NAIST-IS-MT1451023

Master's Thesis

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

Kohei Oshima

March 1, 2016

Graduate School of Information Science
Nara Institute of Science and Technology

A Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Kohei Oshima

Thesis Committee:

Professor Hirokazu Kato	(Supervisor)
Professor Naokazu Yokoya	(Co-supervisor)
Associate Professor Christian Sandor	(Co-supervisor)
Assistant Professor Goshiro Yamamoto	(Co-supervisor)

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

Kohei Oshima

Abstract

Augmented Reality (AR) systems, that 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 having a 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 thesis, 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 also examine the application of dynamic sharpening filters as a straight-forward, system-independent means for mitigating this blurring effect, thereby improving the clarity of defocused AR contents. We assess the utility of our 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. In addition, as a confirmation of practicality, we validated SharpView in an AR matching task.

Our experimental results confirm that dynamic correction schemes are required for adequately addressing the presence of blur in Optical See-Through AR. Also we got a finding that while users are sensitive to the perceptual effects of focal disparity, their performance in an AR matching task was not negatively impacted even in the presence of significant blur. Through the experiments, we validate the ability of our SharpView model to improve the perceived visual clarity of blurred content, with optimal performance at focal differences well-suited for near-field AR applications.

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1451023, March 1, 2016.

Keywords:

Augmented Reality, Optical-See-Through Head-Mounted Displays, Focus Blur, Depth Perception, Sharpening Filter, Point Spread Function

光学シースルー型ヘッドマウントディスプレイにおける焦点ボケ映像の視認性改善*

大嶋 晃平

内容梗概

光学シースルー型ヘッドマウントディスプレイを用いた拡張現実感（AR）のシステムは, すでに消費者レベルにおいて利用可能なものもあり, 一般的になりつつある. しかし, 現代の光学シースルー型デバイスにおける共通の要素として, 仮想コンテンツがデバイスの焦点距離に固定されている点が挙げられる. ビデオシースルー型 AR は仮想コンテンツと現実環境の両方がディスプレイにおいて単一平面に表示されているので, 固定焦点に関して問題がないのに対し, 光学シースルー型 AR は仮想のディスプレイ平面と現実環境との間の奥行きの不一致は避けられない.

本論文において, 我々は異なる奥行きに存在する仮想コンテンツと現実物体を同時に見る際に生じる焦点ボケの問題に関して調査した. またボケの影響を軽減することで, 焦点ボケした AR コンテンツの視認性を向上させる, システムに依存しない動的な鮮鋭化フィルタのアプリケーション SharpView を検討した. 我々は, ユーザーが光学シースルー型ディスプレイとは異なる奥行きに提示された現実環境の画像を注視した状態で, 仮想コンテンツにおける焦点ボケを軽減するために鮮鋭化具合を調節する実験により本手法 SharpView の有用性を評価した. それに加え, 実用性の調査として, 我々は AR のマッチングタスクにおいて SharpView を検証した.

我々は実験結果から, 光学シースルー型 AR における焦点ボケに適切に対処するためには本手法 SharpView の様な動的な補正が必要であることを確認した. また, 我々はユーザが光学シースルー型 AR における焦点ボケに敏感でありながら, マッチングタスクにおいては焦点ボケがパフォーマンスに大きな影響を及ぼさないことを発見した. 実験を通して, 焦点ボケした仮想コンテンツの視認性を向上させる SharpView の有用性を評価し, 結果から SharpView は近距離の AR アプリケーションに適していると考えられる.

キーワード

拡張現実感, 光学シースルーヘッドマウントディスプレイ, 焦点ボケ, 奥行き知覚, 鮮鋭化フィルタ, 点拡がり関数

*奈良先端科学技術大学院大学 情報科学研究科 修士論文, NAIST-IS-MT1451023, 2016 年 3 月 1 日.

Acknowledgements

First of all, I would like to thank Professor Hirokazu Kato for giving me the opportunity to conduct this research. Prof. Kato gave me great training and advice in everyday life. Thank you.

I had the pleasure to work with an associate professor, Christian Sandor. I received much guidance from Prof. Sandor about the research content and direction. The amazing point of Prof. Sandor is the strong desire for new knowledge. Prof. Sandor always get and share latest information on virtual reality or computer vision with lab members. I would like to follow his active style.

I was also supervised by two assistant professors: Prof. Goshiro Yamamoto and Prof. Takafumi Taketomi. Prof. Yamamoto gave me many advice for my research life. Prof. Taketomi gave me some important advice about this research. Also, secretary: Makiko Ueno has supported our research life. I would like to thank them deeply.

Here, I would like also to thank co-workers. This work is helped by people of Mississippi State University: Doctoral student Kenneth Moser and Professor J. Edward Swan II. Kenneth has high programming skill and writing ability. Kenneth helped me when implementing the system and writing papers in English. Prof. Swan is a professional in statistical analysis. Prof. Swan extracted the meaningful facts from the data of my experiments. Without these two people, I cloud not have completed my research. Dr. Sei Ikeda, lecturer of Ritsumeikan University, also supported this research. Dr. Ikeda was my supervisor at my previous university and taught me many things. In addition, I would like to thank Yuta Itoh, Doctoral student in Technical University of Munich. I received a variety of advices from him.

I would also like to thank my co-supervisor Professor Naokazu Yokoya. His comments and suggestions helped me to improve my thesis's overall quality.

I had the pleasure to work with our lab members. They not only gave me advice on the research, but also organized many lab events. So I was able to enjoy everyday.

Finally, I would like to thank my family for their economical and mental support during my whole life.

Contents

1. Introduction

In this Introduction part, first we talk about the current problem of Optical See-Through (OST) Head-Mounted Displays (HMDs) Augmented Reality (AR) in Section1.1, then we introduce the related works and clarify the difference between the our research in Section1.2. Our methodology called SharpView is given in Section 1.3, and the contribution will be discussed in Section 1.4. Finally we describe the overall structure of the thesis in Section 1.5.

1.1 Motivation

AR is a technology that gives a feeling as if a virtual object was present in the actual real environment. AR related to vision is common, presenting a virtual object drawn by computer graphics (CG) on to the real environment seen by users. AR systems that link a real environment and a virtual environment seamlessly, can present information from a computer to users intuitively. We think AR is useful for commercial application in various fields.

Various displays and projectors are common devices for AR. Displays are categorized as head-worn, hand-held and spatial depending on the device position [33]. Recently, OST 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. Moreover, companies have announced future offerings, such as Microsoft describing the HoloLens to be on the horizon. The transparent display technology used in these HMDs affords an unique experience, allowing the user to view on-screen CG content while maintaining a direct view of their environment, a property extremely well-suited for AR systems. Unfortunately, the current generation of consumer-level OST HMDs are only capable of presenting CG content at a single fixed focal distance [4]. 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) illustrates an OST AR system in which a user concurrently observes real and CG images at disparate distances. It is desirable that both the CG and the object in the real world appear to be as clear as in Figure 1 (b). In practice, however, either of both would be blurry because of the difference in depth as in Figure 1 (c). 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 from childhood to adulthood [7], with more rapid decrease from ages 40 to 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.

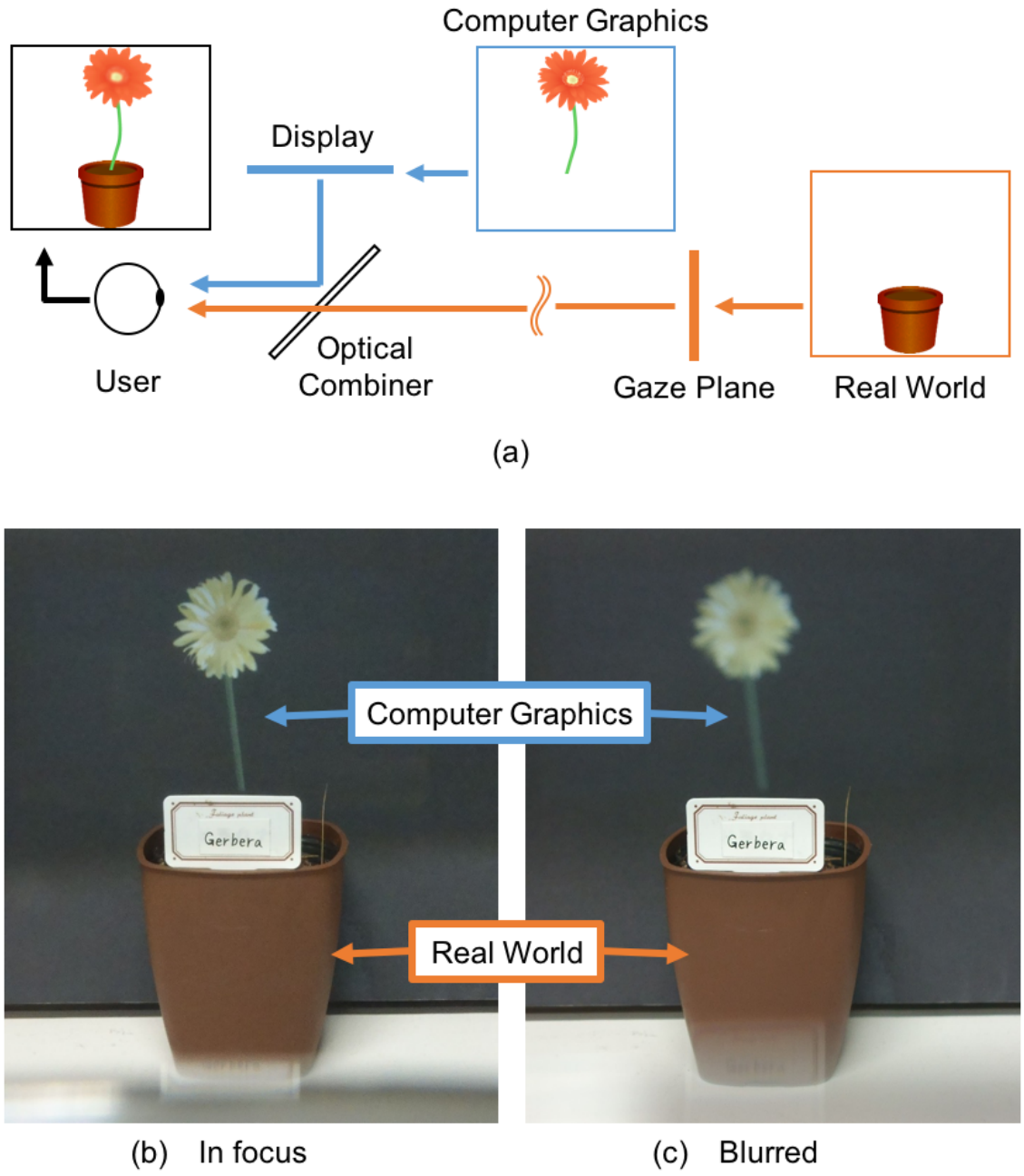


Figure 1. (a): Simplified schematic of an OST AR system. Blurring occurs when the virtual display screen and real world imagery are viewed at unequal focal distances. (b, c): Views through an OST AR system. (b) shows ideal AR, both of the CG and the object in the real world appears to be clear. (c) shows an actual AR example wherein the CG looks blurry. When the user focuss on the object in the real world, the CG cannot avoid being blurred because of the difference in depth.

1.2 Related Work

Prior studies have proposed techniques for directly improving the distortion effects caused by the optical combiners within OST HMDs [14]. 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 [25], 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 [15]. 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.

