

A SharpView Font with Enhanced Out-of-Focus Text Legibility for Augmented Reality Systems

Mohammed Safayet Arefin*
Colorado State University, USA

Carlos Montalto†
University of Costa Rica, Costa Rica
J. Edward Swan II§
Mississippi State University, USA

Alexander Plopski‡
Graz University of Technology, Austria

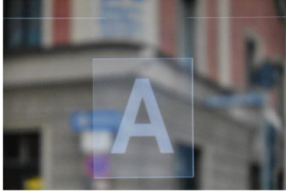

















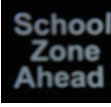

The AR Out-of-Focus Problem		The SharpView Font Solution					
		Standard Font (Arial)	SharpView Font	Standard Font (Arial)	SharpView Font	Standard Font (Arial)	SharpView Font
 Focus on standard AR font (Arial), blurred background  Blurred standard AR font (Arial), focus on the background	Focus on AR Font (20cm)	 $b = 0$		 $b = 0$		 $b = 0$	
	Focus on Background (Simulated 4m)	 $b = 0.80$	 $b = 0.46$	 $b = 0.62$	 $b = 0.35$	 $b = 0.39$	 $b = 0.24$
	Focus on Background (Optical 4m)	 $b = 1.1$	 $b = 0.80$	 $b = 0.64$	 $b = 0.47$	 $b = 0.38$	 $b = 0.26$

Figure 1: (left column) The AR Out-of-Focus problem: A user of an optical see-through AR display is looking at a building (e.g., 10 meter focal distance). (left, top) When focused on the AR display (e.g., 1.5 meter focal distance), AR fonts are in focus, but the background is blurred. (left, bottom) When focused on the background, the background is in focus, but the AR fonts are blurred. (right column) The SharpView font solution: when seen out of focus, a SharpView font looks better (sharper) than a regular font. (right, top row) Focus is on the font (the AR display, at 20 cm); note that the standard font is in good focus. (right, middle row) Focus on the background, with simulated blur at 4m. (right, bottom row) Focus on the background, with optical blur, images taken with an optical camera at 4m. For the last two rows, note that the SharpView font looks better (sharper) than the regular font. b is a quantitative measure of blur, described in the paper (see Section 6.1). Note that b is higher for standard fonts than for SharpView fonts. The values of b are scaled by a factor of 1000. Note that the camera-captured images do not reflect the actual perceptual experience of observing the fonts through the optics with a human eye—the fonts appear brighter than suggested here.

ABSTRACT

In an optical see-through augmented reality system, virtual and real information can be viewed at different distances. This requires users to frequently change their eye focus from one distance to another, and only one piece of information is in sharp focus while the other is out of focus. Previous studies have found that due to out-of-focus virtual information, when integrating information between the distances, users suffer fatigue and miss important information. Therefore, this paper introduces a novel font, termed a *SharpView font*, which looks sharper and more legible than standard fonts when seen out of focus. Our method models out-of-focus blur with *Zernike polynomials and coefficients*, develops a focus correction algorithm based on constrained *total variation* optimization, and proposes a

*e-mail: arefin@acm.org

†e-mail: carlos.montalto@ucr.ac.cr

‡e-mail: alexander.plopski@icg.tugraz.at

§e-mail: swan@acm.org

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This is an author version preprint. The final version is available as: Mohammed Safayet Arefin, Carlos Montalto, Alexander Plopski, and J. Edward Swan II. A SharpView Font with Enhanced Out-of-Focus Text Legibility for Augmented Reality Systems. In *Proceedings of IEEE Virtual Reality and 3D User Interfaces (VR)*, IEEE, 2024, pages 31–41. DOI 10.1109/VR58804.2024.00027

novel gradient-based algorithm to quantify the sharpness of textual information. We have evaluated the *SharpView font* through simulation and optically viewed camera-based measurement. When seen out of focus, our proposed font are significantly sharper than standard fonts, as assessed both visually and quantitatively through simulation (40%–44%), as well as the optics of an augmented reality display (24%–32%).

Index Terms: Augmented Reality—Focal Distance Switching—Out-of-focus Problem—Short AR Text; Depth Based Visual Aberration—Zernike Polynomials & coefficients—; Out-of-focus correction—SharpView Font—Total Variation Image Deconvolution

1 INTRODUCTION

Augmented Reality (AR) is a modern technology that superimposes computer-generated graphical content on a user’s view of the real world to enhance their vision, perception, and understanding of their environment. One way to deploy AR is through partially reflective semitransparent optical combiners, known as an optical see-through (OST) head-mounted display (HMD) [51, 52]. In this system, real and virtual information is located at varying distances from the user; which could potentially cause *context switching* (switching

visual and cognitive attention between information sources) and *focal distance switching* (changing the shape of the eye lens between different focal distances to see information in sharp focus at a new distance) [6, 7, 25]. During focal distance switching, users can only observe one piece of information in focus, and the other becomes blurred (for around 350ms [16]), generating the *out-of-focus* problem. For example, a soldier using an OST AR display on a battlefield may be looking at a real object far away (e.g., a tank) while also viewing virtual textual information (e.g., technical information about the tank) that is at a closer focal distance. This requires the soldier to constantly change the focus of his eye between the two distances, creating *context switching* and *focal distance switching*. Additionally, the soldier can only focus on one set of information, causing the other to become blurred, resulting in the *out-of-focus* problem. This can lead to eye fatigue and the potential to miss important information that could cause unexpected errors [7, 25, 31, 35, 42]. In addition to this scenario, these issues arise in many other OST AR applications, such as surgery, head-up displays in automotive, maintenance, assembly tasks, and others. Therefore, these issues can be considered a fundamental aspect of OST AR, which needs to be better understood. To address this lack of understanding, this paper presents a novel algorithm that renders virtual information in a way that allows it to look sharper and more legible when it is seen out of focus.

The initial motivation for conducting this research came from the previous work of Arefin et al. [7] and Gabbard et al. [25], where they performed a partial replication and extension of a text-based visual search task to investigate the impact of context switching, focal distance switching, and transient focal blur in OST AR systems. Both previous investigations have observed that due to transient out-of-focus blur during focal distance switching, participants' performance declined, leading to missing information in a text-based visual search task. Interestingly, Arefin et al. [7] found that this effect is not specific to AR, but is a general property of visual tasks that require integrating information from multiple displays located at different distances. Because transient out-of-focus blur is endemic to OST AR, this suggests the importance of generating a sharper representation of out of focus virtual information. To accomplish this, we specifically considered short AR text labels. Short AR textual information, for example names on buildings, has been widely used in many AR applications, including AR navigation, notifications, maintenance, education, and others. Furthermore, a recent survey by Gattullo et al. [26] on AR visual assets in maintenance found that, among 348 visual AR assets, short text labels have been the second most widely used. This highlights the importance of textual information in AR research.¹

The main purpose of our research is to create and assess a perceptual image processing-based focus correction algorithm to generate new AR fonts with improved out-of-focus text legibility. We termed this novel font as *SharpView font*, an AR font that is designed to look sharper and more legible when seen out of focus than a standard font (Fig. 1). As a first step, we modeled the blurred retinal *out-of-focus* in the OST AR system using the Zernike polynomial, which are commonly used to model visual deficiencies of human vision. We adopted the constrained total variation (TV) image deconvolution approach for systematically generating the precorrected images. Although vision scientists have used similar techniques for people with refractive vision problems (e.g., problems corrected by glasses or contacts, such as myopia, hyperopia, and astigmatism) [33, 50, 53], these methods have never been previously used to address the out-of-focus problem in OST AR. Additionally, researchers developed a SharpView algorithm for general images (e.g., the Lenna image, rock images, etc.) [39]. However, this algorithm did not consider the depth of the focused and out-of-focus virtual objects while mod-

eling the out-of-focus blur. Furthermore, their algorithm did not show promising results for textual information [21]. In addition, we developed a novel algorithm to measure and quantify font sharpness (Fig. 1). Finally, we evaluated our algorithm with synthetic simulation, and with camera-captured images through the optics of the AR display. The evaluations covered three types of textual information: a letter, a word, and a phrase. For each type of textual information, our proposed *SharpView font* has better visual acuity and appears sharper in out-of-focus situations (Fig. 1), both system visually and quantitatively. Overall, the main contributions of this research are:

- We present novel mathematical formulations for modeling out-of-focus blur with Zernike polynomials, parameterized by the depth and size of the focused and out-of-focus virtual objects for the OST AR system.
- We develop a constrained total variation (TV) optimization-based focus correction imaging algorithm to rectify the out-of-focus issue in the OST AR system. This algorithm generates the novel *SharpView font* for the OST AR system.
- We propose novel mathematical formulations based on image gradients to measure and quantify the blur amount in the font.
- We evaluated our proposed technique with both a synthetic simulation and a physical camera-based setup with an optical see-through AR display that showed improved sharpness and legibility.

