicate the maximum and minimum of these intensity values. The image addition or transparency coefficient $0<\alpha<1$ depends on the transparency of the glasses; in and ideal hypothetical case α=1and the glass is fully transparent with no surface reflection. In practice this coefficient is between 0.3 and 0.7. Equation (2) indicates the augmented image dynamic range D_{AR}. This value is more important than D_{total} since due to uniform attenuation of the environment intensity α with the glasses the dynamic range of the environment image is not impacted by the passive optics. Here I_{ave} is the average intensity of the environment. For example, this average increases significantly at outdoor compared to indoors. As noted, the D_{AR} is reduced notably for a more transparent glass. At completely occluding case where the glasses are dark (α=0) the dynamic range of augmented image reaches the dynamic range of the electronic engine (D_{El}).

So if one starts with an electronic engine with very high dynamic range, the additive nature of passive optics is still expected to reduce that dynamic range significantly. Furthermore, there is a direct relation between dynamic range of augmented image D_{AR}, maximum optical power of the engine P_{ARMax} (engine refers to the light source that feeds the passive optics), the efficiency of the optics η (such as visor or waveguide) to bring engines light to the eye, the transparency coefficient α, environment average brightness I_{ave} , and area of the eye box A_{box} as indicated in equation below.

$$D_{AR} \approx \frac{I_{MaxAR}}{\alpha I_{ave}} = \frac{\eta(\alpha)P_{ARMax}}{\alpha I_{ave}.A_{box}} \leq \frac{P_{ARMax}}{\alpha I_{ave}.A_{box}} \tag{3}$$

Equation 3 shows that dynamic range has a reverse and, in most cases, nonlinear relation with transparency and eye box area. In other words, based on Eq. (3) with the same peak power one can expect a lower dynamic range for a more transparent looking AR glasses or glasses with larger eye box.

As noted in this section, the laws of physics impose a set of fundamental limitations and trade-offs on performance and form factor parameters of AR displays. These limitations are independent of the optical design or fabrication technology used to make the passive optical surfaces. We need to note such guidelines when passive optics is considered for engineering the augmented reality headsets. However, in addition to such guidelines there are some guidelines that are imposed by human visual perception.

3. ULTIMATE SPECIFICATIONS SET BY HUMAN VISUAL PERCEPTION

Human visual perception is rather an extended topic with rich literature focusing on different parameters [3-10]. In most cases its extremely difficult to pinpoint one fixed number that models a certain aspect of human visual perception. For example its difficult to assign one (x,y) resolution number to the eye since eye is really not a fixed camera and the retina sensing mechanism is vastly different from that of an electronic sensor. To make things even more complicated the perception of the image is rather subjective as the signals are processed with brains and both eye and brain processing is known to vary slightly in between population and even with age. However, such complexity does not mean that there is no way to estimate or characterize the visual perception over a defined population. We like to approach visual perception from display perspective rather than phsychophysics perspective. For instance, we would like to know what is the (x,y,z) resolution that would saturate the average eye at age of 30; what is the field of view that would be sufficient for most population, etc. Specifying each of these parameters with indication of their dynamics in one study is not possible so we would use the estimates that has been found in different studies on spatial, temporal, spectral and aberration modeling of human eye [5-10] and just note the numbers briefly here. Table. 1 shows the estimates that would fully saturate the human visual perception for an average healthy eye.

Table 1. Parameters for ultimate augmented and virtual reality experience based on human visual perception and existing standards in display industry. There is not much information on the depth.

Parameter	**Practically sufficient**	**Ultimate saturation of eye perception**
Field of view/eye (θx,θy)	128º×100 º	154 º×120 º with corners rounded
x,y resolution/eye field of view	7680×4320 (33Mpix or 8K/eye)	154×60×120×60 (66.5Mpix)
Spatial acuity (H,V)	60×43pix/degree	1×1 arc min or 60×60pixel/degree
Temporal resolution	240Hz @ RGB	398Hz@B 800Hz@RG
Color	Adobe RGB with 16Bit/channel	Open for debate
Peak dynamic contrast	1000:1	1000,000:1 in the dark
Monocular resolution	?	Varies with age
Stereoscopic acuity	0.5 arc min	0.17 arc min
Binocular depth resolution	2800 levels	Open for debate

As noted in Table 1. while there are at least rough estimates for majority of parameters there is less attention being paid to monocular depth perception specifically depth perception from display perspective. Since majority of light field display modalities have been rather recent [11-15]; there is no clear notion of what is the limit of human eye saturation in depth and what are even elementary lightfield depth levels that should be created in a discrete quantized level to provide the richest experience.

