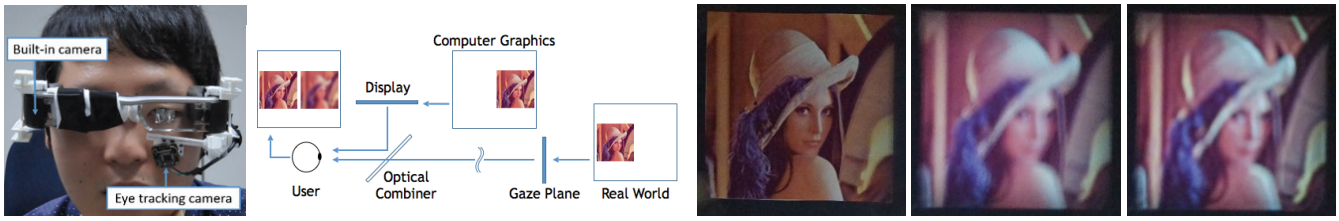
SharpView: Improved Clarity of Defocused Content on Optical See-Through Head-Mounted Displays

Kohei Oshima* † Kenneth R Moser* ‡ Damien Constantine Rompapas† J. Edward Swan II‡ Sei Ikeda§
Goshiro Yamamoto† Takafumi Taketomi† Christian Sandor† Hirokazu Kato†

†Interactive Media Design Laboratory
Nara Institute of Science and Technology

‡Computer Science & Engineering
Mississippi State University

§Mobile Computing Laboratory
Ritsumeikan University



(a) (b) (c) (d) (e)

Figure 1: The cause and effect of focus blur in Optical See-Through (OST) Head-Mounted Display (HMD) systems. (a) A user wearing the OST HMD and related hardware used in our study. (b) Simplified schematic of an OST AR system. Blurring occurs when the virtual display screen and real world imagery are viewed at unequal focal distances. (c), (d), (e): Views through an OST Augmented Reality system, where the real world image (c) is in focus, causing the virtual image (d) to appear blurred; (e) an improved virtual image after application of SharpView.

Abstract

Augmented Reality (AR) systems, which utilize optical see-through head-mounted displays, are becoming more common place, with several consumer level options already available, and the promise of additional, more advanced, devices on the horizon. A common factor among current generation optical see-through devices, though, is fixed focal distance to virtual content. While fixed focus is not a concern for video see-through AR, since both virtual and real world imagery are combined into a single image by the display, unequal distances between real world objects and the virtual display screen in optical see-through AR is unavoidable.

In this work, we investigate the issue of focus blur, in particular, the blurring caused by simultaneously viewing virtual content and physical objects in the environment at differing focal distances. We additionally examine the application of dynamic sharpening filters as a straight forward, system independent, means for mitigating this effect improving the clarity of defocused AR content. We assess the utility of this method, termed SharpView, by employing an adjustment experiment in which users actively apply varying amounts of sharpening to reduce the perception of blur in AR content shown at four focal disparity levels relative to real world imagery.

Our experimental results confirm that dynamic correction schemes are required for adequately addressing the presence of blur in Optical See-Through AR. Furthermore, we validate the ability of our SharpView model to improve the perceived visual clarity of focus blurred content, with optimal performance at focal differences well suited for near field AR applications.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.4.4 [Image Processing and Computer Vision]: Restoration—Wiener filtering

*These authors contributed equally.
†e-mail: {oshima.kohei.of0, damien.rompapas.dk1, goshiro, takafumi-t, sandor, kato} @is.naist.jp
‡e-mail: krm104@acm.org, swan@acm.org
§e-mail: ikeda.sei.jp@ieee.org

1 Introduction

Optical See-Through (OST) Head-Mounted Displays (HMDs) have seen an increase in both popularity and accessibility with the release of several consumer level options, including Google Glass and Epson Moverio BT-200, and announced future offerings, such as Microsoft's HoloLens, on the horizon. The transparent display technology used in these HMDs affords a unique experience, allowing the user to view on-screen computer generated (CG) content while maintaining a direct view of their environment, a property extremely well suited for augmented reality (AR) systems. Unfortunately, the current generation of consumer-level OST HMDs are only capable of presenting CG content at a single fixed focal distance. This inherent limitation becomes problematic as the user attempts to simultaneously view the world and CG objects together, inducing focal rivalry as the eye's optical system must continuously adjust to accommodate both the real and virtual items.

Figure 1 (a) and (b) illustrate an OST AR system in which a user concurrently observes real and CG images at disparate distances. As the eye's accommodation changes to match the focal distance of the real image, Figure 1 (c), the on-screen image appears blurred, Figure 1 (d). Naturally, the amount of blur perceived is directly related to accommodative ability, which varies from person to person and undergoes further changes with age. The majority of individuals experience steady decline in accommodative ability between child and adulthood [8], with more rapid decreases beginning to onset between 40–50, and general loss of sensitivity to focal differences occurring beyond 60 years of age. In order to effectively address such a wide variance in focal ability across users, flexible corrective measures, adaptable to specific user needs, are required.

Prior studies have proposed techniques for directly improving the distortion effects caused by the optical combiners within OST HMDs [16]. However, optical distortion from device specific components only contributes a constant error. Other prior work has applied image filtering and masking algorithms to video see-through

(VST) systems to produce user-specific corrections [28], which allows users with abnormal and diminished vision to effectively view imagery without the need for corrective lenses. A similar approach has likewise been deployed in an OST system, with the aim of enhancing user vision [16]. However, all of these techniques use static correction schemes, and therefore cannot address the blur produced by OST focal rivalry. This blur does not remain constant, but continuously changes as the user fixates on objects throughout the environment. Therefore, addressing OST focal rivalry blur requires a dynamic approach, which continually adjusts to match the changing focal demand of the user.

The Point Spread Function (PSF) [20], used to measure the focal ability of a system, also has application for identifying the refractive power limit of the eye's internal lens [32]. In this work, we present an active focus blur correction method capable of producing dynamic image improvement by continuously updating a PSF model for the user during system runtime. Our technique, designated SharpView, is the first methodology proposed for on-line processing of virtual imagery to counteract focus blur in OST AR.

While it is possible to enhance the sharpness of all virtual objects within sight, this might not always be preferable, particularly in applications where depth cues, due to blurring, are desired. Therefore, the current scope of our SharpView model is to improve the clarity of only the AR content under direct view in OST AR systems. We believe this affords an immediate benefit to systems which employ heads up displays (HUDs) or require the overlay of virtual objects onto physical items of interest, such as in assembly or maintenance tasks. Likewise, we believe that dynamic adjustment of the perceived focus of virtual content may also aid in the reduction of eye strain produced by the accommodative-convergence mismatch that results in binocular stereoscopic systems.

