

12-9-2022

Augmented reality fonts with enhanced out-of-focus text legibility

Mohammed Safayet Arefin
arefin@acm.org

Follow this and additional works at: <https://scholarsjunction.msstate.edu/td>



Part of the [Graphics and Human Computer Interfaces Commons](#)

Recommended Citation

Arefin, Mohammed Safayet, "Augmented reality fonts with enhanced out-of-focus text legibility" (2022).
Theses and Dissertations. 5666.
<https://scholarsjunction.msstate.edu/td/5666>

This Dissertation - Open Access is brought to you for free and open access by the Theses and Dissertations at Scholars Junction. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholars Junction. For more information, please contact scholcomm@msstate.libanswers.com.

Augmented reality fonts with enhanced out-of-focus text legibility

By

Mohammed Safayet Arefin

Approved by:

J. Edward Swan II (Major Professor)
Carlos R. Montalto Cruz (Co-Advisor)
Shahram Rahimi

J. Adam Jones

T. J. Jankun-Kelly (Graduate Coordinator)
Jason M. Keith (Dean, Bagley College of Engineering)

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi

December 2022

Copyright by
Mohammed Safayet Arefin
2022

Name: Mohammed Safayet Arefin

Date of Degree: December 09, 2022

Institution: Mississippi State University

Major Field: Computer Science

Major Professor: J. Edward Swan II

Title of Study: Augmented reality fonts with enhanced out-of-focus text legibility

Pages of Study: 136

Candidate for Degree of Doctor of Philosophy

In augmented reality, information is often distributed between real and virtual contexts, and often appears at different distances from the viewer. This raises the issues of (1) *context switching*, when attention is switched between real and virtual contexts, (2) *focal distance switching*, when the eye accommodates to see information in sharp focus at a new distance, and (3) *transient focal blur*, when information is seen out of focus, during the time interval of focal distance switching. This dissertation research has quantified the impact of context switching, focal distance switching, and transient focal blur on human performance and eye fatigue in both monocular and binocular viewing conditions. Further, this research has developed a novel font that when seen out-of-focus looks sharper than standard fonts. This *SharpView* font promises to mitigate the effect of transient focal blur. Developing this font has required (1) mathematically modeling out-of-focus blur with *Zernike polynomials*, which model focal deficiencies of human vision, (2) developing a focus correction algorithm based on *total variation* optimization, which corrects out-of-focus blur, and (3) developing a novel algorithm for measuring font sharpness. Finally, this research has validated

these fonts through simulation and optical camera-based measurement. This validation has shown that, when seen out of focus, SharpView fonts are as much as 40 to 50% sharper than standard fonts. This promises to improve font legibility in many applications of augmented reality.

Key words: augmented reality, sharpview, out of focus, zernike, point spread function, blur estimation, sharpness, TV deconvolution, focal distance switching, focal blur, optical see-through display.

DEDICATION

Dedicated to my son—Farzad Sanfeer Arefin

ACKNOWLEDGEMENTS

I am highly grateful to my advisor, Dr. Ed Swan, for his guidance, support, suggestions, and encouragement throughout my Ph.D. journey. He is the best advisor anyone could ever have. Still, I remember the days when I joined the Spatial Perception and Augmented Reality (SPAAR) lab at Mississippi State University as a highly motivated newbie graduate student. From that newbie to the end of my Ph.D. journey, Dr. Ed left no stone unturned to inspire and motivate me in everything I intend to do in my research. Besides science, I enjoyed our lab lunch or dinner, where we discussed many topics, including foods, family, culture, and others.

I want to thank and express my gratitude to my co-adviser Dr. Carlos Montalto. Carlos became my co-advisor during a critical phase of my Ph.D. journey. I still remember the days when I spent hour after hour and the day after day understanding and implementing Carlos's approach of focus correction. Finally, I approached Carlos for collaboration on this research. After that, my Ph.D. work got the booster to move to the next level. I am thankful to Carlos for his very well thoughts, suggestions, and help in developing the mathematical models. Besides, I want to thank my other committee members for their insightful suggestions and comments on my research.

I thank all my lab members during my tenure at the Spatial Perception and Augmented Reality (SPAAR) lab. Among all the members, I want to give special thanks to Nate Phillips. What a fantastic person he is! He is the first person who showed me the calibration procedure of the AR haploscope along with Dr. Ed. We spent many productive hours working and discussing research.

I acknowledge the National Science Foundation (NSF) for supporting this research. This research is based upon work supported by the National Science Foundation, under awards IIS-1937565 and IIS-1320909, to J.E. Swan II. Furthermore, I want to acknowledge the Center for Advanced Vehicular Systems at Mississippi State University, where the research was conducted.

I am grateful to my wonderful parents and my younger brother, for their unconditional support from halfway across the globe, in Bangladesh. I have no words to express how blessed I am to have them in my everyday life. From preschool to Ph.D., I have received endless love, support, and prayers from my parents. I believe my parents are pleased and proud of me now, and genuine happiness is when you are the reason for your parent's happiness. Besides, I am also indebted to my in-law's family for their unbelievable love, appreciation, and encouragement.

I am thankful for the unconditional love and support of my outlandish wife, Farzana. We got married in the middle of my Ph.D. program. The support and inspiration from Farzana helped me overcome the ups and downs of remaining part of my Ph.D.. More specifically, she did not make me feel any pressure or distraction in my Ph.D. journey. Thank you for everything.

During the last part of my Ph.D., we were blessed with a baby boy, Farzad Sanfeer Arefin, the best gift of my life. Also, the most significant motivation and encouragement of my life. When writing this sentence of my dissertation, he crawls in front of me and plays with his toys. What a beautiful moment for me! When I come back to our place from the lab, I see his smile, and his smile removes all the tiredness and stress. Thank you, Abbu. Abbu loves you a lot! I can't express my gratitude to my wife, Farzana, once again for taking care of our baby boy when I was busy with my dissertation tasks.

Finally, and most importantly, I acknowledge the Almighty for pouring positive vibes and blessings over me to complete my Ph.D. dissertation.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
I. INTRODUCTION	1
1.1 OST AR interface design issues	4
1.2 Research Goals	8
1.3 Use Cases of Short AR textual Information	10
1.4 Contributions	11
II. CONTEXT SWITCHING, FOCAL DISTANCE SWITCHING AND FOCAL BLUR	13
2.1 The Problem	13
2.2 Phenomenon	14
2.3 Motivation for Vision Science Approach	15
2.4 Background and Related Work	16
2.4.1 Accommodation and Vergence	16
2.4.2 Context Switching and Focal Distance Switching	18
2.5 Experiment	20
2.6 Analysis	24
2.6.1 Context Switching	26
2.6.1.1 Task Performance	26
2.6.1.2 Eye Fatigue	28
2.6.1.3 Discussion	29
2.6.2 Focal Switching Distance	31
2.6.2.1 Task Performance	32
2.6.2.2 Eye Fatigue	36
2.6.2.3 Discussion	36

2.6.3	Viewing	37
2.6.3.1	Task Performance	38
2.6.3.2	Eye Fatigue	38
2.6.3.3	Discussion	39
2.6.4	Transient Focal Blur	40
2.6.4.1	Letter Undercounts	41
2.6.4.2	Discussion	43
2.7	Conclusion	44
III.	OUT-OF-FOCUS VISUAL ABERRATION FOR AR SYSTEM	46
3.1	Background	47
3.1.1	Retinal Image formation	47
3.1.2	Basic Principle of Retinal Image Calculation	48
3.2	Description of Zernike Polynomials	52
3.2.1	Reason Behind using Zernike Polynomials	52
3.2.2	Applications of Zernike Polynomials	52
3.2.3	Limitations of Zernike polynomials:	53
3.2.4	Zernike Polynomials for the Low-Order Aberration	54
3.2.5	Zernike Coefficient for the Low-Order Aberration	55
3.3	Zernike Polynomials and Coefficient for the Out-of-Focus Aberration in AR	57
3.4	Results of Out-of-focus Aberration	64
IV.	OUT-OF-FOCUS CORRECTION TECHNIQUE FOR AR SYSTEM	69
4.1	Background and Related Work	69
4.1.1	Algorithm Based Focus Correction Through Image Processing	69
4.1.1.1	Refractive Human Vision Error Correction	70
4.1.1.2	Out-of-focus Correction in AR	71
4.1.1.3	Challenges with Algorithm Based Focus Correction Approach	72
4.1.2	Custom Hardware Based Focus Correction	74
4.1.2.1	Refractive Human Vision Error Correction	74
4.1.2.2	Near-eye AR Display Hardware	75
4.1.2.3	Challenges with Hardware Based Focus Correction Approach	79
4.2	Dissertation Research's Approach for Out-of-focus Correction in the AR system	79
4.2.1	Constrained Total Variation (TV) Based Out-of-Focus Correction for AR	80
4.2.1.1	Out-of-focus Correction Model Parameters	82
4.3	Initial Results	84
4.4	Final Pre-corrected Image	89
4.5	Analysis of Algorithmic Parameters	91

V.	EVALUATION OF OUT-OF-FOCUS CORRECTION TECHNIQUE FOR AR SYSTEM	98
5.1	Sharpness Measurement Method	99
5.2	Parameter Selection	102
5.3	Evaluation Methods	105
5.3.1	Synthetic Simulation Evaluation	105
5.3.2	Optically Viewed Camera Captured Image Evaluation	109
VI.	CONCLUSION AND FUTURE WORK	120
6.1	Limitations and Future Work	124
	REFERENCES	126

LIST OF TABLES

2.1	Experimental Design.	22
2.2	F -tests for each panel of Figure. 2.4	26
2.3	F -tests for each panel of Figure. 2.5	29
2.4	F -tests for each panel of Figure. 2.7	33
2.5	F -tests for panels of Figure. 2.7	33
2.6	F -tests for panels of Figure. 2.7	34
2.7	F -tests for panels of Figure. 2.8	35
2.8	F -tests for panels of Figure. 2.9	38
2.9	F -tests for panels of Figure. 2.10	38
2.10	F -tests for each panel of Fig. 2.12	43

LIST OF FIGURES

1.1	Optical see-through AR example and a conceptual diagram of an optical see-through AR display	2
1.2	Revolution of OST AR	3
1.3	Example of AR interface design issues.	5
1.4	Out-of-focus problem in OST AR system	7
1.5	Use cases of short AR text information	9
2.1	Vergence-accommodating conflict	18
2.2	Overall experimental setup	20
2.3	The experimental subtask.	21
2.4	Impact of context switching on performance	27
2.5	Impact of context switching on eye fatigue	28
2.6	Linear increase in fatigue in the experiment.	30
2.7	Focal switching distance on performance.	32
2.8	Focal switching distance on eye fatigue.	35
2.9	Impact of viewing on task performance.	37
2.10	Impact of viewing on eye fatigue.	39
2.11	Transient focal blur during focal distance switching.	41
2.12	Evidence of the transient focal blur effect.	42

3.1	Retinal image formation.	46
3.2	Zernike polynomials up to the fourth order and a Zernike polynomial pyramid . . .	54
3.3	Standard eye prescription with the spherical and cylindrical value.	56
3.4	Scenario 1- Retinal image formation during out-of-focus issue in AR.	60
3.5	Scenario 2- Retinal image formation during out-of-focus issue in AR.	60
3.6	Amplitude function calculation for out-of-focus aberration.	63
3.7	Output of the out-of-focus aberration	64
3.8	Pupil diameter in the retinal image quality	65
3.9	Effects of different pupil diameters in the retinal image quality	66
3.10	Effects of different out-of-focus aberration amount in the retinal image quality . . .	68
4.1	Output of the constrained TV-based image deconvolution algorithm by Montalto et al. [88] for the visually impaired individuals.	71
4.2	Output of the previous SharpView algorithm provided by Oshima et al. [64]. . . .	73
4.3	Application of the SharpView algorithm by Oshima et al. [64] on the out-of-focus textual information [31].	73
4.4	Results of the constrained TV-based out-of-focus pre-correction algorithm for the letter.	85
4.5	Results of the constrained TV-based out-of-focus pre-correction algorithm for word.	86
4.6	Results of the constrained TV-based out-of-focus pre-correction algorithm for phrase.	87
4.7	Process of generating final pre-corrected images.	91
4.8	Relationship among CL (%), θ and TV value.	92
4.9	CL(%) and θ for optimal TV-value over a range of pupil diameter.	94
4.10	Impact of pupil size on optimal final pre-corrected images	95

4.11	Mean time required to generate the pre-corrected images.	96
5.1	How can we define that the left image (sharper font) is sharper than the right image (blur font)?	99
5.2	Sharpness measurement method.	102
5.3	Results of the synthetic simulation evaluation	106
5.4	Synthetic simulation evaluation for letter	108
5.5	Synthetic simulation evaluation for word	109
5.6	Synthetic simulation evaluation for phrase	110
5.7	Augmented reality haploscope	111
5.8	Setup of the optically viewed camera captured image evaluation	112
5.9	Results of the optically viewed camera captured image evaluation	114
5.10	Optically viewed camera captured image evaluation for letter	115
5.11	Optically viewed camera captured image evaluation for word	116
5.12	Optically viewed camera captured image evaluation for phrase	117

CHAPTER I

INTRODUCTION

The great Greek philosopher Aristotle discussed the five primary senses (vision, sound, smell, taste, and touch) in his book *On the Soul* [9]. Vision is one of the human body's most potent and complex senses among those five senses. Via the sense of sight, the human can understand and perceive the visual information of his surroundings. So naturally, the human eye observes and perceives different real world objects with different properties (e.g., size, color, and others) at different distances and under lighting conditions in complex real world environments. Therefore, if real world environments have any additional graphical information along with the real world information, it will enhance or augment the visual sense and perception of the real world surroundings. To achieve this expanded vision, we need to regard augmented reality technology.

Augmented reality (AR) Augmented Reality (AR) is an emerging modern technology that superimposes computer-generated virtual information on the user's view of the real-world environment to enhance user vision and understanding of the real world surroundings. This technology can be achieved with head-mounted (HMD) AR displays, handheld devices (e.g., mobile, tablets, and others), a projector, and others. Consider a scenario where a surgeon uses an OST AR display (e.g., Microsoft HoloLens2) to perform a surgery in the operating room (see Figure 1.1a). During the operation, the OST AR display presents different information (individual patient anatomy via

MRI or CT scans) related to the real world patient. Therefore, AR technology allows surgeons to perform their duty efficiently and effectively. This is a single example from hundreds of use cases of the AR system. Due to the dramatic progress of research and the advancement of commercial AR devices, AR has been playing an essential role in many other applications such as medical applications [14, 115], navigation [91, 103], manufacturing & maintenance [38], education [18], and so on. As AR technology has been progressing rapidly and has enormous applications in many potential areas, it is necessary to identify and examine the current limitations of the AR system from the perceptual perspective.

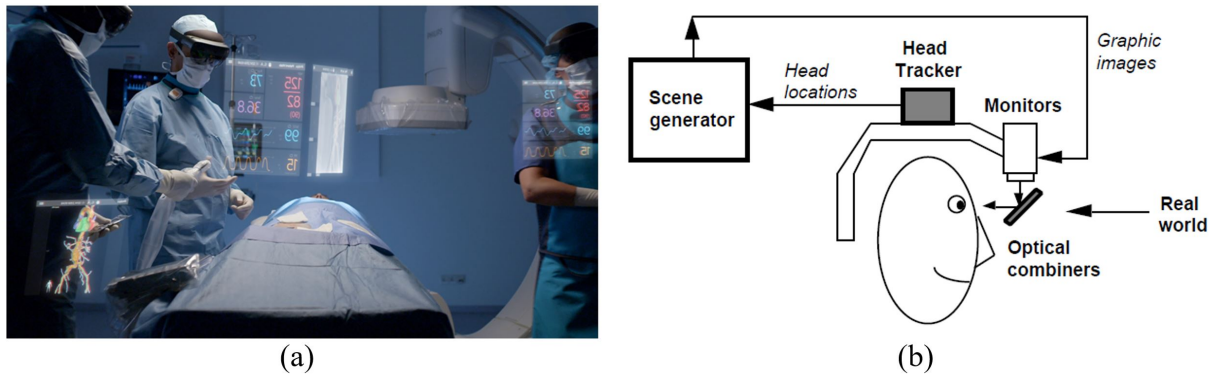


Figure 1.1: (a) Optical see-through AR example (AR-based medical surgery) [40]. (b) A conceptual diagram of an optical see-through AR display [14].

Optical see-through AR display One of the major categories of the AR head-mounted display (HMD) is the optical see-through (OST) AR display. In the OST AR system, users can view the real world environment through the optical combiner with their own eyes. Along with the real world information, users can also observe the virtual information through the semi-transparent partially reflective optical combiners [89, 90]. As both real and virtual information is formed in

the eye's retina, the human eye is the main observational component in the OST AR system. In this case, we do not have access to the underlying pixels of the virtual information formed on the retina. Therefore, it is challenging to understand how users perceive virtual information in the OST AR system. My dissertation research is focused on the perceptual issues of the OST AR display system. A conceptual diagram of the OST AR display is given in Figure 1.1. Figure 1.3a provides a view through the OST AR system as an example.

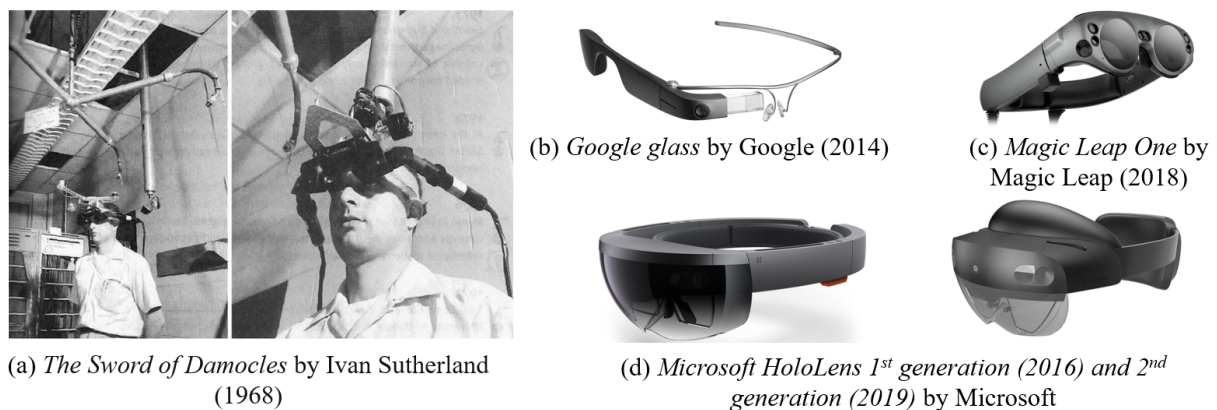


Figure 1.2: (a) First three-dimensional display (*The Sword of Damocles*) by Ivan Sutherland in 1968 [111]. (b) Google glass by Google [110]. (c) Magic Leap One by Magic Leap [106]. (d) Microsoft HoloLens 1st (left) and 2nd (right) generation displays [68].

Revolution of OST AR: AR HMD has a long history which started when Ivan Sutherland built the first three-dimensional display (*The Sword of Damocles*) in 1968 [111], where an image was displayed in such a way that moved with the movement of the user (see figure 1.2a). However, the term "Augmented Reality" was first coined by Boeing researcher Thomas P. Caudell in 1990 after combining head position sensing and real world rendering in the same system to display the virtual information [22]. AR technology's most rapid growth and development occurred at the beginning

of the 20th century when Kato et al. [58] developed software named *ARToolKit* to render virtual information on the real world by tracking the fiducial marker for a multi-user augmented reality video conferencing system.

After decades of steady but slow development by the researchers, the industrial revolution happened when some well-known tech companies (e.g., Google, Microsoft, and others) invested billions of dollars in OST AR hardware and software development. Google introduced their OST AR display *Google Glass* [110] in 2014 (see figure 1.2b). Later, in 2016, Microsoft brought a more advanced version of the OST AR display (Microsoft HoloLens), which has powerful and diverse functionality and interaction techniques [68]. Further, in 2019, Microsoft introduced the 2nd generation of the Microsoft HoloLens with updated software and hardware features (e.g., eye tracking capability) (see figure 1.2d). In 2018, Magic Leap also introduced their first OST AR system—*Magic Leap One*. The rapid growth of the development of software and hardware portions of the OST AR system is easily understandable from the figure 1.2. Therefore, it can be conclusively said that the rapid progress of AR research is still ongoing, and many researchers are uniformly adding new dimensions to the body of AR research.

1.1 OST AR interface design issues

When using OST AR displays, interacting with virtual information requires the user's eyes to focus on the optical depth of the display. Unfortunately, most modern OST AR displays place virtual information at a fixed focal distance (e.g., Microsoft HoloLens, Google Glass). Further, the real world is complex, and real world information can appear at different focal distances from the user. In the OST AR system, information is often distributed between the real world and virtual

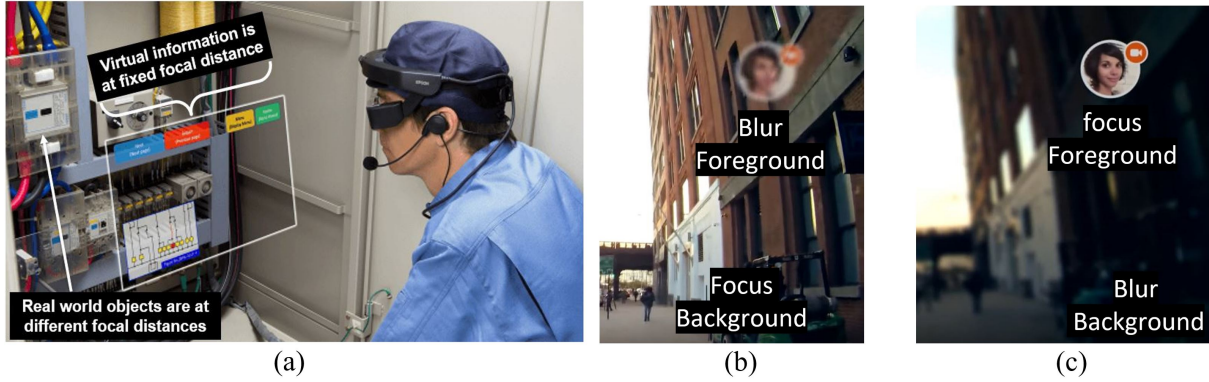


Figure 1.3: Example of AR interface design issues: context switching and focal distance switching. (a) The user needs to switch visual and cognitive attention while using an OST AR display to perform the maintenance task. This brings the issue of context switching. (b) and (c) These images are taken from a 2012 Google Glass concept video on YouTube [85]. When the user focus is on the background, virtual information (upper right) becomes blurry (b). When the user focuses on the virtual information (upper right), the background becomes blurry (c). This is because virtual information and the real-world background are positioned at different focal distances. If the virtual contents and the background appeared at the same focal distance, there would be no focal distance switching, but context switching would still exist. If no information is presented through the AR display, there would be no context switching, but there could still be focal distance switching between real world objects. Throughout this video, only the virtual symbology or the background is in focus at any one time, and the focus constantly switches between them while using the OST AR display. It presents the issue of focal distance switching.

contexts, appearing at varying distances from the user and needs to be precisely located in the real world environment [60, 61, 62]. Therefore, integrating information between real and virtual contexts raises the following issues:

(1) *Context switching*: During context switching, users must switch visual and cognitive attention between information sources. Consider a scenario (Figure 1.3a) where a technician performs a maintenance task using an OST AR display. The OST AR display displays additional graphical information related to the maintenance task. Therefore, the technician needs to continuously switch focus and attention to integrate information from real and virtual contexts. This brings the issue of

context switching. Thus, context switching problems arise when virtual information is presented through an AR display, and users need to integrate real and virtual information.

(2) *Focal distance switching*: Focal distance switching means accommodating (changing the shape of the eye's lens) between different focal distances to see information in sharp focus at a new distance. Consider another real-life scenario (Figure 1.3b) where video frames are taken from a Google Glass concept video on YouTube [85]. In this video, the user is walking around an urban setting, and only the virtual information or the background is in focus at any one time. Both real and virtual information is never in focus at the same time throughout the video. This is because virtual information and the real-world background are positioned at different focal distances. Therefore, the person's eye focus constantly switches between two different focal distances, and one piece of information becomes blurred when focusing on another information at a different focal distance. This brings the issues of *focal distance switching*.

(3) *Out-of-focus problem in OST AR*: One open question is how OST AR research investigates and mitigates the out-of-focus problem during switching focal distances. As mentioned above, during focal distance switching, users observe only one object (either real or virtual) in focus. Other information becomes blurred for a concise amount of time (around 360 milliseconds [21]). This brings the issue of the *out-of-focus problem* in the OST AR system. Let's consider that an AR text label (letter "A") is located at a closer focal distance (e.g., 0.67m) than the real world background (e.g., 4.0m) (see figure 1.4). In this case, if users focus on the AR letter, a sharp version of the letter will be created on the retina, but the background will be blurred, and the retinal image will be formed in front of the retina (see figure 1.4a and b). However, if the users' eyes are focused on the background, the regular AR font will be blurred, and a retinal image will

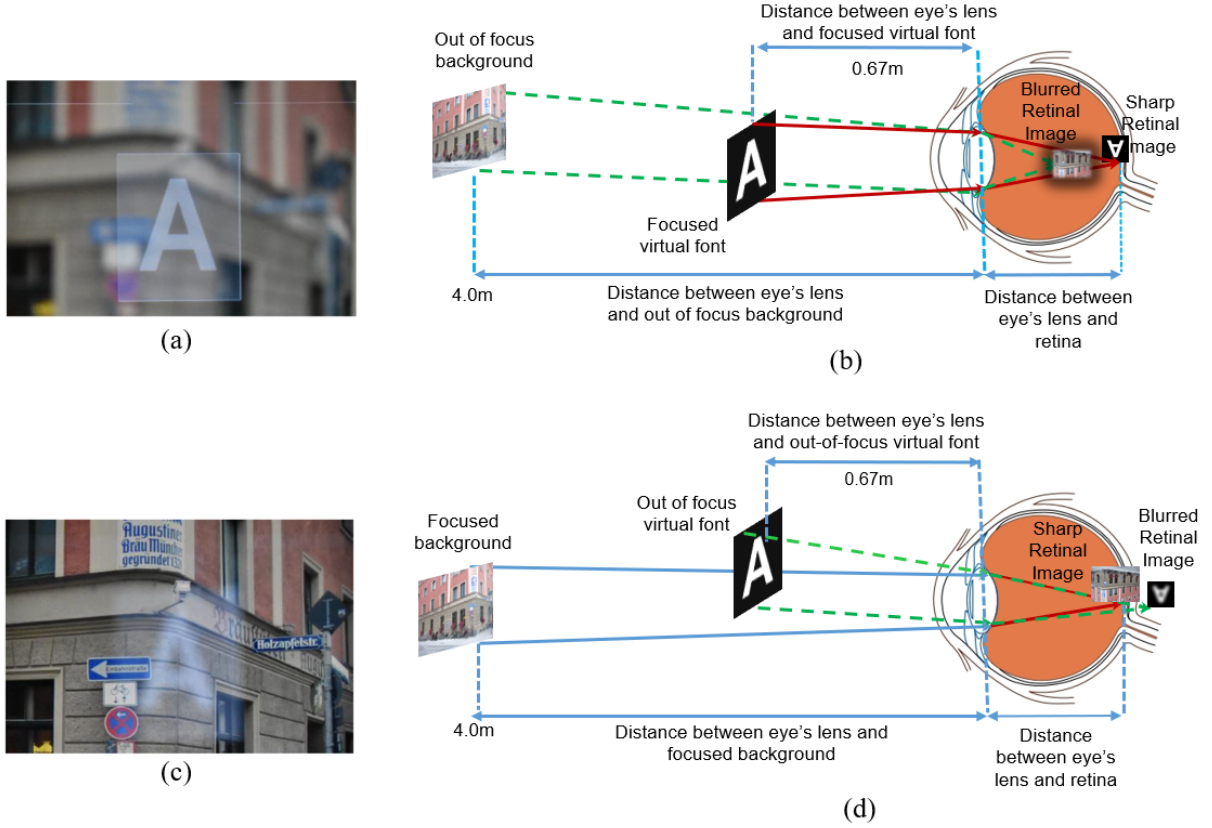


Figure 1.4: Out-of-focus problem in OST AR. An augmented reality (AR) text label (letter “A”) at a closer focal distance (e.g., 0.67m) than the background (e.g., 4.0m). (a) The background is blurred when focusing on the AR display using a regular font. (c) When focused on the background, the text is blurred. (b) and (d) show the corresponding mechanism of the formation of the retinal image during (a) and (c). Figures (a) and (c) are achieved optically.

be formed behind the retina (see figure 1.4c and d). Therefore, this brings the important research question—*How can we develop an OST AR system that can represent a sharper version of the blurred AR font (figure 1.4c) while focusing at a different focal distance?* My complete dissertation research is based on this open AR research problem.

Previous research has found that context switching, focal distance switching, and focal blur have a significant negative impact on human performance, leading to eye fatigue [71, 86] and other perceptual problems [42], including excessive strain on the accommodation mechanism of the eye,

decreasing visual performance [49], increasing the likelihood of missed targets [54], and limiting the ability to fuse binocular images [120]. Among all the previous research, Gabbard et al. [42] investigated AR display context switching and focal distance switching with a text-based visual search task using a monocular Nomad OST AR display. However, their experimental design did not wholly cross the conditions of context switching and focal distance switching and, therefore, could not fully consider how these conditions interact. Further, they expressed the concern that their findings might be specific to the Nomad AR display. Therefore, a more detailed investigation of AR display context switching, focal distance switching, and focal blur is essential.

1.2 Research Goals

The primary goal of my dissertation research was to develop and establish a perceptual image processing system that mitigates the out-of-focus blur issue in OST AR. Therefore, this research leveraged existing work in focus correction through image processing by developing and evaluating AR fonts with enhanced out-of-focus text legibility. This AR font is termed a “*Shaprview font*,” an AR font designed to look better (sharper) when seen out of focus than a regular font. This perceptual imaging algorithm has two main parts: *modeling the out of focus visual aberration of an OST AR system* and *modeling an out-of-focus correction method for an OST AR system*. The research in this document modeled the perceptual *out of focus blur* in the OST AR system using Zernike polynomials and adapted the Total Variation (TV) image deconvolution approach to generate the pre-corrected images. Although vision scientists have used similar techniques for individuals with refractive vision problems (e.g., problems corrected by glasses or contacts, such as myopia, hyperopia, and astigmatism) [88, 95, 52], these methods have never previously been

used to mitigate the out of focus problem in OST AR. Previously, researchers from the Spatial Perception and Augmented Reality (SPAAR) lab developed a SharpView algorithm for general images (e.g., the Lenna image, rock images, etc.) [64]. However, this algorithm did not consider depth while modeling the out-of-focus blur, and based the deconvolution algorithm on Wiener filtering [44]. Further, their algorithm did not show promising results for textual information [31]. This dissertation research overcomes these previous limitations, and develops an AR SharpView font using Zernike polynomials and an TV-based image deconvolution technique, and tests the method for out-of-focus AR textual information.



Figure 1.5: (a) A driver is driving a car while an AR head up display presents AR information that includes short text strings [94]. (b) A person is wearing a Google Glass AR display and walking in an urban street. An AR notification in the form of a short text string is displayed to provide the condition of the subway service [85]. (c) A person is performing a maintenance task while wearing a Microsoft HoloLens AR display. Maintenance information is presented as a short AR text string [112]. In all three pictures, AR textual information is in focus, and the background is blurred. If the focus is changed to the background, then the short AR text string will be out-of-focus, and perceived as blurred. At that moment, according to our hypothesis, the SharpView AR font will be beneficial.