2 RELATED WORK

2.1 Hardware Based Focus Correction

In the last decade, researchers have proposed a variety of hardware-based solutions to reduce the focal switching distance between virtual and real content and improve visual clarity on near-eye displays. Additionally, a large number of custom hardware-based solutions have been developed to address refractive vision issues in vision research.

Refractive Human Vision Error Correction: Pamplona et al. [53] proposed the use of the light field principle to create a customized display for individuals with refractive vision problems. However, the display prototype only covered a small area of the eye's field-of-view (FOV) and had low resolution. Huang et al. [33] then developed a multilayer prefiltering approach to correct refractive aberration for displays such as phones, tablets, laptop screens, workstation monitors, etc. They further improved the technology by creating a computational light field display with a 4D prefiltering algorithm to correct refractive vision problems [34]. This allowed them to generate images with high contrast and resolution for refractive vision correction. Barsky et al. [12] then proposed a multilayer display with an inverse prefiltering method, which successfully reduced the effects of contrast loss and ringing artifacts compared to the single-layer display. All of these methods are applied to creating screen-based displays that observers with visual impairments can see without wearing corrective optics.

Near-eye AR Display Hardware: In the AR research field, a great deal of effort has been put into creating new hardware-based near-eye display technologies to address the issue of accommodation-vergence mismatch. Surveys [36, 40] have discussed various hardware-based approaches, such as multifocal displays [32, 45, 46, 58], varifocal displays [17, 24, 66], light field displays [48, 49], and retinal displays [38, 54, 65], among others. However, these hardware-based focus correction techniques face many difficulties, such as small eye box size, low resolution, low contrast, small field of view (FOV), the need for eye tracker integration, and rendering performance. Notably, most of the hardware-based focus correction approaches

¹Portions of this work are reported in a Doctoral Consortium abstract [4] and a PhD dissertation [5] by the first author.

in the AR-VR research domain, to date, only exist as optical workbench prototypes, and it is uncertain how long it will take for these solutions to become wearable displays.

2.2 Algorithm Based Focus Correction

Algorithm-based focus correction methods use the computational image processing approach to generate the required image for a specific error. When the optical blur is assumed to be space-invariant, then this procedure is known as *deconvolution*, and the rendered image is termed a *precorrected image*. Unlike the standard deconvolution problem of removing blur and noise from the image, the algorithm-based focus correction approach renders the precorrected image in such a way that it looks sharper under the incorrect focus. Sometimes, precorrected images contain negative or very high-intensity values that are not suitable for human visual perception and decrease visual acuity [22, 50]. Based on this, Montalto et al. [50] mentioned three challenges for the image processing-based deconvolution approach for focus correction: stability, bounded pixel values, and ringing sensitivity; the last two do not appear in standard image deconvolution applications.

Refractive Human Vision Error Correction: Alonso and Barreto [1] and Alonso et al. [3] were the first to introduce the use of image-based refractive vision correction for computer screens, which implemented *Wiener filtering* as the focus correction algorithm. Yellott and Yellott [68] then improved the method for reading blurred text with presbyopia, but their technique had low contrast. Montalto et al. [50] developed a novel approach to generate imagery with improved sharpness for visually impaired individuals using the constrained total variation (TV) based image deconvolution method. This algorithm generated an increased sharpness and higher contrast of the precorrected images for refractive vision problems. Furthermore, they provided a thorough comparison between the Wiener filtering and TV-based image deconvolution approaches with the text images synthetically and camera-based testing, and noticed that a constrained TV-based system improved the visual acuity more than Wiener filtering with a Gaussian PSF-based approach. Therefore, the TV-based deconvolution algorithm is established to provide better visual acuity in focus correction through image processing for people with vision impairments. Furthermore, Xu et al. [67] proposed the first fully software-based visual aberration correction for visually impaired people for VR systems, considering Zernike polynomial and TV-based deconvolution techniques. Guzel et al. [30] also considered the Zernike-based PSF modeling and the convolutional neural network-based prescription correction method for visually impaired people in VR.

Out-of-focus Correction in AR: Algorithm-based out-of-focus correction in AR is a relatively less explored topic. To our knowledge, only two papers have addressed the problem of out-of-focus in AR to date [21, 39]. Oshima et al. [39] developed the SharpView algorithm for natural general images [39]. However, they did not appropriately take into account human visual aberration, and considered a standard Gaussian PSF and Wiener filtering deconvolution approach. Later, Trey et al. [21] conducted a psychophysical experiment to examine whether this approach improved text legibility. However, they did not find any promising results for short AR text labels.

2.3 Total Variation (TV) Deconvolution

The total variation (TV) deconvolution technique is a gradient-based image processing approach in which the whole problem is considered as a convex image optimization problem [13, 20]. One of the main advantages of using the convex optimization function is the guarantee of obtaining the global optimum in the solution [19]. Rudin et al. [59] first introduced a total variation-based approach to remove noise from the image by minimizing the total variation norm. Since then, there have been many TV-based algorithms for different

problems and applications [13, 14, 19, 50, 57, 67]. Bioucas-Dias et al. [15] proposed a TV-based image deconvolution algorithm for additive white Gaussian image noise. Their approach showed that the best solution could only be obtained by ensuring a decrease in function values with a few iterations instead of minimizing the objective function to its optimal value. Later, in 2007, Bioucas-Dias et al. [14] introduced a two-step iteration of the TV thresholding 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 an optimization step is performed in the primal domain and another step in the dual domain [18, 19, 57]. Beck and Teboulle [13] derived fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. Their method considered the dual-based method to obtain a faster global convergence rate in the solution. The research of Beck and Teboulle [13] established a novel monotone version of the non-negativity and bounded value constraints of a fast iterative shrinkage/thresholding algorithm (FISTA) for TV-based image deblurring.

3 MODELING VISUAL ABERRATION

To create a sharper and more legible visual representation of virtual information during the out-of-focus period, we first need to model the image formed on the retina at the moment of out-of-focus aberration. To model out-of-focus aberration in the human eye, it is essential to understand the mechanism of retinal image formation.

3.1 Basic Principle of Retinal Image Formation

Retinal Image: The image of an object on the retina is referred to as a *retinal image* I_R . It is formed by the light from the object entering the eye through the pupil and being focused on the retina of the eye by the lens of the eye, whose accommodative power is adjusted by the ciliary muscle [29]. The amount of light entering the eye is controlled by the pupil radius, which can range from 2–4mm in bright light conditions and 4–8mm in low light conditions [9, 29]. By controlling the amount of light entering the eye, the pupil size also affects the amount of observed blur for objects that lie at a depth different from the one focused by the eye lens. The resulting image is also affected by the refractive indices of different parts of the eye, such as the cornea and lens. Given an observed image I , we can describe the corresponding retinal image I_R as

$$I_R = I * k, \quad (1)$$

where $*$ describes a convolution and k is the point spread function (PSF) of the eye that describes the diffraction of light in the pupil, refractive errors, and other optical properties of the human eye.

Point Spread Function of the Eye: The PSF of the human eye can be calculated from the generalized pupil function with complex values [23, 28, 41]

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

where λ is the wavelength of light in a vacuum and (x, y) are the coordinates of the image surface of the pupil plane. $A(x, y)$ is a real-valued circular amplitude function that determines the relative efficiency of light transmission through the pupil. It takes the value 1 within the pupil region and 0 outside. The wavefront aberration function ($W(x, y)$) is the difference between the ideal spherical wavefront and the aberrated wavefront for every point (x, y) on the surface of the pupil plane [8, 9, 60]. The PSF is then defined as

$$k(x, y) = \frac{\mathcal{F}(\mathbb{P})(x, y) \circ \overline{\mathcal{F}(\mathbb{P})(x, y)}}{\|\mathcal{F}(\mathbb{P})(x, y) \circ \mathcal{F}(\mathbb{P})(x, y)\|_{l1}}. \quad (3)$$