In addition to the SharpView algorithm, we also present the results of a user experiment designed to quantify the perception of focus blur in an OST AR system. We provide an analysis, which concentrates on both the change in perceived blur at varying focal disparities and the utility of SharpView for improving virtual image clarity. We close with a final discussion of our experimental results and comments to guide future work.

2 Related Work

Fixed focal distance is an inherent limitation of current consumer level HMD technology including not only VR but also VST and OST AR devices. While a large number of available options do support 3D stereoscopic viewing, the inability to provide correct accommodative cues, to match the user's changing convergence during use, decreases the realism and immersiveness of virtual content and increases the likelihood of simulator sickness and eye strain [21]. Blur and other visual aberrations are also common within single focus systems. Solutions for expanding the focal range of HMDs have included not only contributions from hardware development but also the image processing domain as well.

2.1 Focus Correct HMDs

Investigations by MacKenzie et al. [26] have shown that a minimum of five focal planes will adequately produce an acceptable range of accommodative cues to match user convergence. While using only four focus levels, a stereoscopic multi-focal display developed by Love et al. [25] uses a high speed lens system to present focused content at real-time response rates. Additional displays developed by Akeley et al. [2] and Schowengerdt and Seibel [34] utilize prisms and flexible membranes, respectively, to create variable focus image planes. While able to cover a much larger range, these systems too are only able to create a limited discrete number of focus planes. Liu et al. [24] employ active liquid-lens technology to generate OST imagery over a continuous focal range, though the response rate of the liquid-lens, 75ms, is too slow for use with dynamic virtual content. Light Field Displays (LFDs) have been proposed, [23], as effective alternatives to lens based approaches for generating accommodatively correct virtual content.

LFDs generate images by controlling both the color and direction of each light ray transmitted from the device into the user's eye [19]. Since the path of each ray is independently controlled, the focal distance of virtual content is inherently variable. LFDs commonly employ microlens arrays for image generation, though the volume of light rays required varies across implementation, with Maimone et al. [27] claiming that ultra high-density displays are required for effective near-eye LFD applications. Larger ray volumes however, produce greater processing requirements. Huang et al. [15], for example, present a LFD which requires 20sec to resolve the Light Field projection equation of a single image. An improvement on this earlier model though, [13], has produced a hybrid stereoscopic near-eye LFD comprised of layered LCD panels with diffusers and focusing lenses. This system, however, is only capable of producing VR, or VST AR, content and is not suitable for OST applications. Much like lens based multi-focus systems, the low performance, robustness, and availability to consumers has largely limited the accessibility of current LFD technology, leading to efforts in image processing for the development of more ubiquitous solutions to improving focus perception.

2.2 Focus Correction Through Software

The use of image processing is standard practice for rectifying defocus present in projector based systems caused by misalignment between the projection and image plane. Zhang and Nayar [39] propose that such systems be modeled as a constrained optimization problem in order to minimize artifacts, while utilizing only the available dynamic range of the projector. Often though, a level of blur is desired, to more accurately model depth of field effects, for example. Brown et al. [5] were able to successfully extend the depth of field in their projector based system by applying a pre-filter to the projected image. Daisuke et al. [18] also utilize pre-filtering methods to extend depth of field in conjunction with pre-calculated PSF values obtained by vibrating the projector lens at high speeds. Pre-filtering for focus manipulation has also seen application outside of projector systems.

Alonso and Barreto [3] as well as Yellott and Yellott [38] utilize pre-filtering to generate on-screen images intended to be viewed by visually impaired individuals. Their methodologies employ a pre-calculated PSF function to perform a deconvolution operation in order to distort the image within the rendering pipeline before it is finally presented on the display. Correctly modeling the PSF, to match deficiencies in the user's vision, enables the pre-distortion step to produce images which appear perceptually correct to the user. Subsequent improvements to this technique, by Huang et al. [14], reduce the low contrast side-effect of pre-filtering by presenting images on multi-layer displays. Pamplona et al. [31] further extend the application of these techniques to produce a VST HMD system able to replace corrective lenses for visually impaired users. A study by Montalto et al. [29] applies the same user specific pre-distortion method in order to render visually correct images to OST devices. Similar work by Itoh et al. [17] applies pre-filtered image masks, rendered to OST displays, to reduce blur and improve user vision. These methods, though, ultimately utilize only static correction filters to address constant systematic effects and vision irregularities. We have developed a novel approach, and first attempt, to diminish the visual presence of OST AR focus blur by applying a dynamic PSF model, which modulates the amount of correction provided throughout the user experience.

3 SharpView

As light from the environment passes through the eye, the rays are bent by the cornea and internal lens to ideally focus onto the retina.

The diffraction pattern of the light during this process is modeled by a PSF. If defects are present in the physical structures of the eye, as would be the case in individuals with inhibited vision, a static PSF is sufficient for determining the optical power requirements necessary for corrective lenses able to adjust or redirect the light rays back through the proper focal point within the eye. This same procedure can be analogously expressed, within the image processing domain, as a deconvolution operation between the image focused on the retina and the PSF. Therefore, if a PSF is known in advance, it is possible to restore a degraded image through deconvolution.

Common deconvolution operations include inverse filtering, Wiener filtering, and the iterative Richardson-Lucy algorithm [11]. Difficulty in performing deconvolution with a PSF arises, however, due to inaccuracies in the diffraction pattern model. Okumura et al. [30] have utilized information from captured images of 2D markers to estimate the PSF of a camera based system. Unfortunately, the same principles are not applicable to the human eye, since only the user themselves have access to the perceived image. Additionally, techniques used to measure a user's PSF require expensive or large physical hardware, such as auto-refractometers, and a processing demand which is not suited for real time applications [32]. Our SharpView algorithm overcomes these limitations by employing a simplified PSF model based on three user and system specific parameters: pupil diameter, distance to real world gaze point, and focal distance of the HMD display screen.

3.1 Image Convolution

In general, the convolution of an image o with a point spread function p produces a blurred image b.

$$b = o * p \tag{1}$$

The symbol $*$ represents the convolution operation. Equation (1) can be expressed within the spatial frequency domain as

$$B = O \cdot P \tag{2}$$

where B, P and O represent Fourier transforms of b, p and o, respectively. A blurred image can, likewise, be corrected through application of the inverse PSF, $O' = B/P$. However, the following Wiener Deconvolution is optimal for performing this task with minimal noise.

$$O' = \frac{B \cdot \overline{P}}{|P|^2 + |C|^2} \tag{3}$$

Here, $\overline{P}$ represents the complex conjugate of P. C is a constant to prevent instability of the solution when values of P are small. We apply this same principle to OST AR in order to create a final sharpened on-screen image with improved visibility over the original. We accomplish this by applying the Wiener filter, adjusted using the estimated PSF of the user, to rendered images O and display the resulting sharpened image S to the HMD.

$$S = \frac{O \cdot \overline{P}}{|P|^2 + |C|^2} \tag{4}$$

3.2 PSF Estimation

Effectively counteracting dynamic focus blur, caused by accommodative differences between the display screen and world in OST HMDs, requires the PSF to be determined at run-time. We accomplish this by using a simplified Gaussian function to approximate the PSF, which allows for faster update rates but with a reduction in accuracy.