1.3 Use Cases of Short AR textual Information

As mentioned in the previous section, this research considered textual information as the primary AR component; more specifically, *short AR text labels*. Therefore, it brings the question—“*Who will benefit from this research, and how will they benefit?*” Generally, various types of graphical content can be overlaid in a real-world environment, such as 3D holograms, games, text, and others. Among many AR components, text is one of the most used graphical components. Short AR Textual information has been widely used in many AR applications, including maintenance, education, navigation and driving, and others [4]. For example, we can consider the scenario where a person is driving a car, and short AR text strings are displayed through the AR head-up display (see Figure 1.5a [94]). If we carefully look at this picture, we can see that the driver is focused on the AR information, and the background road is blurred. If we consider that the driver changes their focus to the background road, then the short AR text string will be out-of-focus. At that moment, the font developed by this research (the SharpView font) will be effective and beneficial. This scenario is also true for Figures 1.5b [85] and c [112], where short AR text labels are used for notifications and maintenance instructions. These are only a few examples of short AR text label use cases; additional cases are given by Gattullo et al. [43]. They conducted a systematic survey to investigate what, how, and why visual assets were used in AR maintenance, assembly, and training, from 1997 to 2019. They reviewed 122 papers and found 348 visual AR assets. Among these, 91 were text based visual AR assets. Further, for AR maintenance tasks, the visual asset with the greatest usage is AR text strings. Therefore, this research implies that the SharpView font developed by this research could benefit hundreds of AR applications.

1.4 Contributions

This dissertation research makes four primary contributions.

Contribution 1: The first contribution is that this research systematically investigates the AR interface design issues of context switching, focal distance switching, and focal blur, for both monocular and binocular OST AR. To accomplish this research, a partial replication and extension of the task and experiment reported by Gabbard et al. [42] was conducted on the AR Haploscope [99], a custom-built OST AR display. This part of the research was published in *IEEE Transactions on Visualization and Computer Graphics* [6], published as a poster abstract at *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* [5], and constituted my master's thesis [4]. Chapter II summarizes this research.

Contribution 2: The second contribution is that this research modeled the human visual aberration (focal blur) for the out-of-focus problem in OST AR. The importance of this research is that without modeling the human eye's perceptual blur, the out-of-focus problem cannot be analyzed or corrected. Previous research has considered different image-based aberration modeling approaches, but their purpose was to model the refractive errors of observers viewing images without wearing their corrective optics. For example, Kruger et al. [69] developed a visual aberration technique for individuals with refractive vision problems, including myopia, hyperopia, presbyopia, and astigmatism. They further validated their model with camera-based testing. However, based on my knowledge, I did not find any previous work that modeled the human visual aberration (focal blur) for the out-of-focus problem in OST AR that considered the appropriate eye parameters.

Therefore, this part of the research adds a novel contribution to general academic knowledge (see Chapter III).

Contribution 3: The third contribution is that this research developed an image processing-based focus correction technique for the out-of-focus problem in OST AR. Montalto et al. [88] developed a novel approach to generate imagery with improved sharpness for individuals with refractive vision problems, including myopia, hyperopia, presbyopia, and astigmatism. However, focus correction techniques through image processing developed for the visually impaired has not been tested or validated for viewing out-of-focus virtual content in OST AR. Therefore, this research adopted the total variation-based deconvolution technique from Montalto et. al. [88] and Beck et al. [17] to improve the sharpness of out-of-focus text in OST AR. As a TV-based deconvolution algorithm has never previously been used in OST AR for focus correction, this research contribution adds a new dimension to the body of AR research. This part of the research task is described in Chapter IV.

Contribution 4: The final contribution of this research is that the image processing-based focus correction technique for the out-of-focus problem in OST AR has been evaluated, both synthetically and optically. Chapter V reports these evaluation methods.

CHAPTER II

CONTEXT SWITCHING, FOCAL DISTANCE SWITCHING AND FOCAL BLUR

The first research objective of my dissertation is to systematically investigate the impacts of OST AR display context switching, focal distance switching, and focal blur on human performance and eye fatigue. Based on my knowledge, Gabbard et al. [42] first explored the interaction between context switching and focal distance switching by conducting a text-based visual search task using a Microvision Nomad see-through AR display. This one-eye display uses laser-based retinal scanning technology. Later, to further investigate OST AR display context switching, focal distance switching, and focal blur, researchers at the SPAAR lab of Mississippi State University, including the author of this dissertation, successfully replicated this previous experiment, extended the findings for monocular and binocular viewing conditions, and addressed the limitations of the previous research. This dissertation research objective has been accomplished, was published in *IEEE Transactions on Visualization and Computer Graphics* [6], and constitutes this dissertation's author's master's thesis [4]. This chapter describes this successful research outcome and explains how these findings led to our research objectives.

2.1 The Problem

In OST AR, information is often distributed between the real world and virtual contexts, often appearing at different distances from the user. Therefore, users must repeatedly switch contexts and

refocus their eyes to integrate the information. Here, *context switching* refers to switching visual and cognitive attention between the real world and virtual information. Besides, *focal distance switching* refers to accommodating (changing the shape of the eye's lens) to see, in sharp focus, information at a new distance. During focal distance switching, users observe only one object (either real or virtual content) in focus, and other information becomes blurred for a very short amount of time (around 360 milliseconds [21]). This brings the issue of *focal blur*.

2.2 Phenomenon

Context Switching Context switching, or switching the context of the source of information, is known to be perceptually demanding [42] and to have performance consequences [116]. Because AR can provide overly information on real world contexts, AR information sources can be minimal, and well-designed spatially proximal AR information can cognitively complement real-world information, lowering both the physiological and cognitive overhead of switching contexts. However, despite this promise, very little research has recently quantified context switching in AR. Work done nearly a decade ago by Huckauf et al. [54] and Schwerdtfeger et al. [107] examined an industrial order-picking task and found that context switching between AR and real world displays resulted in decreased visual performance.

Focal Distance Switching and Focal Blur Focal distance switching is required to integrate information between objects at different focal distances. Suppose the users' task required integrating information repeatedly. In that case, focal distance switching leads to eye fatigue [71], biasing distance estimations, decreasing visual performance ([49]), increasing the likelihood of missed targets ([39]), limiting the ability to fuse binocular images ([120]), and many others perceptual issues.

Further, changing accommodation from one focal distance to another takes time. It is expected to take at least 350 to 425 milliseconds to completely accommodate to a new focal distance [21]. While accommodation changes, objects at the new focal distance will be seen with out-of-focus blur. If the task demands performance during this period, this transient focal blur could cause reduced visual performance. Indeed, Gabbard et al. [42] found a negative effect of focal blur on human performance during focal distance switching.

2.3 Motivation for Vision Science Approach

Among the previous research related to context switching and focal distance switching, only Hoffman et al. [49] used a custom-built laboratory display. All the others employed commercial, off-the-shelf OST AR displays, with inherent limitations in the consistency of presented depth cues. For example, Wang Baldonado [116] used standard computer monitors, Huckauf et al. [54] and Gabbard et al. [42] used monitors, Gabbard et al. [42] also used a Microvision Nomad OST AR display, Neveu et al. [92] used television and a Sony Glasstron, Imamov et al. [56] used an HTC Vive Pro, Eiberger et al. [41] used an Epson Moverio BT-100, and Drouot et al. [35] used a Microsoft HoloLens 2. While papers using off-the-shelf displays serve as essential foundations for examining the phenomena of context and focal distance switching, to understand why performance decrements are observed fully, a vision science approach is needed (e.g., [29, 47, 49, 66]). These and related papers have inspired two aspects of the approach reported here: (1) They generally use custom laboratory-built displays, which allow precise control over all relevant optical and visual parameters. And, (2) the experiments often include a monocular condition, where the non-dominant eye is covered. The monocular condition is motivated by the importance of stereo vision for many

human tasks, and the related depth cues of stereo disparity and ocular vergence [33]. Including a monocular condition, therefore, allows stereo vision effects to be separated from other effects, and helps explain experimental findings in the context of the human visual system. Therefore, a unique contribution of this research is that the phenomena of interest were carefully examined using a custom laboratory-built, OST AR display, and under both binocular and monocular viewing conditions. This allows replicating findings seen in previous studies using commercial off-the-shelf displays, and critically allows the findings to be attributed to specific elements of the human visual system. Such results could inform future AR hardware design and practitioners' selection of AR hardware features when considering use cases where focal and context switching demands can be predicted in advance. Accordingly, the purpose of the current experiment was to systematically investigate, in OST AR, the phenomena of context switching, focal switching distance, monocular and binocular viewing, and transient focal blur. This was accomplished by a partial replication and extension of the task and experiment reported by Gabbard et al. [42], on a custom-built optical testbed explicitly designed to examine these issues, i.e., an AR Haploscope (Figure 5.7).

2.4 Background and Related Work

2.4.1 Accommodation and Vergence

Both context switching and focal distance switching strongly correlate with two depth cues: accommodation and vergence. Accommodation means the ability of the eye to observe an object in sharp focus [119]. During accommodation, the eye adjusts its focal length based on the perceived object distance. When the eye accommodates a nearby object, the ciliary muscles contract, causing the lens to assume a more convex shape. In contrast, when a distant object is viewed, the ciliary muscles relax, causing the lens to become thinner and flatter. Along with accommodation, viewing

an object requires rotational vergence eye movements. When fixating on an object closer than where the eyes previously verged, the eyes converge and rotate towards each other. When fixating on an object farther than where the eyes previously verged, the eyes diverge and rotate away from each other. The primary stimulus that drives the vergence response is stereo disparity; when fixating on an object, the eyes verge until the images can be fused into a single image. The link between them is known as the vergence-accommodation reflex. Therefore, any changes in accommodation bring changes in vergence (accommodative vergence), and changes in vergence drive changes in the accommodation (vergence accommodation). However, both accommodation and vergence can be influenced by changing pupil size. When observing an object near accommodation distance, the pupils become smaller, whereas when observing an object at a far accommodation distance, the pupils become bigger. These components are interlinked physiologically and are known as *Near Triad* [109]. Therefore, changes in any element of the near triad drive change in the other two components.

However, to date, most commercial OST AR displays have generally presented virtual objects at a single fixed focal distance (e.g., Microsoft HoloLens (versions 1 and 2) and the Google Glass). Therefore, observing virtual contents through any single or fixed focal plane display, the accommodative depth cue is usually fixed at the focal distance of the display's optics. Consequently, as shown in Figures 2.1b and 2.1b, a viewer can be required to visually fuse an object with a vergence demand which is quite different from its accommodation demand. This inconsistency between the depth cues is called the accommodation-vergence mismatch problem — a pervasive problem AR displays. It can cause eye strain, double-vision, reduced user performance, and increased cognitive load [87, 67, 71, 119].

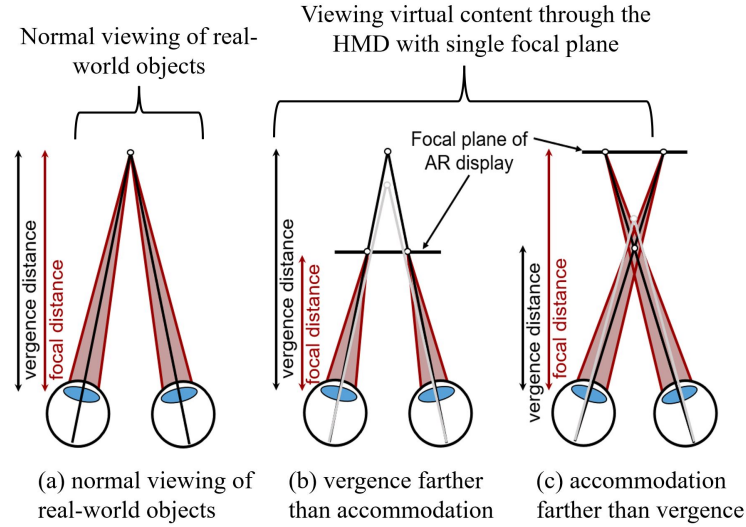


Figure 2.1: The vergence-accommodation conflict pictures adapted from Singh, Ellis, and Swan [109]. (a) During normal viewing, vergence, and focal distance match. (b) Viewing virtual content behind the focal plane of the HMD (vergence distance is farther than focal distance). (c) Viewing virtual content in front of the focal plane of the AR display (focal distance is farther than the vergence distance).

Further, changing accommodation and vergence to fixate on an object at a new distance takes time. Up to age 20, the human eye requires 360 milliseconds to accommodate from far to near and 380 milliseconds to accommodate from near to far. After the age of 20, the time needed to accommodate from near to far remains relatively constant, but the time required to accommodate from far to near increases [48, 21].

2.4.2 Context Switching and Focal Distance Switching

Although switching of context and focal distance in OST AR is frequent, only a few researchers have considered the impact on human performance and eye fatigue. Gabbard et al. [42] was the first to examine the interaction between context and focal distance switching, by conducting a text-based visual search task that required participants to integrate information from both real and

AR sources. As an experimental display, they used a Microvision Nomad see-through AR display. Their study's results revealed that context switching had a negative impact on performance when information was presented at 6 meters, but not at closer distances of 2 or .7 meters. However, context switching resulted in more significant eye fatigue at all three distances. Further, focal distance switching resulted in reduced performance, and additional performance reductions were attributed to the transient focal blur experienced while switching focal distances. However, their experimental design did not wholly cross the conditions of context switching and focal distance switching and, therefore, could not fully consider how these conditions interact. Moreover, their experiment used binocular vision while wearing an AR display covering only one eye. We term this condition *semi-binocular* viewing. In addition, their experiment did not examine monocular or binocular viewing. Eiberger et al. [41] examined the combined effects of context and focal distance switching by simulating environmental viewing with an Epson Moverio BT-100 display, which presents collimated imagery (infinite focal distance) at a stereo disparity of 3.7 meters. This was compared to a projected image at .3 meters (a typical smartwatch distance). They conducted a graphical visual search task. Context and focal distance switching resulted in a higher task completion time and a larger error rate. Recently, Drouot et al. [35] also examined the combined effect of context and focal distance switching, using a Microsoft HoloLens 2 at 1.5 or 2.0 meters disparity distances, and a graphical visual search and target detection task. They found a negative performance effect for distance switching, but no impact for context switching. Two previous studies have examined only AR context switching. The first was Huckauf et al. [54], who used semi-binocular viewing of a one-eye Microvision Nomad display focused at .61 meters and a monitor placed at the same distance. They found that context switching between the displays

reduced performance on several visual tasks. Most recently, Imamov et al. [56] investigated the issue of context switching by displaying information on two interfaces within a VR environment (simulating AR interaction). Their research found that context switching increased task completion time and decreased user comfort.

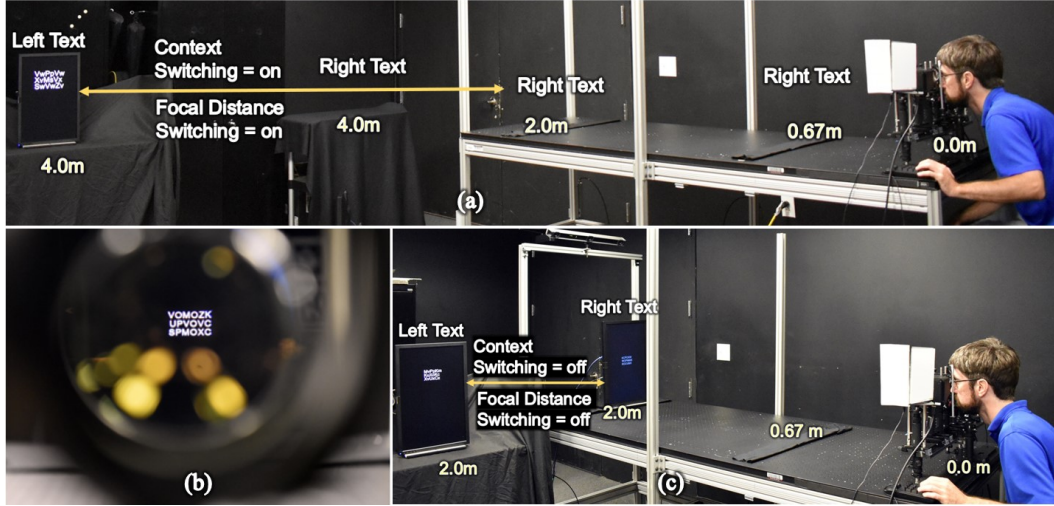


Figure 2.2: Participants performed a text searching task, in an experiment that switched contexts between the real world and augmented reality (AR), at either matched or mismatched focal distances. (a) A participant observed the left text on a monitor at 4 meters distance, and the right text in AR at one of three focal distances: 0.67, 4.0 or 2.0 meters. (b) View of the right text through the custom-built AR Hapscope. (c) The participant observed the left and right text on a monitor at 2 meters.

2.5 Experiment

The purpose of the current experiment was to systematically investigate the phenomena of AR display context switching, focal switching distance, binocular and monocular viewing, and transient focal blur. To accomplish this experiment, we used a custom-made *augmented reality hapscope*—an AR display mounted on an optical workbench (Figure 5.7). A detailed description

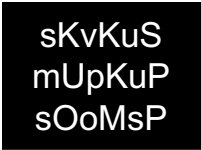
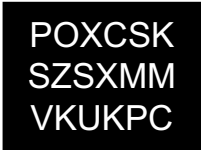
Task Description	Left text	Right text
Participants identified the doubled target letter ‘O’ in the left text and counted the number of target occurrences in the right text. Here, the correct answer is ‘1’.		

Figure 2.3: The experimental subtask. In Gabbard et al. [42], the left text was presented in either AR or the real world, while in the current experiment, the right text was presented in either AR or the real world (Figure 2.2).

of the experimental task, apparatus, setup, variables, setup, design, and procedure are given in Arefin et al. [6, 4]. Here, I summarize the overall experiment.

Experimental Task: To conduct this research, we adapted the visual search task of Gabbard et al. [42]. The main property of this task was that it forced participants to integrate information from two different text blocks. When one of these text blocks was presented in AR and the other in the real world, the task required context switching. Alternatively, the task did not require context switching when both text blocks were presented in the real world. In addition, the text blocks could be presented at different focal distances, or the same focal distance. The absolute difference between these focal distances was the amount of focal switching distance.

During the task, participants observed two side-by-side text blocks, the left text and the right text (Figure 2.3). Each text block comprised three text strings, and each text string contained six letters. The task consisted of a series of subtasks. Each subtask required three actions: (1) Searching the left text for the *target letter*, which was encoded by a pair of side-by-side identical letters, one upper case and the other lower case (e.g., “Oo” in Figure 2.3). In the left text, there was always exactly one target letter. (2) Searching for the target letter in the right text. The target

Table 2.1: Experimental Design. Viewing is encoded by the letter: m (monocular), b (binocular).

Real World to Real World Conditions (<i>Context Switching</i> = no)						
<i>Test Distance (R):</i>						
<i>Reference Dist. (R):</i>	R1 (0.67)	R2 (2.0)	R3 (4.0)	R1 (0.67)	R2 (2.0)	R3 (4.0)
R1 (0.67)	mR1R1	mR1R2	mR1R3	bR1R1	bR1R2	bR1R3
R2 (2.0)	mR2R1	mR2R2	mR2R3	bR2R1	bR2R2	bR2R3
R3 (4.0)	mR3R1	mR3R2	mR3R3	bR3R1	bR3R2	bR3R3
Real World to AR Conditions (<i>Context Switching</i> = yes)						
<i>Test Distance (A):</i>						
<i>Reference Dist. (R):</i>	A1 (0.67)	A2 (2.0)	A3 (4.0)	A1 (0.67)	A2 (2.0)	A3 (4.0)
R1 (0.67)	mR1A1	mR1A2	mR1A3	bR1A1	bR1A2	bR1A3
R2 (2.0)	mR2A1	mR2A2	mR2A3	bR2A1	bR2A2	bR2A3
R3 (4.0)	mR3A1	mR3A2	mR3A3	bR3A1	bR3A2	bR3A3

letter could appear at most once in each line of text, and could appear in total 0, 1, 2, or 3 times.

In Figure 2.3, the target letter “O” appears once, in the first line of text. (3) After counting the number of target letters in the right text, the target letter count was entered on a numeric keypad.

The task was to complete five subtasks within 25 seconds.

Experimental Variables and Design: The independent variables of our experiment were *context switching* (no, yes), *focal switching distance* (0, 1.33, 2, 3.33 meters), and *viewing* (monocular, binocular). Here, focal switching distance was a function of reference distance (.67, 2, or 4 meters) and test distance (.67, 2, or 4 meters). We measured five dependent variables: *number of subtasks completed*, *number of subtasks correct*, *undercount errors*, *overcount errors*, and *eye fatigue*.

We considered a within-subject experimental design, which is shown in 2.1. The upper half shows the real world to real world conditions, where no context switching occurred, while the lower half shows the real world to AR conditions, where context switching occurred. The left half shows

conditions viewed monocularly, while the right half shows conditions viewed binocularly. Within the resulting 2 (*context switching*) \times 2 (*viewing*) design, all 9 combinations of the 3 reference and 3 test distances were presented. The cells where the reference and test distances match are highlighted; these are cells where focal distance switching was not required. The cells in Table 2.1 uniquely label each combination of conditions. For example, cell mR2R3 indicates that participants monocularly viewed the reference text in the real world at a distance of 2 meters, and the test text in the real world at a distance of 4 meters. Here, when looking between the left and right text, the task did not require switching context, but did require switching focal distance by 2 meters. A 4×4 Latin square controlled the presentation order for each participant, and therefore the presentation order was fully counterbalanced for each group of four participants. Therefore, each participant completed 2 (*viewing*) \times 2 (*context switching*) \times 3 (*reference distance*) \times 3 (*test distance*) \times 5 (*repetition*) = 180 tasks, where each task comprised as many as 5 subtasks.

Twenty-four participants from the Mississippi State University community participated in this experiment; 12 were male, and 12 were female. The mean age of the participants was 22.9 years; ages ranged from 18 to 31. The participants had a mean inter-pupillary distance of 63.1 mm; 17 participants were right-eye dominant (71%), and seven left-eye dominant (29%), which agrees with the expected distribution of eye dominance [101]. No corrective vision restriction was provided to filter the participants; 13 participants wore corrective lenses, while 11 did not require correction. Participants were young enough not to exhibit presbyopia [36]. Each session lasted %1.5 hours, within a range of 1 hour for the fastest participants to 2 hours for the slowest.

2.6 Analysis

When one of the independent variables was continuous, data was analyzed by examining the slopes and intercepts of linear equations, and multiple regression was used to determine if the slopes and intercepts significantly differed [30, 96]. We presented the results in scatter plots such as Figure. 2.4. Here, the x axis, the *continuous independent variable*, gives the distance to the reference text, and the y axis, the *dependent variable*, shows the number of subtasks completed (upper row) and correct (lower row). The grey points are the number of subtasks completed for each (x, y) value. Context switching, the *categorical independent variable*, is indicated by the color and position of violin plots, which summarize the point distributions for each level of context switching. As indicated by the caption, each panel displays 144 grey points with substantial overlap. The multiple linear regression procedure, from Pedhazur [96] chapter 12, fits one or two linear regression lines in each panel of each scatter plot. The procedure is separately applied to each panel. Each graph has a corresponding table; Table 2.2 corresponds to Figure. 2.4. Each graph panel has a corresponding row in the associated table; note panels a to f in Figure. 2.4, and rows a to f in Table 2.2.

The multiple linear regression analysis proceeds in four steps:

Step 1: Two linear regressions are generated for each level of the categorical independent variable.

An F -test then determines if the slopes significantly differ. If they do, as in Figure. 2.4a (Table 2.2a: *slope diff*), both linear regressions are reported as the best overall description of the data in the panel. Two lines are drawn, and two linear equations are given. The interaction between the continuous and categorical independent variables is significant.

Step 2: If the slopes do not differ, then the slopes are set to a common value, and an F -test determines if the intercepts significantly differ. If they do, as in Figure. 2.5a (Table 2.3a: *intercept diff*), these two linear regressions are reported as the best overall description. Two lines are drawn, with a common slope, and two linear equations are given. The main effect of the categorical variable is significant.

Step 3: If the intercepts do not differ, as in Figure. 2.4b (Table 2.2b), then a single linear regression is reported as the best overall description. One line is drawn, and one linear equation is given.

Step 4: If the two slopes do not differ, an additional F -test determines if the single slope differs from 0. This can either be the common slope of two regressions, as in Figure. 2.5a (Table 2.3a: *slope 0*, same degrees of freedom as the intercept test), or the slope of a single regression, as in Figure. 2.4b (Table 2.2b: *slope 0*, one degree of freedom larger than the intercept test). If the slope differs from 0, then the main effect of the continuous independent variable is significant.

The multiple regression analysis yields two measures of effect size: (1) R^2 , the overall percentage of variation explained by the linear model, and (2) dR^2 , the percentage of variation explained by the categorical variable. Both R^2 and dR^2 are reported for every panel. In addition, if two linear regressions are reported, then d , the distance between the lines in y axis units, are reported. If the slopes differ (e.g., Figure. 2.4a), signed distances are reported for the leftmost and rightmost data points along the x axis (for Figure. 2.4a, $x = .7$ and 6 meters). If the slopes do not differ (Figure. 2.5a), an unsigned distance is reported. Sometimes the value of the slope, b , is also discussed.

Table 2.2: F -tests for each panel of Figure. 2.4

<i>Number of Subtasks Completed:</i>		
(a)	<i>Semi-Binocular, Nomad:</i>	
	slope diff:	$F_{1,140} = 16.1 \quad p < .001^{***}$
(b)	<i>Monocular, Haploscope:</i>	
	slope diff:	$F_{1,140} < 1$
	intercept diff:	$F_{1,141} = 1.7 \quad p = .19$
	slope 0:	$F_{1,142} = 11.0 \quad p < .01^{**}$
(c)	<i>Binocular, Haploscope:</i>	
	slope diff:	$F_{1,140} < 1$
	intercept diff:	$F_{1,141} < 1$
	slope 0:	$F_{1,142} < 1$

<i>Number of Subtasks Correct:</i>		
(d)	<i>Semi-Binocular, Nomad:</i>	
	slope diff:	$F_{1,140} = 16.4 \quad p < .001^{***}$
(e)	<i>Monocular, Haploscope:</i>	
	slope diff:	$F_{1,140} < 1$
	intercept diff:	$F_{1,141} < 1$
	slope 0:	$F_{1,142} = 5.5 \quad p < .05^*$
(f)	<i>Binocular, Haploscope:</i>	
	slope diff:	$F_{1,140} < 1$
	intercept diff:	$F_{1,141} < 1$
	slope 0:	$F_{1,142} < 1$

2.6.1 Context Switching

Context switching was expected to reduce task performance and increase fatigue (**H1**). Context switching was examined by comparing cells where context switching occurred, but focal distance was held constant: the shaded cells in Table 2.1 were compared between the conditions of context switching = *no* and context switching = *yes*. These cells contain 30% of the collected data.

2.6.1.1 Task Performance

The task performance effects of context switching and reference distance are analyzed in Figure. 2.4. The left-hand column shows the relevant data from Gabbard et al. [42] (display =

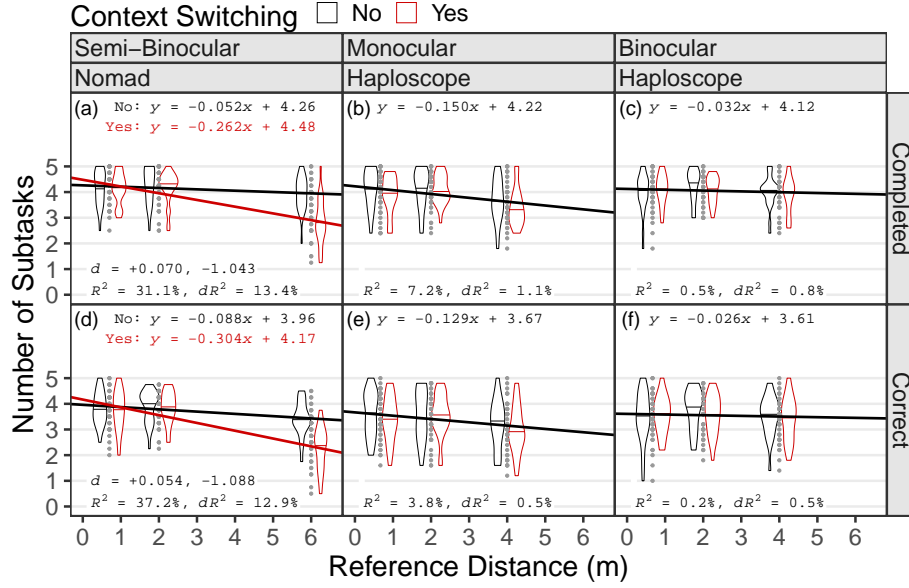


Figure 2.4: When examining data matched in depth, on the Nomad display (Gabbard et al. [42]), context switching reduced performance at the far reference distance. However, in the Haploscope context switching had no effect, and increasing reference distance only reduced performance under monocular viewing. *Nomad*: Data from highlighted cells in Tables 1 and 2 [42]: R1R1, R2R2, R3R3; V1F1R1, V2F2R2, V3F3R3. *Haploscope*: Data from highlighted cells in Table 2.1: mR1R1, mR2R2, mR3R3; mR1A1, mR2A2, mR3A3; bR1R1, bR2R2, bR3R3; bR1A1, bR2A2, bR3A3. Table 2.2 shows the related F -tests. Each panel displays 144 data points.

Nomad), under the *semi-binocular* viewing condition. The center and right columns show the data from the current experiment (display = *Haploscope*), under the viewing conditions of *monocular* and *binocular*. The upper row shows performance in terms of *subtasks completed*, while the lower row shows performance in terms of *subtasks correct*.

On the Nomad, there was a significant interaction between context switching and reference distance (panels *a, d*): at short distances of .7 and 2 meters, context switching had minimal effect, but at the longer distance of 6 meters, $d = 1.043$ (1.088) fewer subtasks were completed (correct). The effect is strong, explaining $dR^2 = 13.4\%$ (12.9%) of the variation. In the current experiment, this effect was not repeated; there was no effect of context switching with either monocular or

binocular viewing. However, there was an effect of reference distance: under monocular viewing, increasing reference distance resulted in reduced performance, at a rate of $b = .150$ (.129) subtasks completed (correct) per meter. While significant, this effect only explains $R^2 = 7.2\%$ (3.8%) of the variation, much less than what is explained for the Nomad. Under binocular viewing, there was no effect of either context switching or reference distance.

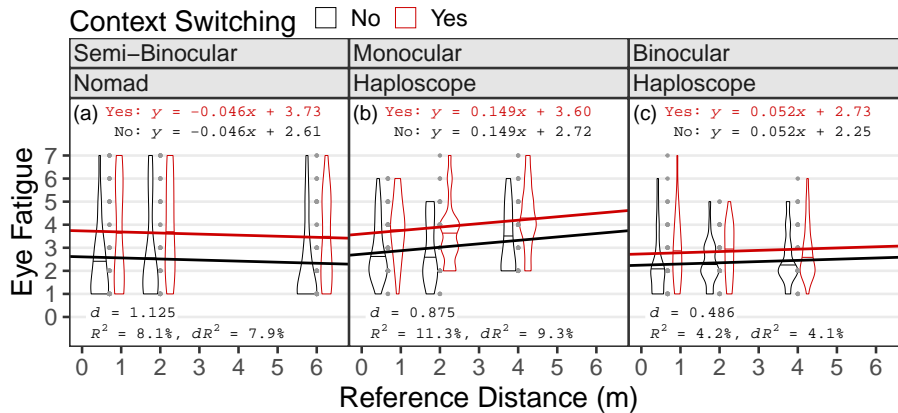


Figure 2.5: When examining data matched in depth, context switching increased eye fatigue for both the Nomad display (Gabbard et al. [42]) and the Haploscope. On the Haploscope, the amount of increased eye fatigue was higher with monocular viewing, compared to binocular viewing. Also, increased reference distance under monocular viewing resulted in greater eye fatigue. *Nomad*: Data from highlighted cells in Tables 1 and 2 [42]: R1R1, R2R2, R3R3; V1F1R1, V2F2R2, V3F3R3. *Haploscope*: Data from highlighted cells in Table 1: mR1R1, mR2R2, mR3R3; mR1A1, mR2A2, mR3A3; bR1R1, bR2R2, bR3R3; bR1A1, bR2A2, bR3A3. Table 2.3 shows the related F -tests. Each panel displays 144 data points.