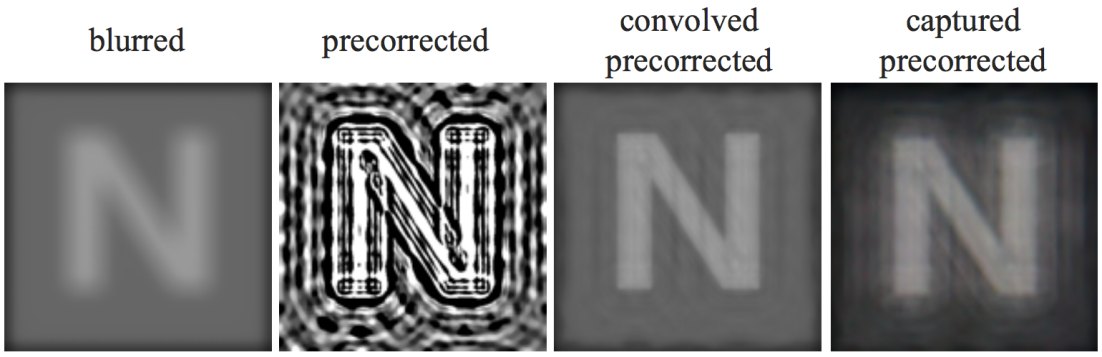


Figure 2. Image filtering to video see-through (VST) systems. A Total Variation Approach for Customizing Imagery to Improve Visual Acuity [25]

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

Figure 3. Enhancing user vision on OST system. (a) A user-perspective image. (b) A degraded user view by optical blur. (c) A compensation image, and (d) An enhanced image created by combining (b) and (c). Vision Enhancement: Defocus Correction via Optical See-Through Head-Mounted Displays [15]

1.3 Methodology

In this Section1.3, first we explain SharpView algorithm as an approach. Then, we introduce a prototype that implements SharpView system. Finally, we describe the experiments that were performed by using the prototype.

1.3.1 SharpView Technique

We propose the SharpView system, which always presents an optimal pre-filtering image to the user by considering a dynamic Point Spread Function (PSF) [18]. PSF is a function that represents how light spreads through a lens. The PSF can also be applied to identifying the refractive power limit of the eye's internal lens [29]. In the case of OST-HMDs, we must consider how environmental light spreads on the retina through the eye lens.

It's possible to calculate the blurred image b from the original image o by using the PSF p as follows:

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

If the PSF p is already known, we can create a pre-sharpening image s which reduces the defocusing influence from an original image. This pre-sharpening image is restored through focus blur process as follows:

$$s * p = r \tag{2}$$

The symbol r refers to the restored image. Figure 4 shows both of the convolution processes: equation (1) and equation (2) by using image examples. We generate a sharpened image by using the Wiener filter in order to suppress the ringing effect. We accomplish the correction by presenting a sharpened image on the OST-HMD. The image is restored by the human eyes through the focus blur in the eye lens. For the image to be restored properly, it is necessary to estimate the PSF of user's eyes.

We propose the on-time PSF estimation by approximating a human eye as a simple lens model and assuming an eye's PSF with a Gaussian function. The Gaussian function, as expressed by the following equation, depends only on the parameter σ which refers to the degree of sharpening.

$$P(x,y) = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{3}$$

The σ value is back-calculated from three parameters in Figure 5: the pupil diameter a, the distance from the eye to the gaze point u and the distance from the eye to the virtual display u'. This calculation is based on the official triangle ratio and the lens equation. Therefore, our SharpView system can always provide suitable sharpening images for users by continuous estimation of the eye's PSF using these three values and pre-filtering according to the PSF.

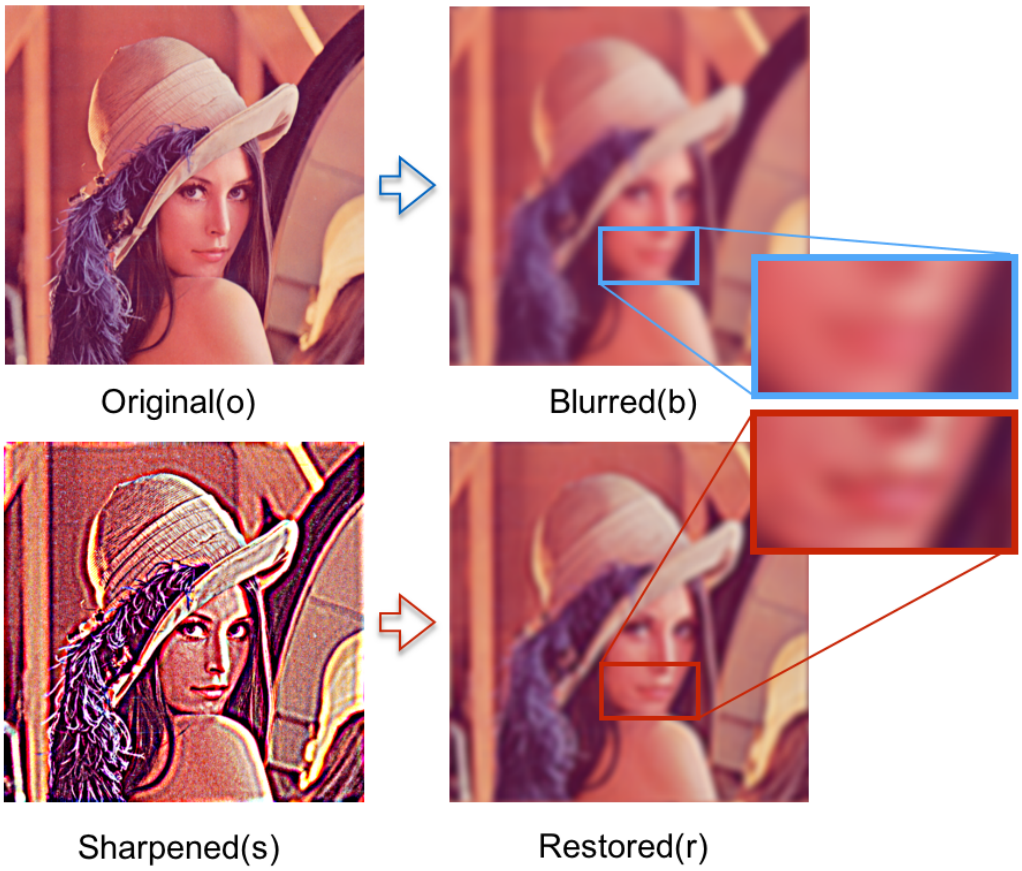


Figure 4. The top pictures show the image deterioration by focus blur. The bottom pictures show the same degradation process for the sharpened image. We can confirm that the restored image has better visibility than the blurred image.

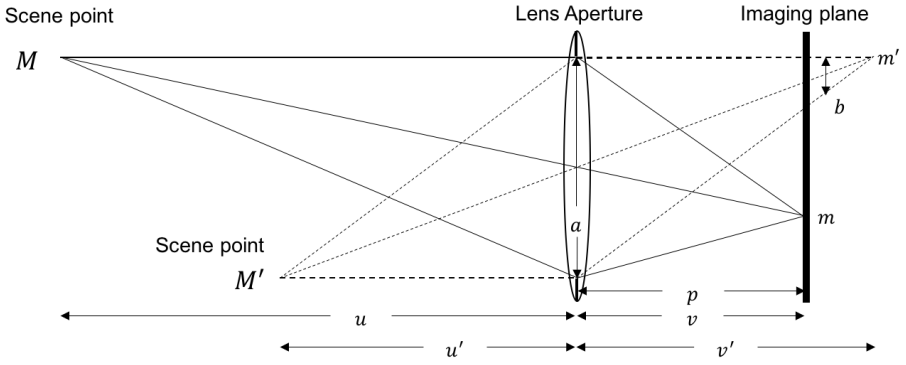
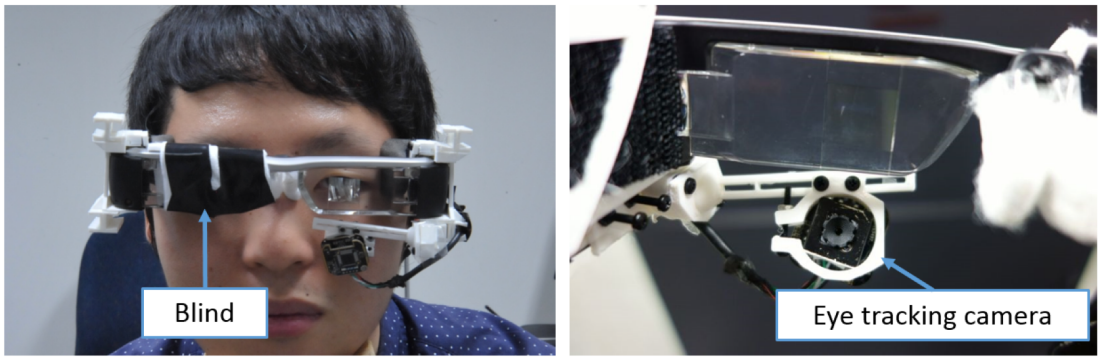


Figure 5. Optical system formed by the user's eye and an OST HMD. The imaging plane corresponds to the user's retina and the lens aperture to the user's pupil.

1.3.2 Prototype

We created a prototype of the SharpView system running on the Epson Moverio BT-2000, an OST-HMD. We then conducted experiments to verify the effect of SharpView. Figure 6 shows our prototype, the display for right eye is covered to prevent binocular parallax. The usb camera mounted below the display is for getting the pupil diameter of the user. In the case of OST-HMDs, the distance to the virtual display is constant. Also, we controlled the distance to the fixation point by instructing to users to focus on the target. By using this prototype, we conducted the experiments.



(a) (b)

Figure 6. Hardware setup: (a) OST HMD viewed from the front. The display of the right eye is covered to provide a monocular view. (b) Inside the OST HMD. The camera is attached under the display. The eye image from this camera is used for gaze tracking and pupil diameter estimation.

1.3.3 Experiments

The objectives of our user experiments are three fold. Firstly, we desire to quantify user perception of focus blur during OST AR tasks. Since focus blur only occurs while fixation is away from on-screen content, it is essential to define the level of awareness users have of the effect. We run Exp I and confirmed that user's pupil diameter changes while engaged in a simple visual AR task. Therefore, active pupil size measures must be taken at run time in order to effectively utilize SharpView.

In addition, if corrective methods, such as SharpView, are to be viable, then user sensitivity to focus improvements must also be measured. From Exp II, we confirmed 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.

Lastly, we wish to obtain performance metrics relating blur level to user accuracy in AR

image identification. A matching task was chosen for this measure due to the inherent need of both screen and world information for image comparison. The utility of SharpView in improving any user performance degradation due to focus blur is also examined. As a result from Exp Ⅲ, we could not confirm the performance improvement in the matching task by SharpView. However, we found that focus blur did not inhibit the natural image recognition ability of users. The experimental details are described in the Section 4 Experiments.

1.4 Contributions

In this thesis, 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. In addition to the SharpView algorithm, we also present the results of user studies designed to quantify the perception of focus blur in an OST AR system. From this result, we proved that SharpView system is effective in the situation when users looking at the depth of 50cm (arm's length) such as a factory task. Also, 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. The analysis led us to the following sub-findings.

- User's pupil diameter changes while engaged in a simple visual AR task. (Exp Ⅰ)
- Focus blur is a noticeable problem in the OST-HMDs. (Exp Ⅱ)
- Focus blur does not obstruct the matching task. (Exp Ⅲ)

1.5 Thesis Structure

The rest of the thesis is structured as follows: in Section 2, related work for focus correction techniques in terms of hardware and software are discussed. Section 3, Methodology, details the SharpView system in two parts of image convolution and PSF estimation. Finally, the work is validated through three experiments in Section 4 and the conclusions are stated in Section 5.