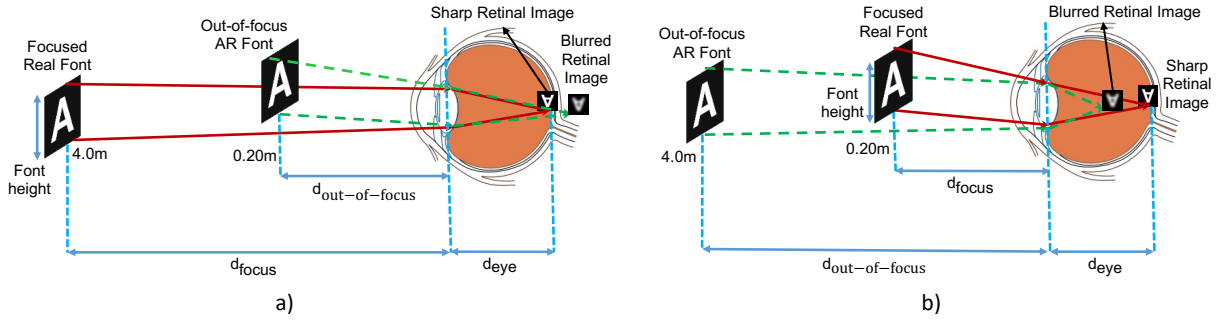


Figure 2: Retinal image formation during the out-of-focus problem in AR. (a) Out-of-focus spherical aberration S is positive. Here, the eye focuses on a real font at 4.0 m and the retina observes a corresponding sharp image. In front of the eye’s focused font, an AR font is placed at 0.20m. Therefore, a correspondence retinal image is placed behind the retina, and the user observes a blurred AR font. (b) 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. The vertical dashed lines indicate the approximate location of the lens of the eye and the retina.

Here, \mathcal{F} is the Fourier transform operator, \circ denotes the Hadamard product (element-wise multiplication), and the bar is the complex conjugate operation.

The impact of focusing distance on the PSF of the eye has been modeled as a Gaussian PSF [21, 39]. However, this is an overly simplified consideration that does not account for the diffraction of light and does not correctly represent the effects of pupil size and focus distance on the retinal image [41]. A Zernike polynomial-based PSF more accurately represents low-order (e.g., defocus, astigmatism) and high-order (e.g., trefoil, coma, quatrefoil, secondary astigmatism) aberrations of the human eye by considering the appropriate properties of eye and light propagation [1, 23, 41, 50, 62, 67]. It is important to note here that out-of-focus observations of the AR text can be modeled as a low-order aberration resulting in defocus. The wavefront aberration function $W(x, y)$ is the standard way to report human visual aberrations by considering the sum of the set of weighted Zernike polynomials in Cartesian coordinates [63]:

$$W(x, y) = \sum_{n, m \in \mathbb{Z}} C_n^m Z_n^m(x, y). \quad (4)$$

Here, x and y are the coordinates relative to the center of the pupil and normalized by the radius of the pupil [63]. The values n and m represent order and frequency, respectively. C_n^m denotes the Zernike coefficient in micrometers and defines the standard aberration weight. In particular, defocus can be described by the $Z_2^0(x, y)$ component [23, 67]:

$$Z_2^0(x, y) = \sqrt{3}(2x^2 + 2y^2 - 1). \quad (5)$$

The magnitude of the aberration of the eye is quantified by the Zernike coefficients. Vision research has made extensive use of Zernike coefficients from eye prescription data [41, 50, 67]. Dai [23] calculated the Zernike coefficients for eye prescription data after analyzing spherocylindrical ocular aberration as:

$$C_2^0 = \frac{R^2(S + C/2)}{4\sqrt{3}}. \quad (6)$$

Here, R is the radius of the pupil in millimeters, S is the value of the sphere in diopters, and C is the value of the cylinder in diopters. For modeling individual refractive vision problems, these values can be obtained from the prescription of glasses [41, 50, 67].

3.2 Modeling Out-of-Focus Aberration in AR

We propose using the Zernike defocus polynomial and mathematical techniques to obtain Zernike coefficients to model out-of-focus aberration in AR. Among low-order aberrations, only *defocus aberration* is related to the accommodative stance of the human eye. Among the

parameters of C_2^0 defined in equation 6, only S is related to defocus ($S < 0$ for myopia and $S > 0$ for hyperopia), while C describes the impact of astigmatism [67]. When observing out-of-focus content with normal, or corrected-to-normal vision, no additional astigmatism effects are introduced. Therefore, we can set $C = 0$. The modified Zernike defocus coefficient for the out-of-focus aberration in an AR system (see equation 7) is thus defined as

$$C_2^0 = -\frac{R^2(S)}{4\sqrt{3}}. \quad (7)$$

Calculating S Value: We developed a mathematical formulation to obtain the S for the out-of-focus aberration in AR without considering eye prescription data. Furthermore, the accommodation of the eyes changes from one distance to another, and therefore S is not constant. Previously, Xu and Li [67] developed the formulation to calculate the S value only from the viewing distance or the focused distance while modeling the lower-order visual aberration of an eye, to improve content visibility on an HMD. Therefore, to our knowledge, this is the first attempt to calculate the S value for out-of-focus aberration in AR based on the focused real object and the out-of-focus virtual object distances.

The focal length of the eye must first be computed. According to the thin-lens formula, the focal length of the eye based on the focused object is:

$$\frac{1}{f_{eye}} = \frac{1}{d_{focus}} + \frac{1}{d_{eye}}, \quad (8)$$

where f_{eye} is the focal length of the eye at the focused distance, d_{focus} is the distance between eye’s lens and the focused object, and $d_{eye} \approx 22\text{mm}$, the distance between eye’s lens and retina [29]. Equation 8 represents an ideal situation when a person is focused on a specific distance. However, this equation is not applicable for other out-of-focus distances, which initiate with changing accommodation of the human eye. Suppose that the distance of the out-of-focus object from the eye lens is $d_{out-of-focus}$. The aberration value of the eye, S , can then be calculated in diopters using the following equation:

$$S = \frac{1}{d_{out-of-focus}} - \frac{1}{d_{focus}}, \quad (9)$$

where S denotes out-of-focus aberration in the eye in diopters and $d_{out-of-focus}$ is the distance between the lens of the eye and the out-of-focus object. From equation 9, 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 this theory, even though there is no aberration, the retinal image will have a slight blur [60], which is known as a *diffraction-limited* retinal image. In

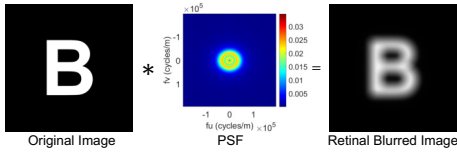


Figure 3: Retinal blurred image is generated for the amount of out-of-focus aberration of $+4.75D$ with a pupil diameter of $5mm$. The color map of PSF shows the intensity/frequency level of the wavefront. Wavefront frequency values closer to zero (blue color) means less or no blur.

addition, the following two scenarios could occur on the basis of the S value.

Potential Scenarios: 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 using equation 8, $f_{eye} = 0.0219m$. Finally, by putting all known values in equation 9, we get the out-of-focus aberration value, $S = +4.75D$. Under this scenario, S is positive. Similarly, S would be negative if the person is focusing on a real font at a close distance ($0.20m$), and an AR font is placed at a far distance ($4.00m$). Schematic diagrams of the potential scenarios are given in Fig. 2.

Wavefront Aberration Function: Using equation 9, the eye's out-of-focus aberration (S) can be computed for different accommodation changes. By putting the S value in equation 7, we get

$$C_2^0 = -\frac{R^2}{4\sqrt{3}} \times \left(\frac{1}{d_{out-of-focus}} - \frac{1}{f_{focus}} \right). \quad (10)$$

Therefore, the *final wavefront aberration function* for out-of-focus in AR is given below, where C_0^2 and $Z_0^2(x, y)$ can be obtained from equations 10 and 5:

$$W(x, y) = C_2^0 \times Z_0^2(x, y). \quad (11)$$

We computed the amplitude function based on object size and corresponding retinal image size by using the formulation of similar triangles and similar rectangles (see Appendix). Therefore, by providing the final wavefront aberration function and amplitude function for the out-of-focus problem in AR in equation 2, the pupil function for the out-of-focus aberration is obtained. Here, $\lambda = 550nm$ as only black and white images are considered (standard monochromatic simulation [23, 41]). To process color images, it is necessary to apply a distinct filtering procedure to each of the color channels using varying λ values for the RGB color. Using the principle of convolution theorem, the final out-of-focus aberrated retinal image is calculated. Fig. 3 shows out-of-focus aberrated retinal image based on our proposed mathematical formulations. The PSF of the retinal image is generated for the out-of-focus aberration of $+4.75D$ with a pupil diameter of $5mm$. We assumed that the image size of the focused object is $250(\text{width}) \times 250(\text{height})$ for all PSF images and their corresponding retinal blurred images.