2.6.1.2 Eye Fatigue

The eye fatigue effects of context switching and reference distance are analyzed in Figure. 2.5. Other than the y axis displaying fatigue, the graph structure is the same as Figure. 2.4. On the Nomad, there was a significant main effect of context switching on fatigue, but no impact of

Table 2.3: F -tests for each panel of Figure. 2.5

Eye Fatigue:

(a) <i>Semi-Binocular, Nomad:</i>			
slope diff:	$F_{1,140} < 1$		
intercept diff:	$F_{1,141} = 12.0$	$p < .01^{**}$	
slope 0:	$F_{1,141} < 1$		
(b) <i>Monocular, Haploscope:</i>			
slope diff:	$F_{1,140} < 1$		
intercept diff:	$F_{1,141} = 14.7$	$p < .001^{***}$	
slope 0:	$F_{1,141} = 3.2$	$p = .08$	
(c) <i>Binocular, Haploscope:</i>			
slope diff:	$F_{1,140} < 1$		
intercept diff:	$F_{1,141} = 5.7$	$p < .05^*$	
slope 0:	$F_{1,141} < 1$		

reference distance: context switching increased eye fatigue by $d = 1.125$ units at all distances. On the Haploscope, this effect was repeated. Under monocular viewing, context switching significantly increased eye fatigue by $d = .875$ units at all distances. In addition, there was a marginally significant main effect of reference distance ($p = .08$), where increasing distance resulted in increased fatigue, at a rate of $b = .149$ units per meter. Under binocular viewing, context switching also significantly increased eye fatigue by $d = .486$ units, with no effect on reference distance. Therefore, context switching increased eye fatigue across all conditions but with different magnitudes.

2.6.1.3 Discussion

It was hypothesized that context switching would decrease performance and increase eye fatigue (**H1**). Figures. 2.4 and 2.5 directly compare the previous findings with those of the current experiment. On the Haploscope, context switching did not affect task performance but increased fatigue. Therefore, the current results partially support hypothesis **H1**.

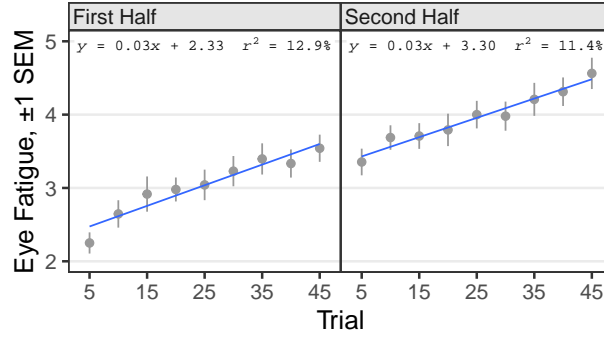


Figure 2.6: Across all sessions, there was a linear increase in fatigue as the experiment progressed. In each session, the level of context switching changed after 45 tasks, separating the session into the first and second halves. Data from all cells in Table 2.1.

As described by Gabbard et al. [42], on the Nomad, the reason for reduced performance was blurry vision, especially at the far distance of 6 meters. Likely reasons for this blurry vision were the distance from the resting point of accommodation (for most participants, less than 50 centimeters [55]), laser speckle in the Nomad display, and smaller font size at the far distance. The Nomad uses laser-based retinal scanning technology. All such displays exhibit laser speckle, which reduces image quality and can be particularly problematic for text and graphics with a small visual footprint [28]. In contrast, on the Haploscope, the text size in terms of visual angle was constant regardless of distance, the image generators did not exhibit laser speckle, and the display resolution was 1920×1080 pixels, compared to 800×600 for the Nomad. Although the maximum tested distance was 4 meters, instead of the 6 meters tested on the Nomad, for most participants, 4 meters is still very far from their resting point of accommodation, so this explanation for the different results on the Nomad and Haploscope seems unlikely. Instead, the most likely reason for the increased performance is improved AR image quality. On the Haploscope, increasing reference distance reduced performance, but only under monocular viewing.

Context switching caused greater eye fatigue on the Nomad, and as hypothesized (**H1**), it also caused greater fatigue on the Haploscope. Context switching was more fatiguing at every distance. Given the replication of this effect on two different display devices, this finding is consistent with the hypothesis that switching cognitive and visual attention between real and AR objects (context switching) causes eye fatigue in all OST AR systems. The effect was stronger on the Nomad than on the Haploscope. The effect was also stronger for monocular viewing than for binocular viewing, and in addition, fatigue increased with increasing distance under monocular viewing.

In addition, as shown in Figure. 2.6 (covering all of the data), there was a linear increase in reported eye fatigue as each experimental session progressed. Participants attended two experimental sessions on different days, with the viewing condition changing between sessions. In Figure. 2.6, each experimental session is broken into two halves, where the level of context switching changed at the halfway point. During this transition, the participant closed their eyes while the equipment was moved, which lasted several minutes. After this transition period, eye fatigue declined slightly but again began steadily increasing. For both experimental half, the growth of fatigue occurred at a constant rate of $b = .03$ units per trial. Both slopes significantly differ from 0 (*first half*: $F_{1,214} = 31.8, p < .001^{***}$; *second half*: $F_{1,214} = 27.5, p < .001^{***}$).

2.6.2 Focal Switching Distance

Larger focal switching distances were expected to reduce performance and increase fatigue (**H2**). The previous section examined context switching when there was no focal distance switching: when *focal switching distance* = 0. Gabbard et al. [42] examined focal distance switching as a binary variable by comparing cells in which focal distance switching did not occur to cells where it did.

In the current experiment, this would compare cells with *focal switching distance* = 0 to cells with *focal switching distance* \neq 0. The current analysis instead analyzes focal switching distance as a continuous variable, resulting in more experimental power [96]. In addition, the analysis examines switching effects over distance, and the interaction between context switching and focal distance switching. This section analyzes all of the data in Table 2.1.

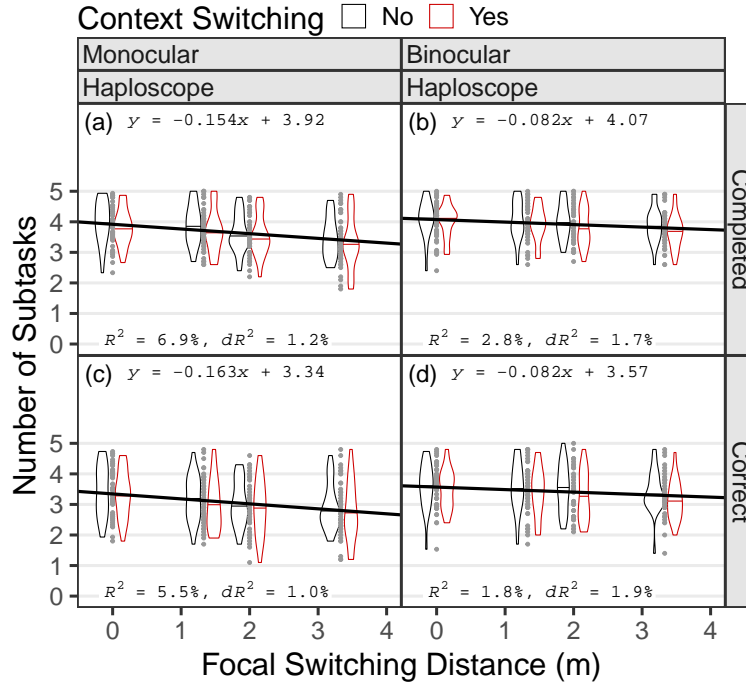


Figure 2.7: As focal switching distance increased, performance decreased. Monocular viewing decreased performance. Context switching had no effect. Data from all cells in Table 2.1. Tables 2.4, 2.5 show the related F -tests. Each panel displays 192 data points.

2.6.2.1 Task Performance

The task performance effects of context switching and focal switching distance are analyzed in Figure. 2.7. Here, the x axis shows focal switching distance, at 0, 1.33, 2, and 3.33 meters. The

Table 2.4: *F*-tests for each panel of Figure. 2.7

Number of Subtasks Completed:

- (a) *Monocular, Haploscope:*
slope diff: $F_{1,188} < 1$
intercept diff: $F_{1,189} = 2.5$ $p = .11$
slope 0: $F_{1,190} = 14.1$ $p < .001^{***}$
- (b) *Binocular, Haploscope:*
slope diff: $F_{1,188} < 1$
intercept diff: $F_{1,189} = 3.0$ $p = .085$
slope 0: $F_{1,190} = 5.5$ $p < .05^*$

Number of Subtasks Correct:

- (c) *Monocular, Haploscope:*
slope diff: $F_{1,188} < 1$
intercept diff: $F_{1,189} = 2.0$ $p = .16$
slope 0: $F_{1,190} = 11.1$ $p < .01^{**}$
- (d) *Binocular, Haploscope:*
slope diff: $F_{1,188} < 1$
intercept diff: $F_{1,189} = 3.4$ $p = .068$
slope 0: $F_{1,190} = 3.5$ $p = .064$

Table 2.5: *F*-tests for panels of Figure. 2.7

(a,b) *Number of Subtasks Completed:*

$$\begin{aligned} \text{Binocular:} & \quad y = -.118x + 4.13 \\ \text{Monocular:} & \quad y = -.118x + 3.86 \\ \\ \text{slope diff:} & \quad F_{1,188} < 1.1 \quad p = .30 \\ \text{intercept diff:} & \quad F_{1,189} = 10.9 \quad p < .01^{**} \\ \text{slope 0:} & \quad F_{1,189} = 11.7 \quad p < .01^{**} \\ \\ d = 0.273 & \quad R^2 = 10.7\% \quad dR^2 = 5.7\% \end{aligned}$$

(c,d) *Number of Subtasks Correct:*

$$\begin{aligned} \text{Binocular:} & \quad y = -.122x + 3.64 \\ \text{Monocular:} & \quad y = -.122x + 3.28 \\ \\ \text{slope diff:} & \quad F_{1,188} < 1 \\ \text{intercept diff:} & \quad F_{1,189} = 11.7 \quad p < .01^{**} \\ \text{slope 0:} & \quad F_{1,189} = 7.9 \quad p < .01^{**} \\ \\ d = 0.358 & \quad R^2 = 9.4\% \quad dR^2 = 6.0\% \end{aligned}$$

Table 2.6: *F*-tests for panels of Figure. 2.7

(a,b) <i>Eye Fatigue:</i>		
<i>Monocular:</i>	$y = .261x + 3.74$	
<i>Binocular:</i>	$y = .261x + 2.49$	
slope diff:	$F_{1,188} < 1.7$	$p = .19$
intercept diff:	$F_{1,189} = 111.1$	$p < .001^{***}$
slope 0:	$F_{1,189} = 27.8$	$p < .001^{***}$
$d = 1.251$	$R^2 = 42.4\%$	$dR^2 = 34.4\%$

columns show the results for *monocular* (left) and *binocular* (right) viewing. The upper row shows performance in terms of *subtasks completed*, while the lower row shows performance in terms of *subtasks correct*.

There was no interaction or main effect of context switching on performance in any panel (Table 2.4). However, as focal switching distance increased, performance significantly decreased. Under monocular viewing, performance decreased at a rate of $b = .154$ (.163) subtasks per meter completed (correct), while under binocular viewing, performance decreased at a smaller rate of $b = .082$ (.082) subtasks per meter completed (correct). The negative effect of focal switching distance on performance was larger for monocular viewing, $R^2 = 6.9\%$ (5.5%), than for binocular viewing, $R^2 = 2.8\%$ (1.8%).

Table 2.5 analyzes the effects of viewing and focal switching distance on subtasks completed (correct). Unlike the analysis in the above paragraph, which examines categorical differences within each panel of Figure. 2.7, this analysis examines categorical differences between panels: between monocular and binocular viewing. There was a significant main effect of viewing, where binocular viewing resulted in $d = .273$ (.358) additional subtasks completed (correct), compared to monocular viewing. These d values are the distances between the lines in Figure. 2.7. This

analysis also finds the main effect of focal switching distance, but the model fits the slopes $b = -.118 (-.122)^1$.

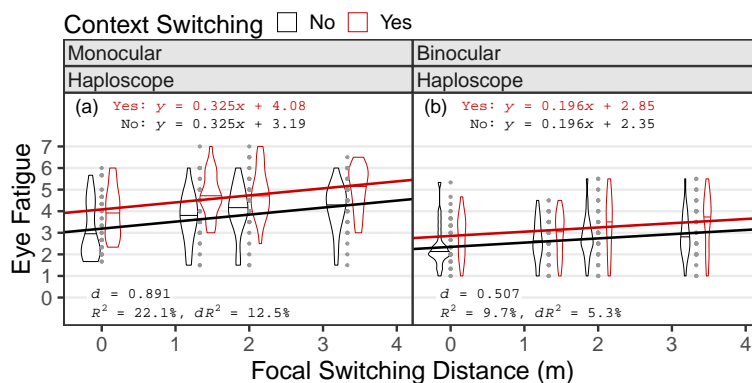


Figure 2.8: Context switching increased eye fatigue for both monocular viewing and binocular viewing. Monocular viewing was more fatiguing. Increasing focal switching distance resulted in greater eye fatigue. Data from all cells in Table 2.1. Tables 2.7, 2.6 show the related F -tests. Each panel displays 192 data points.

Table 2.7: F -tests for panels of Figure. 2.8

Eye Fatigue:

Monocular, Haploscope:

slope diff: $F_{1,188} < 1$
intercept diff: $F_{1,189} = 30.38$ $p < .001^{***}$
slope 0: $F_{1,189} = 23.34$ $p < .001^{***}$

Binocular, Haploscope:

slope diff: $F_{1,188} < 1$
intercept diff: $F_{1,189} = 10.95$ $p = .001^{**}$
slope 0: $F_{1,189} = 9.45$ $p = .002$

¹Note that these are the means of the slopes in panels a , b and c , d in Figure. 2.7.

2.6.2.2 Eye Fatigue

The eye fatigue effects of context switching and focal switching distance are analyzed in Figure. 2.8. Other than the y axis displaying fatigue, the graph structure is the same as Figure. 2.7. In both monocular and binocular viewing, there was a significant main effect of context switching on fatigue and a significant main effect of focal switching distance (Table 2.7). Under monocular viewing, context switching increased fatigue by $d = .891$ units, and increasing focal switching distance increased fatigue at $b = .325$ units per meter. Under binocular viewing, context switching increased fatigue by $d = .507$ units, and increasing focal switching distance increased fatigue at $b = .196$ units per meter, both smaller amounts. The model explains $R^2 = 22.1\%$ of the variation under monocular viewing, much higher than for binocular viewing, $R^2 = 9.7\%$. Monocular viewing was $d = 1.251$ units more fatiguing than binocular viewing (Table 2.6).

2.6.2.3 Discussion

In the experiment, participants had to first accommodate the distance of the left text. Then, if the focal switching distance was greater than 0, participants had to change accommodative distance to visually scan the right text. It was therefore hypothesized (**H2**) that increasing focal switching distance would decrease performance and increase eye fatigue. The results support this hypothesis. During focal distance switching, the eye's ciliary muscles change accommodation to bring information into sharp focus, and the eye's vergence muscles change vergence. Therefore, continuously shifting eye focus between different focal distances tires these muscles, leading to eye fatigue and reduced performance. In addition, for most people, the resting point of accommodation and vergence is about 0.5 meters [114]. To accommodate and verge away from the resting point,

the eye muscles contract, while when returning to the resting point, the eye muscles relax [46]. Therefore, integrating information closer to the resting point is less exerting. As a result, as the amount of focal switching distance increased, eye fatigue increased, and performance decreased.

2.6.3 Viewing

Viewing has been analyzed in the previous sections and has been shown to have effects in each case. However, in Section 2.6.1, while the effects of viewing, context switching, and reference distance were examined (Figures. 2.4 and 2.5), that analysis only covers the 30% of the collected data where focal distance was held constant (focal switching distance = 0). This section examines viewing in the context of reference distance, covering all of the data. The structure of the graphs here, Figures. 2.9 and 2.10, are the same as Figures. 2.4 and 2.5, except that viewing is now analyzed within each panel.

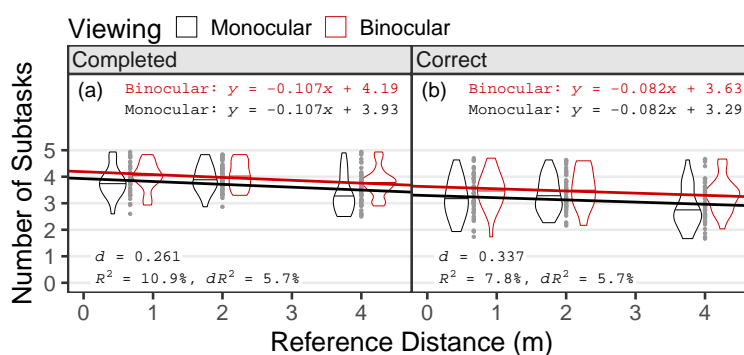


Figure 2.9: Task performance was higher under binocular viewing and closer reference distances. Data from all cells in Table 2.1. Table 2.8 shows the related F -tests. Each panel displays 144 data points.

Table 2.8: *F*-tests for panels of Figure. 2.9

(a) <i>Number of Subtasks Completed:</i>			
slope diff:	$F_{1,140} = 1.4$	$p = .2$	
intercept diff:	$F_{1,141} = 7.6$	$p < .01^{**}$	
slope 0:	$F_{1,141} = 9.7$	$p < .01^{**}$	
(b) <i>Number of Subtasks Correct:</i>			
slope diff:	$F_{1,140} < 1$		
intercept diff:	$F_{1,141} = 8.2$	$p < .01^{**}$	
slope 0:	$F_{1,141} = 3.7$	$p = .06$	

2.6.3.1 Task Performance

The task performance effects of viewing and reference distance are analyzed in Figure. 2.9. There was a significant effect of viewing on subtask completion (Figure. 2.9a) and subtask accuracy (Figure. 2.9b). Monocular viewing decreased performance by $d = .261$ (.337) subtasks, describing $dR^2 = 5.7\%$ (5.7%) of the variation. In addition, increasing reference distance resulted in reduced performance, at a $b = .107$ (.082) subtasks per meter. The overall model explains $R^2 = 10.9\%$ (7.8%) of the performance variation.

Table 2.9: *F*-tests for panels of Figure. 2.10

<i>Eye Fatigue:</i>			
slope diff:	$F_{1,140} < 1$		
intercept diff:	$F_{1,141} = 73.8$	$p < .001^{***}$	
slope 0:	$F_{1,141} = 9.2$	$p < .01^{**}$	

2.6.3.2 Eye Fatigue

The eye fatigue effects of viewing and reference distance are analyzed in Figure. 2.10. There was a significant effect of viewing on fatigue. Monocular viewing increased fatigue by $d = 1.211$

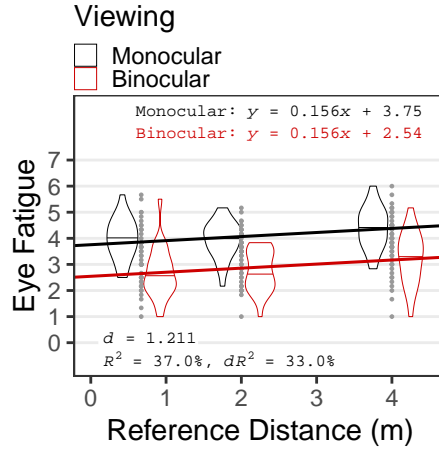


Figure 2.10: Monocular viewing was more fatiguing than binocular viewing. Increasing reference distance resulted in greater fatigue. Data from all cells in Table 2.1. Table 2.9 shows the related F -tests. The panel displays 144 data points.

units, a large effect that describes $dR^2 = 33.0\%$ of the fatigue variation. In addition, increasing reference distance resulted in increased fatigue, at a rate of $b = .156$ units per meter. The overall model explains $R^2 = 37.0\%$ of the fatigue variation.

2.6.3.3 Discussion

When analyzed over all of the data, viewing had the same effects as when it was analyzed over the data where focal distance was held constant (Section 2.6.1): monocular viewing resulted in lower performance and higher fatigue. As reference distance increased, performance declined, and fatigue increased. However, the viewing effect was constant and did not interact with reference distance. Previous work has found that binocular viewing through an HMD provided more accurate accommodation than monocular viewing [66]. As previously discussed in Section 2.4.1, the primary stimulus that drives the accommodative response is a blur gradient [47]. However, under binocular viewing, stereo disparity additionally drives vergence eye movements, which in turn

drive vergence accommodation [59]. So in this experiment, under monocular viewing, there was only one accommodative stimuli (blur gradient), while under binocular viewing there were two accommodative stimuli (blur gradient and vergence accommodation). This suggests that under binocular viewing, changing accommodation should be more efficient. In addition, during monocular viewing, participants covered their non-dominant eye with an eye patch, which could have resulted in additional pressure and discomfort. As a result of all of these factors, under monocular viewing performance decreased, and fatigue increased.

During the post-experiment informal interview, participants did not complain about image quality at any specific distance. Still, at or after the experiment's midpoint, 7 out of 24 participants complained about general visual fatigue in monocular viewing. In contrast, in binocular viewing, the number that complained was 2 out of 24. Therefore, participants subjectively reported that monocular viewing was more fatiguing.

2.6.4 Transient Focal Blur

A performance reduction during focal distance switching, attributed to the transient focal blur effect by Gabbard, Mehra, and Swan [42], was expected to replicate under different conditions of context switching and viewing (**H3**). As discussed in Section 2.4.1, changing accommodation from one focal distance to another can be expected to take at least 350 milliseconds, and possibly as long as 425 milliseconds. While accommodation changes, objects at the new focal distance will be seen with out-of-focus blur. If the task demands performance during this period, then this *transient focal blur* could cause reduced visual performance. A visual representation of the transient focal blur problem is given in Figure 2.11.

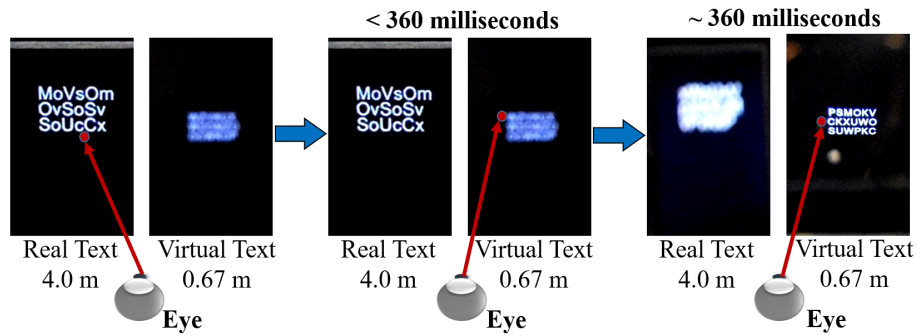


Figure 2.11: Focus blur due to switching accommodation distances while gathering information both from the real and virtual text. In the figure, real and virtual text are displayed at two focal distances: real text at 4.0m and AR text at 0.67m. To accomplish the experimental task, participants first focused on the real text to identify the target letter (left). After identifying the target letter, participants switched their gaze to the virtual text (middle), and began to change their accommodative distance to bring it into focus. However, changing accommodation takes time, and they began scanning the first line of the virtual text for a target letter before their eyes had finished accommodating (middle). This made the text blurry. Assuming participants scanned the text in left-right-right top-to-bottom order, they were more likely to miss a target letter in the first line of text. After the completion of accommodation, participants could observe the virtual text in sharp focus (right).

2.6.4.1 Letter Undercounts

Gabbard et al. [42] hypothesized that during the transient focal blur time period, participants would be more likely to miss (undercount) target letters in the first line of text. Assuming that participants scan the right text in the standard reading direction of left-to-right, top-to-bottom, the most likely target letters to encounter during this time period would be the letters in the first line of text (see Figure 2.11). Gabbard et al. [42] found that when focal distance switching was required, these letters were significantly more likely to be undercounted. Figure. 2.12a replicates the related graph. Significantly more letters were undercounted per participant when focal distance switching was required, and when a target letter was in the first line of text (Table 2.10). In addition, there was a significant interaction, where the most letters were undercounted when a target letter

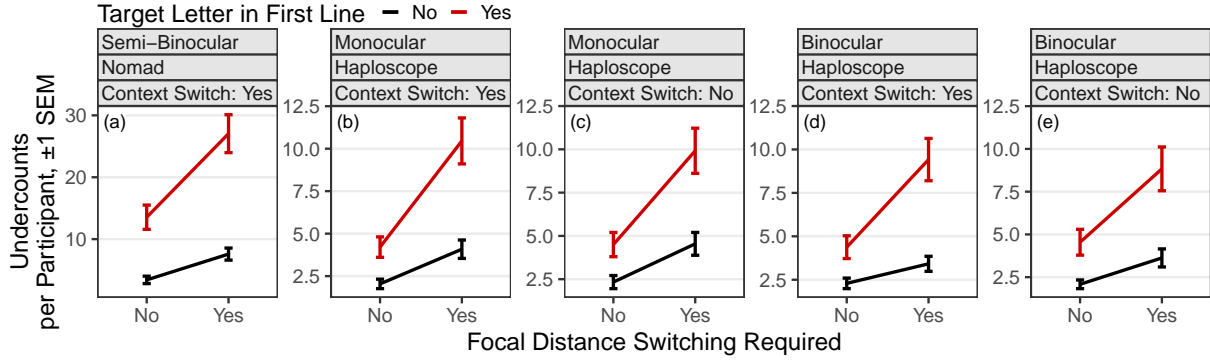


Figure 2.12: Evidence for the *transient focal blur effect*: Participants undercounted more letters when a target letter appeared in the first line of the text and when focal distance switching was required. In addition, these factors interacted; more letters were undercounted when the target letter was in the first line, and participants switched focal distances. In this case, participants tried to read the letter in the first line while that line was still out of focus. (a) The transient focal blur effect found by Gabbard et al. [42]. (b–e) There was a strong transient focal blur effect for every combination of viewing and context switching. *Nomad*: Data from Table 2 (Gabbard et al. [42]): Focal Distance Switching Required: *No* = hatched cells; *Yes* = remaining cells. *Haploscope*: Data from all cell in Table 2.1: Focal Distance Switching Required: *No* = shaded cells; *Yes* = remaining cells. Table 2.10 shows the related *F*-tests.

was in the first line of text and focal distance switching was required ($p < .001$ for all effects). Gabbard et al. [42] hypothesized that this interaction could be explained by the fact that the task was time-pressured, which caused participants to begin scanning the right text during the transient focal blur period, when their eyes were still accommodating to the new distance.

Hypothesis **H3** is based on the idea that this transient focal blur effect is not specific to AR, but instead is a general property of visual tasks that require integrating information from different displays, possibly located at different focal distances. If correct, this hypothesis predicts that the interaction shown in Figure. 2.12a will replicate under different context switching and viewing conditions. The current experiment strongly supported this hypothesis: as shown in

Table 2.10: F -tests for each panel of Fig. 2.12

Undercounts Per Participant:

(a)	<i>Semi-Binocular, Nomad, Context Switching = Yes:</i>		
	Focal distance switching:	$F_{1,23} = 46.2$	$p < .001^{***}$
	Target letter in the first line:	$F_{1,23} = 68.3$	$p < .001^{***}$
	Focal distance switching \times		
	Target letter in the first line:	$F_{1,23} = 25.0$	$p < .001^{***}$
(b)	<i>Monocular, Haploscope, Context Switching = Yes:</i>		
	Focal distance switching:	$F_{1,23} = 42.1$	$p < .001^{***}$
	Target letter in the first line:	$F_{1,23} = 37.3$	$p < .001^{***}$
	Focal distance switching \times		
	Target letter in the first line:	$F_{1,23} = 27.3$	$p < .001^{***}$
(c)	<i>Monocular, Haploscope, Context Switching = No:</i>		
	Focal distance switching:	$F_{1,23} = 32.6$	$p < .001^{***}$
	Target letter in the first line:	$F_{1,23} = 35.8$	$p < .001^{***}$
	Focal distance switching \times		
	Target letter in the first line:	$F_{1,23} = 11.5$	$p < .001^{***}$
(d)	<i>Binocular, Haploscope, Context Switching = Yes:</i>		
	Focal distance switching:	$F_{1,23} = 40.1$	$p < .001^{***}$
	Target letter in the first line:	$F_{1,23} = 29.9$	$p < .001^{***}$
	Focal distance switching \times		
	Target letter in the first line:	$F_{1,23} = 13.4$	$p < .001^{***}$
(e)	<i>Binocular, Haploscope, Context Switching = No:</i>		
	Focal distance switching:	$F_{1,23} = 23.2$	$p < .001^{***}$
	Target letter in the first line:	$F_{1,23} = 23.0$	$p < .001^{***}$
	Focal distance switching \times		
	Target letter in the first line:	$F_{1,23} = 10.1$	$p < .001^{***}$

Figures. 2.12*b, c, d, e*, and Table 2.10, both main effects and their interaction was significant for every combination of context switching and viewing ($p < .001$).

2.6.4.2 Discussion

Although Gabbard et al. [42] found the transient focal blur effect, their experiment only examined the condition of context switching and semi-binocular viewing of a one-eyed display. This left the possibility that the effect was somehow specific to this set of conditions. However,

the replication of the effect under both context switching and viewing conditions suggests that the transient focal blur effect is more general.

When context switching, participants integrated information between a monitor in the real world and an AR display. In contrast, when not context switching, both sources of information were on monitors in the real world. The transient focal blur effect is replicated in both conditions, which is consistent with the hypothesis that the transient focal blur effect is indeed not specific to AR, but is a general property of visual tasks that require integrating information from multiple displays located at different distances.

In addition, the transient blur effect was replicated under both binocular and monocular viewing conditions. As discussed in the previous section, vergence accommodation should make changing accommodation more efficient under binocular viewing. There is some evidence for this: when focal distance switching is required, the magnitude of the undercounts is lower with binocular viewing (Figure. 2.12*d, e*) than with monocular viewing (Figure. 2.12*b, c*). This effect is related to the increased performance and decreased fatigue for binocular viewing discussed in the previous section (2.6.3). Despite this increased efficiency, the transient focal blur effect was just as statistically strong for binocular viewing ($p < .001$) as it was for monocular viewing ($p < .001$).

2.7 Conclusion

This experiment examined the effects of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur. The visual search task required integrating information distributed between real and virtual contexts. The experiment was conducted on a custom-built AR Haploscope, which allowed an accurate representation of focal distances and ver-

gence angles. The experiment partially replicated and extended a previous investigation conducted on a Microvision Nomad; a one-eye display viewed semi-binocularly [42]. The primary findings are:

- *Context switching* did not reduce task performance, but did increase eye fatigue.
- As *focal switching distance* increased, performance decreased, and eye fatigue increased.
- Compared to *binocular viewing*, *monocular viewing* resulted in reduced performance and increased eye fatigue.
- *Transient focal blur* resulted in reduced task performance under all combinations of context switching and viewing. This validates the importance of exploring the out-of-focus problem in OST AR. It also suggests that having a sharper or improved visual representation of the virtual information during the period of focal blur will lead to increased performance, and perhaps reduced fatigue as well.