2. Related Work

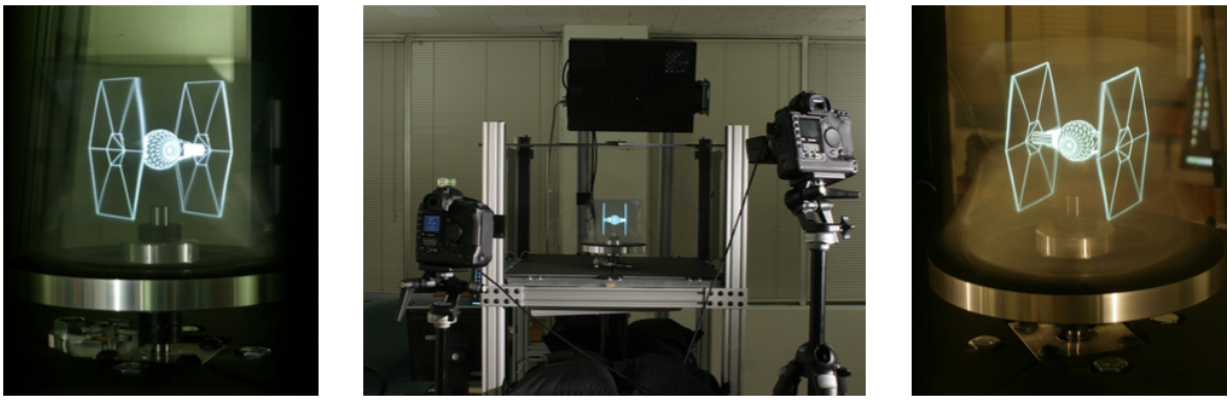
Fixed focal distance is an inherent limitation of current consumer level HMD technology including not only VR but also VST and OST AR devices. While a large number of available options do support 3D stereoscopic viewing, the inability to provide correct accommodative cues, to match the user's changing convergence during use, decreases the realism and immersiveness of virtual content and increases the likelihood of simulator sickness and eye strain [19]. 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 Correction through Hardware

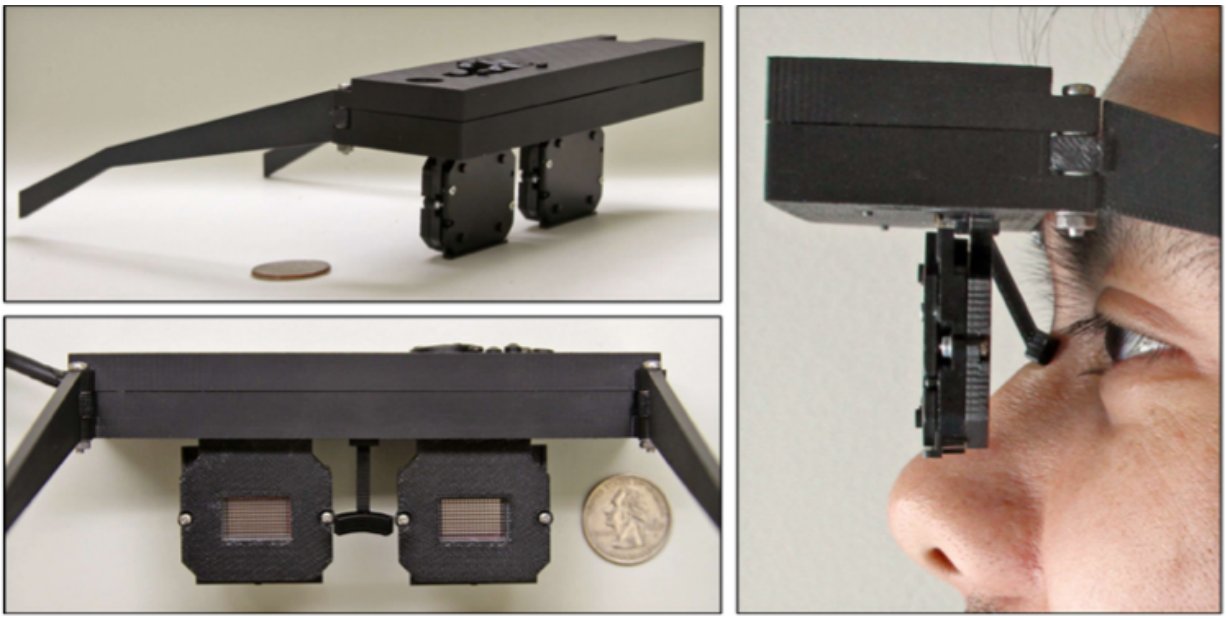
Investigations by MacKenzie et al. [23] have shown that a minimum of five focal planes will adequately produce an acceptable range of accommodative cues to match user convergence. While using only four focus levels, a stereoscopic multi-focal display developed by Love et al. [22] uses a high speed lens system to present focused content at real-time response rates. Additional displays developed by Akeley et al. [2] and Schowengerdt and Seibel [31] 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. [21] 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, [17, 20, 28], 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 [17]. 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. [24] 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. [13], 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, [11], 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.



(a)



(b)

Figure 7. Studies of Light Field Displays. (a) Omnidirectional Light Field Display is realized by three devices: the high-speed projector, spinning mirror, and synchronized motor. [17] (b) Light Field Display enables thin, lightweight near-eye display. [20]

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 [36] 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. [16] 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 [35] 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. [12], reduce the low contrast side-effect of pre-filtering by presenting images on multi-layer displays. Pamplona et al. [28] 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. [26] applies the same user specific pre-distortion method in order to render visually correct images to OST devices. Similar work by Itoh et al. [15] 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. Methodology

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. This structure is similar to the camera as shown in Figure 8. We can consider to replace the pupil with the aperture, the crystalline lens with the lens, the retina with the imaging plane. In the case of camera, the external light passes through the lens and forms an image on the imaging plane. When that, camera controls the amount of light passing through the lens by adjusting the aperture size. Therefore, in this thesis, we consider human eye as a simple camera model, and think about the focus blur on the retina based on that premise. 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.

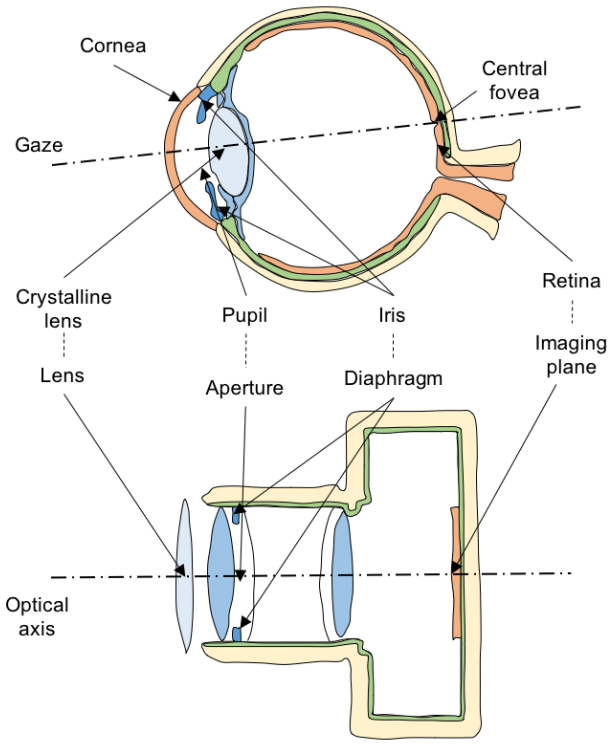


Figure 8. Model of eye and camera

Common deconvolution operations include inverse filtering, Wiener filtering, and the iterative Richardson-Lucy algorithm [9]. Difficulty in performing deconvolution with a PSF arises, however, due to inaccuracies in the diffraction pattern model. Okumura et al. [27] 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 [29]. 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.

$$b = o * p \tag{4}$$

The symbol $*$ represents the convolution operation. Figure 9 (1) shows an example of the image convolution by PSF. Equation (4) can be expressed within the spatial frequency domain as

$$B = O \cdot P \tag{5}$$

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.

$$O' = \frac{B \cdot \overline{P}}{|P|^2 + |C|^2} \tag{6}$$

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. Figure 9 (2) shows an example of the image correction. 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.

$$S = \frac{O \cdot \overline{P}}{|P|^2 + |C|^2} \tag{7}$$

Figure 9 (3) shows an example of the image sharpening. This sharpened image s is restored on the retina by focus blur through the eye lens. The process is expressed as following equation.

$$r = s * p \tag{8}$$

r represents the restored image. Figure 9 (4) shows an example of the image restoring of the sharpened image.

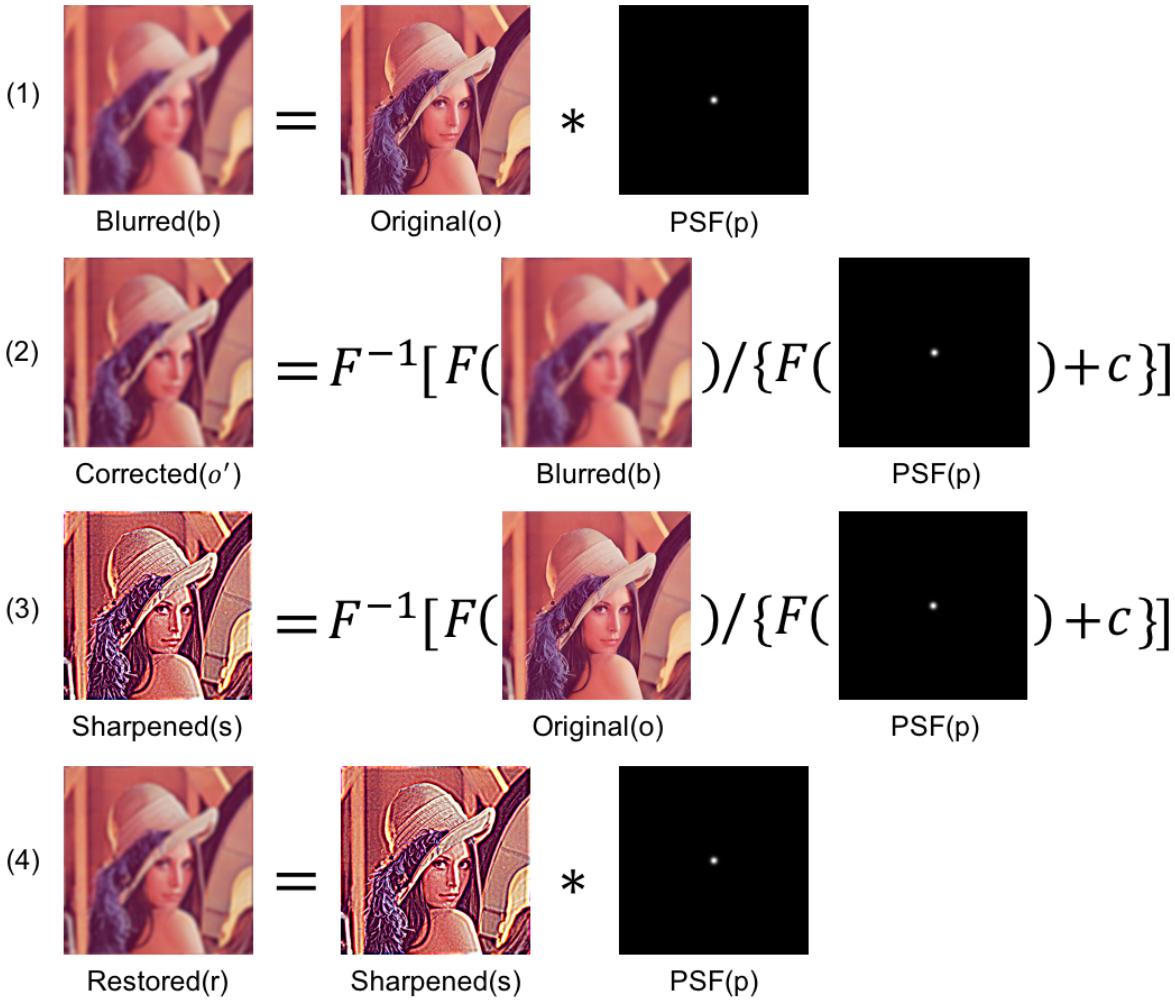


Figure 9. The convolution processing by using image examples. (1) shows the image degradation by convolving with the PSF. (2) shows the image correction by deconvolution processing. (3) shows how to generate a sharpened image from the original image. (4) shows the restoring of the sharpened image by convolving with the PSF.