Monocular depth perception has been pondered over in psychophysics, human perception, neuroscience literature with completely different approach [16-19]. Although informative, these psychophysical studies are human centric with no specific compatibility or interest in digital augmentation of human perception with emerging display technologies. For example, while there are scattered studies on human eye depth of field for different age groups and pupil conditions, there is no general or scientifically grounded conception of a technical guideline or tool to design a VR or AR displays that would satisfy such perception acuity.

In order to quantize the monocular depth levels one has to consider eye diopter range, depth of field and its relation to pupil size. Each of these parameters varies with lighting condition and age and can dramatically change the number of depth levels that are distinguishable. If the largest diopter range reported in the literature for very young eye (15 diopter [20]) is divided by the shallowest depth of field reported in the literature (0.15D full width half maximum [21]) then the 100 depth levels are the absolute maximum number that human eye at age of 10 can distinguish with 6-8mm pupil size. However, if one assumes an average of 6D range for adults with 0.15D depth of field, this maximum number of depth levels reduces rapidly to 40 levels which is a more practical estimation for average young eye. The depth of field significantly varies with pupil size [21] and the accommodation range significantly varies with age [23]; therefore, its essential to consider these two parameters in laying out the physical localizations of these depth levels. Here we use an iterative method to localize the depth levels up to 10 meters.

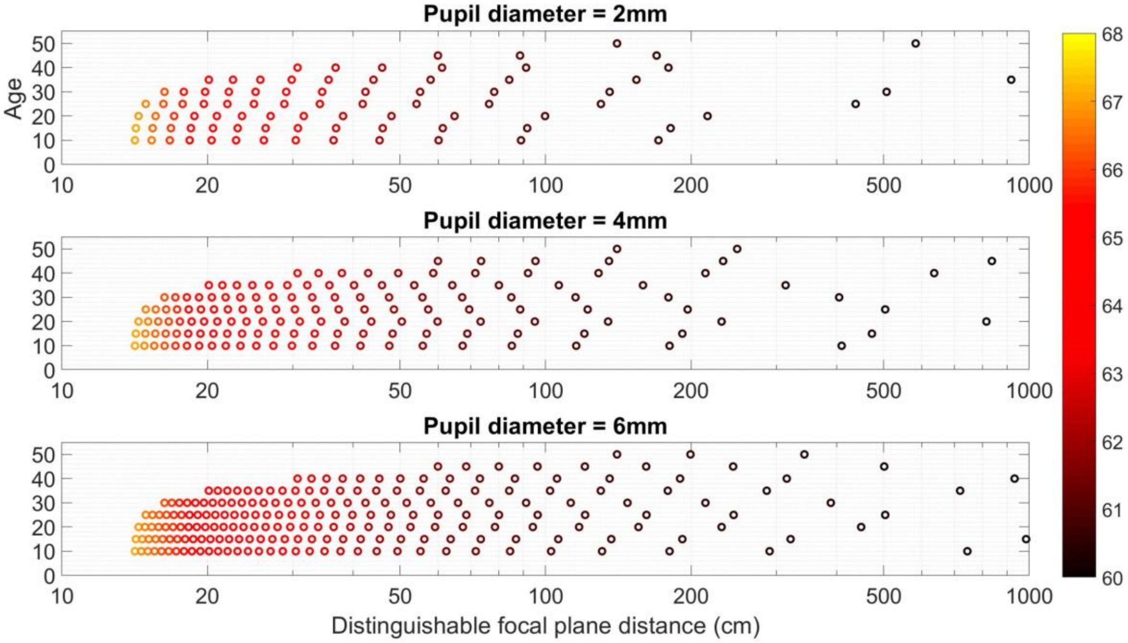

Figure 2, Monocular distinguishable focal planes with age and pupil diameter variations. Focal plane distances are within 10 meters. Color bar shows the corresponding total eye diopter (60 from a relaxed eye+accommodation). As one grows older, the distance to the nearest focal plane becomes larger and number of total focal planes decreases, e.g.. with pupil size equal to 2mm, one is able to distinguish 13 focal planes at age of 10 but can distinguish only 2 focal planes at 50.