CHAPTER III

OUT-OF-FOCUS VISUAL ABERRATION FOR AR SYSTEM

In chapter II, research showed that transient focal blur is caused by focal distance switching, resulting in an additional reduction in task performance. The study concluded that it is necessary to create and present a sharper or more legible visual representation of the virtual information during the out-of-focus period. To achieve this, we first need to obtain the retinal blur formed on the retina. This is an essential step, as without the proper modeling of the human visual out-of-focus aberration, it is not possible to create a sharper image, or enhanced the legibility of virtual information. Therefore, this chapter describes the modeling of the out-of-focus visual aberration for OST AR.

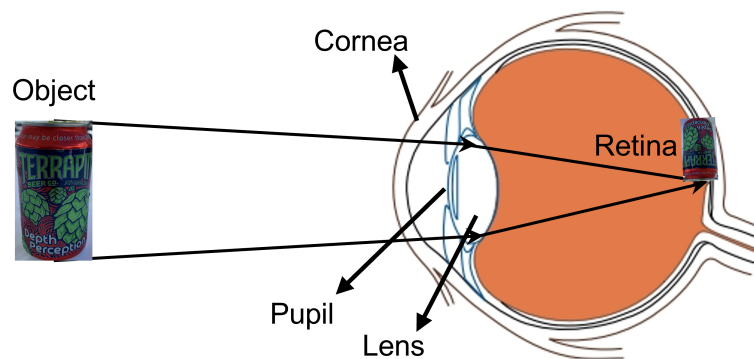


Figure 3.1: Schematic diagram of retinal image formation.

3.1 Background

To model the out-of-focus aberration of the human eye, it is essential to have an understanding of the mechanism of retinal image formation.

3.1.1 Retinal Image formation

The human visual system is a complicated part of the brain, and the primary optical element is the eye. Figure 3.1 shows a simple schematic diagram of retinal image formation. The size of the eye is approximately 24mm in diameter [12]. When the eye focuses on an object, light from that object enters the eye through the cornea. As the refractive index of the cornea is greater than the refractive index of air, lights get bent after hitting the curved surface of the cornea. The amount of light that gets into the eye is controlled by the pupil. The pupil is the black central opening in the iris through which light passes. The size of the pupil depends, in part, on the amount of light in the environment. The pupil size becomes bigger in low light conditions, varying from 4 to 8mm in diameter in adults [11, 46]. Under bright light conditions, the pupil size becomes smaller, varying from 2 to 4 mm in adults [11, 12, 46]. Therefore, it is conclusively said that pupil size significantly affects the quality of the retinal image.

After going through the cornea and controlled by the pupil, light rays pass through the lens of the human eye. The lens is biconvex in shape. The zonule fibers and ciliary muscle attach the lens to the eye behind the iris. Based on the accommodative demand, the ciliary muscle changes the lens's shapes so that the light rays' direction changes in a finely controlled manner, creating a sharp representation of the object on the retina. Both the cornea and lens are the refractive elements of the eye. After the lens refracts the light rays, they reach the retina and form an upside-down image.

The human retina consists of the fovea, blind spot, and optic disk [46]. The fovea is the small and central part of the retina where the image is observed with the highest resolution. Conversely, no light is detected in the blind spot [46, 4]. From the optic disk, neural nerve fibers exit the retina, transferring information to the brain.

3.1.2 Basic Principle of Retinal Image Calculation

Pupil Function: The visual perception of an object depends on the image quality of that object formed on the retina. The analysis of the retinal image can be done by considering the properties of light propagation. Light propagates through space in the form of waves. When a light pulse is emitted from a point source, the wavefront expands in spherical shells. All points on the wavefront are equidistant from the point source, and have traveled an equal distance [105]. This implies that all the points on the wavefront are in *phase* mode. Supporting this theory, when light passes through the eye, it converges as a spherical wavefront to a focal point on the retina, forming the retinal image. The image of a point source that is formed on the retina is known as the “**Point Spread Function (PSF)**” of the human eye. The PSF of the human eye can be computed from the complex-valued generalized pupil function [45, 34, 69], which can give complete information about the retinal image:

$$P(x, y) = A(x, y) \exp^{-i \frac{2\pi}{\lambda} W(x, y)} \quad (3.1)$$

Here, λ is a wavelength of light in a vacuum, and (x, y) are the coordinates of the pupil plane image surface.

Amplitude Function: In the pupil function, $A(x, y)$ is a real-valued circular amplitude function that determines the relative efficiency of light transmission through the pupil. It may be defined

as a binary function, taking the value 1 within the pupil region, and 0 outside the pupil region. The pupil region of the focused image can be computed from the focused object's size and its corresponding retinal image size.

Wavefront Aberration Function: Another essential component of the pupil function is $W(x, y)$, the eye's wavefront aberration function. At the wavefront surface, light waves are perpendicular at the same phase. As the retinal image depends on the light waves, it is, therefore, useful to represent the amount of aberration on the human eye as the amount of wavefront aberration, expressed either in micrometers or number of wavelengths. This can be characterized by the two-dimensional “**Wavefront Aberration Function**”. For every point (x, y) over the pupil plane surface, this function presents the difference between the ideal spherical wavefront and the aberrated wavefront [10, 11, 105].

Aberration on the human eye can be caused by several reasons, including refractive errors, diffraction of light over the pupil, light sensitivity, and others [12, 69]. An eye with aberration generally produces an asymmetric retinal image, whereas an eye with no aberration creates a symmetric retinal image [11]. It is important to note that there will be no aberration when the eyes are in perfect focus. However, the retinal image would still contain a slight blur. The reason is the diffraction of many wavelengths of light as it passes through the pupil. According to the fundamental nature of the wave, diffraction happens when it passes through an aperture. Therefore, even though there is no aberration on a perfectly focused eye, there will be a slight blur on the retinal image, and the diffraction pattern of a point source of light appears as a bright spot in the center of the retina [1, 105]. This particular situation is known as *diffraction limited* (no aberration), and

the retinal image (PSF) in this situation is known as the *Airy disk* [70, 34], which is surrounded by faint concentric rings.

Optical Transfer Function The optical transfer function (OTF) is one the most powerful tools to describe an optical system's imaging performance, including the human eye. Besides, Fourier optics is a great tool to analyze and measure the OTF. By considering the principle of Fourier optics and knowledge of the wavefront aberration function ($W(x,y)$) and pupil function ($P(x,y)$), OTF can be computed by performing the convolution of the pupil function, $P(x,y)$, with its complex conjugate $P^*(-x,-y)$ [105, 1]. It is important to note that PSF and OTF have a relationship from the Fourier transform perspective, which means that if the imaging operations are performed in the frequency domain, then it is called OTF; otherwise, PSF [1].

$$OTF(fx, fy) = P(x, y) \otimes P^*(-x, -y) \quad (3.2)$$

According to the convolution theorem, a similar operation can be done in the frequency domain by performing element-wise multiplication.

$$F[OTF(fx, fy)] = F[P(x, y)] \times F[P^*(-x, -y)] \quad (3.3)$$

Imaging in the Human Eye To analyze human visual perception, we need to obtain the retinal image for a given source image based on the properties of the eye optics [117]. To get the final aberrated image on the retina, a 2-D convolution process needs to be performed in the spatial domain 3.4 or an element-wise multiplication operation needs to be done in the frequency domain 3.5. Let's consider that $I(x, y)$ is an ideal source image with no aberration and diffraction,

and we have the PSF (x, y) . Therefore, the corresponding retinal image $(R(x,y))$ can be obtained by performing the convolution of $I(x, y)$ and $PSF(x, y)$ in the spatial domain or element-wise multiplication of $F[I(x, y)]$ and $OTF(x, y)$ in the frequency domain, where F is the Fourier transform operator.

$$\text{Spatial Domain: } R(x, y) = I(x, y) \otimes PSF(x, y) \quad (3.4)$$

$$\text{Frequency Domain: } F[R(x, y)] = F[I(x, y)] \times OTF(x, y) \quad (3.5)$$

Previously, researchers in the OST AR domain considered the Gaussian PSF for the natural images to compute the retinal aberrated image for the out-of-focus problem [64, 31]. However, visual aberration of the human eye is a depth-dependent phenomena [69]. It means that the amount of visual blur on the human eye varies with the change of the distances between the focused and out-of-focus object from the human eye. Therefore, it is important to consider depth while modeling the aberration of the human eye. Otherwise, the image formed on the retina will be wrong. The most commonly and widely used PSF by the Vision scientists is Zernike-based PSF. They used the Zernike polynomials to model the *low-order aberration* (e.g., defocus, astigmatism) and *high-order aberration* (e.g., trefoil, coma, quatrefoil, secondary astigmatism) of the human eye [34, 69, 1, 88, 113, 123].

After exploring the properties and mechanism of the retinal image theoretically and examining the previous work related to retinal blur, I considered the Zernike polynomial-based PSF to compute the blurred vision that formed on the retina due to out-of-focus aberration. A detailed description of the general mathematical formulation of Zernike PSF and the mathematical formulation of Zernike PSF for the out-of-focus aberration is given in the below sections.

3.2 Description of Zernike Polynomials

Zernike polynomials are a sequence of polygonal that are continuous and orthogonal over a circle of unit radius. Zernike polynomials were first discovered by the Dutch optical physicist Fritz Zernike in 1934 [34]. He developed this theory to describe the diffracted wavefront in phase-contrast imaging with two circular mirrors. This theory became popular and widely used in optical imaging, optical aberrations, optical testing, visual aberration, astronomy and satellite imagery, and many other research fields.

3.2.1 Reason Behind using Zernike Polynomials

Scientists are using the Zernike polynomial extensively because of the Zernike mathematical model's properties [1]. Here are a couple of reasons provided below based on the previous literature [34, 70]:

1. The polynomials are orthonormal over circular pupils, and derivatives are continuous.
2. Produced balanced aberrations between higher-order polynomials and lower-order polynomials so that image intensity can be optimized when the amount of aberration is low.
3. The polynomials efficiently represent many common errors in general optics (e.g., spherical aberration). Besides, effectively perform the error calculation of the ocular wavefront.
4. Efficiently represents both low order aberrations (myopia, hyperopia, astigmatism, and presbyopia) and high order aberrations (e.g., trefoil, coma, quatrefoil, secondary astigmatism) of the human eye.
5. It can define the magnitude and characteristics between the optical system's image and the original object.

3.2.2 Applications of Zernike Polynomials

Vision scientists widely used Zernike polynomials to model the different lower order and higher order visual aberrations of the human eye [73, 69, 88, 1, 117]. Along with vision research, many

other research domains use Zernike polynomials widely. Some of the different applications of the Zernike Polynomials are given below:

1. Zernike polynomials are widely used to measure image moments (weighted average of the image pixels' intensities), also known as Zernike moments. Zernike moments are beneficial for extracting features from an image and describing the shape of an object [77]. Zernike moments significantly contribute to cancer research; by analyzing and quantifying different cancer cells' shapes.
2. It is widely used in astronomy, fitting the wavefront distorted by the atmospheric turbulence [93].
3. Zernike polynomials can also be used in satellite imagery analysis to fit different distorted wavefront [15].

3.2.3 Limitations of Zernike polynomials:

Zernike polynomials are not entirely perfect. Lakshminarayanan and Fleck [70] mentioned some of the limitations of the Zernike polynomials. They are:

1. Zernike coefficients are only valid for the specific pupil diameter they are determined for. Therefore, computing the Zernike coefficient for every pupil diameters are computationally expensive.
2. Zernike polynomials are orthogonal over the circle of the unit radius. Therefore, these polynomials are not ideal for noncircular or irregular (e.g., cone-shaped cornea) optical systems.

Therefore, any system with circular representation (e.g., the human eye, telescope, microscope, etc.) is most suitable for the Zernike polynomials. After performing an extensive literature review, the dissertation research considered the Zernike polynomial to represent the out-of-focus aberration of the human eye for the AR system. Next, a detailed description of how Zernike polynomials characterize lower-order aberrations of the human eye is given. After that, the description of how my dissertation research modified and utilized the Zernike polynomial to model the out-of-focus aberration is given.

3.2.4 Zernike Polynomials for the Low-Order Aberration

i	n	m	Zernike polynomials	Name
0	0	0	1	piston
1	1	-1	$2\rho \sin \theta$	y -tilt
2	1	1	$2\rho \cos \theta$	x -tilt
3	2	-2	$\sqrt{6}\rho^2 \sin 2\theta$	y -astigmatism
4	2	0	$\sqrt{3}(2\rho^2 - 1)$	defocus
5	2	2	$\sqrt{6}\rho^2 \cos 2\theta$	x -astigmatism
6	3	-3	$\sqrt{8}\rho^3 \sin 3\theta$	y -trefoil
7	3	-1	$\sqrt{8}(3\rho^3 - 2\rho) \sin \theta$	y -coma
8	3	1	$\sqrt{8}(3\rho^3 - 2\rho) \cos \theta$	x -coma
9	3	3	$\sqrt{8}\rho^3 \cos 3\theta$	x -trefoil
10	4	-4	$\sqrt{10}\rho^4 \sin 4\theta$	y -quadrafoil
11	4	-2	$\sqrt{10}(4\rho^4 - 3\rho^2) \sin 2\theta$	y -secondary astigmatism
12	4	0	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$	spherical aberration
13	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2) \cos 2\theta$	x -secondary astigmatism
14	4	4	$\sqrt{10}\rho^4 \cos 4\theta$	x -quadrafoil

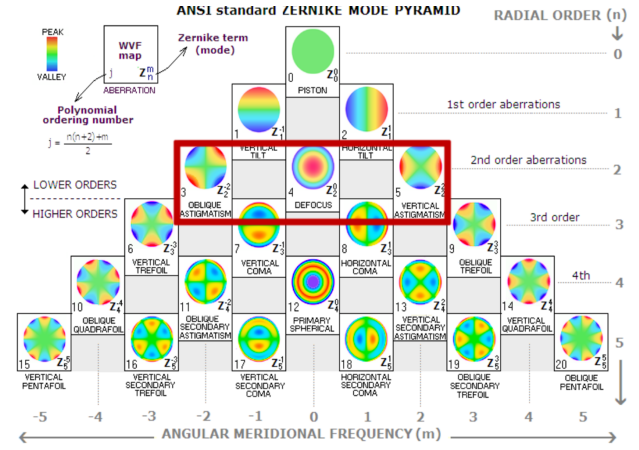


Figure 3.2: Zernike polynomials up to the fourth order and a Zernike polynomial pyramid [34]

Human visual aberrations can be divided into two types: low-order visual aberration (myopia, hyperopia, astigmatism, and presbyopia), corrected through the corrective glass/ lens, and high-order aberration (trefoil, coma, quadrafoil, secondary astigmatism) [34, 69]. The Zernike polynomials can approximately model the low-order and high-order human visual aberration. Zernike polynomials up to the fourth order and a Zernike polynomial pyramid are given in the figure 3.2. The characterization of the human visual aberrations in Zernike polynomials can be expressed either in Cartesian (x, y) or polar (r, θ) coordinates. Wavefront aberration function $(W(x, y))$ is the standard way to report the human visual aberrations by considering the sum of the set of the weighted Zernike polynomials 3.6 [117].

$$W(x, y) = \sum_{n,m} C_n^m Z_n^m(x, y) \quad (3.6)$$

Here, x and y are the coordinates relative to the pupil center and normalized by the pupil radius [117]. n and m represent the order and frequency, respectively. C_n^m denotes the Zernike coefficient in micrometers and defines the standard weight of aberrations. In ophthalmology and vision science, the order (n) is the basis for classifying the different visual aberrations. Such as:

- Low-order aberrations, with the Zernike order $n < 3$.
- High-order aberrations, with Zernike order $n \geq 3$.

Vision scientists mostly considered 2nd order Zernike polynomials to model the vision of individuals with refractive vision problems (myopia, hyperopia, and astigmatism) [69, 123, 88, 117]. My dissertation research is highly related to these aberrations as these are related to the changes in accommodation or focus problem. The wavefront aberration function with the required 2nd order Zernike polynomials is given below [34, 123]:

$$W(x, y) = \sum_{i=-1}^1 C_2^{2i} Z_2^{2i}(x, y) \quad (3.7)$$

$$\text{Oblique Astigmatism (y-astigmatism): } Z_2^{-2}(x, y) = 2\sqrt{6}xy \quad (3.8)$$

$$\text{Defocus: } Z_2^0(x, y) = \sqrt{3}(2x^2 + 2y^2 - 1) \quad (3.9)$$

$$\text{Vertical Astigmatism (x-astigmatism): } Z_2^2(x, y) = \sqrt{6}(x^2 - y^2) \quad (3.10)$$

3.2.5 Zernike Coefficient for the Low-Order Aberration

So far, I have discussed the Zernike polynomial. Nevertheless, the Zernike polynomial is not complete without the Zernike coefficient, which provides the magnitude of eye aberration. Zernike coefficients are independent and are not constant. All the vision research extensively uses Zernike coefficients as “Zernike coefficient for prescription data” [88, 69, 123]. After analyzing

spherocylindrical ocular aberration, Dai [34] was able to compute the Zernike coefficient for prescription data of the human eyes. Zernike coefficient for prescription data are given below:

$$\text{Oblique Astigmatism (y-astigmatism) coefficient: } C_2^{-2} = \frac{R^2 C \sin 2A}{4\sqrt{6}} \quad (3.11)$$

$$\text{Defocus coefficient : } C_2^0 = \frac{R^2(S + C/2)}{4\sqrt{3}} \quad (3.12)$$

$$\text{Vertical Astigmatism (x-astigmatism) coefficient: } C_2^2 = \frac{R^2 C \cos 2A}{4\sqrt{6}} \quad (3.13)$$

where,

- R = Radius of the pupil in mm
- S = Sphere value in diopters.
- C = Cylinder values in diopters.
- A = Cylinder axis expressed in degree.

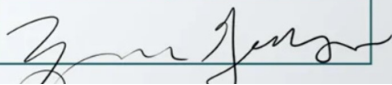
Rx	SPH	CYL	AXIS	PRISM	BASE
O.D.	-2.00	-1.00	100		
O.S.	-2.75	-0.75	90		
PD 61		REMARKS Exp. 01/01/2021			
					

Figure 3.3: Standard eye prescription with the spherical and cylindrical value. Vision scientists considered values from the eye prescriptions to obtain the Zernike coefficients.

While working with Zernike polynomials and coefficients for individuals with refractive vision problems, vision scientists directly measure the S, C, and A values from the person's eyeglass

prescription (see figure 3.3) [88, 69, 123]. After calculating the Zernike polynomials (3.8, 3.9, 3.10) and the Zernike coefficients (3.11, 3.12, 3.13), vision scientists place the values in the equation 3.7 to get the corresponding wavefront aberration function of the human eye. The final PSF of the eye's optical system can be obtained by considering the generalized pupil function (see equation 3.1) and optical transfer function (see equation 3.2) of the human eye [1]. Performing convolution (see equation 3.4) between the human eye's PSF and the display focus image gives the human perceived blurred image [88, 123]. Though vision scientists are working with different eye aberrations, based on my knowledge, I did not find any research using Zernike polynomials and coefficients for the bifocal lenses. It is still unknown how to deal with the Zernike parameters in this scenario.

The formulation of Zernike polynomials and the coefficient for the out-of-focus aberration in AR is theoretically related to the Zernike formulation for the lower-order aberration, more specifically, the *defocus aberration*. Therefore, my dissertation research considered the defocus Zernike polynomials and modified coefficient version to formulate the model for out-of-focus aberration in the AR system. To achieve this, this research developed the parameters of the equations based on the focused and out-of-focus distances rather than prescription data.

3.3 Zernike Polynomials and Coefficient for the Out-of-Focus Aberration in AR

Among the parameters of 2nd order Zernike coefficients (see equations 3.11, 3.12, 3.13), only S (sphere value) is related to the defocus aberration ($S < 0$ for myopia and $S > 0$ for hyperopia), while the C and A is related to astigmatism aberration [123]. Further, among the low-order aberrations, only *defocus aberration* is related to the human eye's accommodative stance. In *myopia aberration*, a person can sharply focus on an object at a near distance but not at a far distance. Conversely, in

hyperopia aberration conditions, a person can perceive an object far away, not at a near distance. Similarly, during the out-of-focus condition, users can only sharply focus at a specific distance, either near or far, and information at the out-of-focus distance will be blurred. Therefore, the human eye's accommodation changing is related to the defocus aberration rather than any other aberration types. For this reason, this research considered the *defocus Zernike polynomial* to investigate out-of-focus aberration in AR and ignored everything related to the astigmatism aberration. Only the S value (related to defocus aberration) is considered, and the C and A values (related to astigmatism aberration) are ignored for the defocus Zernike coefficients. After ignoring the C and A values, the defocus Zernike coefficient for the out-of-focus aberration in the AR system is given below:

$$\text{Defocus Zernike coefficient for out-of-focus aberration in AR : } C_2^0 = -\frac{R^2(S)}{4\sqrt{3}} \quad (3.14)$$

Calculating *S value* based on the focused and out-of-focus distances As this research is not considering any eye prescription data, it is necessary to develop a method to calculate the *S value* for the out-of-focus aberration in AR. Further, eye accommodation changes from one distance to another; therefore, S is inconsistent. Previously, Xu and Li. [123] developed the formulation to calculate the *S value* only from the viewing distance or focused distance while modeling the lower-order visual aberration for HMD, not for the out-of-focus aberration. Getting motivated by the previous work, this research developed the method to calculate the *S value* for the out-of-focus aberration in AR. Based on the research's technique, the focal length of the eye must be computed first. According to the thin lens formula, the focal length of the eye based on the focused object is:

$$\frac{1}{f_{eye}} = \frac{1}{d_{focus}} + \frac{1}{d_{eye}} \quad (3.15)$$

where,

- f_{eye} = Focal length of the eye at focused distance
- d_{focus} = Distance between eye's lens and focused object.
- d_{eye} = Distance between eye's lens and retina, which is approximately 22mm [46].

Equation 3.15 presents an ideal scenario when a user is focused at a specific distance. Except for the focused distance, this equation is not suitable for all other out-of-focus distances, which brings the issue of out-of-focus aberration due to the changing accommodative stance of the human eye.

Let us consider the out-of-focus object distance is $d_{out-of-focus}$ from the eye's lens. Therefore, the eye will observe out-of-focus aberration, and the eye's aberration value, S , can be computed in diopters using the following formula:

$$S = \frac{1}{d_{out-of-focus}} - \frac{1}{d_{focus}} \quad (3.16)$$

By placing the $\frac{1}{d_{focus}}$ value from the equation 3.15 in equation 3.16, we get the following:

$$S = \frac{1}{d_{out-of-focus}} - \left(\frac{1}{f_{eye}} - \frac{1}{d_{eye}} \right) \quad (3.17)$$

where,

- S = Out-of-focus spherical aberration on the eye in diopters.
- $d_{out-of-focus}$ = Distance between eye's lens and out-of-focus object.

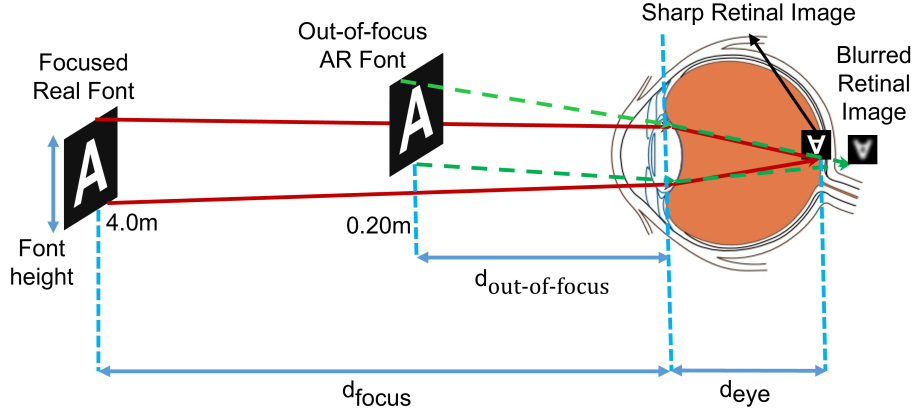


Figure 3.4: Scenario 1- Retinal image formation during out-of-focus issue in AR. S is positive. Here, the eye focuses on a real font at 4.0m, and a retina observes a corresponding sharp image. In front of the eye's focused font, an AR font is placed at 0.20m. Therefore, a correspondence retinal image is placed behind the retina, and the user observes a blurred AR font.

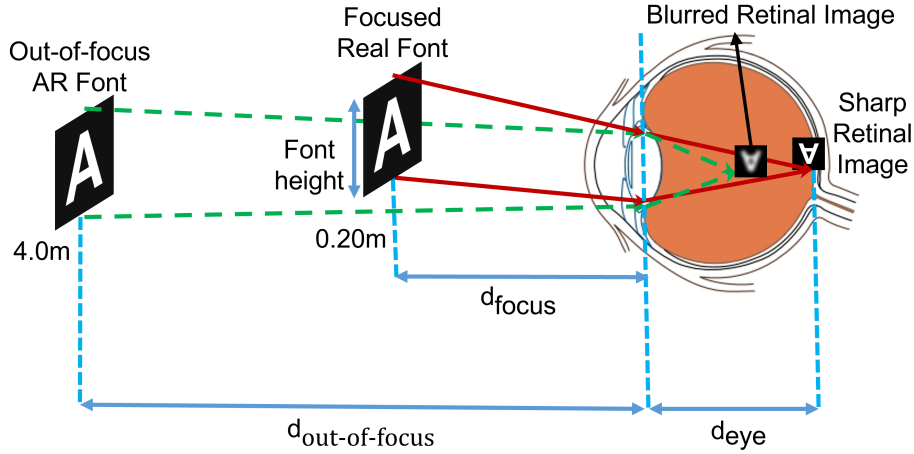


Figure 3.5: Scenario 2- Retinal image formation during out-of-focus issue in AR. S is negative. Here, the eye focuses on a real font at 0.20m, and a retina observes a corresponding sharp image. Behind the eye's focused font, an AR font is placed at 4.0m. Therefore, the user observes a correspondence retinal image placed in front of the retina and blurred.

From the above equation 3.17, it is understandable that the value of S can be positive or negative. However, $S \neq 0$ as the model supports the diffraction theory of light waves. According to the fundamental theory of light waves, diffraction will happen when light passes through an

optical system's aperture. Therefore, even though there is no aberration, the retinal image will have a slight blur [105], which is known as a "Diffraction limited" retinal image. My dissertation's model fully supports this theory. Besides, the following two scenarios could occur based on the S value.

Scenario 1: When S is positive Let's assume that a person is focusing on a real font at a far distance (4.00m) and an AR font is placed at a near distance (0.20m). Therefore, we can calculate the focal length of the eye while focusing the far distance by using equation 3.15, $f_{eye} = 0.0219m$. Finally, by putting all the known values on equation 3.17, we can get the out-of-focus aberration value, $S = +4.75D$. A schematic diagram of the scenario is given in figure 3.4.

Scenario 2: When S is negative Again, let's assume that a person is focusing on a real font at a near distance (0.20m), and an AR font is placed at a far distance (4.0m), which is out-of-focus. Therefore, we can calculate the focal length of the eye while focusing the near distance by using equation 3.15, $f_{eye} = 0.0198m$. Finally, by putting all the known values on equation 3.17, we can get the out-of-focus aberration value, $S = -4.75D$. A schematic diagram of the scenario is given in figure 3.5.

Final out-of-focus wavefront aberration function Using the equation 3.17, the eye's out-of-focus aberration for different accommodation changes can be computed. By putting S value in equation 3.14, we get,

$$C_2^0 = -\frac{R^2}{4\sqrt{3}} \times \left(\frac{1}{d_{out-of-focus}} - \left(\frac{1}{f_{eye}} - \frac{1}{d_{eye}} \right) \right) \quad (3.18)$$

Therefore, the *final wavefront aberration function* for out-of-focus problems in AR is:

$$W(x, y) = C_0^2 \times Z_0^2(x, y) = \left(-\frac{R^2}{4\sqrt{3}} \times \left(\frac{1}{d_{out-of-focus}} - \left(\frac{1}{f_{eye}} - \frac{1}{d_{eye}}\right)\right)\right) \times (\sqrt{3}(2x^2 + 2y^2 - 1)) \quad (3.19)$$

Here, x and y describe the coordinates relative to the pupil plane surface. The x and y coordinates depend on the focused real object size and corresponding retinal image size.

Amplitude function calculation In this research, the pupil plane surface for the amplitude function was completed based on the object size and corresponding retinal image size by using the formulation of similar triangles (see figure 3.6a) and similar rectangles (see figure 3.6b). The geometry in Figure 3.4 shows how to obtain the dimension of the retinal image. Let assume that the focused font height is $h_{focused}$, width is $w_{focused}$ distance from the eye is d_{focus} , retinal image height is $h_{retinal}$, width is $w_{retinal}$ and the distance between the eye lens and retina is $d_{eye} = 0.022m$. Therefore, according to the principle and geometry of similar triangle from Figure 3.6a, retinal image's height ($h_{retinal}$) was calculated. Similarly, according to the principle and geometry of similar rectangle from Figure 3.6b, retinal image's width ($w_{retinal}$) was calculated. The equations are given below:

$$\frac{h_{focused}}{d_{focused}} = \frac{h_{retinal}}{d_{eye}} \quad (3.20)$$

$$\frac{w_{focused}}{h_{focused}} = \frac{w_{retinal}}{h_{retinal}} \quad (3.21)$$

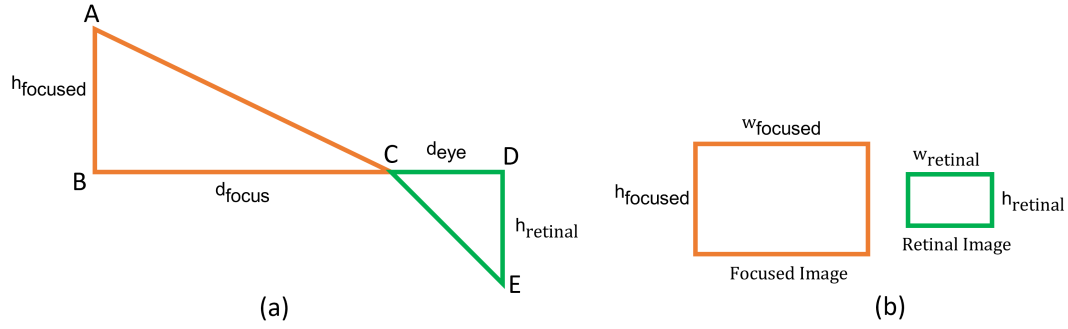


Figure 3.6: Amplitude function calculation for out-of-focus aberration using the principles of similar triangles and rectangles.

After calculating the pupil plane surface, the research followed the general principle of amplitude function mentioned before in the sub section 3.1.2.

As mentioned in sub-section 3.1.2, the wavefront aberration function is an important component of the pupil function (see equation 3.1). Therefore, by providing the final wavefront aberration function and amplitude function for the out-of-focus problem in AR in equation 3.1, the pupil function for the out-of-focus aberration is obtained. Here, $\lambda = 570\text{nm}$ as only black and white images are considered for the dissertation research. By using the principle of the optical transfer function (see equation 3.2) and convolution theorem (see equation 3.4), the final out-of-focus aberrated retinal image is calculated.

Therefore, this dissertation's model considers the following parameters to compute the out-of-focus aberration in AR.