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.

$$P(x,y) = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{9}$$

In equation (9), 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 10, 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

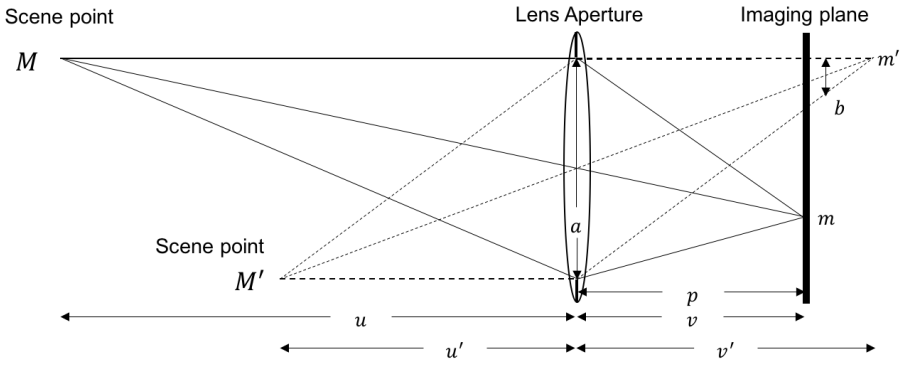


Figure 10. Optical system formed by the user's eye and an OST HMD. The imaging plane corresponds to the user's retina and the lens aperture to the user's pupil.

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 (9). This σ can be derived from the basic triangle ratio and lens

equations as follows:

$$\sigma = \frac{av}{2}(\frac{1}{u'} - \frac{1}{u}) \tag{10}$$

Once obtained, it is then necessary to scale σ 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.

$$\sigma : \sigma_d = v : u' \tag{11}$$

Here, σ_d is directly obtainable from equations (10) and (11),

$$\sigma_d = \frac{a}{2}(1 - \frac{u'}{u}) \tag{12}$$

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 [1, 10, 30, 32] 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. We formally examine the efficacy of SharpView, and the application of sharpening filters in general, for mitigating focus blur, through a user study.

4. Experiments

The objectives of our user experiments are two fold. Firstly, we desire to quantify user perception of focus blur during OST AR tasks. Since focus blur only occurs while fixation is away from on-screen content, it is essential to define the level of awareness users have of the effect. In addition, if corrective methods, such as SharpView, are to be viable, then user sensitivity to focus improvements must also be measured. Lastly, we wish to obtain performance metrics relating blur level to user accuracy in AR image identification. A matching task was chosen for this measure due to the inherent need of both screen and world information for image comparison. The utility of SharpView in improving any user performance degradation due to focus blur is also examined. Since our SharpView model relies on user pupil diameter measures, we first conducted a pilot study to investigate the need for real time pupil measurement over a static approximation, before implementing the algorithm in the subsequent experiments.

4.1 Exp I: Pupil Variability during OST AR Tasks

Pupil dilation and contraction due to fluctuations in light intensity is a commonly known physical response. It is also well documented, within the psycho-physical domain, that cognitive load and emotional stress [19] prompt changes in pupil state as well. While it is possible to manage the illumination level of an experimental environment, we cannot predict the effect that the required mental focus of an AR task will have on user's pupillary responses. The objective of our first experiment is to simply confirm whether the variation in user's pupil diameter, while engaged in a simple visual AR task, is sufficiently large to require active measurement of the pupil during future experiments, or if a single pre-experimental acquisition of pupil diameter size will suffice. We obtained measures for rate of pupil diameter change over time through the following pilot test.

4.1.1 Participants

Three males, between the ages of 23 and 26, were recruited for participation in this experiment. All three were confirmed to have normal vision, using a simple acuity test, and did not require the use of glasses or contact lenses. Due to the abundance of literature on pupil size effects due to mental load, three subjects is an adequate sampling for this experiment, intended only as a confirmation that pupil variance is enough to warrant the added complexity of live measurements during the remaining experimental tasks.

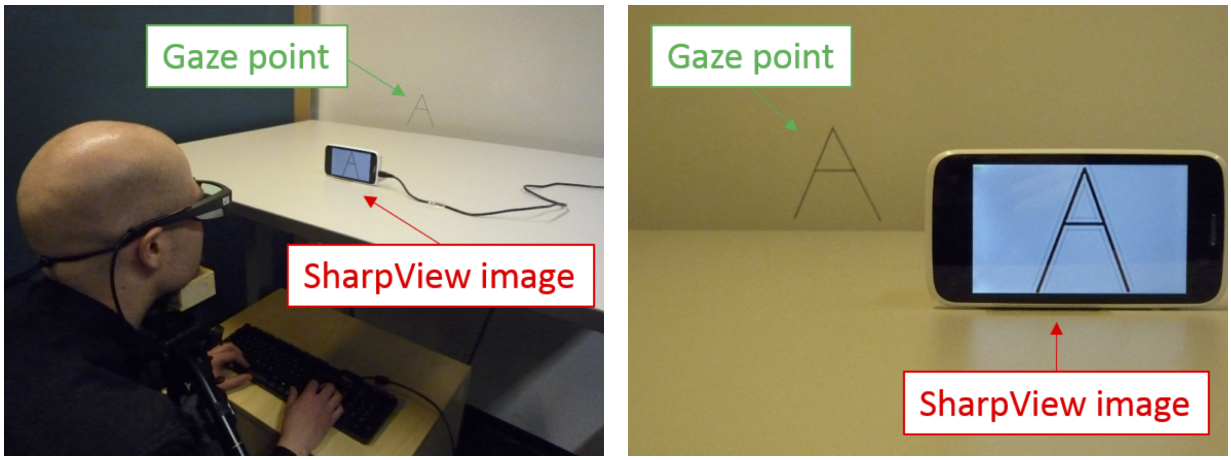
4.1.2 Experimental Platform

The open source computer vision library, OpenCV, was used for performing the sharpening operation. A Covia's FLEAZ F4s, 4" screen at 480×800 resolution and running Android 4.4, was used to display the sharpened images. A Windows 7 PC was conducted to the device for use as a second display and also for processing keyboard presses by the users. The head worn eye tracker, used for both gaze determination and pupil diameter measurement, was an iViewETG manufactured by SMI Corporation, with a recording rates from 50 Hz to 1250 Hz and resolution of 1280×1024.

4.1.3 Task and Procedure

Subjects were tasked with determining which of two CG images, displayed on a mobile device, appeared with the best visual clarity or amount of focus. Users were sat facing a whiteboard 1.0m away and instructed to fixate and maintain their gaze upon a physical character 'A' written on the board. Each subject was then fit with a head-mounted gaze tracker, and instructed to stabilize their head upon a rest and examine images displayed on a mobile Android device placed 0.5m in front of them, but positioned so as not to overlap their view of the physical 'A'. The complete experiment environment is provided in Figure 11 (a), with the subject's view depicted in Figure 11 (b).

During each trial, two different levels of sharpening, ranging between -0.6% and 0.6% in 0.2% intervals, were selected and applied to an image of a capital 'A' shown on the Android device. Using a keyboard, subjects were able to switch between the two sharpening levels, until they were able to determine which sharpening degree provided the best visual quality, using the



(a) (b)

Figure 11. (a) Environment for Experiment I. (b) User's view.

whiteboard 'A' as a ground truth guide. Subjects responded by pressing the enter key while the prefered sharpening level was visible. All combination of sharpening degrees, except same sharpening degree pair, resulting in 42 total trials per subject. We deter subjects from looking at the Android screen directly by using the head worn gaze tracking and switching off the display screen when the user is no longer focused on the intended target.

The only metric recorded for this test was the participants' pupil size. Subject responses for each trial were not recorded, since the task itself was designed merely to induce a noticeable cognitive load on the user. Also, while an OST HMD was not explicitly used for this task, the focal disparity created from simultaneously viewing the whiteboard and mobile device adequately mimics that which would be present in a similar AR task.

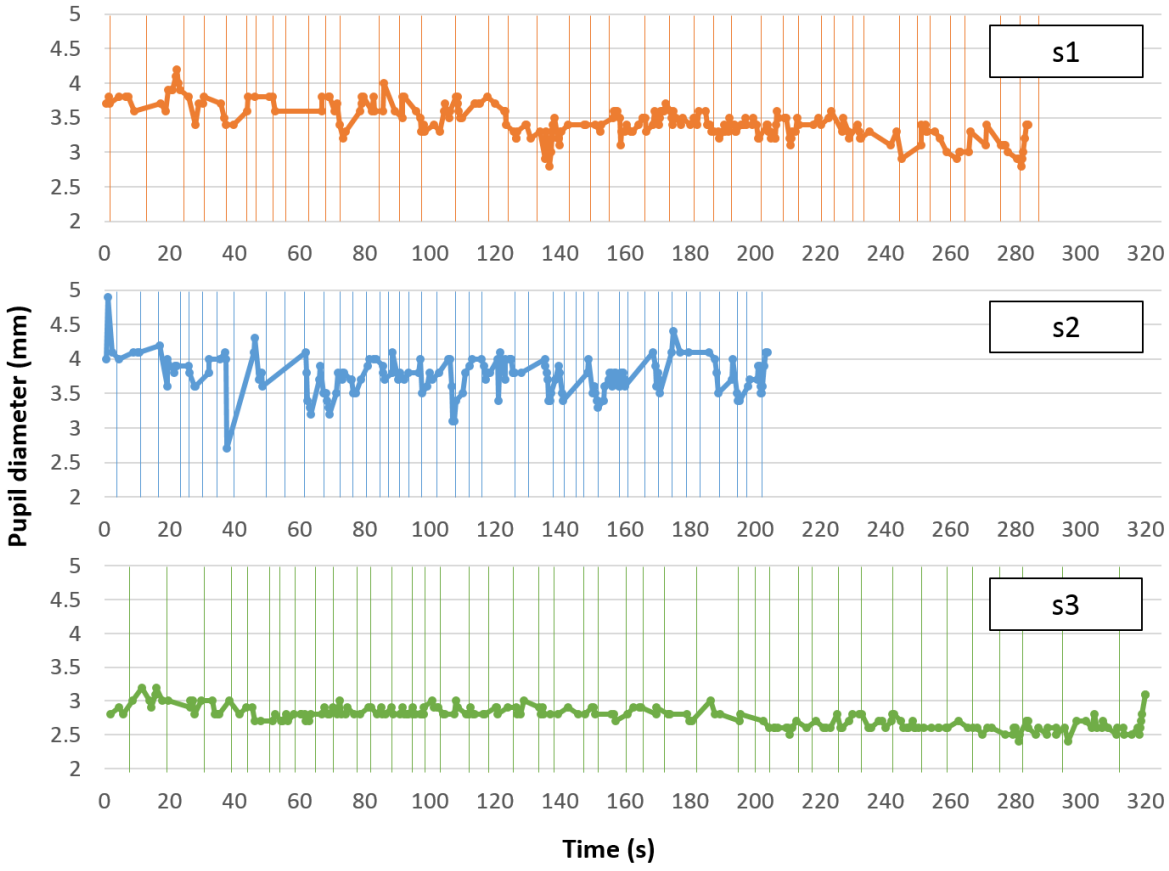


Figure 12. Experiment I results: The horizontal axis represents time and the vertical axis represents pupil diameter. For each subject, we can observe a significant change in pupil diameter during the image comparison task.