We present the results of our modeled out-of-focus aberration for various pupil sizes and out-of-focus aberrations (see Appendix). This is in agreement with previous research on retinal images with different pupil diameters [10, 11, 43], which showed that a larger pupil size (e.g., $7mm$) produces more blur than a smaller pupil size (e.g., $2mm$), although the amount of aberration is the same. Furthermore, Liang and Williams [43] reported that the amount of visual aberration was roughly the same for both the left and right eyes. Accordingly, our defined out-of-focus aberration model is applicable to both eyes.

4 OUT-OF-FOCUS ABERRATION CORRECTION IN AR

We consider a complete image-based out-of-focus correction methodology for the out-of-focus problem in AR. In this methodol-

ogy, we implement the constrained total-variation (TV)-based image deconvolution method to render virtual text for the out-of-focus condition in AR. We called the final precorrected image a *SharpView font*, a font that looks sharper when seen out of focus. We rendered the font in white on a black image. All the original images (without precorrection) in this research are Arial fonts because, compared to other fonts, the Arial font has simple geometry.

In this research, to generate the precorrected 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 [13], which is a constrained TV algorithm. Taking into account this algorithm, the optimal precorrected image always has constrained or bounded pixel values between 0 and 1. Furthermore, the methodology for calculating contrast reduction was adapted from Montalto et al. [50]. According to Beck and Teboulle [13] and Montalto et al. [50], the general model (see Equation 12) for the constrained TV-based image deconvolution method to generate precorrected images is:

$$p(\theta, I_c) = \arg \min_{0 \leq \|p\| \leq 1} (\|k * p - I_c\|_{L^2} + \theta \|\nabla p\|_{L^1}). \quad (12)$$

Here, I is the original image, I_c is the rescaled version of contrast of the original image, p is the precorrected image, k is the PSF of the out-of-focus visual aberration, $*$ defines the convolution operation, θ is the regularization parameter, and ∇ is the gradient. Equation 12 attempts to compute the precorrected image (p) in such a way that the convolved precorrected image ($k * p$) is similar to the original image, $(k * p) \sim I_c$. The constrained TV-based model (equation 12) can be divided into two terms:

1. $\|k * p - I_c\|_{L^2}$ is the *deconvolution term* of the model. This term confirms that the convolved precorrected image is visually similar to the original image in the L^2 sense [50].
2. $\theta \|\nabla p\|_{L^1}$ is the *regularization term* of the constrained TV algorithm, which is the TV norm of the precorrected image in the L^1 sense. This term estimates, controls, and monitors the amount of ringing in the precorrected image. Furthermore, it also contributes to generating sharp edges in the convolved precorrected image of the model [50].

Contrast Rescaled Version of the Original Image (I_c): Contrast rescaled version of the original image within the contrast range of lower contrast level ($C_{low} = 0$) to higher contrast level ($C_{high} = 1$). Following the contrast reduction technique of Montalto et al. [50], the reduced contrast version of the original image (I) is

$$I_c = I \times (C_{high} - C_{low}) + C_{low}. \quad (13)$$

The percentage of contrast loss CL is defined as

$$CL = (C_{low} + (1 - C_{high})) \times 100. \quad (14)$$

Regularization Parameter (θ): The regularization parameter behaves like a weighted value of the regularization term, where $\theta > 0.0$. If the θ value is high, then more weight contributes to the regularization term ($\theta \|\nabla p\|_{L^1}$), and the precorrected image will have less ringing and be smoother. On the other hand, if the θ value is small, then the contribution of the regularization parameter to the regularization term will be small, and therefore the precorrected image will have a larger amount of ringing and be less smooth. Therefore, the regularization parameter allows us to control the desired level of ringing in the precorrected image based on the visual acuity.

Focus distance = 4.0m, Out of focus distance = 0.20m, Out of focus Aberration (D) = +4.75D, Pupil Diameter = 5mm					
CL (%)	10	30	50	70	90
Contrast Adjusted Original Image (I_c)					
Initial Pre-corrected Image	$\theta = 0.01$				
	$\theta = 0.00001$				
	$\theta = 0.0000001$				
Final Pre-corrected Image	$\theta = 0.01$				
	$\theta = 0.00001$				
	$\theta = 0.0000001$				

Figure 4: Results of the constrained TV-based out-of-focus pre-correction algorithm for the letter. The precorrected images are generated with five different CL(%) and three different θ levels under out-of-focus aberration of $+4.75D$ over a $5mm$ pupil. The final precorrected images are generated by following the description of section 4.2.

4.1 Initial Precorrected Image

In our research, we considered three types of textual information: *letter*, *word* and *phrase*. The result of the out-of-focus constrained TV-based pre-correction algorithm for the letter is shown in Fig. 4 (see the supplemental files for word (Fig. 12) and phrase (Fig. 13)). The sub-rows under *initial precorrected image* contain the initial result of our algorithm. The precorrected images are generated for the out-of-focus aberration of $+4.75D$ with a pupil diameter of $5mm$. We considered five different amounts of contrast loss—10%, 30%, 50%, 70%, and 90%, and three different levels of regulation parameters (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 textual information is rendered in different resolutions of images based on the amount of information, for the letter: 500×500 , for the word: 1100×500 , and for the phrase: 1400×1200 resolution image.

For a particular CL (%) amount, the original image is adjusted to that CL (%) amount and then supplied to compute the out-of-focus aberration and the precorrected image. The row of *initial precorrected image* from Fig. 4 shows that as the amount of CL (%) increases, a larger amount of ringing appears in the precorrected image for each level of θ . When comparing different levels of θ , we found that the precorrected image gains more ringing with smaller θ values. These characteristics in the precorrected images for the values of θ and different amounts of CL (%) are also true for images with words and phrases (see Appendix: Figs. 12 and 13).

4.2 Final Precorrected Image

One of the major challenges of the constrained TV-based deconvolution approach is the ringing artifacts (unwanted pixel intensities) on the precorrected image. Alonso et al. [2] stated that not all the in-

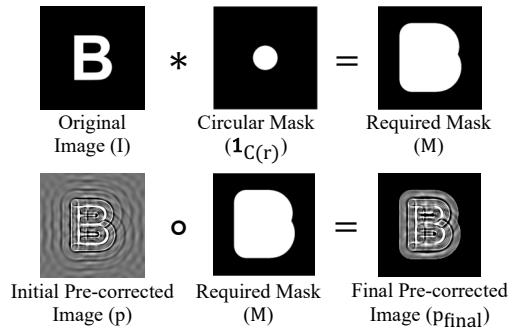


Figure 5: Process of generating final precorrected images. By performing the convolution between the original image I and the circular mask $\mathbb{1}_{C(r)}$, we obtained the required mask M . The final precorrected image is generated based on the Hadamard product between the precorrected image p and a mask M . The process followed the description of section 4.2.

formation on the precorrected image from the image deconvolution method is necessary and extracted only the necessary information from the initial precorrected image. We followed a similar principle for the precorrected images of the out-of-focus correction method. We observed that the ringing waves uniformly spread throughout the precorrected image based on the different amounts of CL (%) and θ (see Fig. 4). Human visual perception is highly sensitive to ringing [33, 50] and can misperceive the image. Perceptually, enhanced out-of-focus text legibility highly depends on the edge region of the precorrected textual image. Therefore, only a finite ringing region is needed around the edge of the textual information to perceive the information with improved visual acuity during an out-of-focus situation. If no ringing is present, the pixel values around the edge will ultimately not appear sharper because a constant amount of ringing reduces the constrained bounded-pixel value and allows more freedom to form a perceptually sharper edge. Therefore, we considered up to two waves of ringing in the final precorrected image.

The final precorrected image is generated based on the Hadamard product between the precorrected image p and a mask M . The mask contains zeros where the information is unnecessary and ones where the information needs to be extracted (see Fig. 5). To obtain the required mask region, we first need to construct a circular mask defined as the indicator function on the circle centered at $(0,0)$ with radius r , denoted by $C(r)$

$$\mathbb{1}_{C(r)}(x,y) = \begin{cases} 1, & \text{if } \sqrt{x^2+y^2} \leq r \\ 0, & \text{otherwise} \end{cases}$$

We take the radius r to be a percentage α of the radius of the PSF r_{PSF} , that is $r = \alpha \cdot r_{PSF}$. By performing the convolution between the original image $I(x,y)$ and the circular mask $\mathbb{1}_{C(r)}$ we obtain the required mask $M(x,y) = (I * \mathbb{1}_{C(r)})(x,y)$. Finally, the Hadamard product of between the precorrected image $p(x,y)$ and required mask $M(x,y)$ provides the final precorrected image $p_{final}(x,y) = p(x,y) \circ M(x,y)$. This process is illustrated in Fig. 5. Note that by changing the value of α , the circular mask's size will change, which eventually determines the amount of ringing in the precorrected image that we can select. Here we consider $\alpha = 1.5$ to generate one wave of ringing around the precorrected textual images. We obtained this value using a trial-and-error approach.