- Eye pupil size in diameter
- Focused real object distance in meter
- Out-of-focus AR object distance in meter
- Focused real object image size (width and height)

It is essential to notice that if we consider Zernike polynomials and coefficients, we need only one human-specific parameter (*pupil radius*). Other parameters vary and depend on the specific scenario or experimental conditions and are measured according to the mathematical equations and simulation. Further, data from the person's eyeglass prescription is unnecessary for getting the wavefront aberration function of the human eye for the out-of-focus aberration modeling in AR.

3.4 Results of Out-of-focus Aberration

Figure 3.7 shows a retinal image for the out-of-focus aberration based on the above description. In this figure, according to the convolution theorem (see equation 3.4) or element-wise multiplication (see equation 3.5) between the original sharp image and the eye's PSF, the blurred retinal image is generated for the out-of-focus aberration of $+4.75D$ with pupil diameter of $5mm$. The dissertation research made the assumption that the focused object image size is $250(width) \times 250(height)$ for all the PSF images and its corresponding retinal blurred images in whole dissertation.

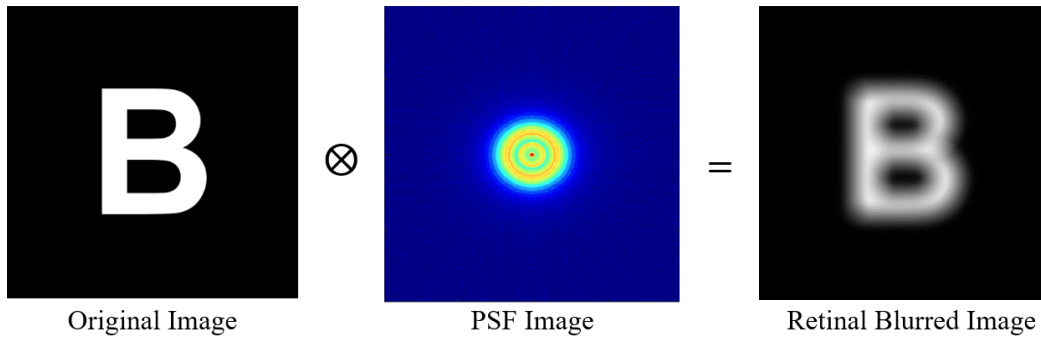


Figure 3.7: Retinal blurred image is generated for the amount of out-of-focus aberration of $+4.75D$ with a pupil diameter of $5mm$.

Impact of different amounts of pupil size on out-focus aberration As mentioned above, one of the mandatory parameters of the developed out-of-focus aberration model is the human eye's *pupil size*. While discussing the quality of the retinal image, Artal [12] mentioned that the retinal image's quality is significantly affected by pupil size (see figure 3.8). From the figure 3.8, we can observe that the larger pupil size (e.g., $7mm$) generates more blurred than the smaller pupil size (e.g., $3mm$), though the amount of aberration is the same. This dissertation's research also supports this claim.

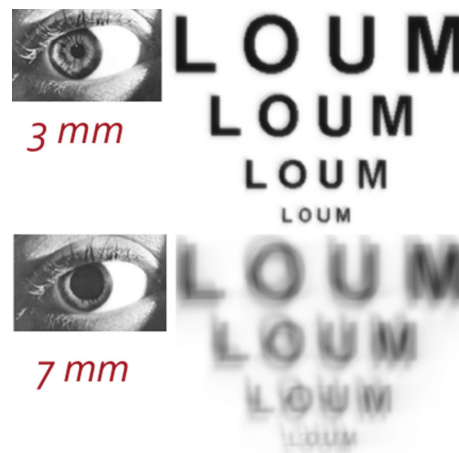


Figure 3.8: This image is adapted from the Artal [12]. The picture shows the effect of the larger pupil diameter on the retinal image quality than the smaller pupil diameter.

Figure 3.9 shows the table of different PSF images and its corresponding retinal blurred image for a range of pupil diameter ($2mm$ to $8mm$) under the out-of-focus aberration of $+4.75D$. The PSF image shows how much the wavefront intensity/frequency contributes to the blur. The color map shows the intensity/frequency level of the wavefront. Wavefront frequency values closer to zero (Blue color) means less or no blur. As the wavefront frequency values increase, the amount

<p>Focus distance = 4.0m Out of focus distance = 0.20m Out of focus Aberration (Diopter) = +4.75D</p>	Pupil Diameter (mm)	PSF Image	Simulated retinal blur image (Letter)	Simulated retinal blur image (Word)	Simulated retinal blur image (Phrase)
	2				
	3				
	4				
	5				
	6				
	7				
	8				

Figure 3.9: Effects of different pupil diameters (ranges from 2mm to 8mm) in the retinal image quality under +4.75D of out-of-focus aberration in AR.

of blur increases. Further, the wider the circle, the more waves are away from the center and the higher the blur. It means that more of the wavefront at this point is contributing to the retinal blur.

From these results, we can see that, for different sizes of pupil diameters, retinal image quality is different and have a different amount of blur. However, the amount of out-of-focus aberration is the same. More specifically, blurred on retinal image increases with pupil diameter for a constant amount of out-of-focus aberration or focal difference in AR. These findings support the previous research regarding the retinal image with different pupil diameters[13, 12, 74]. Further, Liang and Williams [74] stated that "aberrations are relatively similar between the left and right eyes of the same observer." Therefore, my dissertation model is not eye-specific. Using the same methodology, the dissertation's developed model can generate out-of-focus aberration in AR for both left and right eyes.

Impact of different amounts of out-focus aberration in the retinal image Figure 3.10 shows the table of different PSF images and their corresponding retinal blurred image for a range of out-focus aberration amounts ($-4.75D$ to $+4.75D$) under the pupil diameter of $5mm$. A positive sign (+) on the out-of-focus aberration means that the focused real object is located at a far distance, and the out-of-focus AR object is displayed at a near distance (e.g., figure 3.4), and vice versa for the negative sign (-) (e.g., 3.5). Supporting the theory of aberration, the result shows that as the amount of out-of-focus aberration or focal switching distance increases in diopters, more of the wavefront contributes to the retinal blur and the amount of blur increases in the final retinal images.

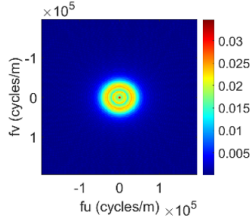



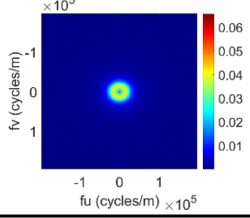



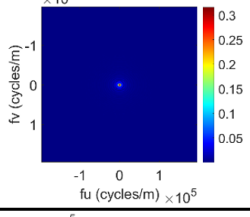



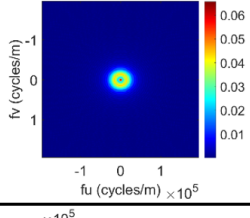



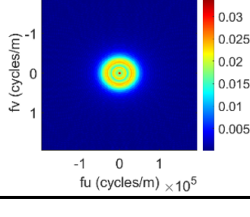
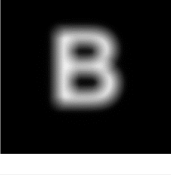


Pupil Diameter = 5mm	Out-of-focus Aberration (Diopter)	PSF Image	Simulated retinal blur image (letter)	Simulated retinal blur image (word)	Simulated retinal blur image (phrase)
	-4.75				
	-2.75				
	0.0				
	+2.75				
	+4.75				

Figure 3.10: Effects of different out-of-focus aberration amount in the retinal image under the pupil diameter of 5mm.

CHAPTER IV

OUT-OF-FOCUS CORRECTION TECHNIQUE FOR AR SYSTEM

The previous chapter described how to model the out-of-focus aberration based on the focal switching distance in AR for text-based information. This chapter provides a detailed description of the out-of-focus aberration correction technique with proper related work and background. Further, this chapter presents the image processing-based research of developing AR fonts with enhanced text legibility when seen at the time out-of-focus.

4.1 Background and Related Work

4.1.1 Algorithm Based Focus Correction Through Image Processing

Algorithm-based correction through image processing means modeling an image with the computational procedure such that the image looks sharper and more legible than the original aberrated image when seen in incorrect focus. This computational procedure is known as "*Deconvolution*", and the rendered image from the computational approach is termed as "*Pre-corrected image*". The previous chapter shows that any particular eye aberration (e.g., defocus, astigmatism, out-of-focus) can be modeled using the convolution theorem. Generally, the inverse process of convolution is the deconvolution approach. One of the essential and challenging parts of working with image processing-based focus correction is finding an appropriate and effective deconvolution algorithm.

4.1.1.1 Refractive Human Vision Error Correction

The goal of the image-based focus correction technique for refractive vision problems is to allow observers who usually wear glasses or contacts to view images without wearing their corrective optics. According to this methodology, the deconvolution algorithm generates the pre-corrected image in a controlled way such that when viewed by the observer without wearing their corrective optics, it appears perceptually correct and sharper. Through many deconvolution algorithms available in the image processing research domain, researchers use only a few specific algorithms to correct human visual aberration (individuals having refractive vision problems, including myopia, hyperopia, presbyopia, and astigmatism). Alonso and Barreto [1] in 2003 and Alonso et al. [3] in 2004 first described the image-based vision correction technique and provided the theoretical foundation for individuals with refractive vision problems. Their research aimed to pre-compensation digital images before they are displayed to low-vision users on a computer screen to compensate for their visual limitations. To conduct this research, they considered *Wiener filtering* as the deconvolution algorithm. Yellott and Yellott [124] then improved the method for reading blurred text by presbyopia but pointed out that the technique fundamentally suffers from a loss of contrast. Montalto et al. [88] developed a novel approach to generate imagery with improved sharpness for visually impaired individuals using the constrained Total Variation (TV)-based image deconvolution method (see figure 4.1). This algorithm yielded improved sharpness, higher contrast and introduced a novel parameter that allows controlling the tradeoff between sharpness, ringing, and contrast of the pre-corrected images for the refractive vision problems. Further, they provided a detailed comparison between the wiener filtering, and TV-based image deconvolution approaches with the text images synthetically and camera-based testing. Both

synthetically and camera-based testing showed that a constrained TV-based system improved the visual acuity more than Wiener filtering with a Gaussian PSF-based approach under $-2.5D$ of a blur. Further, their user study found that 86% of participants read TV-based pre-corrected text while only 9% were able to read Wiener filtering-based pre-corrected text. Therefore, it is established that the TV-based deconvolution algorithm provides better visual acuity in focus correction through image processing for individuals with vision impairments.

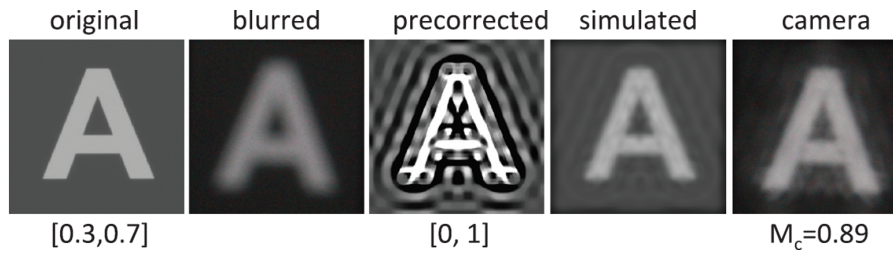


Figure 4.1: Output of the constrained TV based image deconvolution algorithm by Montalto et al. [88] for the visually impaired individuals. They considered Zernike-based PSF and TV-based deconvolution algorithms for the pre-corrected image. Their pre-corrected image showed high visual acuity both in simulation and through the camera. The blur and its corresponding pre-corrected image are generated under $-2.5D$ of blur in this image.

4.1.1.2 Out-of-focus Correction in AR

The primary principles of algorithm-based focus correction through image processing for visually impaired people are related to the accommodation mismatch and out-of-focus problems in AR and VR systems. However, successful vision correction algorithms for visual impairment methods have not been tested or validated for viewing out-of-focus virtual content in the AR system. Further, algorithm-based out-of-focus correction in AR is a less studied topic. Based on my knowledge, only three papers considered the accommodation mismatch and out-of-focus

problem in AR and VR to date [64, 31, 123]. Oshima et al. [64] developed SharpView algorithm for general images [64] (See figure 4.2). However, this work did not appropriately consider human visual aberration, considering Gaussian PSF and Wiener filtering deconvolution approach. Further, they did not evaluate how their algorithm behaves for the AR textual information. Later, Trey et al. [31] conducted a psychophysiology experiment to examine whether applying SharpView algorithm [64] to virtual text presents improved text legibility. However, they did not find any promising results for short AR text labels and symbols [31] (See figure [31]). Xu et al. [123] proposed the first fully software-based visual aberration correction for VR systems that considers Zernike polynomial and TV-based deconvolution techniques. This is the only paper that utilized the principle of refractive vision correction technique in the VR research domain. They computed the pre-corrected images based on the specific refractive error of a user in VR. However, their research goal was to correct the refractive vision aberration so that users with visual impairments can experience the VR environment without their corrective glasses or lens, not to mitigate the out-of-focus aberration.

4.1.1.3 Challenges with Algorithm Based Focus Correction Approach

Generally, the problem of rendering pre-corrected images for an algorithm-based focus correction approach seems like the standard deconvolution approach. However, the methods are different. In the traditional deconvolution approach, the goal of the deblurred algorithms is to remove the blur of the blurred image to look similar to the original image. Therefore, the blurred and original image pixel values do not change. In the focus correction-based deconvolution algorithm, the pre-corrected and original images are not the same, but the pre-corrected images look sharper under the

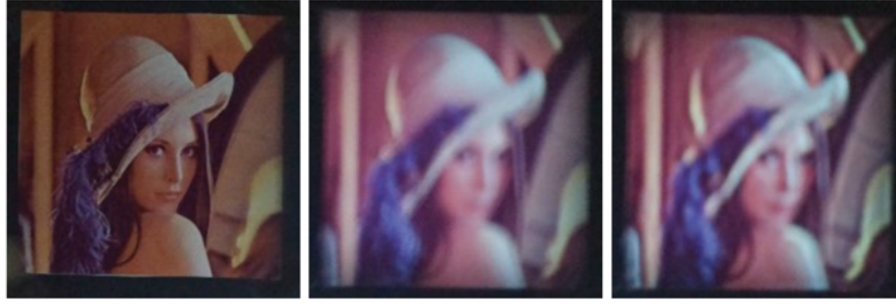


Figure 4.2: Output of the previous SharpView algorithm provided by Oshima et al. [64]. This approach is considered Gaussian-based PSF and Weiner filtering-based deconvolution approach. All three images were taken through the view of the OST AR system. The left image shows the real image in focus. Due to the focus on the real image, the virtual image becomes blurred (Middle image). An improved version of the blur virtual image after applying their sharpview algorithm is shown in the rightmost image.

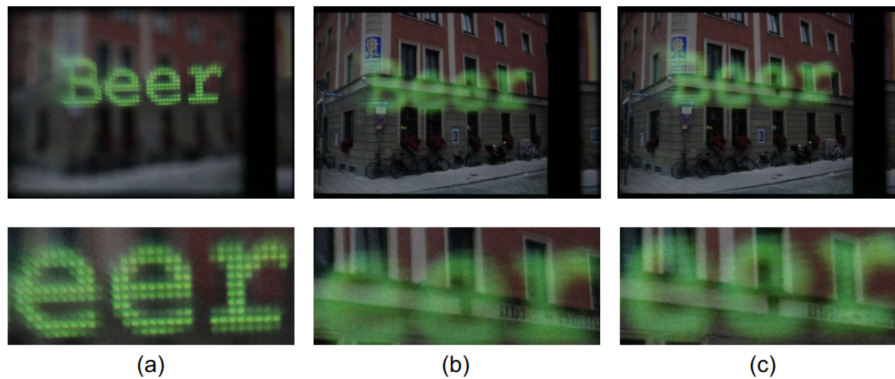


Figure 4.3: [Application of the SharpView algorithm by Oshima et al. [64] on the out of focus textual information [31]. In (a), the camera is focused on the AR text, which makes the background building out of focus. In (b), the camera is focused on the building, and the text is now out of focus. In (c), SharpView is applied to the text, and the edges of the letters have increased contrast compared to (b). Note that a camera cannot completely capture how this scene appears to a human eye. However, their research showed that the SharpView algorithm by Oshima et al. [64] did not show promising results for the AR textual information.

incorrect focus. They have improved visual acuity with incorrect focus on the retina. Therefore, the pixel values between the blurred and pre-corrected images differ. Sometimes, pre-corrected images contain negative or very high-intensity values that are unsuitable for human visual percep-

tion and decrease visual acuity [88, 32]. According to Montalto et al. [88], this problem can be solved by considering the whole precorrection approach in two steps: (1) obtain the pre-corrected image with deconvolution algorithms and (2) restrict the pixel range of the pre-corrected images within a positive valid bounded pixel value (e.g., [0, 1]). However, considering these two steps to restrict the pixel values could cause unwanted ringing artifacts and contrast reduction in the pre-corrected images [32]. Further, human visual perception is susceptible to ringing artifacts and contrast. Montalto et al. [88] developed an image processing-based deconvolution algorithm capable of controlling ringing and contrast in the pre-corrected images. Therefore, while working on the algorithm-based focus correction approach, it is necessary to consider these additional complications in the calculations.

4.1.2 Custom Hardware Based Focus Correction

Along with the image processing-based approach, the researchers have proposed many hardware-based solutions over the last decade to minimize the switching distance between virtual and real content and improve the visual acuity in the near-eye display. Further, several custom hardware-based solutions have been introduced to correct refractive vision problems in vision research. Few previous research considered both the algorithmic computation and extra optical elements in the hardware-based solution. A detailed description of different hardware-based methods from the vision and AR research is given in the following subsections.

4.1.2.1 Refractive Human Vision Error Correction

Pamplona et al. [95] presented the tailored display technology using the light field principle to compensate for human visual aberration and provide better visual acuity to individuals with

refractive vision problems. The display has the dynamic capability to adjust itself for people with specific visual aberration, including cataracts and refractive effects. However, their display prototype utilized only a small portion of the eye's field-of-view (FOV), and the perceived information has low resolution. Huang et al. [52] presented a multi-layer pre-filtering approach to correct the refractive aberration of the eye. Their method increases visual acuity so that individuals with vision impairments can perceive images on displays such as phones, tablets, laptop screens, workstation monitors, or others in sharp focus without eye-wears. Further, they mitigated the existing limitations of the single-layer display, eliminated the ringing artifacts, and increased image contrast by developing a multi-layer pre-filtering algorithm. However, their research required controlled viewer tracking to model the PSF in the display plane. Later in 2014, Huang et al. [53] developed a computational light field display with the 4D pre-filtering algorithm to correct the refractive vision problems. The authors could generate images with high contrast and resolution for refractive vision correction by combining the light field display technology and computational pre-filtering technique. Further, Barsky et al. [16] developed a multilayer display with an inverse pre-filtering method and successfully mitigated the effects of contrast loss and ringing artifacts compared to the single layer display. The methods listed above are all applied to creating screen-based displays, which observers with visual impairments can see without wearing corrective optics.

4.1.2.2 Near-eye AR Display Hardware

In the AR research domain, along with the image processing-based approach, the researchers have proposed many hardware-based solutions over the last decade to minimize the switching distance between virtual and real content and improve the visual acuity in the near-eye display.

While discussing the accommodation and vergence (see section 2.4.1), the research indicated that accommodation and vergence mismatch is a major issue for the current OST AR displays as most of the currently available OST AR displays have a single fixed focal plane. Theoretically, the out-of-focus problem is related to the accommodation mismatch issue. Therefore, there has been an intense flurry of activity in developing new hardware-based near-eye display technologies to mitigate the accommodation-vergence mismatch issue in the AR domain. Several recent surveys and review papers [57, 65, 126, 125] discussed different hardware-based approaches such as multifocal displays [79, 50, 78, 102], varifocal displays [37, 122, 23], light field displays [82, 84], retinal displays [120, 63, ?, 97] and others.

Multifocal displays Multifocal displays provide multiple image planes, each at different focal depths, allowing users to refocus between different available focal planes [125]. Multiple focal planes reduce the maximum focal distance difference between a real and virtual object. MacKenzie et al. [80] investigated accommodation control to multifocal plane displays. Their experiment revealed that three focal planes with five focal plane separation distances in the monocular condition is good enough for providing accurate stimulus to accommodation. Further, in 2012, MacKenzie et al. [79] conducted another study and found that maximum image plane separation for the binocular condition could be up to 0.6D for providing near-accurate accommodation. Recently, Change et al. [27] added a new dimension to the multifocal plane display research. They developed occlusion aware multifocal plane display prototype that enables occlusion cues between the focal planes of the multifocal plane display.

Varifocal displays Another interesting hardware-based solution is the Varifocal displays, where the display can dynamically change a single focal plane with an adjustable focal depth. Varifocal displays required a precise eye tracking method [65] that required a different method to estimate the depth [8]. Dunn et al. [37] demonstrated a prototype with a wide field of view and supported the range of focal distances from infinity to less than 10cm. Further, their prototype preserved the quality of the AR images with latency. Later, Xia et al. [122] proposed a novel near-eye display design where a tunable focus lens was adjusted to provide the accurate accommodation demand for the AR display. Under this prototype, the virtual image can be rendered from 0.33m to infinity and the tunable lens can adjust to correct accommodation for different accommodation demands. Thus, the vergence-accommodation conflict in the OST AR display can be mitigated. Further, according to this research, prescription eyeglasses won't be needed for observers with visual aberration to observe the virtual information.

Light field and Holographic displays To provide near accurate accommodation demand and mitigate the out-of-focus issue in the retina, a different light field near AR displays has been proposed [51, 82, 72]. The goal of these displays is to utilize the series of light rays to generate the virtual information with accurate accommodation and vergence demand. Lanman et al. [72] developed a microlens array-based light field near the display that can provide nearly correct depth cues such as convergence, accommodation, binocular disparity, and retinal defocus. Further, Hua and Javidi [51] integrated the freeform optical technology and microscopic integral imaging method to utilize the light rays for the reconstruction of AR 3D scene to mitigate the effects of vergence-accommodation conflict. Further, Maimone and Fuchs [82] demonstrated a light field

near-eye display prototype with a stack of light modulators to support a broad field view of the AR displays. Another near-eye AR display technology is Holographic near-eye display. Several researchers believed that holographic display technology is the ideal approach to provide a wide field of view without the vergence-accommodation conflict [122, 121]. In holographic near-eye display technology, light wavefronts are reconstructed by recording both the amplitude and phase of the display scene. Xia et al. [121] developed a prototype of the holographic near-eye display based on a lenslet array to provide a wide field of view and an expanded eye box without moving parts. Furthermore, a real-time hologram computation method that enabled near-eye holographic display was presented by Maimone et al. [83]. Besides, they extended the prototype for the vision correction capabilities and allowed a wide field of view with high image resolution.

Retinal projection displays Retinal projection displays always render the image directly on the retina. Therefore, retinal project displays do not have any focal distance, virtual information is always in focus, and it triggers depth perception only with the vergence cue. The theory and mechanism of the retinal projection display are based on the Maxwellian view [97, 76]. Under the retinal projection display, the pupil is considered a pinhole model. All the light rays are emitted from the spatial light modulator and forced to converge at the center of the pupil. Finally, a sharp, clear image formed on the retina. Therefore, this technology is suffering from the smaller eye box size, which depends on pupil size. For more information about the retinal projection display technology, I refer the reader to a recent review paper by Lin et al. [76].

4.1.2.3 Challenges with Hardware Based Focus Correction Approach

In the previous paragraphs, I have discussed some recent works of hardware-based focus correction techniques from both vision research and AR-VR research perspective. Now, some challenges of working with the hardware-based focus correction approach are given below:

1. In the refractive vision correction system, most screen-based approaches restrict the user's movement, preventing real-life practical application.
2. Most screen-based refractive vision correction techniques face numerous engineering challenges, such as contrast loss and chromatic aberration.
3. Hardware-based focus correction approaches in the AR-VR research domain, to date, only exist as optical workbench prototypes and face engineering challenges regarding eyebox size, resolution, field of view(FOV) eye tracker integration, rendering performance, and shrinking the optical and electronic components to head-worn dimensions. Some hardware-based approaches also need to handle the extra computation complexity.
4. Most prototypes exist as workbench prototypes, and it is unknown whether or how long it will take for these solutions to mature into wearable displays.

4.2 Dissertation Research's Approach for Out-of-focus Correction in the AR system

Most hardware-based methods are still workbench prototypes and face enormous engineering challenges and computational complexity. Further, the image-based approach is an integral part of some hardware-based systems, which means the efficiency of the image-based process is also crucial for the hardware-based approach. Therefore, finding an effective and appropriate image processing method to mitigate the out-of-focus problem in AR would be a novel and challenging research avenue. Mitigating the out-of-focus problem in AR means providing the nearly accurate accommodation demand of the out-of-focus information when the user focuses on information with another accommodation demand.

After examining all the advantages and challenges of both image-based and hardware-based focus correction methods, my dissertation considered fully **image-based out-of-focus correction**

methodology. Under this methodology, this research considered *constrained total-variation (TV) based image deconvolution* approach to rendering the virtual text for the out-of-focus condition in AR. This research termed the pre-corrected image as *SharpView font*, a font that looks sharper when seen in the out-of-focus. It is important to note that the font is rendered white on a black image. All the original fonts (without pre-processing) in this research are Arial fonts because, compared to other fonts, Arial font has simple geometry. Further, in the AR system, black is considered transparent. Therefore, only the white font will be displayed through the optics. Additionally, no other information exists in the image except the font.

4.2.1 Constrained Total Variation (TV) Based Out-of-Focus Correction for AR

The total variation (TV) deconvolution technique is a gradient-based image processing approach where the whole problem is considered a convex image optimization problem [26, 17]. Mathematically optimization is the process of obtaining an optimal solution for a specific function from a range of values by considering a list of constraints. The optimization process involves tuning a set of parameters to either maximize or minimize the objective function. It has various applications (e.g., image processing, circuit manufacturing, networking, machine learning, and others), but this research considers the optimization from the image processing perspective. Furthermore, the objective function of our image optimization problem is the convex optimization function. Mathematically, we can consider a function is convex if we draw a straight line between any two points of the function, then all points between those two points will be below the line. Further, the convex function could refer to every local minimum as the global minimum. Therefore, one of the main advantages of using the convex optimization function is the guarantee to obtain the global

optimum in the solution [25]. Moreover, according to Chambolle and Pock [25], the TV-based technique on convex optimization for imaging considers the sharp discontinuities in the solution, which preserve and represent essential properties of the edges, which is an important feature of this research.

Rudin, Osher, and Fatemi [104] first introduced a total variation-based approach for removing noise from the image in 1992. They minimized the total variation norm in the numerical algorithm to get the optimal solution for the denoised image. Since then, many variations of TV-based algorithms have existed for different problems and applications [19, 17, 100, 25, 88, 123]. Bioucas-Dias et al. [20] proposed a TV-based image deconvolution algorithm for the additive white Gaussian noise in the image. Their approach showed that only the best solution could be obtained by assuring the decreased function values with a few iterations instead of minimizing the objective function to its optimal value. Later, in 2007, Bioucas-Dias et al. [19] introduced two steps iterating thresholding TV algorithm with a much faster convergence rate than the traditional TV algorithm for image restoration. Another TV-based algorithm is the primal-dual approach, where one step of the optimization is performed in the primal domain and another step in the dual domain [24, 100, 25]. Beck and Teboulle [17] derived fast gradient-based algorithms for constrained total variation image denoising and deblurring Problems. Their method considered the dual-based method to get a faster global convergence rate in the solution. The research of Beck and Teboulle [17] established non-negativity and bounded value constraints novel monotone version of a fast iterative shrinkage/thresholding algorithm (FISTA) for TV-based image deblurring, which is so far the faster and better algorithm than the existing approaches.

In this research, to generate the pre-corrected image, the TV algorithm minimizes a convex objective function by tuning a set of parameters to achieve an optimal solution. We adopted the monotone version of the fast iterative shrinkage/thresholding algorithm (MFISTA) by Beck and Teboulle [17], which is a constrained TV algorithm. Considering this algorithm, the optimal pre-corrected image always has constrained or bounded pixel values between 0 and 1. Further, the methodology to calculate the contrast reduction was adapted from Montalto et al. [88]. This research also introduced a ringing control parameter to remove the unwanted ringing waves around the pre-corrected image. According to Beck and Teboulle [17] and Montalto et al., [88], the general model (see equation 4.1) for the constrained TV-based image deconvolution method to generate pre-corrected images is given below:

$$p(\theta, I_c) = \arg \min_{0 \leq \|p\| \leq 1} (\|k \otimes p - I_c\|_{L^2} + \theta \|\nabla p\|_{L^1}) \quad (4.1)$$

Here, the equation 4.1 attempts to compute the pre-corrected image (p) such a way that convolved pre-corrected image ($k \otimes p$) similar to the original image, $(k \otimes p) \sim I_c$. The constrained TV-based model (equation 4.1) can be divided into two terms:

1. *Deconvolution term*: $\|k \otimes p - I_c\|_{L^2}$ is the *deconvolution term* of the model. This term confirms that convolved pre-corrected image is visually similar to the original image in the L^2 sense [88].
2. *Regularization term*: $\theta \|\nabla p\|_{L^1}$ is the *regularization term* of the constrained TV algorithm which is the TV-norm of pre-corrected image in the L^1 sense. This term estimates, controls, and monitors the amount of ringing in the pre-corrected image. Further, it also contributes to generating sharp edges in the convolved pre-corrected image of the model [88].

4.2.1.1 Out-of-focus Correction Model Parameters

Out-of-focus PSF (k) Out-of-focus visual aberration (PSF), k , is obtained from Chapter III.

Convolution Operation (\otimes) Convolution operation (\otimes) based on the image current domain, either spatial or frequency domain.

Contrast Rescaled Version of the Original Image (I_c) Contrast rescaled version of the original image within the contrast range of lower contrast level (C_{low}) to higher contrast level (C_{high}). Let's assume that the original image is I . By following the contrast reduction technique of Montalto et al. [88], the reduced contrast version of the original image (I) is calculated using the following formula:

$$I_c = I \times (C_{high} - C_{low}) + C_{low} \quad (4.2)$$

$$\text{Percentage of contrast loss, } CL = (C_{low} + (1 - C_{high})) \times 100 \quad (4.3)$$

Here, t is the original image with no contrast loss ($C_{low} = 0$ and $C_{high} = 1$). Contrast reduction of the original image is a major part of our algorithm. Previous research by Montalto et al. [88] observed that by decreasing the original images' contrast range, more ringings were produced in the pre-corrected image and provided more sharpness and resolution in the convolved pre-corrected image ($k \otimes p$). Further, Huang et al. [52] observed that lower contrast images had more sharpness than high contrast images. Therefore, in this research, we considered the re-scaled version of the original image with a different amount of contrast loss; contrast reduction is performed until $C_{low} = C_{high}$.

Regularization parameter (θ) Regularization parameter, where $\theta > 0.0$. Here, θ behaves like a weighted value to the regularization term. If the θ value is high, then more weight contributes to the regularization term ($\theta \|\nabla p\|_{L^1}$), and the pre-corrected image will have less ringing and more

smooth. But, a very large θ value could wash out the pre-corrected image. On the other side, if the θ value is small, then the contribution of the regularization parameter to the regularization term will be small. Therefore, the pre-corrected image will have a larger amount of ringing and be less smooth. Most importantly, the regularization parameter allows us to control the desired level of ringing in the pre-corrected image based on the visual acuity.

Gradient (∇) Gradient of the image is a 2D variable function that computes the derivatives for a given matrix (e.g., the provided image) in the horizontal and vertical directions.