Table 1. Experiment I: Pupil size over time by subject

	s1	s2	s3
Median (mm)	3.4	3.8	2.8
Maximum (mm)	4.2	4.9	3.2
Minimum (mm)	2.8	2.7	2.4
Ratio of trials with negligible change	14/42	16/42	29/42

4.1.4 Results

Figure 12 provides each participant's pupil diameter recorded at each trial. The times indicate when the measurement was taken, at the time of user response. Table 1 shows the representative values of the pupil size for each subject. As expected, our results reveal that pupil diameter varied throughout the task and the rate of variation differs between individuals. In order to determine if the variation rate is small enough to be presumed static, we examine the 10% confidence interval for each subject.

The diameter change between trials is expressed as a percentage and the 2% error from the gaze tracker, as stated by the manufacturer, is added to each. The number of trials with less than 10% deviation in value are then determined. Subjects 3's pupil diameter changes were minimal during the task. We can trust 29 out of 42, an 69% confidence, measurements will not vary by 10%. However, subjects 1 and 2 produced 14 and 16 trials deviating by more than 10%,

4.1.5 Discussion

This means that if we used a single value to produce sharpening for image comparison tasks, more than half of the results would be unreliable due to pupil sizes changes. Therefore, these results confirm that in order to effectively utilize SharpView, active pupil size measures must be taken at run time.

4.2 Exp II: Perception of Focus Blur in OST AR

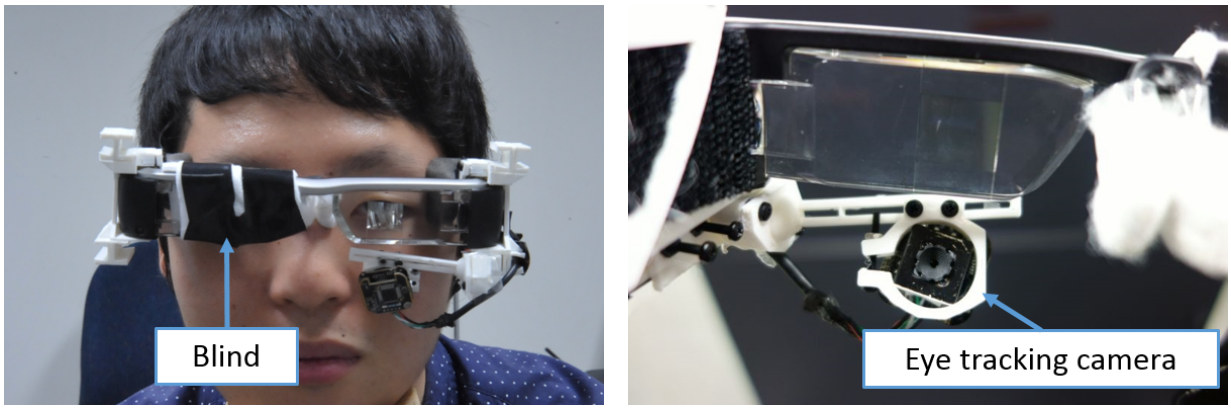
In order to develop an effective improvement strategy for correcting focus blur in OST AR, measures for the amount of blur perceived by users must be obtained. Since focus blur occurs during accommodation rivalry between physical and on-screen content, it is essential to define the level of awareness 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 with a two fold objective. First, we desire to quantify user perception of focus blur within an OST AR system. Secondly, if corrective methods, such as SharpView, are to be viable, then user sensitivity to focus improvements must also be measured. We utilize an adjustment task to not only obtain measures for user preference of blur correction but to also compare the accuracy of SharpView correction across users.

4.2.1 Participants

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. 9 of the 13 subjects stated that they possessed normal uncorrected vision. The remaining 4 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.



(a) (b)

Figure 13. 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.

4.2.2 Experimental Platform

An Epson Moverio BT-200 OST HMD is used as the primary display worn by each subject. The Moverio is a binocular display capable of presenting stereo imagery at 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 13 (a). We attached a Microsoft LifeCam HD-6000 webcam, set to a resolution of 1280×720 at 30fps, beneath the left eye piece, Figure 13 (b). The camera itself is modified by removing the infrared and adding a visible light filter along with a pair of IR LED's. 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. Accuracy of the gaze tracking was confirmed to within 1°. 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.

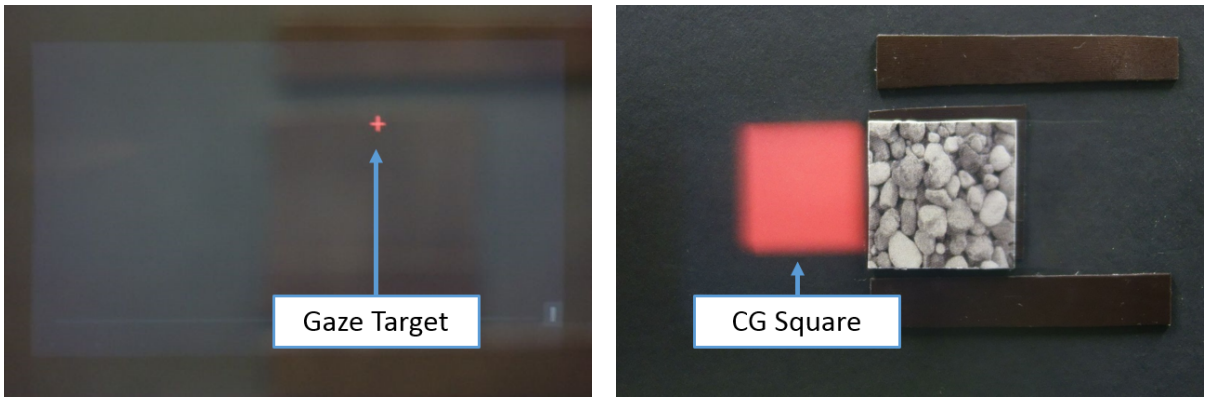
In order to induce and maintain the presence of focus blur during our experimental task, subjects must be forced to fixate their gaze onto physical objects in the world and also be prevented from shifting their focus onto on-screen content. We incorporate an eye tracking process into our OST HMD system 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, are presented one at a time to the subject, Figure 14 (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 PupilLabs 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 14 (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.



(a) (b)

Figure 14. 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.

4.2.3 Task and Procedure

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.

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 square, black and white. asymmetric stone patterns were selected for use during the experiment itself, Figure 15, 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.

We presented the reference images at 4 distance levels, 25cm, 50cm, 100cm, and 500cm, as measured from the subject's eye. Figure 16 shows the location of reference images relative to subjects at the 25cm and 500cm distance levels. As previously noted, the Epson Moverio BT-200 has a fixed focal distance of 7m, or 700cm. Therefore, the corresponding differences in focal distance between the reference and on-screen image are 675cm, 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. A lower focal demand corresponds 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

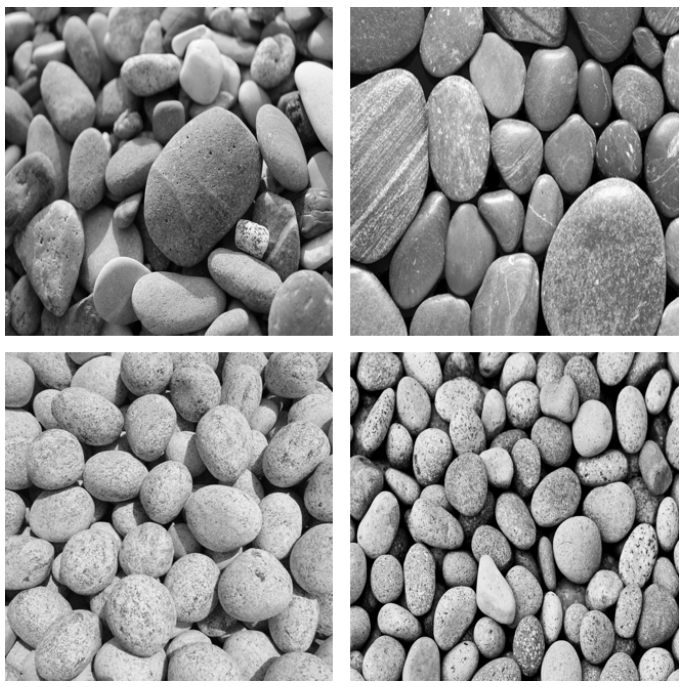
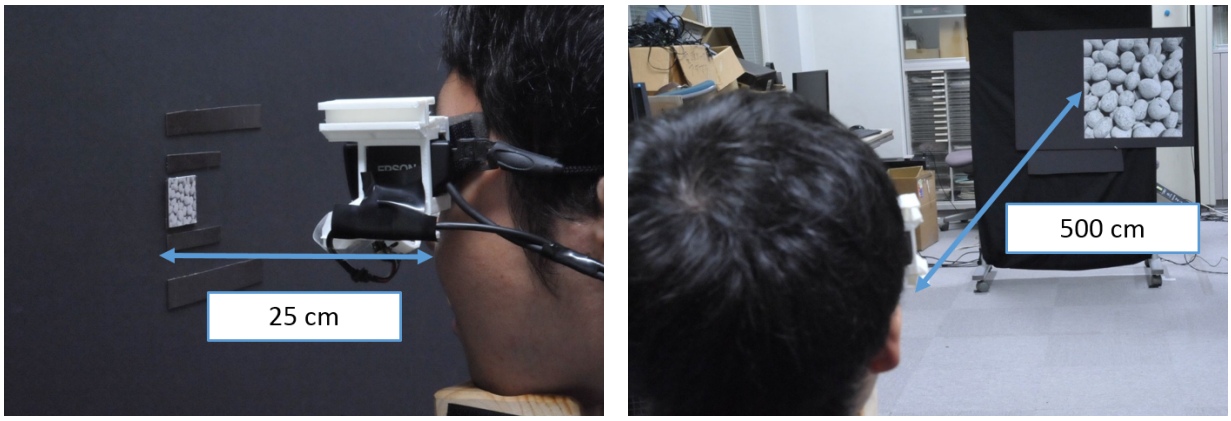


Figure 15. The four black and white stone patters used as both reference and on-screen imagery during the adjustment task.



(a) (b)

Figure 16. Location of subjects relative to reference images placed at 25 cm (a) and 500 cm (b) from the subjects' eyes.

selection of an optimal σ relative to that primary grouping. Each gauge zone reflects a 13 pixel change in σ. Figure 17 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.

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.