The results of the final precorrected images are shown in the *final precorrected image* rows of Fig. 4 (see Appendix for word (Fig. 12) and phrase (Fig. 13)) for different combination of θ and CL(%) under the out-of-focus aberration of $+4.75D$ and pupil diameter of $5mm$. Therefore, we only removed the unnecessary ringing artifacts from the precorrected images with mathematical formulations, and all

other TV-based algorithm parameters, characteristics, and behavior remain the same.

5 IMPLEMENTATION

The Matlab program for modeling out-of-focus aberration and TV-based out-of-focus correction algorithms was executed on an MSI laptop with an Intel Core i7 processor, 32 GB RAM, NVIDIA GeForce RTX 2070 and 64-bit Windows 10 Pro Operating System. Each precorrected image was created with a unique combination of θ , CL (%), pupil diameter, focus distance and out-of-focus distance. The overall mean time for generating a precorrected image for *letter* was 11.92 seconds, for *word* was 16.97 seconds, and for *phrase* was 69.74 seconds. The precorrected image for *phrase* required more time because of the larger image resolution and greater amount of information.

6 EVALUATION

The goal of our evaluation is to quantify how sharp and legible our constraint TV-based SharpView font is compared to the standard Arial font during out-of-focus situations in an AR system. Therefore, to quantify the sharpness of an image, we developed a novel image gradient-based *sharpness measurement technique* for textual information by considering the presence of blur in the edge region of the font. Following the steps of our sharpness measurement method, we evaluated *precorrected images* through *synthetic simulation evaluation* and *optically viewed camera captured evaluation*, and determined the combination of CL (%) and the regularization parameter (θ) that is suitable for +4.75D out-of-focus aberration and the diameter of the pupil of 5mm. However, we have not compared our method with any existing deconvolution algorithms because Montalto et al. [50] have already shown that the TV-based deconvolution method increases visual acuity and legibility for textual information more than any other method, such as Weiner filtering, for visual aberrations.

6.1 Sharpness Measurement Method

The main goal of our sharpness measurement method is to process the image in such a way that the image only contains the pixels that contribute to blur. Let us 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. The mathematical formulations to estimate the presence of blur amount in the $t(x, y)$ image are:

Step 1: In this step we compute an image B_t that contains the blurry regions of the target image t . We start by computing an image of the magnitude of the gradient of the image t by using the Sobel method [27, 37]

$$\delta_t(x, y) = \sqrt{t_x^2 + t_y^2}.$$

We use δ_t as a map to decide which pixels of t are actually blurry and we want to keep. We keep pixels (x, y) in the image t where the value $\delta_t(x, y)$ are within a percentage range $[\alpha, \beta]$ of all values of δ_t . The lower bound allows us to remove the noise of the optical system and the upper bound takes care of removing the edges of the region. The resulting image is defined as

$$B_t(x, y) = \begin{cases} t(x, y), & \text{if } \alpha \leq \frac{\delta_t(x, y)}{\max \delta_t - \min \delta_t} \leq \beta \\ 0, & \text{Otherwise} \end{cases}. \quad (15)$$

For all our images, we chose $[\alpha, \beta] = [0.015, 0.9]$.

Step 2: To handle different amounts of CL (%) of the precorrected images in the sharpness measurement technique, we normalize the image B_t with the l_2 norm:

$$B_n = \frac{B_t}{\|B_t\|_{l_2}}. \quad (16)$$

Previously, Wee et al. [64] performed a similar normalization step while measuring the sharpness of the image using eigenvalues. By performing this normalization step, our image sharpness measurement method becomes less dependent on image contrast effects.

Step 3: In the final step, the sharpness measurement method quantifies the amount of blur in an image t by dividing the l_1 norm of the image ($\|B_n\|_{l_1}$) with the number of pixels of the image B_n (size(B_n)). This gives a real number b , called the blur amount of the the image t :

$$b_t = \frac{\|B_n\|_{l_1}}{\text{size}(B_n)}. \quad (17)$$

The values of blur amount b_t are reported in Fig. 1.

Using the blur amount we can compare the sharpness between two images. If one image is t_1 and the other image is t_2 , then we could determine the sharper image using the formula

$$\text{Sharp}(t_1, t_2) = \frac{b_{t_1} - b_{t_2}}{b_{t_1}} \times 100. \quad (18)$$

Here, if $\text{Sharp}(t_1, t_2) > 0$, then t_2 is sharper than t_1 ; similarly if $\text{Sharp}(t_1, t_2) < 0$, then t_1 is sharper than t_2 .

6.2 Parameter Selection

One of the essential parts of our proposed method is to obtain the pre-correction algorithm's parameter that improves the out-of-focus issues. It is unknown which combination of CL (%) and regularization parameter(θ) of the pre-correction algorithm is appropriate and valid to improve out-of-focus text legibility with respect to a particular out-of-focus aberration and pupil diameter.

Previously, while investigating image-based correction for visually impaired people, Montalto et al. [50] proposed a novel relative total variation term (τ) to control the trade-off between contrast loss (%) and ringing (θ), and considered this as the performance measurement indicator in their research. A single τ value can be obtained from different combinations of CL (%) and θ for a particular visual aberration. However, we wanted to focus on sharpness improvement regardless, and we did not know the optimal τ value for the images that we studied. Therefore, we considered nine different amounts of CL (%) (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90%), and three values of θ ($1e - 02$, $1e - 04$, $1e - 06$) for our evaluation.

Furthermore, we performed an analysis of the parameters of our algorithm to understand the relationship between CL (%), the regularization parameter (θ), and the TV value. The TV value is the TV norm of the precorrected image ($\|\nabla p\|_{L1}$), which defines the amount of energy gained in the precorrected image from optimization. This analysis was performed only for the pupil diameter of 5mm under an aberration out-of-focus of +4.75D. Based on our observations, small values of θ yield a high TV value (more ringing) in the precorrected image, compared to large values of θ in each type of textual information for a specific amount of CL (%). We also observed that for a particular combination of CL (%) and the regularization parameter (θ), *letter* has a much lower TV value than *word* and *phrase*. Therefore, the TV values also depend on the resolution of the image and the amount of information in it.

6.3 Synthetic Simulation Evaluation

Method: This evaluation method examines whether the synthetically convolved precorrected images show sharper and better visual acuity compared to the synthetically simulated out-of-focus blurred image. The upper portion of Fig. 6 shows the steps of the evaluation of the synthetic simulation under +4.75D out-of-focus aberration and 5mm diameter pupil size for the single letter B . Let us assume that the precorrected image is p , the original image is I , and the PSF of the out-of-focus aberration is k . Therefore, the synthetically convolved precorrected image is $k \otimes p$ (see Fig. 6 upper-right), and

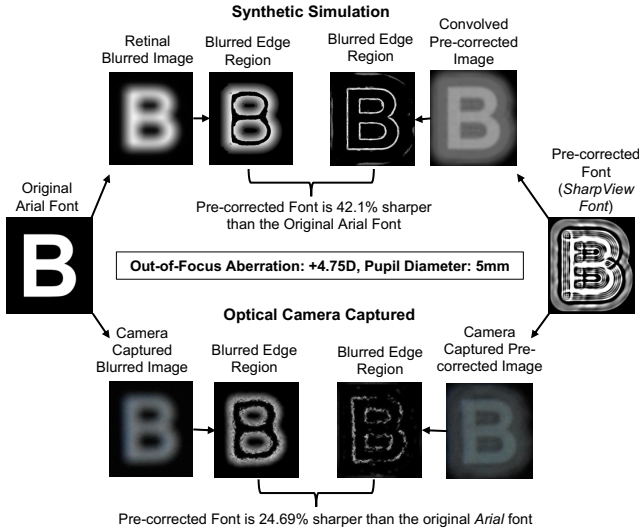


Figure 6: Sharpness comparison in the synthetic simulation and the optical camera captured evaluation methods for the single letter *B*. The precorrected font (SharpView Font) is generated for the out-of-focus aberration of $+4.75D$ and pupil diameter of $5mm$. The top part of the figure displays the evaluation results of the synthetic simulation approach, while the bottom part shows the camera-captured result through the optics of the benchtop AR display. Comparing the blurred edge regions from the standard Arial font with SharpView font, it is apparent that our generated SharpView Font has a sharper representation than the Standard Arial font both synthetically and optically with the camera captured. Also, see Fig. 1 for the word and phrase.

the synthetically simulated out-of-focus retinal blurred image is $k \otimes I$ (see Fig. 6 upper-left). Following the steps of the sharpness measurement method, we achieved the blurred edge region. After that, by comparing the blurred edge region of simulated out-of-focus blurred and convolved precorrected images (see equation 18), we found that our generated precorrected image for the letter *B* is 42.1% sharper than the original Arial letter (see upper portion of Fig. 6). Furthermore, the precorrected word and phrase show 43.24% and 39.61% sharper representations than the original image, respectively, through the synthetic evaluation.