We generate our approximation by modeling the intensity of the light rays, emitted from the display screen, intersecting varying points on the retina. The intensity distribution, p, can be represented by the following function.

$$P(x,y) = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{5}$$

In equation (5), σ represents the focal blur size, while x and y represent the pixel position of the emanating screen point. Our simplified PSF model, therefore, requires only one parameter, σ, to be determined at run time.

Consider the simple system, shown in Figure 2, used to model the combined optical system of the eye and OST HMD screen. The center lens, in this system, represents the internal crystalline lens of the eye and the right imaging plane characterizes the retina. As the user's gaze fixates at point M in the real environment, the light traveling from this point creates an image at point m on the eye's imaging plane, after passing through the lens. Similarly, light emitted from point M' on the HMD screen is imaged at point m' on the retina. While the user's focus is centered on the real environment, all of the light rays from the display will be improperly focused by the lens causing on-screen content to appear blurred. The radius of this blur is equal to σ in our Gaussian function, equation (5). This σ can be derived from the basic triangle ratio and lens equations as follows:

$$\sigma = \frac{av}{2}(\frac{1}{u'} - \frac{1}{u}) \tag{6}$$

Once obtained, it is then necessary to scale from the eye's image plane back to the virtual display. If the radius of display blur is expressed as σ_d, the ratio between the eye's image plane and screen is expressed as follows.

$$\sigma : \sigma_d = v : u' \tag{7}$$

Here, σ_d is directly obtainable from equations (6) and (7),

$$\sigma_d = \frac{a}{2}(1 - \frac{u'}{u}) \tag{8}$$

where a is pupil diameter, u is distance from the eye to the real world gaze point, and u' represents the distance from eye to HMD image plane. When performing the actual convolution between the filter and screen image, generally, σ_d may be converted to pixel size, scaling the values based on the dot matrix of the specific display used. Our complete SharpView algorithm simplifies blur correction into a single convolution operation, where the filter is constructed from an estimated PSF, requiring only three parameters, all measurable at runtime.

Further simplification might be considered by presuming a constant pupil size during usage. However, numerous studies [12, 1, 35, 33] have shown that pupil size is not only dependent on lighting but also mental load and task concentration, making live measurement of the pupil advisable. Additional system complexity may

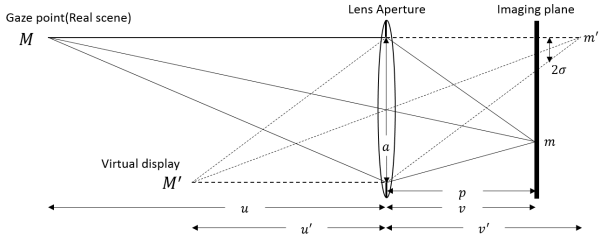
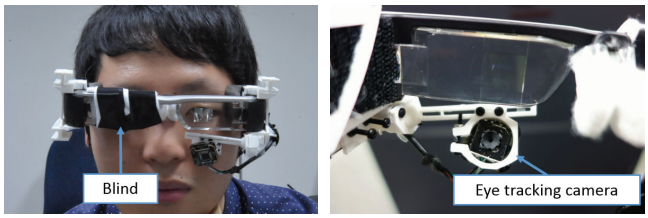


Figure 2: Optical system formed by the user's eye and an OST HMD. The imaging plane corresponds to the user's retina and the lens aperture to the user's pupil.



(a) (b)

Figure 3: The OST HMD system used in our experiment. (a) An example subject wearing the HMD and eye tracking combination. The right eye-piece is covered to produce a monocular view. (b) Closer view of the attached camera used for tracking subject gaze and measuring pupil size.

arise from the necessity to record the fixation distance between the user's eye and real object under focus. While it is possible to calculate the user's fixation through on-line vergence angle measurements provided by stereoscopic eye-tracking [7, 9, 37], an alternative method leveraging known 3D environmental layout information, compiled off-line using a depth sensor or camera such as the Microsoft Kinect, may be more suited for current consumer systems. Correlating the intersection of the user's gaze with this pre-rendered environment model, would facilitate rough estimation for the current focal demand of the user.

4 Experimental Design

In order to develop an effective improvement strategy for correcting focus blur in OST AR, measures for the amount of blur perceived by users must be obtained. Since focus blur occurs during accommodation rivalry between physical and on-screen content, it is essential to define the level of awareness users have to the presence of blur in rendered images. We developed SharpView based on the premise that perceived focus blur varies according to the focal disparity between the display and world. Therefore, if a physical object is viewed at a distance approximately equal to the focal distance of the HMD, very little focus blur should occur. However, if the distance to the physical object greatly differs from the HMD focal distance, a large amount of focus blur should be observed. This implies that a static correction scheme will not suffice, but rather a dynamic range of sharpening levels is required.

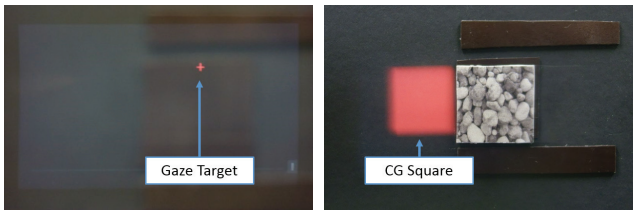
Based on this presumption, we designed and implemented a user study investigation to quantify user perception of focus blur within an OST AR system. Our experimental hypotheses are two-fold. First, we believe preferential sharpening levels for each subject will vary according to the magnitude of focal disparity, with more sharpening desired at greater disparities, and very low sharpening degrees chosen for the low focal disparity conditions. Secondly, we anticipate the users to select sharpening values which will follow a similar trend as that predicted by our SharpView model over-all.

4.1 Subjects

13 participants (10 male, 3 female), between the ages of 22 and 33, participated in our study. All subjects were university students recruited through email solicitation. Each subject was required to provide written consent before proceeding with the study and was provided monetary compensation for their time. 8 of the 13 subjects stated that they possessed normal uncorrected vision. The remaining 5 required the use of corrective contact lenses, which were worn throughout the duration of the task. Subjects were provided a thorough explanation of the experimental hardware and procedure before beginning any aspect of the experiment.