DOF (Depth-of-field) for 2mm, 4mm and 6mm pupil diameter has been previously studied in [21,23]. Anderson et al. [24] has used the objective method to measure the accommodative amplitude in a wide age range of individuals, and has given a sigmoidal function fit to the measured data. The function was used to find the max accommodative amplitude. We started at the nearest focal plane given by this max accommodation and iteratively find the next focal plane at a step of DOF from [24]. The iteration stops when the focal plane distance is larger than 10m (Fig. 2.)

As noted in Fig. 2 the ultimate number of depth levels that would saturate the monocular perception varies from 40 at darkest environment (pupil diameter of 6mm) down to only 2 or 3 depth levels at bright environment for elderly eye. To find an estimate of number of levels needed in average population in average condition we considered the average display brightness of 250nits that would contract the pupil to 3mm [23] and calculated the depth locations based on age as in Fig. 3. As noted for average age of 30-40, which includes the median age of majority of countries, the maximum number of depth levels that is distinguishable is only 10-12 levels. This number can be substituted for the question mark on Table 1.

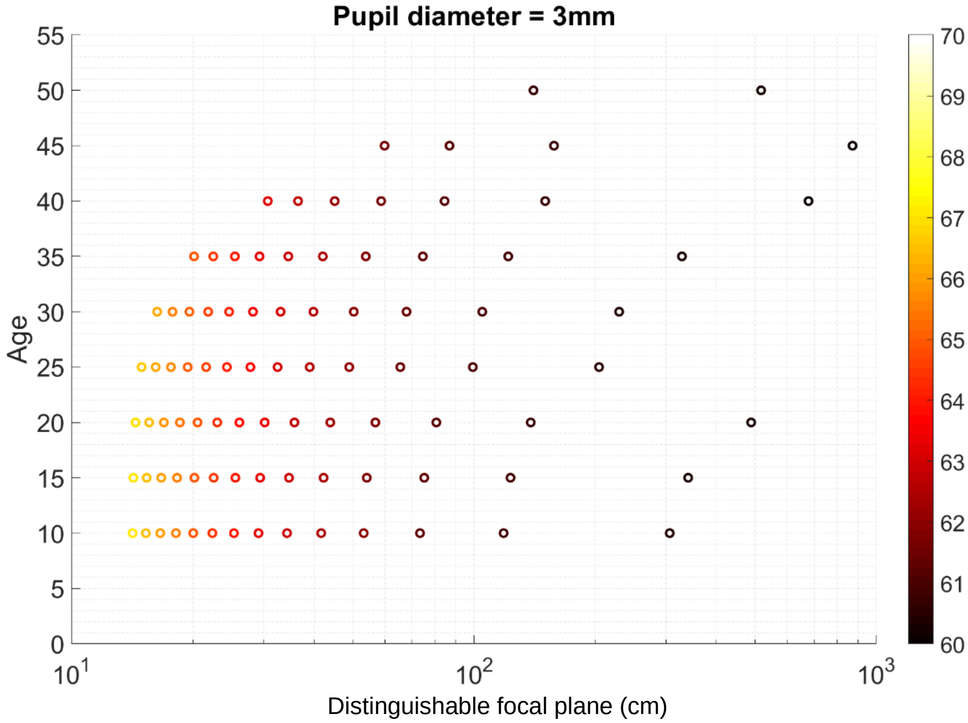

Figure 3, Monocular distinguishable focal planes with age at 3mm pupil diameter. Focal plane distances are within 10 meters. Color bar shows the corresponding total eye diopter (60 from a relaxed eye+accommodation).

Given the statistics of eye diopter across different ages and based on daily task operations one can use optimization to find the priority of allocating depth levels. This is significant in design of 3D displays as the bandwidth is limited and only limited number of monocular depth levels can be rendered. We show this optimized localization of the depth levels in our future studies.

In conclusion, through this brief study we have highlighted seven fundamental limitations that is imposed by physics of passive optics and highlighted the ultimate specifications that is needed based on the saturation of human visual perception. We found that only about 10 depth levels are enough to fully saturate the human monocular depth perception at average display brightness.