Results: The results for our simulation are shown in the top row of Fig. 7, where three panels show the results of the three different textual information (letter, word, and phrase). The x -axis shows the amount of contrast loss (%). The y -axis shows the percentage of sharpness of the precorrected image compared to the retinal out-of-focus blurred image under $+4.75D$ out-of-focus aberration and $5mm$ diameter of the pupil. Each panel has 30 data points representing a unique combination of CL (%) and regularization parameter (θ). Three different color lines represent three different values of θ . The *simulated* row panel in Fig. 7 shows that as the amount of CL (%) increases, the sharpness (%) increases for each value of θ for each textual information. Precorrected images with maximum visual acuity and legibility for each textual information in the synthetic simulation are obtained with CL = 90% and $\theta = 1e - 06$ for the $+4.75D$ of out-of-focus aberration and $5mm$ pupil diameter, and the visual representations are given in Fig. 1.

6.4 Optically Viewed Camera Captured Evaluation

In this evaluation stage, we examined whether the constrained TV-based out-of-focus correction method produces sharper and improved visual acuity through the optics of the OST AR display. In this case, this research compared the camera-captured precorrected image at the moment of out-of-focus with the camera-captured out-of-focus blurred image. All images in this evaluation were taken

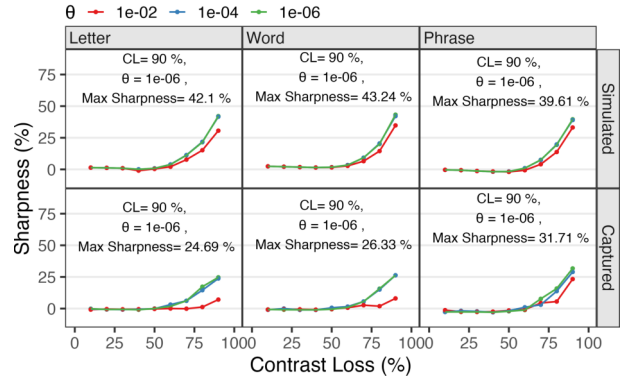


Figure 7: The results of the synthetic simulation and optically viewed camera captured evaluation for the $+4.75D$ out-of-focus aberration and $5mm$ diameter of the pupil size show that as the amount of CL (%) increases, the sharpness increases (%) for each θ value for each textual information. The highest sharpness is achieved with the CL = 90% and $\theta = 1e - 06$ for all textual information and evaluation methods.

through the optics of the AR display with the camera (see Fig. 8).

Apparatus: We used a custom-made tabletop OST AR display as an experimental display, which we call the *AR Haploscope* [7, 55, 56, 61]. In addition, we used a DSLR camera (Nikon D3400) to capture images through the AR optics. 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.

Setup and Procedure: The setup of the camera-based evaluation method is shown in Figs. 8a and b. The DSLR camera was mounted on a tripod to look through the center of the optical combiners of the AR Haploscope. The camera only focused on the real world information (cross) at $4.0m / 0.25D$, which was displayed on a physical monitor (Dell Ultra-sharp Monitor: U2913WM) with a display resolution of 2560×1080 pixels (see Fig. 8a). Each virtual textual information was displayed on the right image generators of the AR Haploscope (Fig. 8b). The image generator (Feelworld F570 5.7" 4K) had a diagonal size of 14.5 cm and a display resolution of 1920×1080 pixels. In our evaluation, we used an accommodation lens with a power of $5.00D$ to place the virtual information at $0.20m$. The camera captured both real and virtual information through the optical combiner of the AR Haploscope, see Fig. 8c, where the left image is virtual and the right image is real (cross). In addition, each precorrected image (virtual information) was randomly presented three times to remove any noise from the camera setting. Furthermore, we used an external remote controller for the camera to capture the image so that no shaking noise contributed to the captured image.

Camera Settings: In our camera-based evaluation, the critical step was to confirm that the camera was carefully focused on the real cross at $4.0m$ and that a sharp real cross image was formed. 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 limit and the DoF far limit. We placed the camera in the manual configuration option. We considered a deep DoF of $3m$ for the camera, where the far DoF limit was $6m$ and the near DoF limit was $3m$ [47], by setting the aperture f-stop to: $\frac{f}{22}$, the lens focal length to: $55mm$, the ISO to: 6400, and the shutter speed to: $\frac{1}{50}$. In total, we captured 10 (CL(%)) $\times 3$ (θ) $\times 3$ (repetitions) = 90 precorrected images for the out-of-focus aberration of $+4.75D$.

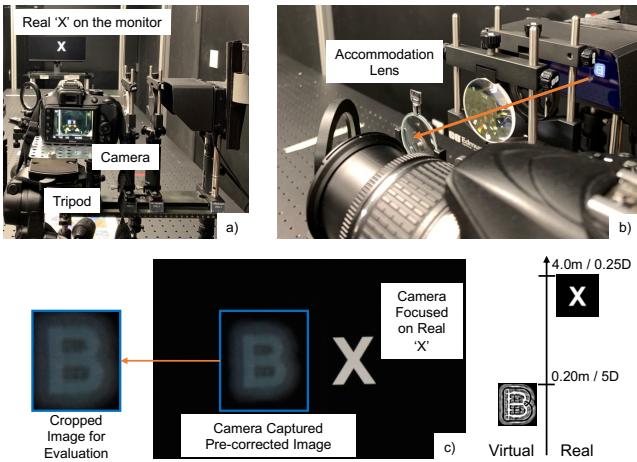


Figure 8: Setup of the optically viewed camera captured image evaluation. (a) A DSLR camera is mounted on a tripod and focused on a real cross ('X') on the physical monitor, which is located at 4.0m. (b) The precorrected image is displayed on the benchtop OST AR display's monitor as virtual content in front of the camera at 0.20m. This precorrected image is generated with $CL = 90\%$ and $\theta = 1e-06$ for the out-of-focus aberration of $+4.75D$ and pupil diameter of $5mm$. The precorrected image is optically out-of-focus through the lens of the camera. (c) The camera captured image where the camera is focused on the real cross, and the precorrected image looks sharper, is more legible, and has improved visual acuity in the out-of-focus AR situation. In addition, the portion of the precorrected image captured by the camera is cropped for evaluation.

Results: The bottom portion of Fig. 6 shows the steps in evaluating the captured images. Unlike the convolved precorrected images of the synthetic simulation approach, the camera-captured images are not noise-free (see Fig. 8c). The noises on the camera captured precorrected images mainly came from the internal camera processing mechanism. Therefore, we performed two additional denoising operations in the sharpness measurement method to reduce noise. In the process, the Wiener filtering method for two-dimensional pixel-wise adaptive noise removal was considered by estimating the local mean and standard deviation of the image [44]. The Wiener filter size for the letter, word, and phrase was 10×10 . We performed this noise removal operation before beginning the sharpness measurement method and before the normalization step. The blurred edge region of the camera captured blurred and precorrected images were obtained by the sharpness measurement method (see subsection 6.1). Comparing the blurred edge region of the captured out-of-focus blurred and precorrected images, we found that our generated precorrected image for the letter 'B' is 24.69% sharper than the original Arial letter (see bottom portion of Fig. 6). Furthermore, the precorrected word and phrase show 26.33% and 31.71% sharper representations than the original image, respectively. The results of the optically camera-captured evaluation are shown in the bottom row (*Captured*) of Fig. 7. Each panel shows the average sharpness value (%) for the unique combination of CL (%) and the regularization parameter (θ). This indicates that as the amount of CL (%) increases, the sharpness (%) increases for each value of θ in each textual information. The precorrected images with maximum visual acuity and legibility for each textual information in the camera captured are obtained with $CL = 90\%$ and $\theta = 1e-06$ for the $+4.75D$ of out-of-focus aberration, and $5mm$ diameter of pupil size. Visual representations are given in Fig. 1.