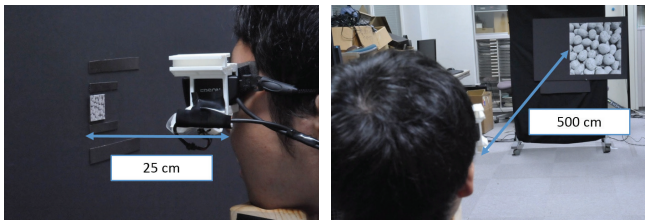
4.2 Experiment Platform

An Epson Moverio BT-200 OST HMD is used as the primary display worn by each subject. The Moverio is a binocular display



(a) (b)

Figure 4: Views of the eye tracking calibration process, as seen through the HMD. (a) An on-screen cross-hair used during the gaze-to-screen mapping phase of calibration. (b) Correct placement of the red square during the position adjustment phase of system calibration.



(a) (b)

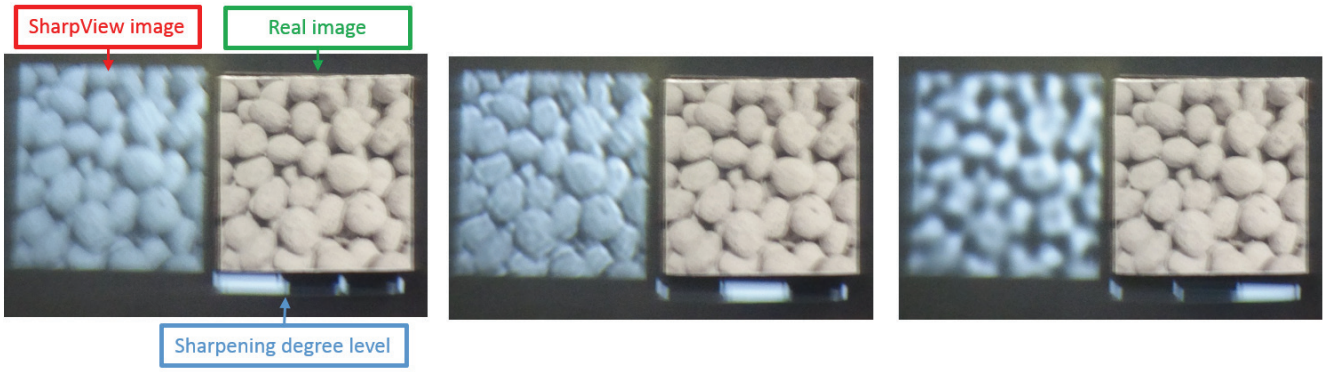
Figure 5: Location of subjects relative to reference images placed at 25 cm (a) and 500 cm (b) from the subjects' eyes.

capable of presenting stereo imagery at 60Hz with a maximum resolution of 960×540 per eye. The focal distance of the display was both independently measured and confirmed by the manufacturer to be approximately 7m. In order to remove effects due to stereoscopy and binocular rivalry, the right eye piece of the display is covered with a patch of black opaque cloth, Figure 3 (a). We attached a Microsoft LifeCam HD-6000 webcam, set to a resolution of 1280×720 at 30fps, beneath the left eye piece, Figure 3 (b). The camera itself is modified by removing the infrared and adding a visible light filter along with a pair of IR LEDs. Measurements for the left eye's pupil diameter, as well as gaze direction, are made using images captured by the HD-6000 and processed via version 0.4.5 of the PupilLabs open-source eye tracking software [22]. Accuracy of the gaze tracking was confirmed to within 1° with measures recorded at the frame rate limitation of the cameras, 30Hz. The OpenCV image processing library is used to perform the deconvolution required by SharpView and generate the on-screen images displayed at run time. A wireless hand held controller is additionally used to record subject adjustments and responses during all calibration and experimental tasks.

4.3 Eye Tracking

In order to induce and maintain the presence of focus blur during our experimental task, subjects must be forced to fixate their gaze onto physical objects in the world and also be prevented from shifting their focus onto on-screen content. We incorporate an eye tracking process into our OST HMD system making it possible to determine the subject's gaze direction and impose countermeasures to prevent accommodation change when their gaze moves toward AR imagery. A brief calibration process is required, though, in order to map the subject's gaze onto screen coordinates of the OST display. We actually employ a two phase procedure to both calibrate our eye-tracking system as well as adjust the placement of AR content within the subject's field of view.

First, we determine the direction of the subject's gaze with relation to the HMD screen. This process is accomplished through a standard gaze mapping routine, commonly used to calibrate commercial and consumer eye-tracking systems. A set of 12 cross-hairs, positioned to uniformly cover the on-screen area of the HMD,



(a) (b) (c)

Figure 6: View through the HMD of a reference and sharpened virtual image pair during the adjustment task. The rendered gauge provides an indication of the amount of sharpening correction applied. (a) Low, (b) moderate, and (c) high amounts of sharpening. Though color and contrast appear to differ, both on-screen and reference images used during the experiment were black and white. The images shown are captured through a digital camera and do not necessarily reflect the actual visual appearance experienced by the subjects during use.

are presented one at a time to the subject, Figure 4 (a). For each cross-hair, the subject is instructed to fixate and maintain their gaze on the center of the cross for approximately 5 seconds. During this fixation period, the gaze angle, in relation to the eye-tracking camera, is recorded. Using the known screen locations of each cross-hair and the recorded gaze angle measurements, a mapping between gaze direction and screen pixels is determined. The Pupil-Labs open-source software provides built-in routines for both collecting the gaze angle data and performing the mapping calculations.

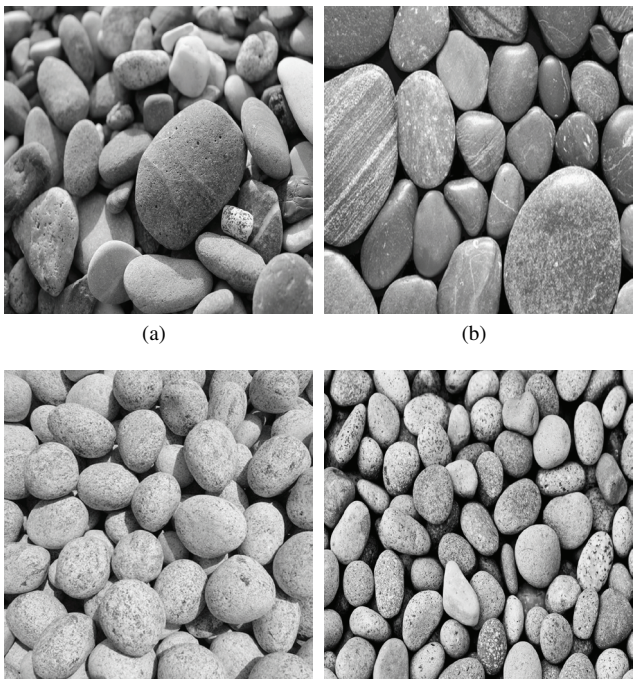
The second phase of calibration is intended to properly position the AR content within the subject's field of view. We perform this operation by rendering a red square to the HMD screen and allowing subjects to manually adjust its location. Using a hand held controller, subjects are allowed to move the square up and down and side-to-side until the right edge of the square is aligned flush with the left edge of a real reference image placed in front of the display, Figure 4 (b). Proper vertical positioning occurs when the center of the square aligns with the center of the reference image.