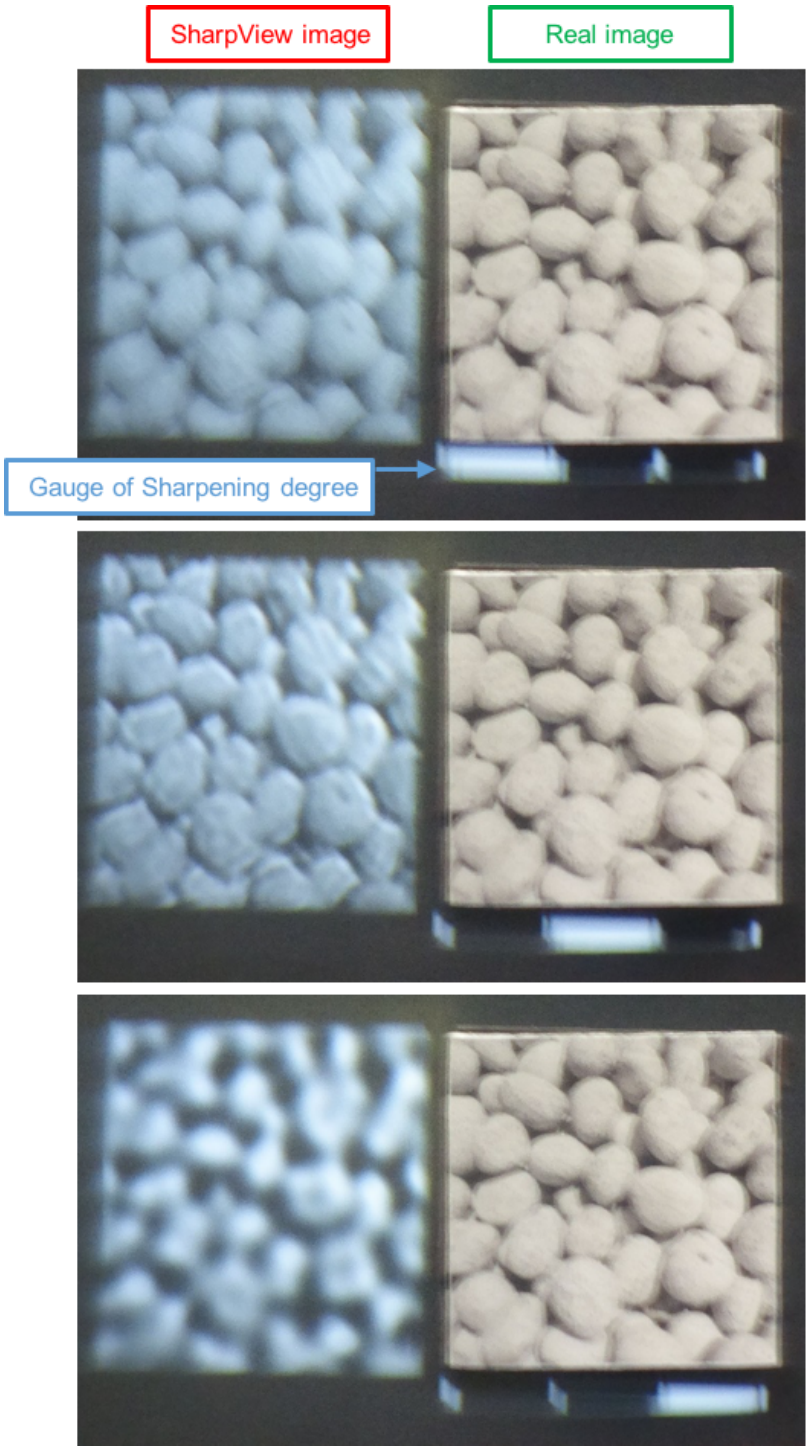


Figure 17. User's view of Experiment II. All subfigures are taken with an SLR camera through our HMD and show: real (right) and virtual image (left). Each virtual image uses a different sharpening degree. Users had to adjust the sharpening degree with a gauge, visible at the bottom of each subfigure: top shows low, middle shows medium, bottom shows high.

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.

4.2.4 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, 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 18 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 18. 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 19 provides a per-subject σ comparison, while Figure 20 shows the comparison with all data combined. In all plots, perfect modeling corresponds to σ pairings lying 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 20, 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 2.

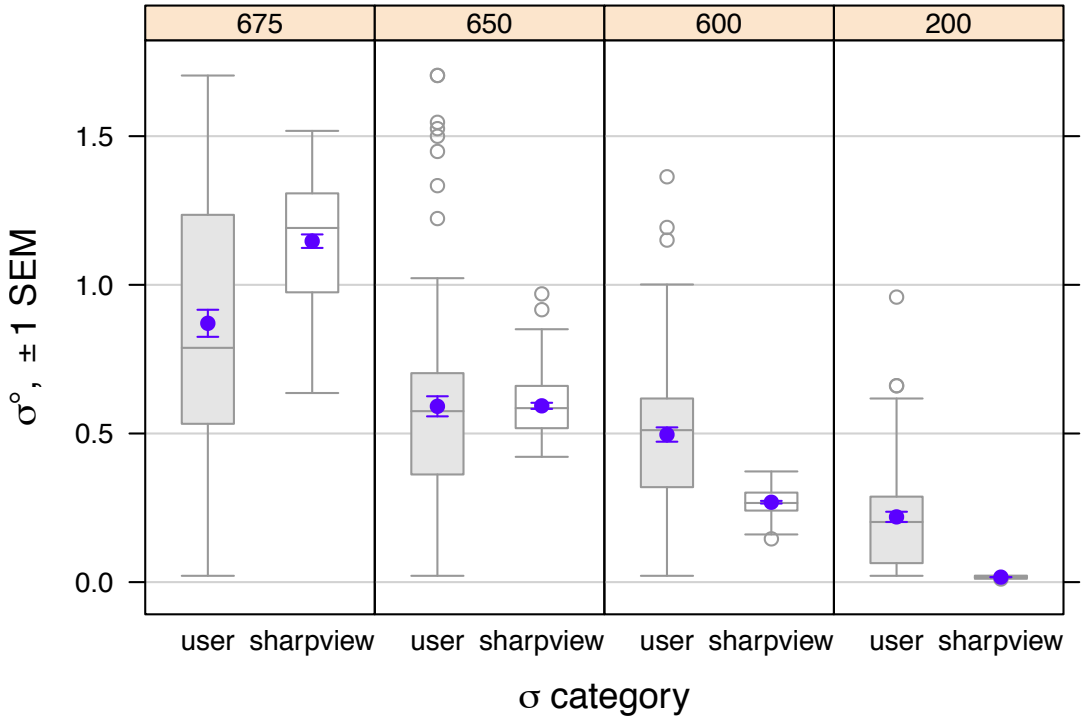


Figure 18. User-σ and SharpView-σ, expressed as visual angle, grouped by focal disparity. The mean and standard error of the mean (SEM) are shown in blue.

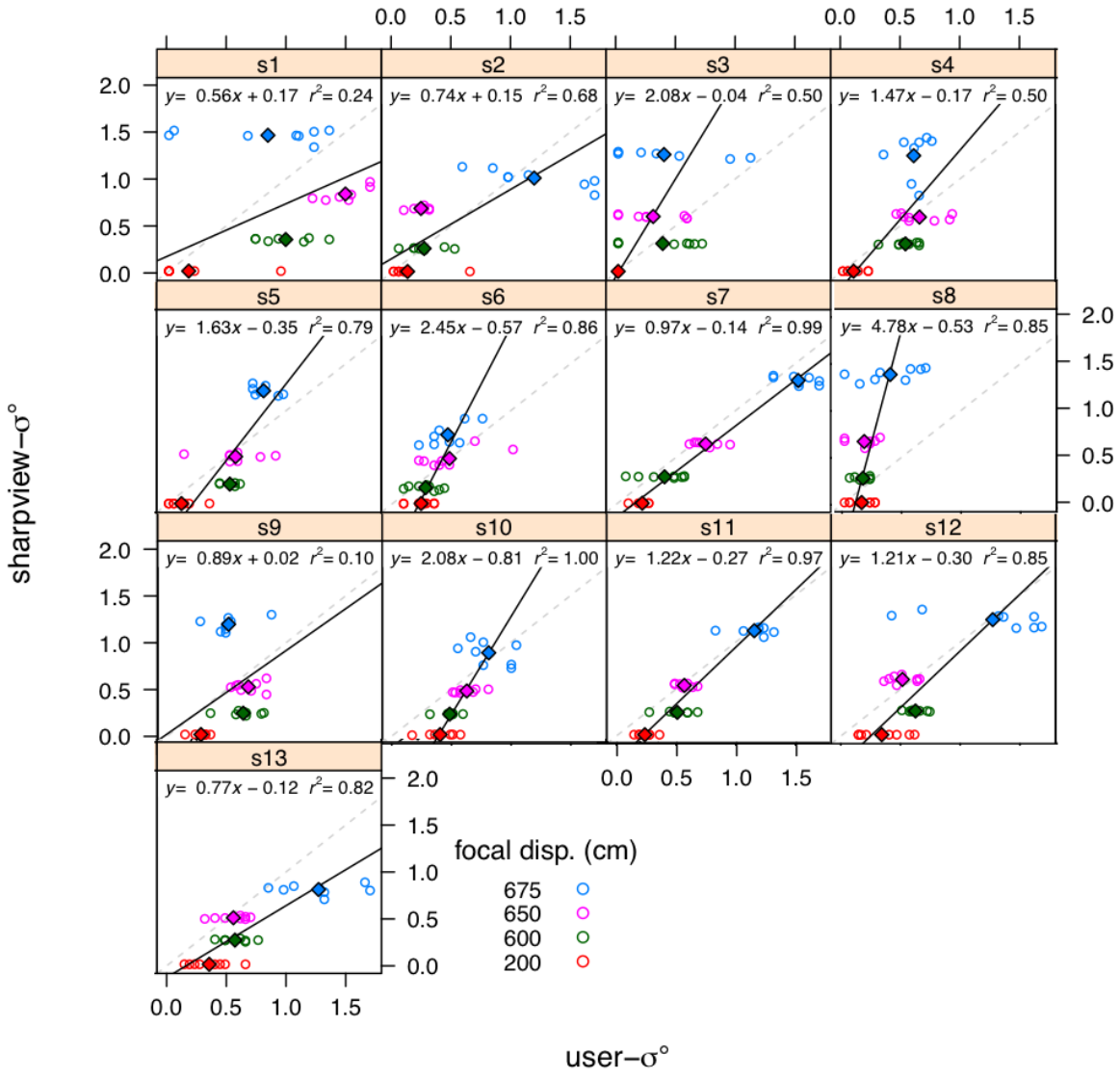


Figure 19. Experiment II results: comparison between user-selected sigma and the sigma predicted by our SharpView algorithm; both values are expressed as degrees of visual angle. The data points are color-coded according to the distance to the real image. The regression line is fit to the centroid (filled diamonds) of each distance.