7 CONCLUSION AND FUTURE WORK

The out-of-focus issue of virtual objects has been relatively unexplored in the field of AR research. However, out-of-focus blur has

been found to reduce performance [7, 25]. Therefore, the objective of this investigation was to create and evaluate a new font, called a *SharpView font*, which would appear sharper and more legible when viewed out of focus. To do this, a mathematical model of out-of-focus blur (retinal blurred image) was created using Zernike polynomials, a perceptual image processing-based focus correction algorithm was developed, and an image gradient-based algorithm was created to measure and quantify font sharpness in the blurred edge region. The results of the synthetic simulation and optical camera-based measurement showed that the constrained TV-based image precorrection algorithm improved visual acuity and legibility for the textual image for the out-of-focus issue in the AR system both visually and quantitatively. Both the simulation and the images captured through the optics revealed that the *SharpView font* had 20-45% more sharpness and improved visual acuity compared to the standard font. The evaluation results also showed that the precorrected image with high visual acuity out of focus was obtained with 90% contrast loss and a regularization parameter of $1e-06$ for the textual information through simulation and camera capture.

Although we successfully implemented the novel *SharpView font* for the problem of out-of-focus in the AR system, some specific limitations still exist that suggest future work.

User Based Evaluation: We evaluated the constrained TV-based out-of-focus correction algorithm through synthetic simulation and an optically viewed camera-captured approach. However, this research did not perform a formal user study. In the future, 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. The experiment should consider an eye tracker to know where the user is focusing to switch on the out-of-focus font rendering. In addition, the eye tracker is needed to get the user's pupil data for the algorithm.

Out-of-Focus Aberrations, Pupil Diameters, and Evaluation Metric: In this research, out-of-focus aberration of $+4.75D$ with a pupil diameter of $5mm$ was evaluated. Furthermore, our sharpness metric did not compare the results with other metrics. 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. In the future, it is important for researchers to also investigate the threshold of the eye's depth of focus. Furthermore, it would be beneficial to explore an alternative modulation transfer function (MTF) in order to compare our evaluation metric for measuring sharpness.

Replicate Transient Focal Blur Experiment: One of the hypotheses behind our out-of-focus correction method is that generated precorrected images would mitigate the transient out-of-focal blur effect in AR. Therefore, one of the potential future studies could include the precorrected 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. [25] and Arefin et al. [6, 7].

Additional Graphical Elements: We have developed and evaluated the constrained TV-based out-of-focus correction method only for textual information. Future research could confirm this algorithm for other fundamental graphical elements of OST AR systems, such as road signs, directions, symbols, and digital notifications.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation, under award IIS-1937565, to J. Edward Swan II. Alexander Plopski was partially sponsored by Snap Inc. This work was conducted at the Center for Advanced Vehicular Systems (CAVS) at Mississippi State University. We acknowledge Christian Sandor for getting us excited about the out-of-focus problem.

REFERENCES

- [1] M. Alonso and A. Barreto. Pre-compensation for high-order aberrations of the human eye using on-screen image deconvolution. In *Engineering in Medicine and Biology Society*, vol. 1, pp. 556–559. IEEE, 2003. doi: 10.1109/IEMBS.2003.1279804
- [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*, 16(8):957–966, 2008. doi: 10.1080/17415970802082823
- [3] M. Alonso, A. Barreto, and J. G. Cremades. Image pre-compensation to facilitate computer access for users with refractive errors. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '04, p. 126–132. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/1028630.1028653
- [4] M. S. Arefin. [dc] sharpview ar: Enhanced visual acuity for out-of-focus virtual content. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 731–732, 2021. doi: 10.1109/VRW52623.2021.00248
- [5] M. S. Arefin. *Augmented reality fonts with enhanced out-of-focus text legibility*. PhD thesis, Mississippi State University, Mississippi State, MS 39762, USA, 2022.
- [6] 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. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 571–572, 2020. doi: 10.1109/VRW50115.2020.00137
- [7] 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*, 28(5):2014–2025, 2022. doi: 10.1109/TVCG.2022.3150503
- [8] P. Artal. Optics of the eye and its impact in vision: a tutorial. *Adv. Opt. Photon.*, 6(3):340–367, Sep 2014. doi: 10.1364/AOP.6.000340
- [9] P. Artal. Image formation in the living human eye. *Annual Review of Vision Science*, 1(1):1–17, 2015. PMID: 28532382. doi: 10.1146/annurev-vision-082114-035905
- [10] P. Artal. *The Eye as an Optical Instrument*. Springer International Publishing, Cham, 2016.
- [11] 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*, 11(1):246–249, Jan 1994. doi: 10.1364/JOSA.A.11.000246
- [12] B. A. Barsky, F.-C. Huang, D. Lanman, G. Wetzstein, and R. Raskar. Vision correcting displays based on inverse blurring and aberration compensation. In *Computer Vision - ECCV 2014 Workshops*. Springer International Publishing, 2014.
- [13] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, Nov 2009. doi: 10.1109/TIP.2009.2028250
- [14] 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*, 16(12):2992–3004, Dec 2007. doi: 10.1109/TIP.2007.909319
- [15] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira. Total variation-based image deconvolution: a majorization-minimization approach. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, pp. II–II, May 2006. doi: 10.1109/ICASSP.2006.1660479
- [16] F. W. Campbell and G. Westheimer. Dynamics of accommodation responses of the human eye. *The Journal of Physiology*, 151(2):285–295, 1960. doi: 10.1113/jphysiol.1960.sp006438
- [17] 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*, 24(11):2906–2916, 2018. doi: 10.1109/TVCG.2018.2868532
- [18] 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.
- [19] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40, 05 2011. doi: 10.1007/s10851-010-0251-1
- [20] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal—dual algorithm. *Math. Program.*, 159(1-2):253–287, Sept. 2016. doi: 10.1007/s10107-015-0957-3
- [21] 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. In *Proceedings of IEEE Virtual Reality (IEEE VR 2018)*, pp. 394–400. IEEE Computer Society, March 2018.
- [22] 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, Purdue University, 2014.
- [23] G.-m. Dai. *Wavefront optics for vision correction*, vol. 179. SPIE, 01 2008. doi: 10.1117/3.769212
- [24] 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*, 23(4):1322–1331, April 2017. doi: 10.1109/TVCG.2017.2657058
- [25] 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*, 25:2228 – 2241, 2018. doi: 10.1109/TVCG.2018.2832633
- [26] 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*, 28(2):1443–1456, 2022. doi: 10.1109/TVCG.2020.3014614
- [27] R. C. Gonzalez and R. E. Woods. *Digital image processing*. Prentice Hall, Upper Saddle River, N.J., 2008.
- [28] J. W. Goodman. *Introduction to Fourier optics*. Englewood, CO: Roberts & Co. Publishers, 2005.
- [29] H. Gross, F. Blechinger, and B. Achnner. *Human Eye*. John Wiley & Sons, Ltd, 2015. doi: 10.1002/9783527699247.ch1
- [30] A. H. Güzel, J. Beyazian, P. Chakravarthula, and K. Akşit. Chromacorrect: prescription correction in virtual reality headsets through perceptual guidance. *Biomed. Opt. Express*, 14(5):2166–2180, May 2023. doi: 10.1364/BOE.485776
- [31] 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*, 8(3):1–30, Jan. 2008. doi: 10.1167/8.3.33
- [32] X. Hu and H. Hua. High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Optics Express*, 22(11):13896–13903, Jun 2014. doi: 10.1364/OE.22.013896
- [33] F. C. Huang, D. Lanman, B. A. Barsky, and R. Raskar. Correcting for optical aberrations using multilayer displays. *ACM Transactions on Graphics*, 31(6), 2012. doi: 10.1145/2366145.2366204
- [34] 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*, 33(4), 2014. doi: 10.1145/2601097.2601122
- [35] 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. In *Proceedings of the ACM Symposium on Applied Perception in Graphics and Visualization*, pp. 41–48. ACM, 2010. doi: 10.1145/1836248.1836255
- [36] 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*, 54(6), july 2021. doi: 10.1145/3453157
- [37] N. Kanopoulos, N. Vasanthavada, and R. L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of solid-state circuits*, 23(2):358–367, 1988.
- [38] S.-B. Kim and J.-H. Park. Optical see-through maxwellian near-to-eye