After successful calibration, We are able to track the subjects' gaze across the HMD screen and detect when fixation is moved toward on-screen content. We simply mute, turn off, all graphics as a counter measure preventing subjects from focusing onto the virtual screen. Once gaze is removed from the AR content's location, graphics are restored and the experiment continued.

In addition to gaze information, the eye-tracking software also provided a rough measure of pupil diameter in terms of image pixels. Presuming a standard iris diameter of 12mm [6, 10], we were able to calculate the distance scale between camera pixels of the eye tracking camera and physical distance. Using this same scale, redetermined per subject at the start of the experiment, the approximate pupil diameter in mm for each subject is estimated at run-time.

4.4 Adjustment Task

We utilize a simple adjustment task to obtain measures for the amount of perceived focus blur and preferred level of sharpening correction for subjects viewing real and virtual content through our OST AR system. Focus blur is induced by making subjects fixate their gaze upon real reference images, placed in front of the display, at distances differing from the actual focal distance of the HMD. A virtual image, identical to the real reference image, is then rendered to the display screen. The subject will consequently perceive the rendered image as blurred, even though it is unaltered, due to the focal disparity between the reference image and HMD.



(a) (b) (c) (d)

Figure 7: The four black and white stone patters used as both reference and on-screen imagery during the adjustment task.

The images used in our experiment were chosen to provide large amounts of high frequency visual information, such as clearly defined edges for example. Four asymmetric stone patterns were selected for use during the experiment itself, Figure 7, with an additional four images used during training sets before data was recorded. The resolution of the images displayed to the HMD, as well as the size of the reference images, was adjusted so that both images subtended a constant 6° of visual angle. In order to minimize effects from color difference between the printed and on-screen images, both the HMD and reference photos were kept black and white.

We presented the reference images at four distance levels, 25cm, 50cm, 100cm, and 500cm, as measured from the subject's eye. These distances were selected in order to obtain measures over a

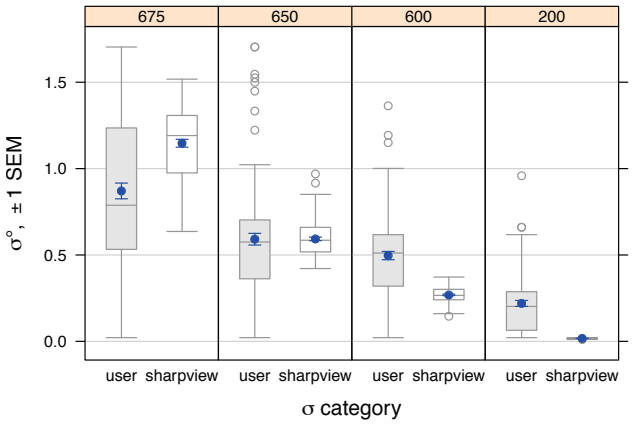


Figure 8: User-σ and SharpView-σ, expressed as visual angle, grouped by focal disparity. The mean and standard error of the mean (SEM) are shown in blue.

plausible viewing range for an indoor AR application. The 50cm condition is intended to roughly mimic arms-length. As previously noted, the accommodative ability of individuals, particularly within the near filed, greatly diminishes with age. Therefore, we chose our minimum distance of 25cm to provide a closer accommodative distance which still falls within the visible limits for older subjects that might suffer from degenerate focal ability. The 100cm condition continues the doubling trend between the 25cm and 50cm condition. Due to space limitation within our experimental environment, it was not possible to acquire an indoor space able to match the 7m focal distance of the Epson Moverio BT-200. Therefore, 500cm, 5m, was chosen to match, as closely as possible, the control condition where HMD and real world object focal distance are approximately equivalent. Figure 5 shows the location of reference images relative to subjects at the 25cm and 500cm distance levels.

In order to increase the applicability of our measures across HMD and system types, we henceforth express our four reference distance levels as the corresponding difference between the reference and on-screen image focal distance. Therefore, using the 7m fixed focal distance of the HMD screen, the 25cm reference image distance produces a focal difference of 675cm. The 50cm, 100cm, and 500cm reference images produce focal difference of 650cm, 600cm, and 200cm, respectively. These values represent the change in accommodative demand required by subjects to change their focus from one image to the other, with a lower focal demand corresponding to less focus blur.

The degree to which each subject perceives blurring of the virtual image is measured by applying our SharpView algorithm and allowing the subject to manually adjust the σ, or radius of sharpening, until the on-screen image appears the most visually clear, or sharp, compared to the reference image. We allowed subjects to increment or decrement σ to values between 1 and 40 pixels, at 1 pixel increments. The starting sharpness is randomly selected at the beginning of each trial. An additional gauge, rendered on-screen beneath the reference image, provides subjects with an indication of where, within the 1–40 pixel range, the current selected value lies. The intention of the gauge is to assist users in their selection by allowing them to first narrow the range of beneficial σ's among three zones, low, medium, or high, and then fine tune their final selection of an optimal σ relative to that primary grouping. Each gauge zone reflects a 13 pixel change in σ. Figure 6 provides a view through the HMD of the task visuals, as seen by each subject. The visual differences in sharpness between the virtual image at σ values in each of the three gauge zones is also shown.

Table 1: Mean and standard deviation for user-σ and SharpView-σ values, expressed in terms of visual angle, at each focal difference level.

Focal Difference (cm)	user-σ		SharpView-σ	
	mean (°)	std	mean (°)	std
675	0.89	0.37	1.14	0.21
650	0.60	0.32	0.61	0.12
600	0.51	0.12	0.27	0.05
200	0.22	0.15	0.02	0.002

4.5 Procedure

Before beginning any trial or training sets, we first fit the subject with our OST HMD and eye-tracker combination, and verified that the attached camera had a clear, unobstructed view of the subject's left eye. The eye-tracking calibration, as previously described, was then performed and the accuracy of gaze measurements verified using the PupilLabs software. A brief training set was also conducted prior to beginning any data recording, in order to verify that the subject had a clear understanding of how to both identify the presence of focus blur, and to properly perform the σ adjustment.