REFERENCES

[1] Hua H., and Javidi B., "A 3D integral imaging optical see-through head-mounted display," Opt. Express 22, 13484-13491 (2014).

[2] Liu C., Qiu J., and Zhao S., "Iterative reconstruction of scene depth with fidelity based on lightfield data," Appl. Opt. 56, 3185-3192 (2017).
[3] Parker A. J., “Binocular depth perception and the cerebral cortex,” Nat. Rev. Neuroscience 8, 379–391 (2007).
[4] Bülthoff I., Bülthoff H., & Sinha P., “Top-down influences on stereoscopic depth-perception,” Nat. Neuroscience 1, 254–257 (1998).
[5] Davis J., Hsieh Y. & Lee H. “Humans perceive flicker artifacts at 500 Hz,” Nat. Sci. Rep. 5, 7861 (2015).
[6] Navarro R., Santamaría J., and Bescós J., "Accommodation-dependent model of the human eye with aspherics," J. Opt. Soc. Am. A 2, 1273-1280 (1985).
[7] Kuppuswamy M., Lakshminarayanan P., Lakshminarayanan V., "Color Vision and Color Spaces," Optics & Photonics News 30, 44-51 (2019).
[8] Paus T., “Location and function of the human frontal eye-field: A selective review,” Neuropsychologia, 34, 475-483 (1996).
[9] Van Nes F.L. and Bouman M. A., "Spatial Modulation Transfer in the Human Eye," J. Opt. Soc. Am. 57, 401-406 (1967).
[10] Hirsch J Curcio C, “The spatial resolution capacity of human foveal retina,” Vision Research 29, 1095-1101 (1989).
[11] Fattal D., Peng Z., Tran T., Vo S., Fiorentino M., Brug J., & Beausoleil R. G., “A multi-directional backlight for a wide-angle, glasses-free 3D display” Nature 495, 348–351 (2013).
[12] Huang H. and Hua, H., "Systematic characterization and optimization of 3D light field displays," Opt. Express 25, 18508-18525 (2017).
[13] MacKenzie K. J., Hoffman D. M., & Watt S. J., “Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control,” J. Vis. 10, 22 (2010).
[14] Lueder E., 3D Displays, John Wiley & Sons, (2012).
[15] Wetzstein, D. Lanman, W. Heidrich, Raskar R., “Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays,” ACM Transactions on Graphics (ToG) 30, 95 (2011).
[16] Tsushima Y., Komine K., Sawahata Y., Hiruma N., “Higher resolution stimulus facilitates depth perception: MT+ plays a significant role in monocular depth perception.” Sci. rep. 4, 6687, (2014).
[17] Fulvio J. M., & Rokers B., “Use of cues in virtual reality depends on visual feedback,” Nat. Sci. Rep. 7, 16009 (2017).
[18] Granrud C., Yonas A., Pettersen L., “A comparison of monocular and binocular depth perception in 5- and 7 month-old infants.” J. of exp. child psychology 38, 19-32 (1984).
[19] Ginis H., Perez G., Bueno J., Artal P., “The wide-angle point spread function of the human eye reconstructed by a new optical method,” J. of Vis. 12, 20-20 (2012).
[20] Mordi J., Ciuffreda K., “Static aspects of accommodation: age and presbyopia.” Vis. Research 38, 1643-1653 (1998).
[21] Marcos S., Moreno E., Navarro R., “The depth-of-field of the human eye from objective and subjective measurements,” Vis. Res. 39, 2039-2049 (1999).
[22] Lo´pez-Gil N., Ferna´ndez-Sa´nchez V., Legras R., Monte´s-Mico´ R., Lara F., et. al. “Accommodation-Related Changes in Monochromatic Aberrations of the Human Eye as a Function of Age.” Invest. Opthal. & Vis. Sci. 49, 1736 (2008).
[23] Watson A., Yellott J., “A unified formula for light-adapted pupil size.” J. of Vis. 12,12-12 (2012).
[24] Anderson H., Hentz G., Glasser A., Stuebing K., Manny R., “Minus-Lens–Stimulated Accommodative Amplitude Decreases Sigmoidally with Age: A Study of Objectively Measured Accommodative Amplitudes from Age 3,” Invest. Opthal. & Vis. Sci. 49, 2919 (2008).