Pre-corrected Image (p) Pre-corrected image within the range of positive bounded pixel values $([0, 1])$. This parameter defines the final output from the algorithm. Theoretically, the pre-corrected image should be visually sharper when seen in the out-of-focus condition in AR.

4.3 Initial Results

The results of the constrained TV-based out-of-focus pre-correction algorithm are shown in Figure 4.4 (for letter), Figure 4.5 (for word) and Figure 4.6 (for phrase). The pre-corrected images for these figures are generated for the out-of-focus aberration of +4.7D with a pupil diameter of 5mm. We consider five different amount of contrast loss- 10%, 30%, 50%, 70%, and 90%, and three different levels of regulation parameter (θ)- large ($\theta = 0.01$), medium ($\theta = 0.0001$) and small ($\theta = 0.000001$).

For this research, the font is 400 pixels in size for the *letter*, *word* and *phrase*. However, each of the textual information is rendered on different resolutions of images based on the quantity of

Focus distance = 4.0m, Out of focus distance = 0.20m, Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
CL (%)		10	30	50	70	90
Contrast Adjusted Original Image						
Initial Pre-corrected Image	$\theta = 0.01$					
	$\theta = 0.0001$					
	$\theta = 0.000001$					
Final Pre-corrected Image	$\theta = 0.01$					
	$\theta = 0.0001$					
	$\theta = 0.000001$					

Figure 4.4: Results of the constrained TV-based out-of-focus pre-correction algorithm for the letter. The pre-corrected images are generated with ten different CL(%) and three different θ levels under out-of-focus aberration of +4.75D over a 5mm pupil. Final pre-corrected images are generated by the following description of section 4.4

the information, for letter 500×500 resolution image, for word 1100×500 resolution image, and phrase 1400×1200 resolution image,

Focus distance = 4.0m, Out of focus distance = 0.20m, Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter (mm) = 5mm					
CL (%)	10	30	50	70	90
Contrast Adjusted Original Image					
Initial Pre-corrected Image	$\theta = 0.01$ 	$\theta = 0.01$ 	$\theta = 0.01$ 	$\theta = 0.01$ 	$\theta = 0.01$
	$\theta = 0.0001$ 	$\theta = 0.0001$ 	$\theta = 0.0001$ 	$\theta = 0.0001$ 	$\theta = 0.0001$
	$\theta = 0.000001$ 	$\theta = 0.000001$ 	$\theta = 0.000001$ 	$\theta = 0.000001$ 	$\theta = 0.000001$
Final Pre-corrected Image	$\theta = 0.01$ 	$\theta = 0.01$ 	$\theta = 0.01$ 	$\theta = 0.01$ 	$\theta = 0.01$
	$\theta = 0.0001$ 	$\theta = 0.0001$ 	$\theta = 0.0001$ 	$\theta = 0.0001$ 	$\theta = 0.0001$
	$\theta = 0.000001$ 	$\theta = 0.000001$ 	$\theta = 0.000001$ 	$\theta = 0.000001$ 	$\theta = 0.000001$

Figure 4.5: Results of the constrained TV-based out-of-focus pre-correction algorithm for word. The pre-corrected images are generated with ten different CL(%) and three different θ levels under out-of-focus aberration of +4.75D over a 5mm pupil. Final pre-corrected images are generated by following the description of section 4.4

A detailed description and discussion of the figure 4.4 is given below:

Contrast Adjusted Original Image This row shows different rescaled versions of the original image based on the different amounts of contrast loss. The calculations are done by following the equations 4.2 and 4.3. For a particular CL(%) amount, the original image is adjusted to that CL(%) amount and then is supplied to compute the out-of-focus aberration (described in Chapter III) and pre-corrected image.










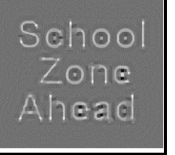














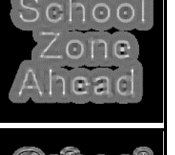




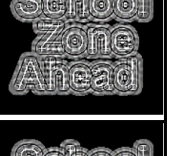





Focus distance = 4.0m, Out of focus distance = 0.20m, Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
CL (%)		10	30	50	70	90
Contrast Adjusted Original Image						
Initial Pre-corrected Image	$\theta = 0.01$					
	$\theta = 0.0001$					
	$\theta = 0.000001$					
Final Pre-corrected Image	$\theta = 0.01$					
	$\theta = 0.0001$					
	$\theta = 0.000001$					

Figure 4.6: Results of the constrained TV-based out-of-focus pre-correction algorithm for phrase. The pre-corrected images are generated with ten different CL(%) and three different θ levels under out-of-focus aberration of +4.75D over a 5mm pupil. Final pre-corrected images are generated by the following description of section 4.4

Pre-corrected Image This row has three sub-rows where each row shows the results for a specific θ value under different amounts of CL(%) for the out-of-aberration of $+4.75D$ and pupil diameter of 5mm.

$\theta = 0.01$: This sub-row shows the effect of large θ values on the pre-corrected image of different CL (%) amount. As the amount of CL(%) increases, a larger amount of ringing appears in the pre-corrected image. However, when the CL(%) is very high (e.g., CL(%) = 90), all ringing oscillations on the pre-corrected image are not visible fully due to the lower amount of contrast.

$\theta = 0.0001$: This sub-row shows the effect of medium θ values on the pre-corrected image of different CL (%) amount. Like before, as the amount of CL(%) increases, a larger amount of ringing appears in the pre-corrected image. Furthermore, the ringing amount is higher than the $\theta = 0.01$ situation. Similarly, when the CL(%) is very high (e.g., CL(%) = 90), not all ringing oscillations on the pre-corrected image are visible in the pre-corrected image due to the presence of a lower amount of contrast.

$\theta = 0.000001$: This sub-row shows the effect of smaller θ values on the pre-corrected image of different CL (%) amounts. Supporting the behavior of large and medium θ values, increasing the amount of CL(%) generates a larger amount of ringing in the pre-corrected image. Compared to the $\theta = 0.01$ and $\theta = 0.0001$ situations, pre-corrected image gains more ringing with smaller θ values. Similarly, during the very high CL(%) situation (e.g., CL(%) = 90), not all ringing oscillations on the pre-corrected image are properly visible in the pre-corrected image as it has a lower amount of contrast.

Both Figure 4.5 (for word) and Figure 4.6 (for phrase) show the similar characteristics in the pre-corrected images for the θ values and different amounts CL (%).

4.4 Final Pre-corrected Image

Alonso et al. [2] stated that not all the information on the pre-corrected image from the image deconvolution method for computer users with visual aberration is necessary. Therefore, they extracted only the necessary information from the initial pre-corrected image. In this research, we considered the same principle for the pre-corrected images of the out-of-focus correction method. We observed that ringing waves uniformly spread over the whole pre-corrected image based on the different amount of CL(%) and θ (see figures 4.4, 4.5 and 4.6). Human visual perception is highly sensitive to ringing [88, 52] and could misperceive the image due to the wrong amount of ringing in the pre-corrected image. Further, only a finite region around the boundary of the textual information is needed to perceive the information with improved visual acuity during the out-of-focus situation. Perceptually, enhanced out-of-focus text legibility highly depends on the edge region of the pre-corrected textual image. Suppose a constant amount of ringing exists around the pre-corrected textual image. In that case, it lessens the constrained bounded pixel value and allows more freedom to form a perceptually sharper edge representation. If no ringing is present, the pixel values around the edge won't ultimately appear sharper. It turns out it could have a negative impact on the perceptually observed pre-corrected images. Therefore, in this research, we have considered up to two waves of ringing in the final pre-corrected image.

The final pre-corrected image is generated based on the convolution between the pre-corrected image (p) and a mask with the required region. Mask with the required area of interest contains zeros

where the information is unnecessary and ones where the information needs to be extracted (see figure 4.7). To obtain the required mask region, we need to perform another convolution operation between the original image and a circular mask. The circular mask ($M_{circular}$) is calculated based on the original image ($I_{x,y}$), original image's center ($Center_{cx,cy}$) and diameter ($M_{diameter}$). Let's assume that α is the ringing control parameter and PSF diameter is $PSF_{diameter}$. Then, the diameter of circular mask $M_{diameter}$ is calculated from the following equation:

$$M_{diameter} = \alpha \cdot PSF_{diameter} \quad (4.4)$$

$$M_{circular}(I_{x,y}, Center_{cx,cy}, M_{diameter}) = \begin{cases} 1, & \text{if } \sqrt{(I_x - Center_{cx})^2 + (I_y - Center_{cy})^2} \leq M_{diameter} \\ 0, & \text{otherwise} \end{cases} \quad (4.5)$$

The circular mask, $M_{circular}$ is calculated by considering the pixels where the square root of the sum of squares from the center to each pixel of the original image falls within the radius value of the mask diameter (see equation 4.5). The formulation and implementation provided by Shoelson [108]. Note that by changing the value of α , the circular mask's size will change, which eventually determines the amount of ringing in the pre-corrected image. This document considers $\alpha = 0.75$ for generating two waves of ringing around the pre-corrected textual images. We obtained this value by performing the trial and error approach. By performing the convolution between the original image ($I(x, y)$) and circular mask ($M_{circular}(x, y)$), the required mask ($M_{required}(x, y)$) region is obtained (see equation 4.6 and figure 4.7). Finally, the convolution operation between the pre-corrected image ($p(x, y)$) and required mask ($M_{required}(x, y)$) provides the final pre-corrected

image ($p_{final}(x, y)$) (see equation 4.7 and figure 4.7)). The output of these two convolution operations is shown in figure 4.7.

$$M_{required}(x, y) = I(x, y) \otimes M_{circular}(x, y) \quad (4.6)$$

$$p_{final}(x, y) = p(x, y) \otimes M_{required}(x, y) \quad (4.7)$$

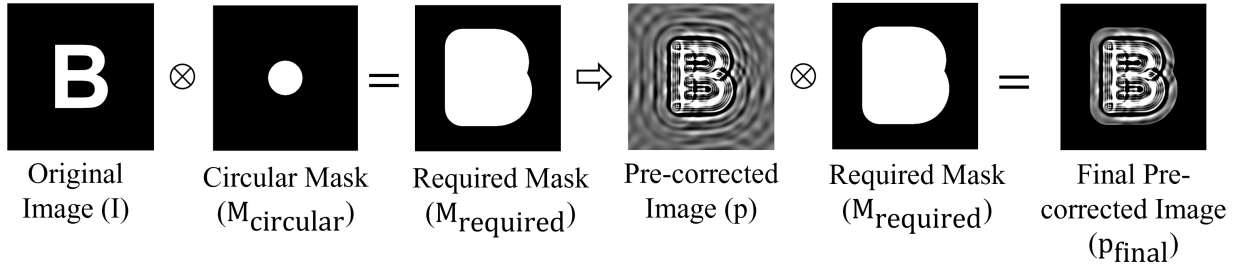


Figure 4.7: Process of generating final pre-corrected images. The process followed the description of section 4.4 and considered the equations 4.4, 4.5, 4.6 and 4.7.

The results of the final pre-corrected images are shown in the *final pre-corrected image* rows of figures 4.4, 4.5 and 4.6 for different combination of θ and CL(%) under the out-of-focus aberration of $+4.75D$ and pupil diameter of 5mm. It is important to note that in this section, we only removed the unnecessary ringing waves from the pre-corrected images to make them tidy. The TV-based algorithm's parameters, characteristics, and behavior will remain the same for each final pre-corrected image.

4.5 Analysis of Algorithmic Parameters

In the previous sections, we have seen the visual effects of our constrained TV-based deconvolution algorithm's parameters on the pre-corrected images for the out-of-focus aberration in AR.

In this section, this research analyses the relationship between the algorithm's parameters, optimal value, and pre-corrected image generation time. The main parameters of the algorithm are CL(%), regularization parameter (θ), pupil diameter, and out-of-focus aberration values. To determine the relationship, we consider the TV value which is the TV norm of pre-corrected image ($\|\nabla p\|_{L^1}$) of the algorithm. This TV value also defines the amount of energy gained in the pre-corrected image from the optimization. Further, we have considered ten different CL(%) amount (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%,) and three θ values ($1e-02$, $1e-04$, $1e-06$)

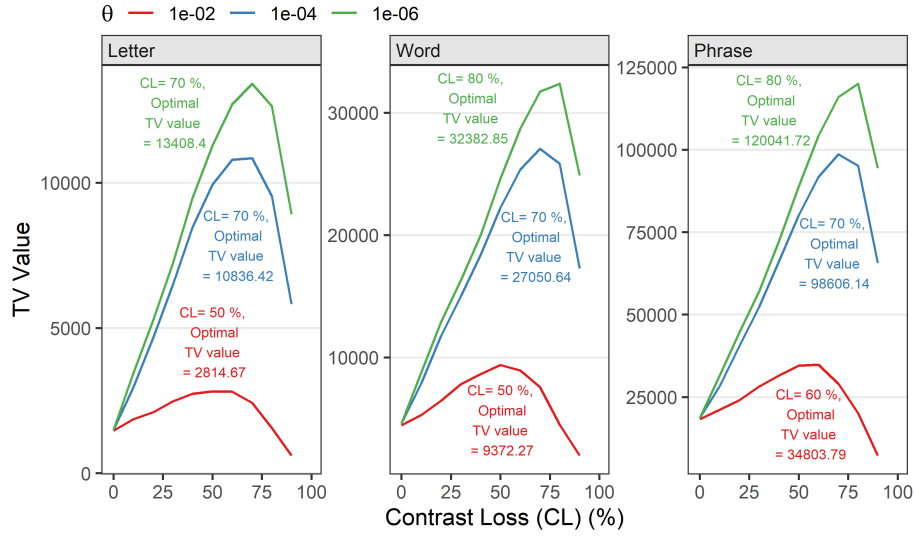


Figure 4.8: Relationship among CL (%), θ and TV value over the pupil diameter of 5mm under out-of-focus aberration of $+4.57D$. X-axis shows different amounts of CL(%) and y-axis denotes the TV value (TV-norm of pre-corrected image(p)). For each of the θ values, as the amount of CL(%) increases, the TV value increases, but after reaching its optimal value, it decreases. Large θ value gains a lower TV value in the pre-corrected image compared to the small θ values

CL (%), θ and TV value Figure 4.8 shows the relationship between the CL (%) and θ value for the pupil diameter of 5mm under out-of-focus aberration of $+4.57D$. Three colors in the figure

denote three θ values where red is for $1e - 02$ (large), blue is for $1e - 04$ (medium) and green is for $1e - 06$ (small) θ values. The observations from this figure are given below:

- For each of the θ values, as the amount of CL(%) increases, the TV value starts increasing, but after reaching its optimal value, it starts decreasing. Based on our investigation -
 - For large θ , optimal TV value is obtained within the range of 50% - 60% of CL for the textual information.
 - For medium θ , optimal TV value is obtained within the range of 70% of CL for the textual information.
 - For small θ , optimal TV value is obtained within the range of 70% - 80% of CL for the textual information.
- TV value depends on the magnitude of θ . For each of the CL(%) amounts under each type of textual information, the small θ value gains a high TV value (more ringing) in the pre-corrected image compared to the large θ values.
- From the optimal values, it is clear that for a particular θ value, *letter* has much smaller optimal TV value than *word* and *phrase*. Supporting this, *word* has larger TV value than a *letter* but a smaller TV value than *phrase*. *Phrase* has the largest TV value than a *letter* and a *word* for a specific θ value. Therefore, we can say that TV values further on the image resolution. The letter contains very little information compared to the phrase, and to render a letter required less space than a phrase. It implies that the image size of a letter is smaller than a phrase, and it gains less TV value from the optimization than the phrase.

Optimal value analysis Figure 4.9 exhibits the CL(%) and θ values for generating the optimal TV values of the algorithm over pupil diameters ranging from 2mm to 8mm. Across all the pupil diameters and textual information (letter, word, and phrase), the optimal TV value is obtained with a small θ value rather than the large and medium θ values. For the pupil diameter of 2mm, 60% CL is enough to obtain the optimal value for the letter, word, and phrase. For the pupil diameters ranging from 3mm to 6mm, the CL(%) goes from 70% to 80% to obtain the optimal value. For the large pupil sizes (7mm and 8mm), a high amount of CL(%) is necessary. Therefore, the takeaway from this analysis is that a constrained TV-based deconvolution algorithm requires more than 50%

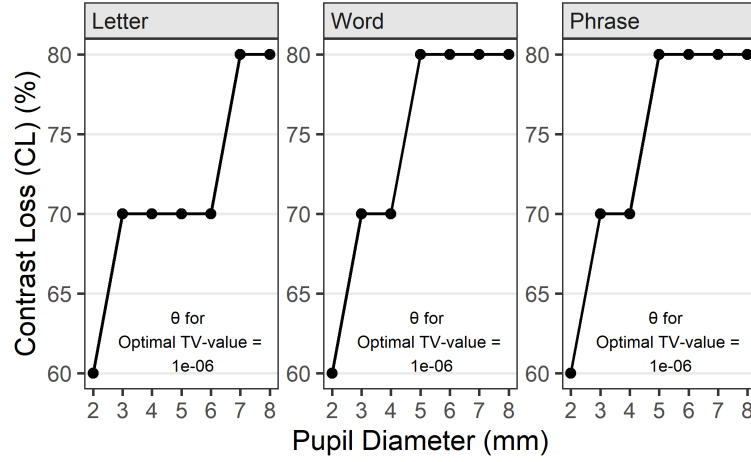


Figure 4.9: CL(%) and θ for optimal TV-value over a range of pupil diameter from 2mm to 8mm. The X-axis shows pupil diameter(mm), and the y-axis denotes the CL(%). Only $\theta = 1e - 06$ is responsible for generating the optimal TV value across all pupil diameters and textual information. The black point represents the (CL%) responsible for obtaining the TV value for each pupil's diameter.

amount of CL and a small θ value ($1e-06$) for achieving the optimal value to generate pre-corrected textual information for different pupil diameters.

Impact of pupil size on optimal pre-corrected images In the above analysis, we have seen the optimal values of the system's pre-corrected images for pupil diameters ranging from 2mm to 8mm with CL(%) and θ under the out-of-focus aberration of $+4.75D$. The optimal pre-corrected images ranges from 2mm to 8mm are given in Figure 4.10. It is important to notice here that our algorithm works well only for pupil diameters ranging from 4mm to 6mm because the pre-corrected image can preserve the shape of the original textual information, and the aberration is balanced within this pupil size range. When the pupil diameter is too small, the algorithm is not performing much pre-correction as the amount of aberration on the blurred image is small (see Chapter III, Figure 3.9). On the other side, a larger pupil diameter (ranges from 7mm to 8mm) causes the











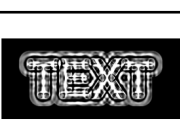

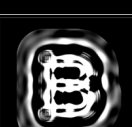
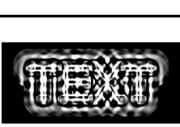

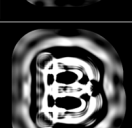


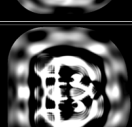


Focus distance = 4.0m, Out of focus distance = 0.20m, Out of focus Aberration (Diopter) = +4.75D	Pupil Diameter (mm)	Optimal Tidy pre-corrected image (Letter)	Optimal Tidy pre-corrected image (Word)	Optimal Tidy pre-corrected image (Phrase)
	2			
	3			
	4			
	5			
	6			
	7			
	8			

Figure 4.10: Impact of pupil size ranges from 2mm to 8mm on optimal final pre-corrected images with a letter, word, and phrase. The algorithm works only for pupil diameters ranging from 4mm to 6mm, not small pupil diameter ranges (2mm to 3mm) and large pupil diameter ranges (7mm to 8mm).

images to be washed out because the blurred image contains unbalanced aberration in the blurred image (see Chapter III, Figure 3.9) and ringing waves are not uniformly distributed.

Image Generation Time Figure 4.11 shows the mean time required to generate the pre-corrected images for pupil diameters ranging from 2mm to 8mm. Here, the mean for each pupil diameter is

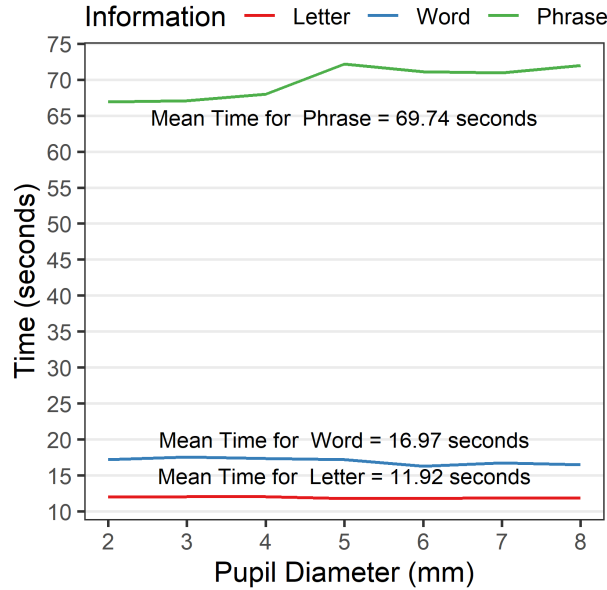


Figure 4.11: Mean time required to generate the pre-corrected images for pupil diameters ranges from 2mm to 8mm. Our approach takes approx. Twelve seconds to generate a pre-corrected image for letter, approx. Seventeen seconds to generate a pre-corrected image for word and approx. Seventy seconds to generate a pre-corrected image for a phrase.

calculated from the image generation time of 30 pre-corrected images. Each pre-corrected image is generated with the unique combination of a θ and CL(%).

The algorithms for generating pre-corrected images are implemented in Matlab. The whole system runs on a MSI Laptop with the following system configuration:

- Processor: Intel Core i7 with 2.60GHz
- RAM: 32 GB
- System type: 64-bit Windows 10 Pro Operating System
- Graphics card configuration: NVIDIA GeForce RTX 2070

Overall mean time of generating a *letter* (11.92 seconds) is smaller than a *word* (16.97 seconds) and a *phrase* (69.74 seconds). Image with *phrase* requires more time to generate the pre-corrected image. This is not surprising as the pre-corrected image generation fully depends on the image's

resolution and information. As the size of the image resolution of the letter is smaller than word and phrase, it takes less time to generate than the word and phrase. In conclusion, we can say that our approach takes approx. Twelve seconds to generate a pre-corrected image for letter, approx. Seventeen seconds to generate a pre-corrected image for word and approx. Seventy seconds to generate a pre-corrected image for a phrase.

CHAPTER V

EVALUATION OF OUT-OF-FOCUS CORRECTION TECHNIQUE FOR AR SYSTEM

The dissertation research aims to design an AR font for the out-of-focus issue in AR, which perceptually looks better (sharper) when seen out of focus than a regular unprocessed font. To accomplish this research goal, in the previous chapters, we have seen the impact of focal distance switching and out-of-focus, method of generating human perceptual out-of-focus blur, and out-of-focus correction algorithm to render pre-corrected textual images. This chapter aims to evaluate the out-of-focus correction algorithm with synthetic simulation and camera-captured images through the optics of the OST AR display, named AR Haploscope. For evaluation purposes, the term *convolved image* means the image is a computer under the synthetic simulation. Further, the term *captured image* implies that a camera captures the image through the optics of the AR Haploscope. For this evaluation, this research developed an image gradient-based *sharpness measurement technique* by comparing the presence of blur in the edge region of the image and the pre-corrected image. By using this method, the amount of out-of-focus blur can be quantified in an image. Further, this research can determine which CL (%) amount and regularization parameter (θ) value are suitable for the $+4.75D$ of out-of-focus aberration under the pupil diameter of 5mm.

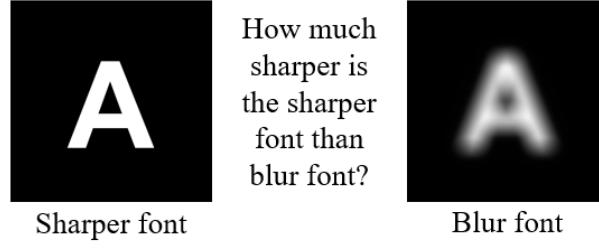


Figure 5.1: How can we define that the left image (sharper font) is sharper than the right image (blur font)?

5.1 Sharpness Measurement Method

In the evaluation stage, one crucial question- *how can we quantify the sharpness or improved visual acuity of our constraint TV-based pre-corrected images in both simulated and camera-captured output?* For example, in figure 5.1, two fonts are given, a sharper font (left image) and a blurred font (right image). It brings the question, *how much sharper is the sharper font than the blurred font?* Therefore, in this dissertation research, the sharpness measurement method determines the sharpness of the textual image based on the presence of blur around the edge region of the font. The main goal of our sharpness measurement method is to process the image in such a way that the image only contains the blur pixels, no sharp edges or pixels. Let's assume that the sharpness measurement method estimates the blur of the grayscale image $t(x, y)$ where x and y are the row and column of the image. Mathematical formulation to estimate the presence of blur amount in $t(x, y)$ image is given below:

1. Image gradient provides the strength of the edges of the image. At first, the method computes the gradient (∇) of the image $t(x, y)$ using the *Sobel* gradient operator and its corresponding gradient magnitude. As the method's goal is to compute only the blur region, we need to obtain certain pixel values which are not contributing to the blur. Therefore, we calculated the range of the magnitude image gradient (δ_G) based on the maximum and minimum magnitude of the image gradient ($\|\nabla t(x, y)\|$) with a certain threshold ($threshold_G$). The threshold is a value (%) that we added to decide the percentage of the range of the image gradient that could contribute to blur. The corresponding equation is given below:

$$\delta_G = (\max \|\nabla t(x, y)\| - \min \|\nabla t(x, y)\|) \times threshold_G \quad (5.1)$$

2. To compute the meaningful blur region, we made everything of the magnitude of the image gradient ($\|\nabla t(x, y)\|$) below the δ_G equal to zero. Therefore, we do not need to account for minimal pixel numbers. The equation is given below:

$$t(x, y) = \begin{cases} 0, & \text{if } \|\nabla t(x, y)\| \leq \delta_G \\ t(x, y), & \text{Otherwise} \end{cases} \quad (5.2)$$

3. From the previous two steps, theoretically, we could obtain the blur region for estimating the sharpness of the image. However, there is still a possibility that the image contains a certain amount of sharp edges and pixels, which are not part of the blur region. Therefore, we computed the edge gradient of the image $t(x, y)$ using the *Sobel* method as this method estimates the edges at those points where the gradient is maximum. Similar to the step 1, we calculated the range of the magnitude image edge gradient (δ_{edge}) based on the maximum and minimum magnitude of the image edge gradient ($\|\nabla t_{edge}(x, y)\|$) with a specific threshold ($threshold_{edge}$). The threshold is a value (%) that we added to decide the percentage of the range of the image edge gradient that could contribute to blur. The corresponding equation is given below:

$$\delta_{edge} = (\max \|\nabla t_{edge}(x, y)\| - \min \|\nabla t_{edge}(x, y)\|) \times threshold_{edge} \quad (5.3)$$

4. To compute the meaningful blur region without any sharp edges and pixels, we made everything of the magnitude of the image edge gradient ($\|\nabla t_{edge}(x, y)\|$) above the δ_{edge} equal to zero. Therefore, we do not need to account for huge pixel numbers. The equation is given below:

$$t(x, y) = \begin{cases} 0, & \text{if } \|\nabla t_{edge}(x, y)\| \geq \delta_{edge} \\ t(x, y), & \text{Otherwise} \end{cases} \quad (5.4)$$

5. As described previously, this research considered the different amounts of CL(%) during the generation of the pre-corrected images. Therefore, it is important to consider this property in the sharpness measurement method before estimating the presence of blur. To handle different amount of CL(%) images in the sharpness measurement technique, we have calculated the contrast invariance by normalizing the image $t(x, y)$ with the L-2 norm of the image $t(x, y)$ (see equation 5.5). Previously, Wee et al. [118] did a similar normalization step while measuring the image sharpness using eigenvalues. By performing this normalization step, our image sharpness measurement method becomes independent of the image contrast effect.

$$\overline{t(x, y)} = \frac{t(x, y)}{\|t(x, y)\|_{L2}} \quad (5.5)$$

6. At the final step, the sharpness measurement method quantifies the blur amount in an image by dividing the L1-norm of the image ($\|t(x, y)\|_{L1}$) with a number of pixels ($x \times y$) contributes to the blur. The hypothesis is that the smaller the value of the blur amount, the higher the sharpness of the textual information in the image. The equation is given below:

$$\text{Blur amount} = \frac{\|t(x, y)\|_{L1}}{(x \times y)} \quad (5.6)$$

As mentioned earlier, we need to compare two images to determine the sharper image with less blur in the edge region. Let's assume that one image is t_1 and its corresponding blur amount is Blur amount_{t_1} . Another image is t_2 and its corresponding blur amount is Blur amount_{t_2} . Therefore, using the following equation, we could determine the sharper image between t_1 and t_2 .

$$\text{Sharpness (\%)} = \frac{\text{Blur amount}_{t_1} - \text{Blur amount}_{t_2}}{\text{Blur amount}_{t_1}} \times 100 \quad (5.7)$$

Here,

- if Sharpness (%) > 0, then t_2 is sharper than t_1 .
- if Sharpness (%) < 0, then t_1 is sharper than t_2 .

Figure 5.2 shows the steps of the sharpness measurement method. Starting from the figure 5.2a, which is the original image, the out-of-focus blurred image was generated based on the description of Chapter III (see figure 5.2b) under $+4.75D$ out-of-focus aberration and 5mm diameter of pupil size. Figure 5.2c shows the blurred edge region and figure 5.2d shows the pixel intensity distribution of the blurred edge region. On the other side, figure 5.2e is the pre-corrected image from the constrained TV-based deconvolution algorithm under $+4.75D$ out-of-focus aberration and 5mm diameter of pupil size (see description in Chapter IV). Figure 5.2f is the convolved pre-corrected image. By following the steps of the sharpness measurement method figure 5.2g is

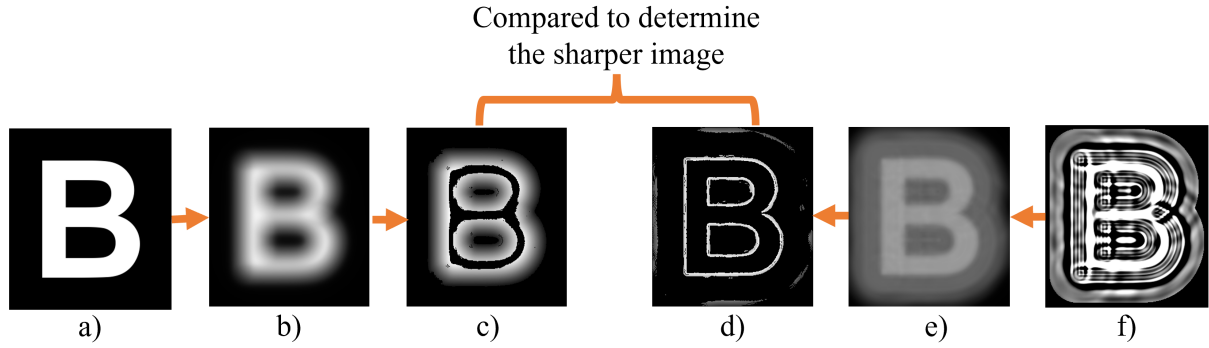


Figure 5.2: Sharpness measurement method. (a) Original image without pre-processing. (b) Blurred image with the out-of-focus aberration of $+4.75D$. (c) Blurred edge region of the out-of-focus aberrated image (b) based on the sharpness measurement method (steps 1-4). (f) Pre-corrected image with $CL=80\%$ and $\theta = 1e - 06$ for out-of-focus aberration of $+4.75D$ and pupil diameter of 5mm. (e) Convolved pre-corrected image. (d) Blurred edge region of the convolved pre-corrected image (e) based on the sharpness measurement method (steps 1-4). Finally, (c) and (d) are compared to determine the sharper image using equations 5.6 and 5.7.

generated to show the blurred edge region. Figure 5.2h shows the pixel intensity distribution of the blurred edge region. By comparing the figures 5.2d and h and using the equation 5.7, we can quantify which image is sharper and how much. In the figure 5.2, the figure 5.2h is 46.47% sharper than the figure 5.2d figure.