During each trial of the experiment, a reference image was first placed in front of the display, at one of the 4 distance levels. The subject was then instructed to fixate their gaze onto the reference image, after which the identical virtual image was rendered to the display with a randomly selected σ, between 1–40 pixels, used to apply initial sharpening. We employed the previously described eye tracking countermeasures to prevent the subject's gaze from drifting onto the virtual image. Subjects were then verbally instructed to adjust σ to first determine which of the three gauge zones provided the best improvement, and then to further refine their choice within that zone until they determined the σ which provided optimal clarity. A wireless hand-held controller was used by the subject to perform the σ adjustment, with one button used to increase and a second to decrease values. A third button was used to confirm the subject's final response. Once the subject had confirmed their response, the selected user-σ value, along with a SharpView-σ calculated using the live measurement of pupil diameter, was recorded.

A total of 8 trials, two for each of the four stone images, were performed at each distance level. After all 8 trials were completed, the process was repeated for the same images displayed at one of the remaining distance levels. Both the distance level and image presentation ordering were permuted to ensure that no two subjects experienced the same sequence of reference images or distance levels. The 8 trials, at each of the 4 distance levels, produced a total of 32 σ measurements per subject.

5 Analysis of Results

We recorded both the subject selected σ, which we will denote as user-σ, and the σ calculated from equation 8 using the on-line pupil measurements and distances to the reference and virtual images, which we will denote as SharpView-σ. Since both σ values represent the radius of perceived blur, we convert the units from pixels into degrees of visual angle, based on the resolution, field of view, and 7m focal distance of the display, which results in 1 pixel subtending approximately 0.043° of visual angle. Additionally, as previously noted, we denote data collected at each of the four reference image distances using the difference in focal distance between the images and the HMD screen, with the 25cm, 50cm, 100cm, and 500cm reference image distances corresponding to 675cm, 650cm, 600cm, and 200cm in focal disparity respectively.

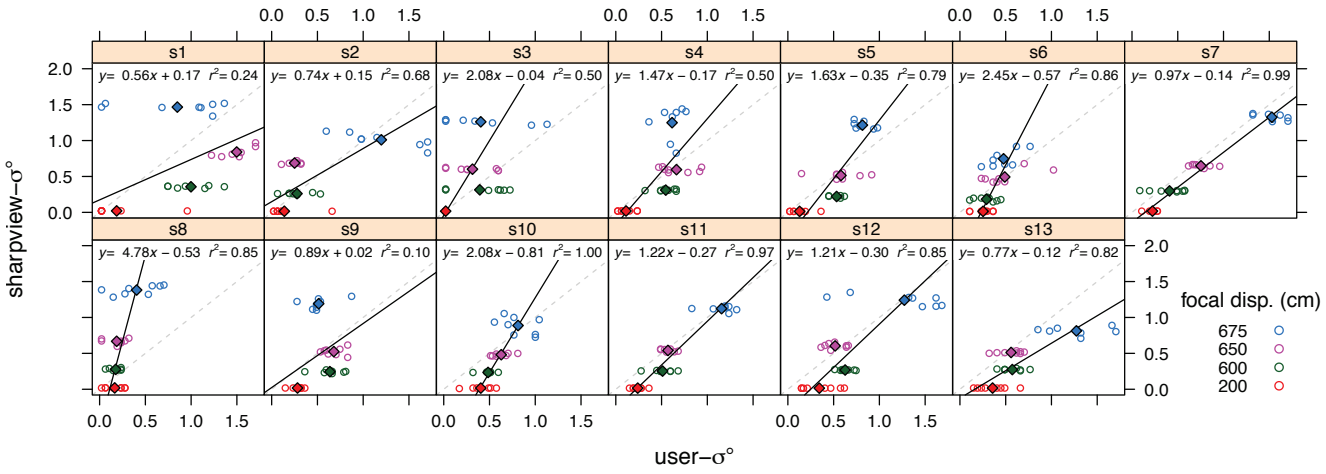


Figure 9: Comparison between user-selected sigma and the sigma predicted by our SharpView algorithm; both values are expressed as degrees of visual angle. The data points are color-coded according to focal disparity. The regression line is fit to the centroid (filled diamonds) of each distance. Ideal correspondence between our SharpView model and user preference is represented by a dashed line along the diagonal.

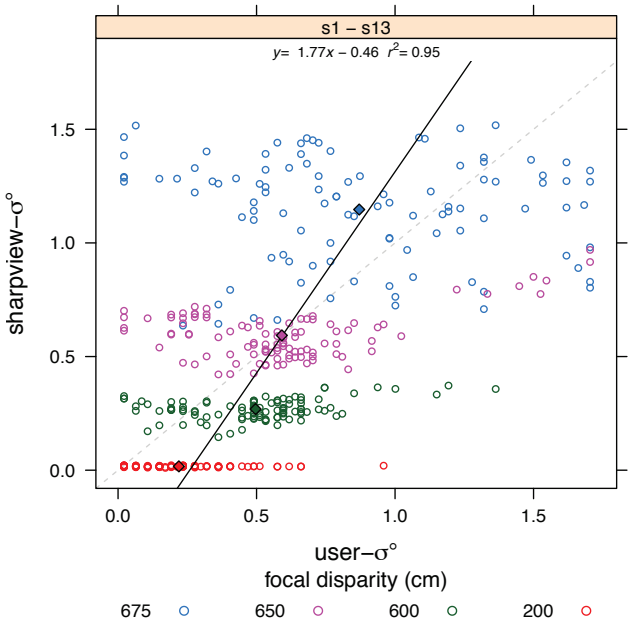


Figure 10: Data from Figure 9, averaged over all users. Ideal correspondence between our SharpView model and user preference is represented by a dashed line along the diagonal.

Figure 8 provides a side by side comparison of the user-σ selections and the modeled SharpView-σ. Visual inspection reveals a noticeable trend of increasing user-σ values with increasing focal distance disparity. The mean and standard error of the mean for the four user-σ groupings are also shown in Figure 8. Repeated measures analysis of variance (ANOVA) across the user-σ groupings, confirms that user-σ significantly differs according to distance ($F(3,36) = 16.7, p < 0.001$). A Ryan REGWQ post-hoc homogeneous subset test additionally indicates that all means are significantly different at $p \leq 0.05$.

The modeled SharpView-σ, likewise, increases with larger focal disparity distance, though at a slightly higher rate than the subject selected values. Paired ANOVA testing across the separate user and SharpView-σ groupings at each disparity level shows that the subject-chosen user-σ values do significantly differ from the modeled SharpView-σ at the 200cm, 600cm, and 675cm focus disparity levels ($F(1,12) = 39.2, p < 0.001$, $F(1,12) = 22.35, p < 0.001$, and $F(1,12) = 4.705, p \approx 0.051$, respectively). However no significance was found between user and SharpView-σ at a disparity of 650cm ($F < 1$).