Table 2. 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

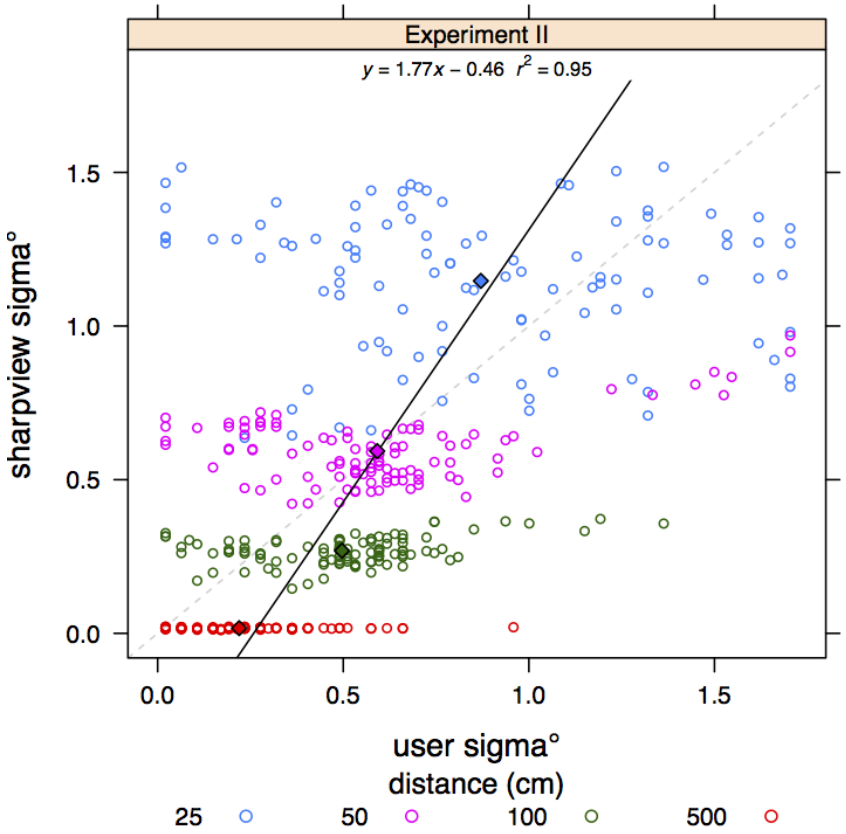


Figure 20. Data from Figure 19, averaged over all users.

4.2.5 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 18, 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 18), 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 19, 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. Since we did not conduct any formal optometrical evaluations of the subjects, we are not able to 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.

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.

4.3 Exp III: Performance Impact of Focus Blur

Even though we have confirmed that sharpening strategies improve user perception of focus blur, we have yet to verify the levels at which blur begins to degrade the usefulness of AR content. We designed and implemented a final user study evaluation to test the hypothesis that high levels of focus blur limit a user's ability to effectively utilize AR information. We chose a simple matching task for our experimental platform, requiring participants to simultaneously compare real and AR images at focal disparity levels.

4.3.1 Participants

16 participants (13 male, 3 female), between the ages of 22 and 33 were recruited from the student population of a university. 13 of these 16 subjects participated in Experiment 2 as well. All participants had no known vision impairments, such as color-blindness. 12 of the 16 subjects possessed normal uncorrected vision. The remaining four wore contact lenses.

4.3.2 Experimental Platform

We utilized identical hardware and software to that in Experiment II.

4.3.3 Task and Procedure

We developed our matching task so that users would be required to make decisions based on information only obtainable from blurred AR content. Using a nearly identical setup to that in Experiment II, participants were shown a real and virtual image, then simply asked to respond if the images were identical or different. Before beginning any trial sets, each subject was fit with the same HMD–eye tracker combination and led through the same gaze calibration procedure described in Experiment II. Eye tracking measures were, once again, used to prevent subjects from focusing on AR content, and also to obtain pupil diameter measures.

During each trial of the experiment, users were shown a real reference image and instructed to maintain their focus upon the center of the picture. An AR image was then rendered to the display screen. This AR image could appear in one of two states: without SharpView correction or with SharpView correction. We will refer to the former case simply as the normal view. Using two buttons on a hand held controller, participants indicated that they believed the images were either the same or different. We informed participants to emphasize accuracy over speed. A total of four reference images were used for this task, those are provided in Figure 21. Also Figure 22 shows an example of user's view during the matching task. The images used are black and white containing different, though highly similar, stone patterns. In order to prevent learning effects, trials were only performed at the 500cm, 50cm, and 25cm distances, and only three of the four images were utilized at each distance level. All 9 possible combinations of the three images were shown to the participants twice, in normal view and in SharpView correction, at each distance level, producing a total of 54 trials per subject.

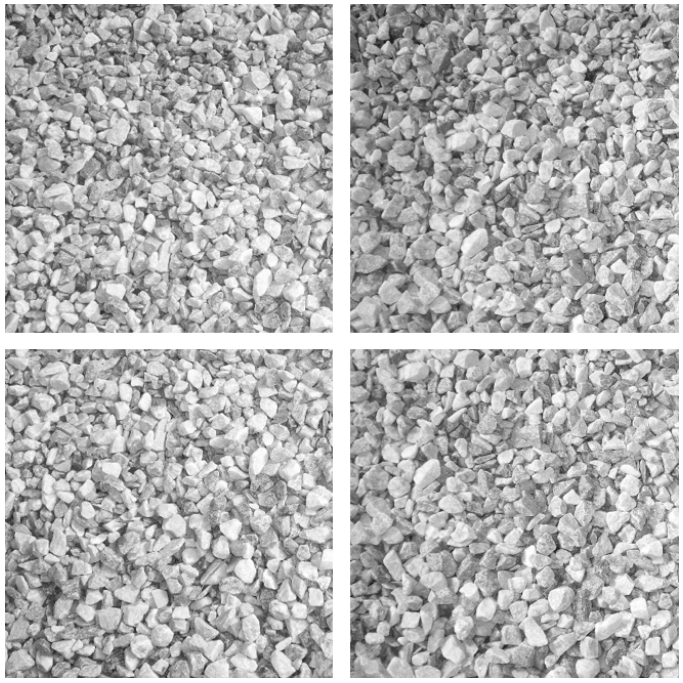
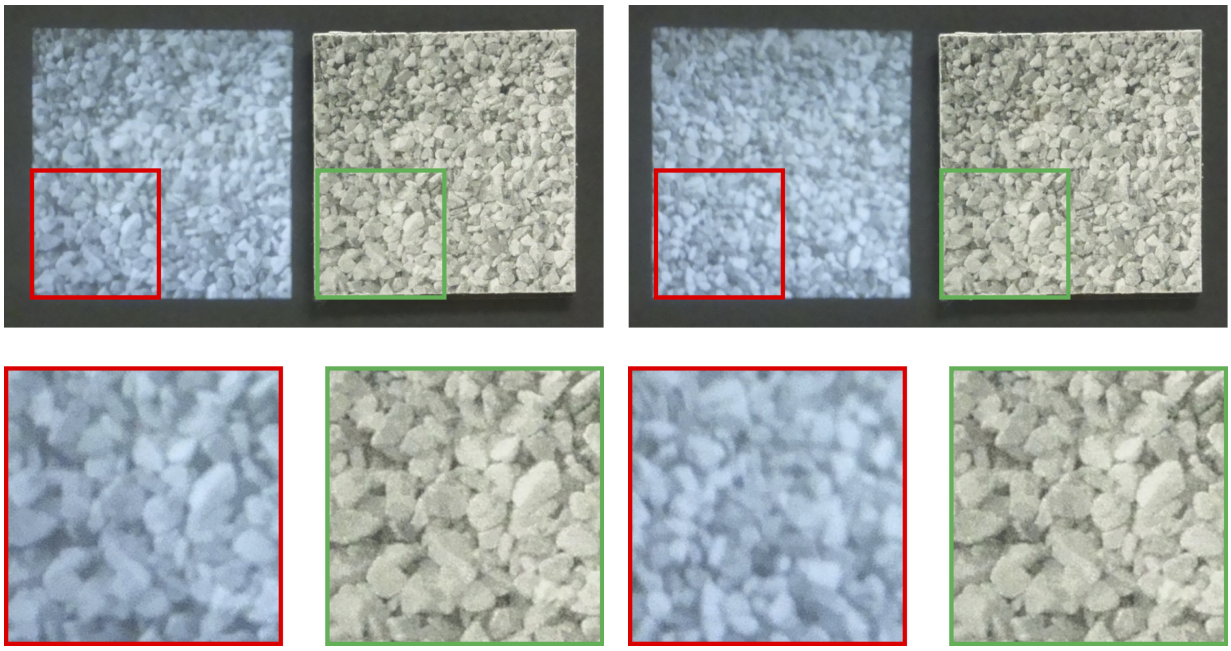


Figure 21. The four black and white stone patters used as both reference and on-screen imagery during the matching task.



(a) (b)

Figure 22. User's view of Experiment Ⅲ. The top pictures are taken through the HMD and show the real image (right), and the virtual image (left). The bottom pictures are the enlarged portion of each image. We used complicated stone images that are difficult to distinguish: (a) The real and virtual images are the same. (b) The images are different.

4.3.4 Results

In order to effectively quantify how focus blur impacted user performance in our matching task, we must first answer the following questions:

Q1: Was observer accuracy better than guessing? Due to the presence of more trials with different images compared to same images, we must first normalize the two groups in order to compare the complete set of contingencies. We created two replications of each trial with same image pairs, which resulted in 48 different and 48 same image pairs. Figure 23 shows the resulting data as a contingency table, providing observer response counts for every combination of the correct answer (same, different), the user's response (same, different), and whether the image pair was shown normally, or with the ShapView method (normal, SharpView). After normalization, 96 total responses are considered. If observers randomly guessed for each image pair, we would expect a count of 48 for each cell (Figure 23). Overall, the panels of the left column test observer skill at determining when the same image appeared in each position. Observers correctly recognized the same image 74% of the time, a rate significantly greater than chance ($\chi^2(1) = 132, p \leq 0.001$). Likewise, the panels of the right column test observer skill at determin-

ing when different images appeared in each position. Observers correctly recognized different images 72% of the time, a rate also significantly greater than chance ($\chi^2(1) = 114, p \leq 0.001$).

Q2: Did SharpView improve user performance? The next question we considered is whether participant performance improved with the SharpView-rendered image pairs over normal image pairs. When observers correctly identified the same image (Figure 23, same/same panel), 50.7% of the responses are SharpView, which does not significantly differ from chance ($\chi^2(1) = 0.085, p = 0.771$). Likewise, when observers correctly identified different images (Figure 23, different/different panel), 51.0% of the responses are SharpView, which also does not significantly differ from chance ($\chi^2(1) = 0.15, p = 0.695$). Therefore, as suggested by Figure 23, performance did not vary according to rendering style.

Q3: Did performance vary significantly across distances? Finally, we considered whether observers had better performance at the control condition distance of 500 cm, versus the test distances of 25 or 50 cm. When observers correctly identified the same image (Figure 23, same/same panel), 34.7% of the responses are at 500 cm, which does not significantly differ from chance ($\chi^2(2) = 0.394, p = 0.821$). Similarly, when observers correctly identified different images (Figure 23, different/different panel), 34.6% of the responses are at 500 cm, which also does not significantly differ from chance ($\chi^2(2) = 0.365, p = 0.833$). Therefore, also as suggested by Figure 23, performance did not vary according to the distance at which the images were rendered.

4.3.5 Discussion

Our initial hypothesis for this experiment was that higher levels of focus blur would negatively impact a user's ability to use AR content. The results of our analysis, however, reveals that participants were able to perform optimally at each distance, regardless of blur. It is possible that the images we used for this task were too easy to identify. We do not believe this to be the case, though. Had the task been too simple, we would expect to see extremely high user accuracy rates close to 100%. Likewise, too difficult a task would produce user accuracy rates of approximately 50%, just guessing. User accuracy rates, overall, were 73%, more or less directly between the two extremes, meaning that improvement is theoretically possible.

We naturally expect users to perform the best at the 500cm reference distance, due to the nearly identical focal levels between the world and the Moverio in this condition. Also, we would similarly expect to see no improvement from SharpView, since sharpness levels are at a minimum for this focal disparity. Since user performance across the remaining two distances match that at 500cm, we can also expect to see no improvement from SharpView at these levels as well. This finding does not indicate a usability issue with SharpView, per se, but rather that there was simply no room for improvement in general.