5.2 Parameter Selection

In chapter III, this document showed that the following parameters are needed to generate the out-of-focus blur in AR:

- Pupil size in diameter
- Focused real object distance in meter
- Out-of-focus AR object distance in meter
- Focused real object

The parameters for the out-of-focus aberration in AR can be obtained or known before generating the corresponding pre-correcting image by following the constrained TV-based deconvolution algorithm described in Chapter IV. Furthermore, for a single out-of-focus aberration, different pre-corrected images can be generated under a specific pupil diameter based on different CL(%) amount and regularization parameter(θ) values (see the description of 4.2.1.1, 4.3, Figures 4.4, 4.5 and 4.6). Therefore, it is unknown which combination of CL(%) and regularization parameter(θ) of the pre-correction algorithm is appropriate and valid for enhancing the out-of-focus text legibility over a particular out-of-focus aberration and pupil diameter.

Previously, while investigating the image-based correction for visually impaired people, Montalto et al. [88] proposed a novel relative total variation term, denoted as τ , for controlling the trade-off between the contrast loss (%) and ringing (θ) for the pre-corrected images. In their evaluation, they measured τ values as an indicator of the performance of their framework. However, the issue is that their approach can get only the optimal τ values. But, a single τ value can be obtained from different combinations of CL(%) and θ for a particular visual aberration. For example:

1. $\tau = 15$, CL(%) = 40 and $\theta = 1e - 02$
2. $\tau = 15$, CL(%) = 50 and $\theta = 1e - 02$
3. $\tau = 15$, CL(%) = 60 and $\theta = 1e - 02$

According to Montalto et al. [88], the system only considered the first combination of CL(%) and θ (CL(%) = 40, $\theta = 1e - 02$) for $\tau = 15$, and ignored the rest of the combination of CL(%) and θ . But, there is a possibility that other combinations of CL(%) and θ could improve better than the optimal one perceptually, synthetically, and optically. Therefore, this research did not consider the relative TV term (τ) for the evaluation. Instead, we considered the different amounts of CL(%)

and θ values and determined the appropriate combination of CL(%) and θ that provides the best text legibility under a specific out-of-focus aberration.

This dissertation research considered the following parameters for the evaluation:

Textual information: This research considered three types of textual information: *a font* (e.g., 'B'), *a word* (e.g., 'TEXT') and *a phrase* (e.g., 'School Zone Ahead'). While evaluating with the camera, the real information will be a cross (e.g., 'X'). All information was displayed by maintaining a constant visual angle of 1.60° .

Out-of-focus aberration: A $+4.75D$ of out-of-focus aberration was considered in the evaluation. To achieve this aberration, this research considered the focused real information distance at 4.0m based on our experimental room size. Further, out-of-focus virtual text distance was at 0.20m within arm-length distance. Therefore, a total amount of 3.8m focal switching distance was evaluated in this research.

Pupil diameter (mm): In general, pupil diameter ranges from 2mm to 8mm based on the lighting conditions. This research considered the average range from 2mm to 8mm, which is 5mm in both simulated and camera-based evaluation methods.

Contrast loss (CL)(%): Ten different CL(%) amount (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%,) were considered in the evaluation.

Regularization parameter (θ): The regularization parameter (θ) controls the amount of ringing in the pre-corrected image. The smaller value of the regularization parameter indicates high ringing,

and the higher value indicates less ringing in the pre-corrected images. Here, three θ values ($1e - 02$, $1e - 04$, $1e - 06$) were considered for the evaluation.

5.3 Evaluation Methods

In section 5.2, the document mentioned that the dissertation research considered three different textual information (a letter, a word, and a sentence), $+4.75D$ out-of-focus aberration, ten different levels of CL (%), three different levels of regularization parameter (θ) and pupil diameter of 5mm for the evaluation. Further, section 5.1 discussed the sharpness measurement method to quantify the sharpness of an image, which is an integral part of the whole evaluation process. Therefore, by considering the parameters and following the steps of the sharpness measurement method, this research evaluated the *pre-corrected images* of the TV-based out-of-focus correction method in two ways: *synthetic simulation evaluation* and *optically viewed camera captured evaluation*.

5.3.1 Synthetic Simulation Evaluation

This evaluation method examines whether the pre-corrected images show sharper and improved visual acuity in synthetic simulation. In this case, this research compared the synthetically convolved pre-corrected image with the synthetically simulated out-of-focus blurred image to determine whether the pre-corrected image mitigates the out-of-focus blur issue.

Procedure: Let's assume that the pre-corrected image is p and the PSF of the out-of-focus aberration is k . Then, the synthetically convolved pre-corrected image is generated by performing the convolution operation between the (k) and (p), which is $k \otimes p$. Similarly, let's consider the original image is I and the PSF of the out-of-focus aberration is k . Then, the synthetically simulated

out-of-focus retinal blurred image is generated by performing the convolution operation between the (k) and (I), which is $k \otimes I$. The process of obtaining the PSF of out-of-focus aberration (k) in AR is described in Chapter III. The method of generating the constrained TV-based pre-corrected image (p) is provided in Chapter IV. For determining the sharpness of the image, we considered the sharpness measurement method described in Section 5.1.

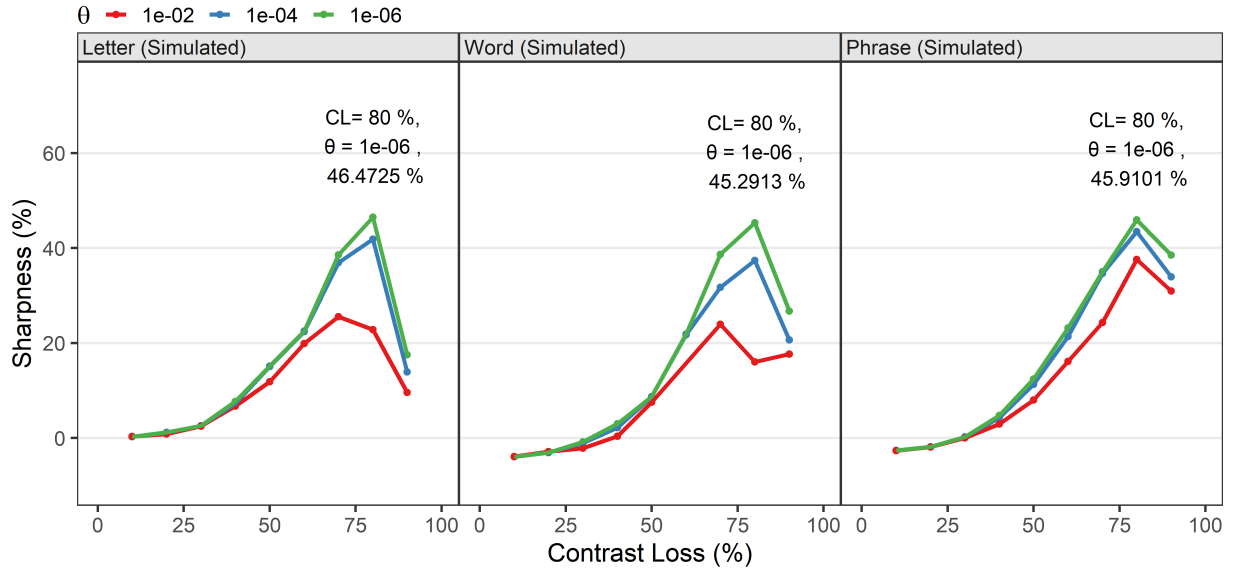


Figure 5.3: Results of the synthetic simulation evaluation. Each panel contains 30 data points representing a unique combination of CL (%) and regularization parameter (θ). The pre-corrected image parameters of the highest sharpness (%) are reported for each panel.

Results: The results of the synthetic simulation evaluation are shown in figure 5.3. Figure 5.3 has three panels to show the results of the three different textual information, the left panel for the letter, the middle panel for the word, and the right panel for the Phrase. The X-axis shows the amount of contrast loss (%). The y-axis shows the percentage of sharpness of the pre-corrected image compared to the retinal out-of-focus blurred image under $+4.75D$ out-of-focus aberration

and 5mm diameter of pupil (see the first row of figures 5.4, 5.5 and 5.6). Each panel has 30 data points representing a unique combination of CL (%) and regularization parameter (θ). Three different color lines represent three different θ values. Figure 5.3 interprets that as the amount of CL (%) increased, sharpness (%) increased for each θ value under each textual information in the synthetic simulation. However, when the amount of CL (%) is huge ($CL(\%) > 80$), the sharpness starts getting decrease, though it can maintain better resolution. The reason is that due to the severe CL (%), the pixel intensity values cannot contribute to the sharper edge region. Furthermore, when the amount of CL (%) is small ($CL(\%) \leq 25$), the pre-corrected images showed hardly any improvement for any textual information.

Figures 5.4 (for letter), 5.5 (for word) and 5.6 (for phrase) denotes the best 5 pre-corrected images out of 30 pre-corrected images of the system for the out-of-focus aberration of $+4.75D$ and pupil diameter of 5mm based on the sharpness (%) values under the synthetic simulation evaluation. Each row of CL (%) and θ are compared with the first row to compute the sharpness (%) according to the sharpness measurement method. The blurred edge region column is obtained based on the sharpness measurement method, and the pixel intensity column provides the distribution of pixel values of the blurred edge region. The main findings of the synthetic simulation evaluation are given below:

- Considering all panels of the figure 5.3, pre-corrected image with $CL = 80\%$ and $\theta = 1e - 06$ shows the best result for the $+4.75D$ of out-of-focus aberration and 5mm diameter of pupil size for all the textual information (letter, word and phrase).
- Best pre-corrected image ($CL = 80\%$ and $\theta = 1e - 06$) for *letter* presents 46.47% more sharper representation than the out-of-focus blurred image in the synthetic simulation.
- Best pre-corrected image ($CL = 80\%$ and $\theta = 1e - 06$) for *word* presents 45.29% more sharper representation than the out-of-focus blurred image in the synthetic simulation.




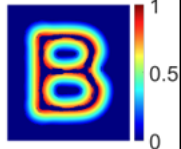



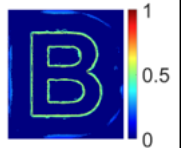



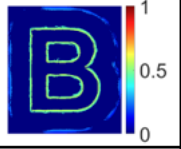



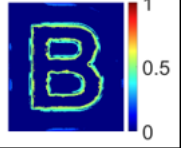



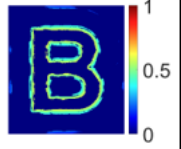



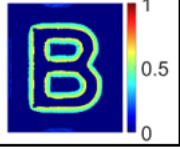
Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
Synthetic Simulation Evaluation for Letter (Best 5 pre-corrected images)		Original Image	Retinal Blurred Image	Blurred Edge Region	Pixel Intensity	
						
CL (%)	θ	Pre-corrected Image	Convolved Pre-corrected Image	Blurred Edge Region	Pixel Intensity	Sharpness (%)
80	$1e-06$					46.47
80	$1e-04$					41.86
70	$1e-06$					38.59
70	$1e-04$					36.91
70	$1e-02$					25.56

Figure 5.4: Synthetic simulation evaluation for letter. The best five pre-corrected images of the system for the out-of-focus aberration of +4.75D and pupil diameter of 5mm based on the sharpness (%) values under the synthetic simulation evaluation are listed.

- Best pre-corrected image ($CL = 80\%$ and $\theta = 1e - 06$) for *phrase* presents 45.91% more sharper representation than the out-of-focus blurred image in the synthetic simulation.

























Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
Synthetic Simulation Evaluation for Word (Best 5 pre-corrected images)		Original Image	Retinal Blurred Image	Blurred Edge Region	Pixel Intensity	
						
CL (%)	θ	Pre-corrected Image	Convolved Pre-corrected Image	Blurred Edge Region	Pixel Intensity	Sharpness (%)
80	1e-06					45.29
70	1e-06					38.65
80	1e-04					37.39
70	1e-04					31.75
90	1e-06					26.71

Figure 5.5: Synthetic simulation evaluation for word. The best five pre-corrected images of the system for the out-of-focus aberration of +4.75D and pupil diameter of 5mm based on the sharpness (%) values under the synthetic simulation evaluation are listed.

5.3.2 Optically Viewed Camera Captured Image Evaluation

In this evaluation stage, the dissertation research examines whether the constrained TV-based out-of-focus correction method signifies any sharper and improved visual acuity through the optics of the OST AR display. In this case, this research compared the camera-captured pre-corrected image with the camera-captured out-of-focus blurred image to determine whether the pre-corrected image mitigates the out-of-focus blur issue. All the images in this evaluation are taken through the optics of the AR display with the camera.

Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
Synthetic Simulation Evaluation for Phrase (Best 5 pre-corrected images)		Original Image	Retinal Blurred Image	Blurred Edge Region	Pixel Intensity	
CL (%)	θ	Pre-corrected Image	Convolved Pre-corrected Image	Blurred Edge Region	Pixel Intensity	Sharpness (%)
80	1e-06					45.91
80	1e-04					43.45
90	1e-06					38.50
80	1e-02					37.57
70	1e-06					35.01

Figure 5.6: Synthetic simulation evaluation for phrase. The best five pre-corrected images of the system for the out-of-focus aberration of +4.75D and pupil diameter of 5mm based on the sharpness (%) values under the synthetic simulation evaluation are listed.

Apparatus To achieve this, we have considered the tabletop custom-made OST AR Haploscope [109, 99, 6, 98] as the experimental display. Unlike the traditional OST AR displays, AR Haploscope can provide virtual information at different focal distances by changing the accommodative power of the accommodating lens of the optical configuration. A detailed description of design, calibration and optical configurations are given in [109] (see figure 5.7) and figure 5.8b).

Further, a DSLR camera (Nikon D3400) was used to capture images through the optics under out-of-focus. The DSLR camera was mounted on a tripod and placed behind the beam splitter of the AR Haploscope. A remote camera controller was used to reduce the displacement error from clicking the DSLR camera's button. A physical monitor was used to display real world information.

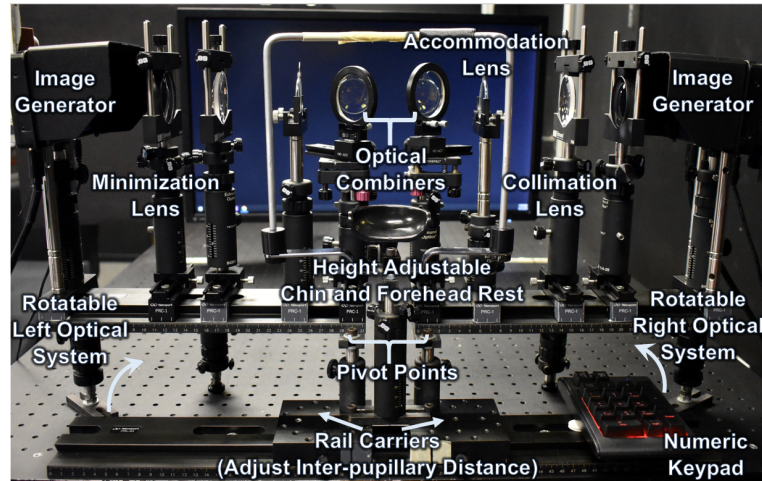


Figure 5.7: The augmented reality haploscope precisely presents virtual information at a particular focal distance.

Setup and Procedure: The setup of the camera-based evaluation method is shown in figure 5.8a and b. The DSLR camera was mounted on the tripod to look through one of the optical combiners of the AR Haploscope. The camera's lens was placed as close as to the behind of the optical combiner (15% reflectivity) so that it could focus through the center of the optical combiner. The camera focused on the real world information (cross ('X')) which was displayed on a physical monitor (Dell Ultra-sharp Monitor: U2913WM) with display resolution of 2560×1080 pixels (see figure 5.8a) at $4.0\text{m} / 0.25\text{D}$.

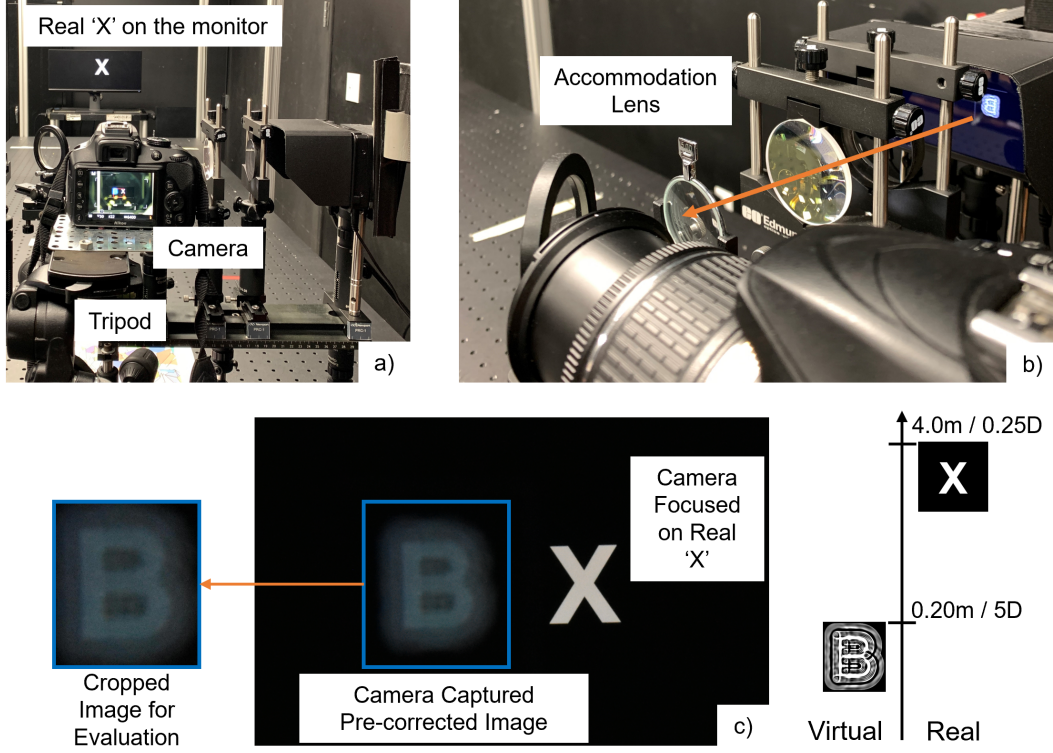


Figure 5.8: Setup of the optically viewed camera captured image evaluation. (a) A DSLR camera is mounted on a tripod and focused on a real cross ('X') on the physical monitor, which is located at 4.0m. (b) Pre-corrected image is displayed on the Haploscope's monitor. This image passes through the optics of the Haploscope and is displayed as virtual content in front of the camera at 0.20m. The pre-corrected image is out-of-focus of the camera. This particular pre-corrected image is generated with $CL=80\%$ and $\theta = 1e - 06$ for out-of-focus aberration of $+4.75D$ and pupil diameter of 5mm. (c) The camera captured an image where the camera is focused on the real cross, and our pre-corrected image looks sharper. It has improved visual acuity in the out-of-focus AR situation. Further, the portion of the camera captured pre-corrected image is cropped for evaluation.

Each virtual textural information was displayed on the AR Haploscope's right image generators. This information was passed through the optics of the Haploscope and displayed as virtual content in front of the camera (figure 5.8b). The AR Haploscope's right image generator (Feelworld F570 5.7" 4K) had a diagonal size of 14.5 cm and a display resolution of 1920×1080 pixels. AR haploscope encoded the virtual information distance by using the accommodation lens. In our

evaluation, we used 5D power of accommodation lens which placed the virtual information at 0.20m. Please note that virtual information was always out-of-focus of the camera.

The camera captured both real and virtual information through the optical combiner of the AR Haploscope, see figure 5.8c, where the left image is virtual and the right image is real (cross). Using an external remote controller's button, the camera captured the image so that no shaking noise contributed to the captured image.

Camera Settings: In our camera-based evaluation, the critical step is to confirm the camera is perfectly focused on the real cross at 4.0m and that a sharp real cross image is formed in the camera. To achieve this, this research considered the camera's concept of *depth of field (DoF)*. The DoF of the camera is the distance or area between the DoF near the limit and the DoF far limit of the camera. Any objects within the DoF of the camera formed an acceptably sharp image, and any things beyond the DoF of the camera appeared blurred. Further, we can have deep or narrow DoF based on the camera parameter settings. Deep DoF allows the camera to focus on more scene information, whereas the narrow DoF is used to focus on a specific object. The camera was under the manual setup option. Therefore, in our evaluation, we considered the deep DoF of the camera by considering the following camera setting:

- Aperture f-stop number: $\frac{f}{22}$
- Lens focal length: 55mm
- ISO: 6400
- shutter speed: $\frac{1}{50}$

According to the camera's DoF calculator by Mah and Alambra [81], the above camera setting achieved the DoF of 3m where the DoF far limit is 6m, and DoF near limit is 3m. According

to Krueger [69], pupil diameter and aperture size could be different, and a specific eye model is needed to make those equal, which is beyond this dissertation research.

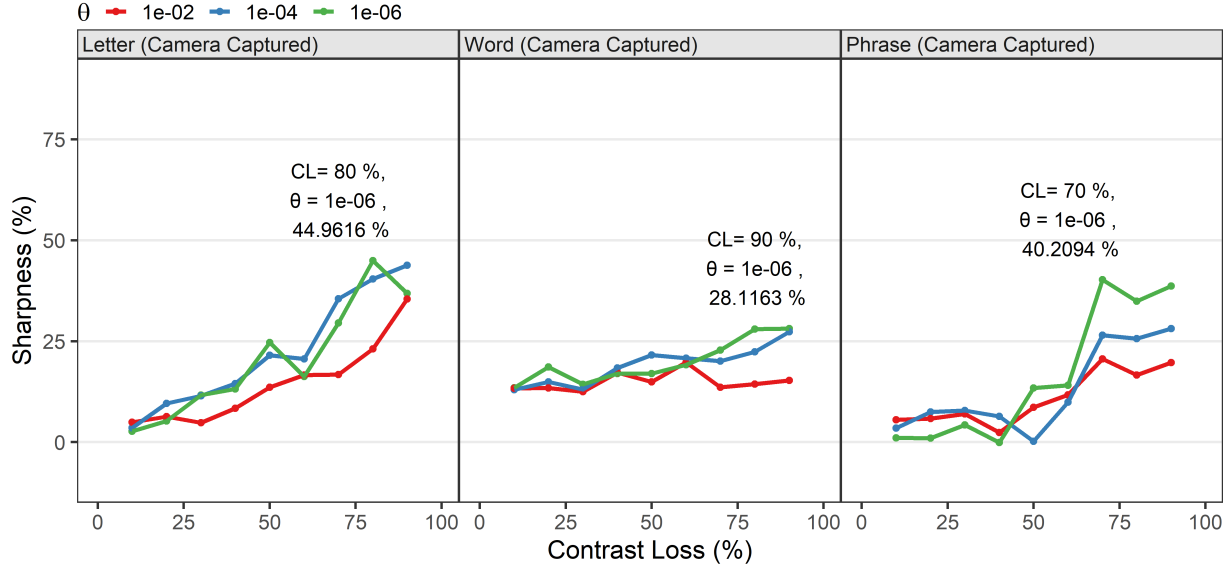


Figure 5.9: Results of the optically viewed camera captured image evaluation. Each panel contains 30 data points representing a unique combination of CL (%) and regularization parameter (θ). The pre-corrected image parameters of the highest sharpness (%) are reported for each panel.

Results: For this dissertation, 540 images are captured through the camera, where each letter, word, and phrase has 180 images. However, for this evaluation, a total of 183 images are considered, where each letter, word, and phrase has 61 images. Among those 61 images of letters, words, or phrases, 30 images are considered when the camera focused on the real cross, and pre-corrected images were presented as virtual content, termed as "*Camera captured pre-corrected image*". Another 30 images are considered when the camera directly focuses on the virtual pre-corrected images, termed as "*Camera captured focused pre-corrected image*". Furthermore, a "Camera captured




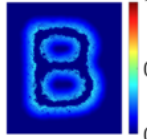



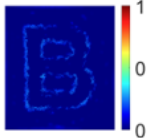



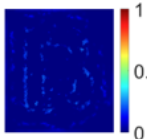



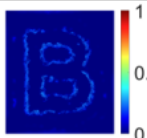



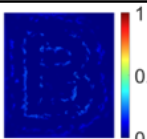


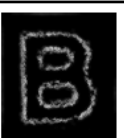
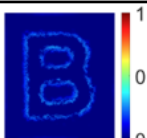
Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
Optically Viewed Camera Captured Evaluation for Letter (Best 5 pre-corrected images)		Original Image	Camera Captured Blurred Image	Blurred Edge Region	Pixel Intensity	
						
CL (%)	θ	Camera Captured Focused Pre-corrected Image	Camera Captured Pre-corrected Image	Blurred Edge Region	Pixel Intensity	Sharpness (%)
80	1e-06					44.96
90	1e-04					43.81
80	1e-04					40.40
90	1e-06					36.87
70	1e-04					35.45

Figure 5.10: Optically viewed camera captured image evaluation for the letter. Best five pre-corrected images of the system for the out-of-focus aberration of +4.75D and pupil diameter of 5mm based on the sharpness (%) values under the optically viewed camera captured image evaluation are listed.

blurred image" is considered when the camera is focused on the real cross, and the original image without any pre-correction was presented as virtual information.

To measure the Sharpness (%), this evaluation step followed the steps provided by the sharpness measurement method (see section 5.1). Unlike the convolved pre-corrected images of the synthetic







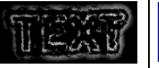

















Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
Optically Viewed Camera Captured Evaluation for Word (Best 5 pre-corrected images)		Original Image	Camera Captured Blurred Image	Blurred Edge Region	Pixel Intensity	
						
CL (%)	θ	Camera Captured Focused Pre-corrected Image	Camera Captured Pre-corrected Image	Blurred Edge Region	Pixel Intensity	Sharpness (%)
90	1e-06					28.12
80	1e-06					28.02
90	1e-04					27.35
70	1e-06					22.81
80	1e-04					22.35

Figure 5.11: Optically viewed camera captured image evaluation for word. Best five pre-corrected images of the system for the out-of-focus aberration of +4.75D and pupil diameter of 5mm based on the sharpness (%) values under the optically viewed camera captured image evaluation are listed.

simulation approach, the camera-captured images are not noise-free (see figure 5.8c). The noises on the camera captured pre-corrected images mainly came from the internal camera processing mechanism. Therefore, we performed two additional steps of denoising operation in the sharpness measurement method to reduce the noise. The two-dimensional pixel-wise adaptive noise-removal Wiener filtering method was considered in the process. This method filters the image by estimating the local image mean and standard deviation [75]. We performed this noise-removal operation before the beginning of the sharpness measurement method and before the normalization steps of

Out of focus Aberration (Diopter) = +4.75D, Pupil Diameter = 5mm						
Optically Viewed Camera Captured Evaluation for Phrase (Best 5 pre-corrected images)		Original Image	Camera Captured Blurred Image	Blurred Edge Region	Pixel Intensity	
CL (%)	θ	Camera Captured Focused Pre-corrected Image	Camera Captured Pre-corrected Image	Blurred Edge Region	Pixel Intensity	Sharpness (%)
70	$1e-06$					40.21
90	$1e-06$					38.64
80	$1e-06$					34.88
90	$1e-04$					28.10
70	$1e-04$					26.47

Figure 5.12: Optically viewed camera captured image evaluation for phrase. Best five pre-corrected images of the system for the out-of-focus aberration of +4.75D and pupil diameter of 5mm based on the sharpness (%) values under the optically viewed camera captured image evaluation are listed.

the sharpness measurement method. In our evaluation, the Wiener filter size for letter is 20×20 , for word is 20×20 and for sentence is 15×15 .

The results of the optically camera captured evaluation are shown in figure 5.9. Figure 5.9 has three panels to show the results of the three different textual information, the left panel for the letter, the middle panel for the word, and the right panel for the Phrase. The X-axis shows the amount

of contrast loss (%), and the y-axis shows the percentage of sharpness of the camera captured pre-corrected image compared to the blurred image under $+4.75D$ of out-of-focus aberration and 5mm diameter of pupil size. Each panel has 30 data points representing a unique combination of CL (%) and regularization parameter (θ). Three different color lines represent three different θ values. Figure 5.9 indicates that as the amount of CL (%) increased, sharpness (%) increased for each θ values under each textual information. Unlike the synthetic simulation approach, when the amount of CL (%) is huge ($CL(\%) > 80$), the sharpness does not start getting decrease. But, when the amount of CL (%) is small ($CL(\%) \leq 25$), the pre-corrected images showed little improvement in visual acuity for any textual information.

Figures 5.10 (for letter), 5.11 (for word) and 5.12 (for phrase) represent the best 5 pre-corrected images out of 30 pre-corrected images of the system for the out-of-focus aberration of $+4.75D$ and pupil diameter of 5mm based on the sharpness (%) values under the optically camera captured evaluation. By considering the sharpness measurement method, each row of CL (%) and θ compared with the first row to compute the sharpness (%). The blurred edge region column is obtained based on the sharpness measurement method, and the pixel intensity column provides the distribution of pixel values of the blurred edge region. The main findings of the optically camera captured evaluation are given below:

For letter: Under the $+4.75D$ of out-of-focus aberration and 5mm diameter of pupil size, pre-corrected image with $CL = 80\%$ and $\theta = 1e - 06$ shows the best improved visual acuity in the optically camera captured evaluation.

- Best pre-corrected image ($CL = 80\%$ and $\theta = 1e - 06$) shows 44.96% more sharper representation than the out-of-focus camera captured blurred image.

For word: Under the $+4.75D$ of out-of-focus aberration and 5mm diameter of pupil size, pre-corrected image with $CL = 90\%$ and $\theta = 1e - 06$ shows the best improved visual acuity in the optically camera captured evaluation.

- Best pre-corrected image ($CL = 90\%$ and $\theta = 1e - 06$) shows 28.12% more sharper representation than the out-of-focus camera captured blurred image.

For phrase: Under the $+4.75D$ of out-of-focus aberration and 5mm diameter of pupil size, pre-corrected image with $CL = 70\%$ and $\theta = 1e - 06$ shows the best improved visual acuity in the optically camera captured evaluation.

- Best pre-corrected image ($CL = 70\%$ and $\theta = 1e - 06$) shows 40.21% more sharper representation than the out-of-focus camera captured blurred image.

CHAPTER VI

CONCLUSION AND FUTURE WORK

AR is an emerging technology, and it has all the promises of becoming the next generation's everyday usable digital medium. AR experiences with OST HMD technologies are growing in prominence, mainly due to the increase in lower-cost commercially available devices, such as the Microsoft Hololens2, Microsoft Hololens1, Magic Leap, Google Glass, Epson Moverio, BT-300, and Meta 2. Due to the dramatic progress of research and the advancement of commercial AR devices, AR is rapidly progressing to provide an unprecedented richer user experience in various applications. As AR system has a wide range of applications, and commercial AR devices are advancing daily, it is essential to consider AR display technologies' current limitations and issues from the human perspective. Although AR interface design issues such as context switching, focal distance switching, and out-of-focus blur in AR are prevalent, only a few research works have been conducted to explore these issues. Further, it is established that improving the clarity of out-of-focus graphical contents during focal disparity in AR is an open research question in the AR research domain. Therefore, this research investigated AR display context switching, focal switching distance, and focal blur effects on human performance and eye fatigue under both monocular and binocular viewing conditions. Further, this research proposed a novel *SharpView font* to compensate for the effect of ubiquitous out-of-focus blur in AR, a font that looks sharper

when seen at out-of-focus, by modeling the Zernike-based out-of-focus aberration and developing a novel TV based deconvolution algorithm for the textual information. Several groups are pursuing similar goals using many hardware-based implementations- multi-focal, varifocal, light field, and holography AR display promise to eventually mitigate this problem. However, most of these displays currently exist as workbench prototypes, facing enormous engineering challenges and computational complexity. It is unknown whether or how long it will take for these solutions to mature into wearable displays. Therefore, this part of the research adds a new AR capability to the existing AR research.