We obtain a second metric for quantifying the modeling accuracy of our SharpView algorithm by plotting user-σ versus SharpView-σ. Figure 9 provides a per-subject σ comparison, while Figure 10 shows the comparison with all data combined. While both the SharpView model and accommodative functions of the user are both exponential in nature, our correspondence plots are intended to visualize the correlation between our sharpening prediction and user preference. Therefore, an ideal modeling between our SharpView and user selections will result in a linear trend, expressed in each plot as a dashed line along the diagonal.

The regression lines, shown in black, are fit to the centroids of the data clusters associated with recordings from each of the four focal disparity levels. Inspection of per-subject comparisons, once again, indicates that the modeled SharpView-σ values increase at a larger rate than the subject-selected σ values. The complete data comparison, Figure 10, confirms that this is indeed the case. However, it is clearly evident that the modeled SharpView-σ is most accurate at the 650cm focus disparity level, matching the ANOVA findings. Additionally, the clustering exhibited by data points within each distance level also reveals that subjects had the most consistent agreement on σ choices at the 200cm, 600cm, and 650cm disparities. Mean and standard deviation values for both the user and SharpView-σ results at each distance level are provided in Table 1.

6 Discussion

The impact of our experimental results is two-fold. Foremost, the significant difference in user-σ values across focus disparities, as exhibited in Figure 8, confirms that subjects are not only able to perceive the effects of focus blur on AR content, but also the improvement in clarity afforded by sharpening. Secondly, the significant differences in user selected correction validates our claim that focus blur mitigation requires a dynamic solution, such as our SharpView algorithm.

Since the focal disparity is lowest at the 500cm reference image distance—200cm focal difference—we would also expect the effects from focus blur to be lowest during trials at this level. Conversely, as the disparity rises, the amount of focus blur perceived in

AR content should also increase. This is indeed the trend we see in the user-σ results. We are able to, likewise, verify the positive impact of sharpening using similar reasoning. If the use of sharpening did not afford any improvement, or noticeable change in image appearance, we would expect to see overlapping σ distributions. This is, of course, not the case for our results (Figure 8), but instead our statistical analysis shows that mean user-σ values, between disparity levels, are not equivalent, but rather follow a similar trend to that predicted by our SharpView model.

Both the visual inspection and statistical analysis of our calculated SharpView-σ values show that the SharpView model is able to most accurately produce sharpening correction for focus blur at a focus disparity level of 650cm. For our OST HMD system, which has a focal distance of 700cm, this means that SharpView performs optimally when subjects view AR content in relation to real objects that are located at approximately 50cm, or arms length, away. This means that applications requiring the use of AR for near field, or near vision, tasks, such as maintenance, 3D modeling, or medical procedures, will experience the largest benefit from SharpView compensation.

The individual σ correlation plots, Figure 9, show that optimal blur correction is quite subject specific. It is not surprising, however, that this is the case. Our SharpView model is intended to be a lightweight solution, and subsequently uses a very simplified PSF model, which ignores user specific criteria. We do not, for example, consider influences from existing vision impairments, including myopia or presbyopia. The simplified PSF employed by SharpView, instead, presumes an ideal system, allowing us to reduce the number of independent parameters to three, pupil diameter of the user, focal distance to the display screen, and focal distance to world objects.

A pre-experiment questionnaire, provided to each subject prior to performing the experimental task, did record a base level of subject specified visual acuity. S1, S3, S6, S9, and S10 all required the use of corrective contact lenses. S2 and S4 described their vision as better than 20/20, with the remaining subjects claiming normal 20/20 visual acuity. Since we did not conduct any formal optometrical evaluations of the subjects though, we are not able to accurately investigate the possible impact of visual impairment, or heightened acuity, present in multiple participants. We would encourage future investigations to consider including simple vision tests, using Snellen or LogMAR charts, which may identify vision-related trends and aid in the development of correction factors for SharpView, and similar methods, based on known vision requirements. Furthermore, the σ values from Figures 9 and 10 also do not visually show the impact of pupil diameter across trials or subjects. While the distance between the HMD screen and reference image target conditions was held constant between subjects, the pupil diameter between participants was not controllable. Of course, variation in pupil diameter alone is not sufficient to explain or model the variability across subjects, it is still important to remember this factor when considering the sharpening responses.

An additional improvement to our experiment methodology includes expanding the number of disparity levels examined, or perhaps utilizing depth sensors or cameras to directly measure the distance to the reference images, or objects. The promised availability of such hardware in forthcoming consumer OST HMDs, such as the Microsoft HoloLens, not only makes the implementation of such a procedure more accessible to the research community at large, but also showcases the versatility and ease to which our SharpView algorithm can be adapted across display systems.

7 Conclusions and Future Work

In this work, we have presented a method for mitigating the effects of focus blur, caused by unequal accommodative demand, inherent to current consumer OST displays. Our SharpView technique utilizes a simplified PSF model of the human visual system, which enables real-time generation of sharpening filters, able to mitigate the dynamic visually degrading effects from focal disparity. We also performed a formal user investigation to confirm the utility of image enhancement methods for improving the clarity of blurred AR content.

Our experimental results confirm that users of OST AR are able to not only perceive the presence of focus blur, but also the improvements afforded by sharpening on-screen imagery. In addition, the optimal amount of blur reduction desired by subjects varied across focal disparity levels, verifying the need for dynamic blur reduction solutions. We were also able to validate the effectiveness of SharpView in mitigating the presence of blur at focus differences optimally suited for near field AR systems and applications.

However, further work is required. In particular, studies examining the relationship between visual acuity and blur perception could identify the correction factors needed to improve the robustness of SharpView and generalize its utility across a larger user base. Subject specific visual acuity data, specifically, will be beneficial in modeling the complete ideal PSF, providing additional insight into the accuracy of our own simplified SharpView and its ability to properly predict sharpening correction on a per user basis. Nevertheless, we believe that the SharpView methodology will be readily extendable to forthcoming consumer OST devices. A direct benefit from these next generation HMDs is the utilization of on-board depth sensors and cameras to provide real-time measurements of the focal distance to world objects, allowing the active correction of focus blur over a continuous range of object distances, as opposed to a discrete number of focal levels as employed by our study.

Furthermore, subsequent investigations extending SharpView to stereoscopic applications will be able to quantify the possible benefits in eye strain reduction due to alleviation of accommodation-convergence mis-match from binocular viewing. The identification of currently unknown side effects, positive or negative, must also be considered, along with the limitations of sharpening correction in systems where focus blur may be desirable to provide essential depth cues of virtual objects.