- display with an enlarged eyepiece. *Optics Letters*, 43(4):767–770, Feb 2018. doi: 10.1364/OL.43.000767
- [39] 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. In *IEEE Symposium on 3D User Interfaces (3DUI 2016)*, pp. 173–181. IEEE, Greenville, South Carolina, USA, Mar. 2016. doi: 10.1109/3DUI.2016.7460049
- [40] G. A. Koulieris, 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*, 38(2):493–519, 2019. doi: 10.1111/cgf.13654
- [41] 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*, pp. 64–71, 2017. doi: 10.1109/SIBGRAPI.2016.018
- [42] M. Lambooj, M. Fortuin, I. Heynderickx, and W. IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):030201–030201–14, Jan. 2009. doi: 10.2352/J.ImagingSci.Technol.2009.53.3.030201
- [43] J. Liang and D. R. Williams. Aberrations and retinal image quality of the normal human eye. *Journal of the Optical Society of America A*, 14(11):2873–2883, Nov 1997. doi: 10.1364/JOSAA.14.002873
- [44] J. S. Lim. *Two-Dimensional Signal and Image Processing*. Prentice-Hall, Inc., USA, 1990.
- [45] 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*, 17(18):15716–15725, 2009. doi: 10.1364/OE.17.015716
- [46] 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*, 21(1):011002–1, 2012. doi: 10.1117/1.JEI.21.1.011002
- [47] J. J. Mah and K. Alambra. Depth of field calculator. MATLAB Central File Exchange.
- [48] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 29–38, Oct 2013. doi: 10.1109/ISMAR.2013.6671761
- [49] 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*, 33(4):89:1–89:11, July 2014. doi: 10.1145/2601097.2601141
- [50] 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*, 34(3):28:1–28:16, May 2015. doi: 10.1145/2717307
- [51] 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. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 21–30, 2018. doi: 10.1109/ISMAR.2018.00025
- [52] 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. In *Poster Abstracts, Proceedings of IEEE Virtual Reality (IEEE VR 2018)*, pp. 641–642. IEEE Computer Society, March 2018. doi: 10.1109/VR.2018.8446429
- [53] V. F. Pamplona, M. M. Oliveira, D. G. Aliaga, and R. Raskar. Tailored displays to compensate for visual aberrations. *ACM Transactions on Graphics*, 31(4), jul 2012. doi: 10.1145/2185520.2185577
- [54] E. Peillard, Y. Itoh, G. Moreau, J.-M. Normand, A. Lécuyer, and F. Argelaguet. Can retinal projection displays improve spatial perception in augmented reality? In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 80–89, 2020. doi: 10.1109/ISMAR50242.2020.00028
- [55] N. Phillips, K. Massey, M. S. Arefin, and J. E. Swan. Design and calibration of an augmented reality haploscope. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 75–76, 2018. doi: 10.1109/ISMAR-Adjunct.2018.00037
- [56] N. Phillips, K. Massey, M. S. Arefin, and J. E. Swan II. Design, assembly, calibration, and measurement of an augmented reality haploscope. In *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)*, pp. 1770–1774. IEEE, Osaka, Japan, Mar. 2019. doi: 10.1109/VR.2019.8798335
- [57] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the mumford-shah functional. In *2009 IEEE 12th International Conference on Computer Vision*, pp. 1133–1140, 2009. doi: 10.1109/ICCV.2009.5459348
- [58] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs. An extended depth-of-field volumetric near-eye augmented reality display. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2857–2866, Nov 2018. doi: 10.1109/TVCG.2018.2868570
- [59] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. doi: 10.1016/0167-2789(92)90242-F
- [60] T. O. Salmon. *Corneal contribution to the wavefront aberration of the eye*. Phd dissertation, Indiana University, 1999.
- [61] 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*, pp. 1–14, 2018. doi: 10.1109/TVCG.2018.2869729
- [62] L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb. Standards for reporting the optical aberrations of eyes. *Journal of Refractive Surgery*, 18(5):652–660, 2002.
- [63] A. B. Watson. Computing human optical point spread functions. *Journal of Vision*, 15(2):1–25, 2015. doi: 10.1167/15.2.26
- [64] C.-Y. Wee and R. Paramesran. Image sharpness measure using eigenvalues. In *2008 9th International Conference on Signal Processing*, pp. 840–843, 2008. doi: 10.1109/ICOSP.2008.4697259
- [65] 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*, 26(9):11553–11567, Apr 2018. doi: 10.1364/OE.26.011553
- [66] 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*, 25(11):3114–3124, 2019. doi: 10.1109/tvcg.2019.2932238
- [67] F. Xu and D. Li. Software based visual aberration correction for hmds. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 246–250, 2018. doi: 10.1109/VR.2018.8447557
- [68] J. I. Yellott and J. W. Yellott. Correcting spurious resolution in defocused images. In T. N. P. Bernice E. Rogowitz and S. J. Daly, eds., *Human Vision and Electronic Imaging XII*, vol. 6492. International Society for Optics and Photonics, SPIE, 2007. doi: 10.1117/12.698240

APPENDIX

A AMPLITUDE FUNCTION

Let us assume that the focused font height is $h_{focused}$, the width is $w_{focused}$ the distance from the eye is d_{focus} , the retinal image height is $h_{retinal}$, the width is $w_{retinal}$ and the distance between the lens of the eye and the retina is $d_{eye} = 0.022m$. The $\angle ACB$ is closely equal to the $\angle DCE$, so according to the principle and geometry of a similar triangle from Fig. 9a, the retinal image's height ($h_{retinal}$) was calculated. Similarly, according to the principle and geometry of a similar rectangle from Fig. 9b, the retinal image's width ($w_{retinal}$) was calculated. The equations are given below:

$$\frac{h_{focused}}{d_{focused}} \approx \frac{h_{retinal}}{d_{eye}} \quad (19)$$

$$\frac{w_{focused}}{h_{focused}} \approx \frac{w_{retinal}}{h_{retinal}} \quad (20)$$

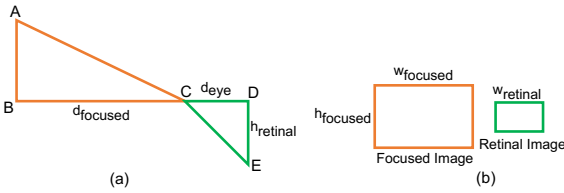


Figure 9: Amplitude function calculation for out-of-focus aberration using the principles of similar triangles and rectangles.

B RETINAL IMAGES OF DIFFERENT AMOUNTS OF PUPIL SIZE AND OUT-FOCUS ABERRATIONS

Fig. 10 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 $+4.75D$. Furthermore, Fig. 11 shows the table of different PSF images and its corresponding retinal blurred image for different out-of-focus aberrations under the pupil diameter of $5mm$.

Pupil Diameter (mm)	PSF Image	Simulated Retinal Blur Image		
		Letter	Word	Phrase
2		B	TEXT	School Zone Ahead
3		B	TEXT	School Zone Ahead
4		B	TEXT	School Zone Ahead
5		B	TEXT	School Zone Ahead
6		B	TEXT	School Zone Ahead
7		B	TEXT	School Zone Ahead
8		B	TEXT	School Zone Ahead

Figure 10: 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. 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) mean less or no blur. As the wavefront frequency values increase, the amount of blur increases. Furthermore, the wider the circle, the more waves are away from the center and the higher the blur. This means that more of the wavefront at this point is contributing to the retinal blur.

Pupil Diameter = 5mm	Out-of-focus Aberration (D)	PSF Image	Simulated Retinal Blur Image		
			Letter	Word	Phrase
	-4.75				
-2.75					
0.0					
+2.75					
+4.75					

Figure 11: Effects of different out-of-focus aberration values in the retinal image quality under 5mm diameter of the pupil size. The retinal blur amount changes according to the magnitude of the out-of-focus aberration. Positive and negative signs appear based on the described scenarios in 3.2.

C PRECORRECTED IMAGES FOR WORD AND PHRASE

The initial and final precorrected images for word and phrase (see Subsections 4.1 and 4.2) are displayed in Figs. 12 and 13. As the CL(%) increases, more ringing appears in the precorrected image for each level of θ , similar to the letter results.

Focus distance = 4.0m, Out of focus distance = 0.20m, Out of focus Aberration (D) = +4.75D, Pupil Diameter (mm) = 5mm					
CL (%)	10	30	50	70	90
Contrast Adjusted Original Image (I ₀)					
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 12: Results of the constrained TV-based out-of-focus pre-correction algorithm for word. The precorrected 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 precorrected images are generated by following the description of section 4.2.

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 (I ₀)					
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 13: Results of the constrained TV-based out-of-focus pre-correction algorithm for phrase. The precorrected 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 precorrected images are generated by the following description of section 4.2.