In Chapter II, this research described the investigation of AR interface design issues from the human perception perspective. It has uncovered that context switching and focal distance switching are general OST AR interface design issues. The experiment partially replicated and extended the text-based visual search task reported by Gabbard et al. [42] on a custom-built OST AR display, AR Haploscope. Further, motivated by the vision science research, the research considered both monocular and binocular viewing conditions in the experiment and presented the findings related to the effects of context switching, focal distance switching, and out-of-focus blur in the context of the human visual system. The research found that context switching did not reduce performance but increased eye fatigue. As the amount of the focal switching distance increased, performance decreased, and eye fatigue increased. Furthermore, out-of-focus blur during the focal distance switching negatively affected the performance reduction. Further, as most currently available OST AR displays have a fixed or single focal plane, there is always a mismatch between the accommodation demand and vergence demand. Therefore, it is necessary to have a sharper or enhanced visual representation of the virtual information to provide the need of accommodation

required during the period of out-of-focus aberration. Chapter III and chapter IV addressed this research paradigm by modeling the out-of-focus aberration for AR and developing a constrained TV-based out-of-focus correction method for textual information in AR.

In Chapter III, the dissertation research modeled the out-of-focus visual aberration for the AR system. To model the out-of-focus PSF, we considered the Zernike polynomials for defocus aberration and modified the defocus Zernike co-efficient for the eye's accommodation changes. Therefore, the final parameters of the Zernike-based out-of-focus aberration are - (1) pupil size, (2) focused real object distance, (3) out-of-focus virtual object distance, and (4) focused real object size. Without properly modeling out-of-focus aberration and its corresponding retinal image, it would be very difficult to correct the out-of-focus aberration. Further, this research investigated the impact of pupil size and different amounts of out-focus aberrations in the retinal image. Under the same amount of out-focus aberration, retinal image quality differs for different pupil sizes. Though modeling both low-order and high-order human visual aberration with Zernike polynomials and co-efficient are common in vision research, based on my knowledge, I did not find any research discussed and modeled the Zernike-based out-of-focus aberration by examining the appropriate model parameters for the out-of-focus AR problem in AR research domain.

After generating the out-of-focus aberrated image from Chapter III, this research considered a constrained TV-based out-of-focus correction method for the textual information in AR. This part of the research is presented in Chapter IV. This correction method is an entirely software-based image processing approach where the algorithm's goal is to pre-correct the image so that it looks sharper and more legible when seen out-of-focus. Further, the optimal pre-corrected image always has bounded pixel values between 0 and 1. This research only considered the required area of

the region of the pre-corrected image and removed the unnecessary information, which makes the pre-corrected images tidy and compact. A detailed analysis of the algorithmic parameter was also performed, and research found that at least 50% contrast loss is necessary for obtaining the optimal TV value. A constrained TV-based algorithm was previously used to generate imagery with improved sharpness for individuals with refractive vision problems, including myopia, hyperopia, presbyopia, and astigmatism. However, this approach has never been tested or implemented to improve the visual acuity of out-of-focus virtual content in AR research. Therefore, this algorithm and successful investigation will add a new dimension to the body of AR-VR research.

Chapter V describes the objective evaluation of the TV-based focus correction technique for the out-of-focus problem in the AR system. We conducted the proof-of-concept evaluation with synthetic simulation and optically viewed a camera-based approach to show that the developed method improves the visual acuity of textual information under out-of-focus. We developed a gradient-based sharpness measurement method to evaluate to find a sharper image. Both synthetic simulation and camera-based approach revealed that pre-corrected images from our TV-based out-of-correction method provide nearly 50% more sharpness of the textual information than the blurred image under the $+4.75D$ out-of-focus aberration in AR. It means that if the system uses our generated pre-corrected font instead of general font, it will appear at most about 50% sharper or text legibility when seen out-of-focus. However, to develop the pre-corrected image with the best performance, we need to consider a certain amount of contrast loss, ranging between 70% to 90%, and a small regularization parameter.

6.1 Limitations and Future Work

Although the dissertation research has successfully investigated the OST AR interface design issues and implemented the constrained TV-based out-of-focus correction method for the AR system, some specific limitations still exist. The limitation of my dissertation research leads to the idea of future work. The limitations and future work of the dissertation research are given below:

User Based Evaluation In the dissertation research, we evaluated the constrained TV-based out-of-focus correction algorithm through synthetic simulation and optically viewed camera-captured approach. However, this research did not perform any formal user study to assess the parameters of the focus correction algorithm. Therefore, it is necessary to consider human-based research to determine how much and what values of the algorithm parameters are responsible for improving the text legibility under out-of-focus with our developed out-of-focus correction method.

Other Out-of-Focus Aberrations and Pupil Diameters In this research, out-of-focus aberration of $+4.75D$ with pupil diameter of $5mm$ was evaluated both synthetically and optically through camera. A potential future research could consider a range of out-of-focus aberration values with a range of pupil diameters to quantify and measure the effectiveness of the focus correction method.

Replicate Transient Focal Blur Experiment Our research found that due to the transient focal blur, participants missed information during the text-based visual search task. Further, one of the hypotheses behind the constrained TV-based out-of-focus correction method is that generated pre-corrected images would mitigate the transient out-of-focal blur effect in AR. However, this research did not perform any formal experiment to validate or verify this hypothesis. Therefore, one of the potential future studies could include the pre-corrected image from the constrained

TV-based out-of-focus correction method only during the transient focal blur period by replicating the text-based visual search task reported by Gabbard et al. [42] and Arefin et al. [4, 5, 6].

Integration of Eye Tracking The assumption behind the transient out-of-focus blur explanation assumes that participants scan the right text in the standard reading direction of left-to-right, top-to-bottom. The experiment should be replicated with an eye tracker, which could verify that the eye gaze moves in the predicted pattern. This would allow testing the hypothesis that the participant's eye gaze is on the first line of text during the transient focal blur time period. Eye movement data would also enrich understanding of context switching and focal switching distance effects.

Other Graphical Components We have developed and evaluated the constrained TV-based out-of-focus correction method only for textual information. Future research could validate this algorithm for other fundamental graphical components of OST AR systems, such as road signs, directions, symbols, and digital notifications. The hypothesis is that this algorithm will compensate for the out-of-focus problem in OST AR for other fundamental graphical contents.

Background Effects Throughout the dissertation research, we have considered the textual AR information on the black background. However, the real world is complex and dynamic. Therefore, one potential research direction would be replicating the dissertation research experiment outside or under different real-world natural conditions [7].

REFERENCES

- [1] M. Alonso and A. Barreto, “Pre-compensation for high-order aberrations of the human eye using on-screen image deconvolution,” *Engineering in Medicine and Biology Society*. 2003, vol. 1, pp. 556–559, IEEE.
- [2] M. Alonso, A. Barreto, and M. Adjouadi, “Digital image inverse filtering for improving visual acuity for computer users with visual aberrations,” *Inverse Problems in Science and Engineering*, vol. 16, no. 8, 2008, pp. 957–966.
- [3] M. Alonso, A. Barreto, and J. G. Cremades, “Image Pre-Compensation to Facilitate Computer Access for Users with Refractive Errors,” New York, NY, USA, 2003, Association for Computing Machinery.
- [4] M. S. Arefin, *Impact of Context Switching and Focal Distance Switching on Human Performance in all Augmented Reality System*, master’s thesis, Mississippi State University, Mississippi State, MS 39762, USA, 2020.
- [5] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan, “Impact of AR Display Context Switching and Focal Distance Switching on Human Performance: Replication on an AR Haploscope,” *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 571–572.
- [6] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan, “The Effect of Context Switching, Focal Switching Distance, Binocular and Monocular Viewing, and Transient Focal Blur on Human Performance in Optical See-Through Augmented Reality,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, 2022, pp. 2014–2025.
- [7] M. S. Arefin, N. Phillips, A. Plopski, and J. E. Swan, “Effects of a Distracting Background and Focal Switching Distance in an Augmented Reality System,” *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2021, pp. 96–99.
- [8] M. S. Arefin, J. E. Swan II, R. A. Cohen Hoffing, and S. M. Thurman, “Estimating Perceptual Depth Changes with Eye Vergence and Interpupillary Distance Using an Eye Tracker in Virtual Reality,” *2022 Symposium on Eye Tracking Research and Applications*, New York, NY, USA, 2022, ETRA ’22, Association for Computing Machinery.
- [9] Aristotle, *On the Soul*, translation: J. A. Smith, The Internet Classics Archive, MIT, Retrieved 2 February 2016, 350 B.C.E.

- [10] P. Artal, "Optics of the eye and its impact in vision: a tutorial," *Adv. Opt. Photon.*, vol. 6, no. 3, Sep 2014, pp. 340–367.
- [11] P. Artal, "Image Formation in the Living Human Eye," *Annual Review of Vision Science*, vol. 1, no. 1, 2015, pp. 1–17, PMID: 28532382.
- [12] P. Artal, *The Eye as an Optical Instrument*, Springer International Publishing, Cham, 2016.
- [13] P. Artal and R. Navarro, "Monochromatic modulation transfer function of the human eye for different pupil diameters: an analytical expression," *Journal of the Optical Society of America A*, vol. 11, no. 1, Jan 1994, pp. 246–249.
- [14] R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoper. Virtual Environ.*, vol. 6, no. 4, Aug. 1997, pp. 355–385.
- [15] S. Bará, M. Nievas, A. S. de Miguel, and J. Zamorano, "Zernike analysis of all-sky night brightness maps," *Applied Optics*, vol. 53, no. 12, Apr 2014, pp. 2677–2686.
- [16] B. A. Barsky, F.-C. Huang, D. Lanman, G. Wetzstein, and R. Raskar, "Vision Correcting Displays Based on Inverse Blurring and Aberration Compensation," *Computer Vision - ECCV 2014 Workshops*. 2014, Springer International Publishing.
- [17] A. Beck and M. Teboulle, "Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, Nov 2009, pp. 2419–2434.
- [18] M. Billinghurst and A. Duenser, "Augmented Reality in the Classroom," *Computer*, vol. 45, no. 7, July 2012, pp. 56–63.
- [19] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, Dec 2007, pp. 2992–3004.
- [20] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total Variation-Based Image Deconvolution: a Majorization-Minimization Approach," *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, May 2006, vol. 2, pp. II–II.
- [21] F. W. Campbell and G. Westheimer, "Dynamics of accommodation responses of the human eye," *The Journal of Physiology*, vol. 151, no. 2, 1960, pp. 285–295.
- [22] T. P. Caudell and D. W. Mizell, "Augmented reality: an application of heads-up display technology to manual manufacturing processes," *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, Jan 1992, vol. ii, pp. 659–669 vol.2.

- [23] P. Chakravarthula, D. Dunn, K. Aksit, and H. Fuchs, “FocusAR: Auto-focus augmented reality eyeglasses for both real world and virtual imagery,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, 2018, pp. 2906–2916.
- [24] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock, “An introduction to Total Variation for Image Analysis,” working paper or preprint, Nov 2009.
- [25] A. Chambolle and T. Pock, “A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, 05 2011.
- [26] A. Chambolle and T. Pock, “On the Ergodic Convergence Rates of a First-order Primal—dual Algorithm,” *Math. Program.*, vol. 159, no. 1-2, Sept. 2016, pp. 253–287.
- [27] J.-H. R. Chang, A. Levin, B. V. K. V. Kumar, and A. C. Sankaranarayanan, “Towards Occlusion-Aware Multifocal Displays,” *ACM Trans. Graph.*, vol. 39, no. 4, jul 2020.
- [28] K. V. Chellappan, E. Erden, and H. Urey, “Laser-based displays: a review,” *Appl. Opt.*, vol. 49, no. 25, Sep 2010, pp. F79–F98.
- [29] S. A. Cholewiak, G. D. Love, and M. S. Banks, “Creating correct blur and its effect on accommodation,” *Journal of Vision*, vol. 18, no. 9, 09 2018, pp. 1–1.
- [30] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied multiple regression/correlation analysis for the behavioral sciences*, 3 edition, Lawrence Erlbaum Associates Publishers, 2003.
- [31] T. Cook, N. Phillips, K. Massey, A. Plopski, C. Sandor, and J. E. Swan II, “User Preference for SharpView-Enhanced Virtual Text During Non-Fixed Viewing,” *Proceedings of IEEE Virtual Reality (IEEE VR 2018)*. March 2018, pp. 394–400, IEEE Computer Society.
- [32] C. R. M. Cruz, *Stability analysis of three inverse problems: The study of the hyperbolic inverse boundary value problem, current density impedance imaging and image based visual correction*, Phd dissertation, 2014.
- [33] J. E. Cutting and P. M. Vishton, “Chapter 3 - Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth,” *Perception of Space and Motion*, W. Epstein and S. Rogers, eds., San Diego, 1995, Handbook of Perception and Cognition, pp. 69–117, Academic Press.
- [34] G.-m. Dai, *Wavefront optics for vision correction*, vol. 179, SPIE, 01 2008.
- [35] M. Drouot, N. Le Bigot, J.-L. de Bougrenet, and V. Nourrit, “Effect of Context and Distance Switching on Visual Performances in Augmented Reality,” *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 476–477.

- [36] A. Duane, “Normal values of the accommodation at all ages,” *J. of the American Medical Association*, vol. LIX, 1912, pp. 1010–1013.
- [37] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs, “Wide field of view varifocal near-eye display using see-through deformable membrane mirrors,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, April 2017, pp. 1322–1331.
- [38] F. Echter, F. Sturm, K. Kindermann, and G. Klinker, “The Intelligent Welding Gun: Augmented Reality for Experimental Vehicle Construction,” 01 2003.
- [39] G. K. Edgar, J. C. D. Pope, and I. R. Craig, “Visual accommodation problems with head-up and helmet-mounted displays?,” *Displays*, vol. 15, no. 2, 1994, pp. 68–75.
- [40] M. Editors, “Philips Envisions Use of Augmented Reality in Operating Room (Video),”, 2019.
- [41] A. Eiberger, P. O. Kristensson, S. Mayr, M. Kranz, and J. Grubert, “Effects of Depth Layer Switching between an Optical See-Through Head-Mounted Display and a Body-Proximate Display,” *Symposium on Spatial User Interaction*, New York, NY, USA, 2019, SUI ’19, Association for Computing Machinery.
- [42] J. L. Gabbard, D. G. Mehra, and J. E. Swan II, “Effects of AR display context switching and focal distance switching on human performance,” *IEEE transactions on visualization and computer graphics*, vol. 25, 2018, pp. 2228 – 2241.
- [43] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. L. Gabbard, “What, How, and Why are Visual Assets Used in Industrial Augmented Reality? A Systematic Review and Classification in Maintenance, Assembly, and Training (From 1997 to 2019),” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, 2022, pp. 1443–1456.
- [44] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*, Pearson Education International, 2007.
- [45] J. W. Goodman, *Introduction to Fourier optics*, Englewood, CO: Roberts & Co. Publishers, 2005.
- [46] H. Gross, F. Blechinger, and B. Aichtner, *Human Eye*, John Wiley & Sons, Ltd, 2015.
- [47] R. T. Held, E. A. Cooper, J. F. O’Brien, and M. S. Banks, “Using Blur to Affect Perceived Distance and Size,” *ACM Transactions on Graphics*, vol. 29, no. 2, Mar. 2010, pp. 19:1–16.
- [48] G. Heron and B. Winn, “Binocular accommodation reaction and response times for normal observers,” *Ophthalmic and Physiological Optics*, vol. 9, 1989.

- [49] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, “Vergence-accommodation conflicts hinder visual performance and cause visual fatigue,” *Journal of Vision*, vol. 8, no. 3, Jan. 2008, pp. 1–30.
- [50] X. Hu and H. Hua, “High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics,” *Optics Express*, vol. 22, no. 11, Jun 2014, pp. 13896–13903.
- [51] H. Hua and B. Javidi, “A 3D integral imaging optical see-through head-mounted display,” *Optics Express*, vol. 22, no. 11, Jun 2014, pp. 13484–13491.
- [52] F. C. Huang, D. Lanman, B. A. Barsky, and R. Raskar, “Correcting for optical aberrations using multilayer displays,” *ACM Transactions on Graphics*, vol. 31, no. 6, 2012.
- [53] F. C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar, “Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays,” *ACM Transactions on Graphics*, vol. 33, no. 4, 2014.
- [54] A. Huckauf, M. H. Urbina, J. Grubert, I. Böckelmann, F. Doil, L. Schega, J. Tümler, and R. Mecke, “Perceptual issues in optical-see-through displays,” *Proceedings of the ACM Symposium on Applied Perception in Graphics and Visualization*. ACM, 2010, pp. 41–48.
- [55] J. H. Iavecchia, H. P. Iavecchia, and I. Stanley N Roscoe, “Eye accommodation to head-up virtual images,” *Human Factors*, vol. 30, no. 6, 1988-01-01, pp. 703–712.
- [56] S. Imamov, D. Monzel, and W. S. Lages, “Where to display? How Interface Position Affects Comfort and Task Switching Time on Glanceable Interfaces,” *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 851–858.
- [57] Y. Itoh, T. Langlotz, J. Sutton, and A. Plopski, “Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-Mounted Displays,” *ACM Computing Surveys*, vol. 54, no. 6, july 2021.
- [58] H. Kato and M. Billinghurst, “Marker tracking and HMD calibration for a video-based augmented reality conferencing system,” *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR’99)*, Oct 1999, pp. 85–94.
- [59] D. Kersten and G. E. Legge, “Convergence accommodation,” *Journal of the Optical Society of America*, vol. 73, no. 3, Mar 1983, pp. 332–338.
- [60] F. A. Khan, M. S. Arefin, N. Phillips, and J. E. Swan, “A Replication Study to Measure the Perceived Three-Dimensional Location of Virtual Objects in Optical See Through Augmented Reality,” *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022, pp. 796–797.

- [61] F. A. Khan, V. V. R. M. K. R. Muvva, D. Wu, M. S. Arefin, N. Phillips, and J. E. Swan II, "Measuring the Perceived Three-Dimensional Location of Virtual Objects in Optical See-Through Augmented Reality," *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2021, pp. 109–117, IEEE Computer Society.
- [62] F. A. Khan, V. V. R. M. K. R. Muvva, D. Wu, M. S. Arefin, N. Phillips, and J. E. Swan II, "A method for measuring the perceived location of virtual content in optical see through augmented reality," *Abstracts and Workshops Proceedings, IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2021)*. March 2021, pp. 657–658, IEEE Computer Society.
- [63] S.-B. Kim and J.-H. Park, "Optical see-through Maxwellian near-to-eye display with an enlarged eyebox," *Optics Letters*, vol. 43, no. 4, Feb 2018, pp. 767–770.
- [64] Kohei, Oshima, K. R. Moser, D. C. Rompapas, J. E. Swan II, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato, "SharpView: Improved clarity of defocused content on optical see-through head-mounted displays," *IEEE Symposium on 3D User Interfaces (3DUI 2016)*, Greenville, South Carolina, USA, Mar. 2016, pp. 173–181, IEEE.
- [65] G. A. Koulouris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality," *Computer Graphics Forum*, vol. 38, no. 2, 2019, pp. 493–519.
- [66] G.-A. Koulouris, B. Bui, M. S. Banks, and G. Drettakis, "Accommodation and Comfort in Head-Mounted Displays," *ACM Transactions on Graphics*, vol. 36, no. 4, jul 2017.
- [67] G. Kramida, "Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, July 2016, pp. 1912–1931.
- [68] B. C. Kress and W. J. Cummings, "Invited paper: Towards the ultimate mixed reality experience: HoloLens display architecture choices," *SID Symposium Digest of Technical Papers*, vol. 48, no. 1, 2017, pp. 127–131.
- [69] M. L. Krueger, M. M. Oliveira, and A. L. Kronbauer, "Personalized visual simulation and objective validation of low-order aberrations of the human eye," *Proceedings - 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2016*, 2017, pp. 64–71.
- [70] V. Lakshminarayanan and A. Fleck, "Zernike polynomials: A guide," *Journal of Modern Optics - J MOD OPTIC*, vol. 58, 04 2011, pp. 1678–1678.
- [71] M. Lambooi, M. Fortuin, I. Heynderickx, and W. IJsselstein, "Visual discomfort and visual fatigue of stereoscopic displays: a review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, Jan. 2009, pp. 030201–030201–14.

- [72] D. Lanman and D. Luebke, “Near-eye Light Field Displays,” *ACM Transactions on Graphics*, vol. 32, no. 6, Nov. 2013, pp. 220:1–220:10.
- [73] J. Liang, B. Grimm, S. Goelz, and J. F. Bille, “Objective measurement of wave aberrations of the human eye with the use of a Hartmann–Shack wave-front sensor,” *Journal of the Optical Society of America A*, vol. 11, no. 7, Jul 1994, pp. 1949–1957.
- [74] J. Liang and D. R. Williams, “Aberrations and retinal image quality of the normal human eye,” *Journal of the Optical Society of America A*, vol. 14, no. 11, Nov 1997, pp. 2873–2883.
- [75] J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, Inc., USA, 1990.
- [76] J. Lin, D. Cheng, C. Yao, and Y. Wang, “Retinal projection head-mounted display,” *Frontiers of Optoelectronics*, vol. 10, 11 2016.
- [77] M. Liu, Y. He, and B. Ye, “Image Zernike moments shape feature evaluation based on image reconstruction,” *Geo-spatial Information Science*, vol. 10, no. 3, 2007, pp. 191–195.
- [78] G. D. Love, D. M. Hoffman, P. J. W. Hands, J. Gao, A. K. Kirby, and M. S. Banks, “High-speed switchable lens enables the development of a volumetric stereoscopic display,” *Optics Express*, vol. 17, no. 18, 2009, pp. 15716–15725.
- [79] K. J. MacKenzie, R. A. Dickson, and S. J. Watt, “Vergence and accommodation to multiple-image-plane stereoscopic displays: “real world” responses with practical image-plane separations?,” *Journal of Electronic Imaging*, vol. 21, no. 1, 2012, pp. 011002–1.
- [80] K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, “Accommodation to multiple focal plane displays: Implications for improving stereoscopic displays and for accommodation control,” *Journal of Vision*, vol. 10, no. 8, 07 2010, pp. 22–22.
- [81] J. J. Mah and K. Alambra, “Depth of Field Calculator,” MATLAB Central File Exchange.
- [82] A. Maimone and H. Fuchs, “Computational augmented reality eyeglasses,” *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Oct 2013, pp. 29–38.
- [83] A. Maimone, A. Georgiou, and J. S. Kollin, “Holographic Near-Eye Displays for Virtual and Augmented Reality,” *ACM Transactions on Graphics*, vol. 36, no. 4, jul 2017.
- [84] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs, “Pinlight displays: Wide field of view augmented reality eyeglasses using defocused point light sources,” *ACM Transactions on Graphics*, vol. 33, no. 4, July 2014, pp. 89:1–89:11.
- [85] MarioCon, “Google Project Glass: Official Concept Walkthrough Video, “One Day” [HD],”, 2012.
- [86] R. J. Miller, R. G. Pigion, M. F. Wesner, and J. G. Patterson, “Accommodation fatigue and dark focus: The effects of accommodation-free visual work as assessed by two psychophysical methods,” *Perception & Psychophysics*, vol. 34, no. 6, 1983, pp. 532–540.

- [87] M. Mon-Williams and J. R. Tresilian, “Ordinal depth information from accommodation?,” *Ergonomics*, vol. 43, no. 3, 2000, pp. 391–404.
- [88] C. Montalto, I. Garcia-Dorado, D. Aliaga, M. M. Oliveira, and F. Meng, “A total variation approach for customizing imagery to improve visual acuity,” *ACM Transactions on Graphics*, vol. 34, no. 3, May 2015, pp. 28:1–28:16.
- [89] K. R. Moser, M. S. Arefin, and J. E. Swan II, “Impact of alignment point distance and posture on SPAAM calibration of optical see-through head-mounted displays,” *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2018, pp. 21–30.
- [90] K. R. Moser, M. S. Arefin, and J. E. Swan II, “Impact of Alignment Point Distance Distribution on SPAAM Calibration of Optical See-Through Head-Mounted Displays,” *Poster Abstracts, Proceedings of IEEE Virtual Reality (IEEE VR 2018)*. March 2018, pp. 641–642, IEEE Computer Society.
- [91] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wiegardt, H. Hörtnner, and C. Lindinger, “Augmented reality navigation systems,” *Universal Access in the Information Society*, vol. 4, 03 2006, pp. 177–187.
- [92] C. Neveu, T. Blackmon, and L. Stark, “Evaluation of the effects of a head-mounted display on ocular accommodation,” *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, 1998, pp. 278–289.
- [93] R. J. Noll, “Zernike polynomials and atmospheric turbulence*,” *J. Opt. Soc. Am.*, vol. 66, no. 3, Mar 1976, pp. 207–211.
- [94] A. Oagana, “2015 Mercedes-Benz C-Class W205 To Come With Head-Up Display And Airmatic,” 2013.
- [95] V. F. Pamplona, M. M. Oliveira, D. G. Aliaga, and R. Raskar, “Tailored Displays to Compensate for Visual Aberrations,” *ACM Transactions on Graphics*, vol. 31, no. 4, jul 2012.
- [96] E. J. Pedhazur, *Multiple Regression in Behavioral Research (2nd ed.)*, Holt, Rinehart, & Winston, 1982.
- [97] E. Peillard, Y. Itoh, G. Moreau, J.-M. Normand, A. Lécuyer, and F. Argelaguet, “Can Retinal Projection Displays Improve Spatial Perception in Augmented Reality?,” *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020, pp. 80–89.
- [98] N. Phillips, K. Massey, M. S. Arefin, and J. E. Swan, “Design and Calibration of an Augmented Reality Haploscope,” *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2018, pp. 75–76.

- [99] N. Phillips, K. Massey, M. S. Arefin, and J. E. Swan II, "Design, assembly, calibration, and measurement of an augmented reality haploscope," *Proceedings of PERCAR: The Fifth IEEE VR Workshop on Perceptual and Cognitive Issues in AR, IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2019)*, Osaka, Japan, Mar. 2019, pp. 1770–1774, IEEE.
- [100] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, "An algorithm for minimizing the Mumford-Shah functional," *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 1133–1140.
- [101] C. Porac and S. Coren, "The dominant eye," *Psychological Bulletin*, vol. 83, no. 5, 1976, pp. 880–897.
- [102] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs, "An extended depth-at-field volumetric near-eye augmented reality display," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, Nov 2018, pp. 2857–2866.
- [103] U. Rehman and S. Cao, "Augmented-Reality-Based Indoor Navigation: A Comparative Analysis of Handheld Devices Versus Google Glass," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 1, Feb 2017, pp. 140–151.
- [104] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, 1992, pp. 259–268.
- [105] T. O. Salmon, *Corneal contribution to the wavefront aberration of the eye*, Phd dissertation, Indiana University, 1999.
- [106] B. D. Schwab, R. E. Hand, and B. N. S. Vlaskamp, *Display systems and methods for clipping content to increase viewing comfort*, patentus 20190311527A1, Magic Leap Inc., 2019.
- [107] B. Schwerdtfeger, R. Reif, G. Klinker, D. Hamacher, L. Schega, and F. Doil, "Pick-by-vision: a first stress test," *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 2009, pp. 115–124, IEEE.
- [108] B. Shoelson, "createCirclesMask.m," 2022, MATLAB Central File Exchange.
- [109] G. Singh, S. R. Ellis, and J. E. Swan II, "The effect of focal distance, age, and brightness on near-field augmented reality depth matching," *IEEE Transactions on Visualization and Computer Graphics*, 2018, pp. 1–14.
- [110] M. B. Spitzer, X. Miao, and B. Amirparviz, "Method and apparatus for a near-to-eye display,".
- [111] I. E. Sutherland, "A Head-Mounted Three Dimensional Display," *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, New York, NY, USA, 1968, AFIPS '68 (Fall, part I), p. 757–764, Association for Computing Machinery.

- [112] D. Takahashi, “GlobalFoundries uses PTC augmented reality to transform chip manufacturing,” 2019.
- [113] L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb, “Standards for reporting the optical aberrations of eyes,” *Journal of Refractive Surgery*, vol. 18, no. 5, 2002, pp. 652–660.
- [114] D. R. Tufano, “Automotive HUDs: The Overlooked Safety Issues,” *Human Factors*, vol. 39, no. 2, 1997, pp. 303–311, PMID: 9302890.
- [115] D. Van Krevelen and R. Poelman, “A Survey of Augmented Reality Technologies, Applications and Limitations,” *International Journal of Virtual Reality*, vol. 9, no. 2, 2010, pp. 1–20.
- [116] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky, “Guidelines for Using Multiple Views in Information Visualization,” *Proceedings of the Working Conference on Advanced Visual Interfaces*, New York, NY, USA, 2000, AVI ’00, p. 110–119, Association for Computing Machinery.
- [117] A. B. Watson, “Computing human optical point spread functions,” *Journal of Vision*, vol. 15, no. 2, 2015, pp. 1–25.
- [118] C.-Y. Wee and R. Paramesran, “Image sharpness measure using eigenvalues,” *2008 9th International Conference on Signal Processing*, 2008, pp. 840–843.
- [119] R. W. West and D. A. Goss, *Introduction to the Optics of the Eye*, Butterworth-Heinemann, 2001.
- [120] Y. Wu, C. P. Chen, L. Mi, W. Zhang, J. Zhao, Y. Lu, W. Guo, B. Yu, Y. Li, and N. Maitlo, “Design of retinal-projection-based near-eye display with contact lens,” *Optics Express*, vol. 26, no. 9, Apr 2018, pp. 11553–11567.
- [121] X. Xia, Y. Guan, A. State, P. Chakravarthula, T.-J. Cham, and H. Fuchs, “Towards Eyeglass-style Holographic Near-eye Displays with Statically,” *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020, pp. 312–319.
- [122] X. Xia, Y. Guan, A. State, P. Chakravarthula, K. Rathinavel, T.-J. Cham, and H. Fuchs, “Towards a switchable AR/VR near-eye display with accommodation-vergence and eyeglass prescription support,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 11, 2019, pp. 3114–3124.
- [123] F. Xu and D. Li, “Software Based Visual Aberration Correction for HMDs,” *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 246–250.
- [124] J. I. Yellott and J. W. Yellott, “Correcting spurious resolution in defocused images,” *Human Vision and Electronic Imaging XII*, T. N. P. Bernice E. Rogowitz and S. J. Daly, eds. International Society for Optics and Photonics, 2007, vol. 6492, SPIE.

- [125] R. Zabels, K. Osmanis, M. Narels, U. Gertners, A. Ozols, K. Rūtenbergs, and I. Osmanis, “AR Displays: Next-Generation Technologies to Solve the Vergence–Accommodation Conflict,” *Applied Sciences*, vol. 9, no. 15, 2019.
- [126] T. Zhan, J. Xiong, J. Zou, and S.-T. Wu, “Multifocal displays: review and prospect,” *Photonix*, vol. 1, 12 2020.