Finally, we believe that our results are highly relevant for a wide range of 3D user interfaces. First, enabling users to perceive graphics more clearly improves performance for all universal interaction tasks [4], including selection, manipulation, travel, way-finding, system control, and symbolic input. Second, our results can also be applied to displays other than OST HMDs, as long as there is a disparity between image plane and user focus. For example, to hand-held AR, as demonstrated by Tomioka [36] and colleagues, or projective displays, as demonstrated by Zhang and Nayar [39]; both systems could be enhanced based on our results.

Acknowledgements

Support for this work is provided in part by NSF awards IIS-1320909 and IIS-1018413 to J. E. Swan II, and a NASA Mississippi Space Grant Consortium fellowship, Bagley Graduate Internship and JSPS scholarships to K. Moser, as well as the Japan Society for the Promotion of Science under awards Grant-in-Aid for Scientific Research (B) #15H02737 and Challenging Exploratory Research #15K12084.

References

[1] S. Ahern and J. Beatty. Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, 205(4412):1289–1292, 1979.

[2] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks. A stereo display prototype with multiple focal distances. In *ACM Transactions on Graphics (TOG)*, pages 804–813. ACM, 2004.

[3] J. Alonso, M. 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, volume 1, pages 556–559, 2003.

[4] D. Bowman, E. Kruijff, J. LaViola Jr., and I. Poupyrev. *3D User Interfaces Theory and Practice*. Addison Wesley, 2004.

[5] M. S. Brown, P. Song, and T.-J. Cham. Image pre-conditioning for out-of-focus projector blur. In *Computer Vision and Pattern Recognition*, volume 2, pages 1956–1963, 2006.

[6] P. Caroline and M. Andre. The effect of corneal diameter on soft lens fitting, part 2. *Contact Lens Spectrum*, 17(5):56–56, 2002.

[7] B. C. Daugherty, A. T. Duchowski, D. H. House, and C. Ramasamy. Measuring vergence over stereoscopic video with a remote eye tracker. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 97–100. ACM, 2010.

[8] A. Duane. Studies in monocular and binocular accommodation with their clinical applications. *American Journal of Ophthalmology*, 5(11):865–877, 1922.

[9] A. T. Duchowski, D. H. House, J. Gestring, R. Congdon, L. Świrski, N. A. Dodgson, K. Krejtz, and I. Krejtz. Comparing estimated gaze depth in virtual and physical environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 103–110. ACM, 2014.

[10] J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman. *The eye: basic sciences in practice*. Elsevier Health Sciences, 2015.

[11] R. C. Gonzalez and R. E. Woods. *Instructor's Manual for Digital Image Processing*. Addison-Wesley, 1992.

[12] E. H. Hess and J. M. Polt. Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192, 1964.

[13] D. C. Huang, K. Chen, and G. Wetzstein. The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Transactions on Graphics (TOG)*, 34(4):60, 2015.

[14] F.-C. Huang, D. Lanman, B. A. Barsky, and R. Raskar. Correcting for optical aberrations using multilayer displays. *ACM Transactions on Graphics (TOG)*, 31(6):185, 2012.

[15] 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 (TOG)*, 33(4):59, 2014.

[16] Y. Itoh and G. Klinker. Light-field correction for spatial calibration of optical see-through head-mounted displays. *Visualization and Computer Graphics, IEEE Transactions on*, 21(4):471–480, 2015.

[17] Y. Itoh and G. Klinker. Vision enhancement: defocus correction via optical see-through head-mounted displays. In *Proceedings of the 6th Augmented Human International Conference*, pages 1–8. ACM, 2015.

[18] D. Iwai, S. Mihara, and K. Sato. Extended depth-of-field projector by fast focal sweep projection. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 21(4):462–470, 2015.

[19] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. An interactive 360 light field display. In *ACM Special Interest Group on Computer GRAPHics (SIGGRAPH)*, page 13, 2007.

[20] R. C. Jones. On the point and line spread functions of photographic images. *JOSA*, 48(12):934–937, 1958.

[21] D. Kahneman and J. Beatty. Pupil diameter and load on memory. *Science*, 154(3756):1583–1585, 1966.

[22] M. Kassner, W. Patera, and A. Bulling. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1151–1160. ACM, 2014.

[23] D. Lanman and D. Luebke. Near-eye light field displays. *ACM Transactions on Graphics (TOG)*, 32(6):220, 2013.

[24] S. Liu, D. Cheng, and H. Hua. An optical see-through head mounted display with addressable focal planes. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 33–142. IEEE, 2008.

[25] G. D. Love, D. M. Hoffman, P. J. Hands, J. Gao, A. K. Kirby, and M. S. Banks. High-speed switchable lens enables the development of a volumetric stereoscopic display. *OSA Optics Express*, 17(18):15716–15725, 2009.

[26] K. J. MacKenzie, D. M. Hoffman, and S. J. Watt. Accomodation to multiple focal plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision*, 10(8):22, 2010.

[27] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 29–38, 2013.

[28] 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 (TOG)*, 34(3):28, 2015.

[29] 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 (TOG)*, 34(3):28:1–28:16, 2015.

[30] B. Okumura, M. Kanbara, and N. Yokoya. Augmented reality based on estimation of defocusing and motion blurring from captured images. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 219–225, 2006.

[31] V. F. Pamplona, M. M. Oliveira, D. G. Aliaga, and R. Raskar. Tailored displays to compensate for visual aberrations. *ACM Transactions on Graphics (TOG)*, 31(4):81, 2012.

[32] J. Santamaría, P. Artal, and J. Bescós. Determination of the point-spread function of human eyes using a hybrid optical–digital method. *JOSA A*, 4(6):1109–1114, 1987.

[33] T. Schaefer Jr, J. B. Ferguson, J. A. Klein, and E. B. Rawson. Pupillary responses during mental activities. *Psychonomic Science*, 12(4):137–138, 1968.

[34] B. T. Schowengerdt and E. J. Seibel. True 3-d scanned voxel displays using single or multiple light sources. *Journal of the Society for Information Display*, 14(2):135–143, 2006.

[35] H. Simpson and S. M. Hale. Pupillary changes during a decision-making task. *Perceptual and Motor Skills*, 29(2):495–498, 1969.

[36] M. Tomioka, S. Ikeda, and K. Sato. Approximated user-perspective rendering in tablet-based augmented reality. In *International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 21–28. IEEE, 2013.

[37] R. Wang, B. Pelfrey, A. T. Duchowski, D. H. House, et al. Online gaze disparity via bioncular eye tracking on stereoscopic displays. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 184–191. IEEE, 2012.

[38] J. I. Yellott and J. W. Yellott. Correcting spurious resolution in defocused images. In *Electronic Imaging*, pages 64920O–64920O. International Society for Optics and Photonics, 2007.

[39] L. Zhang and S. Nayar. Projection defocus analysis for scene capture and image display. In *ACM Transactions on Graphics (TOG)*, pages 907–915. ACM, 2006.