Our goal for this experiment was to confirm if focus blur reduced AR usability. We can only say, at this point, that focus blur did not inhibit the natural image recognition ability of users. This is actually a very important, and noteworthy, finding. Degradation in usability has, to this point, been the driving mechanism behind the steady development of blur correction methods. However, it might be of more utility to examine tangential applications for these methods. Reducing blur to minimize continual refocusing during OST HMD use, would reduce eye strain and levels of visual fatigue, allowing longer session of use and, therefore, improving the AR experience as a whole.

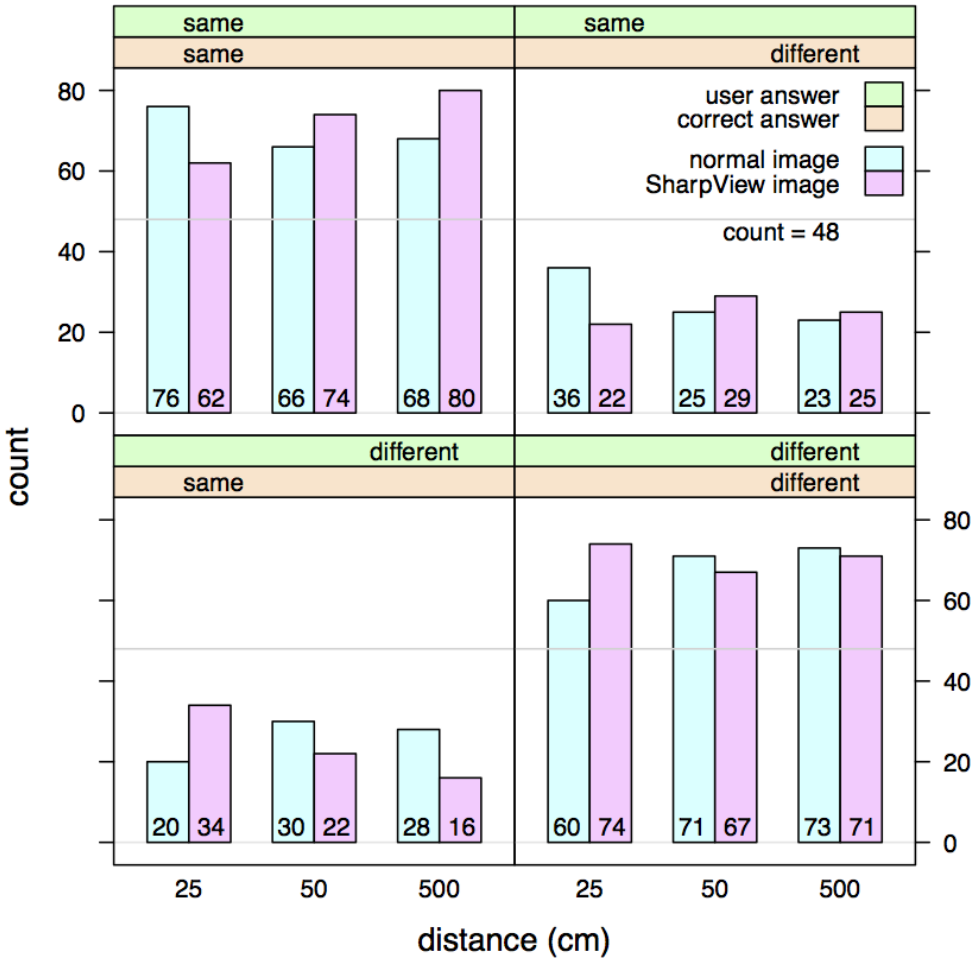


Figure 23. Experiment Ⅲ results: count of participants' responses to image pairs, separated by distance (25, 50, 500 cm), the user's answer (same, different), the correct answer (same, different), and whether the image pair was shown normally, or with the ShapView method (normal, SharpView).

5. Conclusions

This concluding section first summarizes our work and discusses about the experiment results in Section 5.1. Then, possible future directions for enhancing the SharpView system are pointed out in Section 5.2. Finally, I explain about my publications through this research life.

5.1 Overall Discussion

In this work, we have described the focus blur problem inherent to all OST displays. We have also introduced our SharpView algorithm. A simple model, that is a function of user pupil diameter and focal distances to the world and virtual image, designed to generate sharpening filters for correction of blur on OST displays. In addition, we conducted a series of user studies to investigate the impact of focus blur on both the perception of image clarity and AR usability, as well as the utility of our correction method.

Our experiments revealed numerous facts. First, we confirmed that user's pupil diameter changes while engaged in a simple visual AR task from the pilot test, Exp I. It means the dynamic image correction in OST-HMDs such as SharpView system requires the online pupil diameter tracking. Then, the experimental result from Exp Ⅱ have confirmed that users are able to perceive, not only the degradation of image quality due to focus blur, but also the improvement in clarity offered by sharpening techniques. Furthermore, we have found that sharpness requirements are not constant, but rather vary between focal disparities, and even across users themselves. This means that static schemes for blur correction will fail, and that dynamic solutions are required. Also we revealed that the modeled SharpView-σ is most accurate at the 50cm from the user, matching the ANOVA findings. From this finding, we think SharpView system is effective at the close distance tasks such as factories task. We additionally examined the effects of focus blur on an AR dependent matching task at Exp Ⅲ. Though were unable to show any significant effect from focus blur on the accuracy of user responses. However we found that a focus blur does not have a significant impact in the matching task. We need to think about the another practical task for proving the effectiveness of the SharpView system.

5.2 Future Work

There are two parts, the improvement of SharpView technique and the implementation, as future works. The improvement of SharpView technique includes power-up of deconvolution process and change of PSF model. Zhang et al. [36] used the steepest descent method and enabled the generation of high-visibility sharpened image by minimizing the error of the convolution with the original picture. However, it is not suitable for real-time system like SharpView because it takes

much time for the optimization. We need other sharpening techniques to improve the SharpView technique. As the other PSF model, there are various models such as a Zernike function [26] . We need to select the more effective PSF model instead of a gaussian function. Also, in this thesis, we approximated a human eye as a simple lens model and assumed the eye's PSF depends on three parameters: the pupil diameter, the fixing point and the virtual display depth. In practical, the eye's PSF would be related other factors like illusion. We need deeper understanding of the human vision.

The improvement in the implementation part includes the estimation of fixation point and the real-time property. In the case of the prototype, we controlled the fixation point by the indication for users to focus on the specified target. We can imagine a simplified method for estimating fixation depth may leverage known environmental layout information in addition to eye tracking gaze vectors. Through the use of depth sensors and cameras, such as the Microsoft Kinect, a 3D map of the user's environment can be obtained. By determining the intersection of the user's gaze with this model, through application of calibrated eye-tracking, a rough estimation of the necessary focal demand for the user can be determined. Also, a number of studies investigate on-line measurement of vergence angle in stereo systems [6, 8, 34]. In addition, we believe a parallel computation such as GPU improves the real-time property. Currently, it takes about one second to make a sharpened image, of course it also depends on the image size. It requires a high-speed calculation for the real-time SharpView system. Further investigative endeavors will also target a wider selection of OST HMD hardware. Particularly, next generation consumer level offerings, such as the Microsoft HoloLens.

Publication List

Journal Articles

1. Kenneth R Moser, Yuta Itoh, **Kohei Oshima**, J Edward Swan II, Gudrun Klinker and Christian Sandor. Subjective Evaluation of a Semi-Automatic Optical See-Through Head-Mounted Display Calibration Technique, *ACM Transactions on Visualization and Computer Graphics*, Vol.21, No.4, page 491-500, March, 2015.

Peer-Reviewed Conference Publications

1. **Kohei Oshima**, Kenneth R Moser, Damien Constantine Rompapas, J Edward Swan II, Sei Ikeda, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor and Hirokazu Kato. Improved Clarity of Defocussed Content on Optical See-Through Head-Mounted Displays, In *Proceedings of IEEE Symposium on 3D User Interfaces*, 9 pages, Greenville, South Carolina, March, 2016.

2. **Kohei Oshima**, Kenneth R Moser, Damien Constantine Rompapas, J Edward Swan II, Sei Ikeda, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor and Hirokazu Kato. [POSTER] Improved Clarity of Defocussed Content on Optical See-Through Head-Mounted Displays, Poster in *Proceedings of International Conference on Virtual Reality*, 2 pages, Greenville, South Carolina, March, 2016.

Other Conference and Workshop Publications or Presentations

1. **Kohei Oshima**, Damien Constantine Rompapas, Kenneth R Moser, J Edward Swan II, Sei Ikeda, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor and Hirokazu Kato. [DEMO] Improved Legibility of Defocussed Content on Optical See-Through Head-Mounted Displays, Demo in *IEEE International Symposium on Mixed and Augmented Reality*, 2 pages, Fukuoka, Japan, October, 2015.

2. Damien Constantine Rompapas, **Kohei Oshima**, Sei Ikeda, Takafumi Taketomi, Goshiro Yamamoto, Christian Sandor and Hirokazu Kato. [DEMO] EyeAR: Physically-Based Depth of Field through Eye Measurements, Demo in *IEEE International Symposium on Mixed and Augmented Reality*, 2 pages, Fukuoka, Japan, October, 2015.

References

[1] Sylvia Ahern and Jackson Beatty. Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, 205(4412):1289–1292, 1979. 3.2

[2] Kurt Akeley, Simon J Watt, Ahna Reza Girshick, and Martin S Banks. A stereo display prototype with multiple focal distances. In *ACM transactions on graphics (TOG)*, pages 804–813. ACM, 2004. 2.1

[3] Miguel Alonso Jr and Armando B. 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. 2.2

[4] Ronald T Azuma et al. A survey of augmented reality. *Presence*, 6(4):355–385, 1997. 1.1

[5] Michael S Brown, Peng Song, and Tat-Jen Cham. Image pre-conditioning for out-of-focus projector blur. In *Computer Vision and Pattern Recognition*, volume 2, pages 1956–1963, 2006. 2.2

[6] Brian C Daugherty, Andrew T Duchowski, Donald H House, and Celambarasan Ramasamy. Measuring vergence over stereoscopic video with a remote eye tracker. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 97–100. ACM, 2010. 5.2

[7] Alexander Duane. Studies in monocular and binocular accommodation with their clinical applications. *American Journal of Ophthalmology*, 5(11):865–877, 1922. 1.1

[8] Andrew T Duchowski, Donald H House, Jordan Gestring, Robert Congdon, Lech Świrski, Neil A Dodgson, Krzysztof Krejtz, and Izabela 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. 5.2

[9] Rafael Ceferino Gonzalez and Richard E Woods. *Instructor's Manual for Digital Image Processing*. Addison-Wesley, 1992. 3

[10] Eckhard H Hess and James M Polt. Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192, 1964. 3.2

[11] Fu-Chung Huang, Kevin Chen, and Gosrdon 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. 2.1

[12] Fu-Chung Huang, Douglas Lanman, Brian A Barsky, and Ramesh Raskar. Correcting for optical aberrations using multilayer displays. *ACM transactions on graphics (TOG)*, 31(6):185, 2012. 2.2

[13] Fu-Chung Huang, Gordon Wetzstein, Brian A Barsky, and Ramesh Raskar. Eyeglasses-free display: towards correcting visual aberrations with computational light field displays. *ACM transactions on graphics (TOG)*, 33(4):59, 2014. 2.1

[14] Yuta Itoh and Gudrun Klinker. Light-field correction for spatial calibration of optical see-through head-mounted displays. *Transactions on Visualization and Computer Graphics*, 21(4):471–480, 2015. 1.2

[15] Yuta Itoh and Gudrun Klinker. Vision enhancement: defocus correction via optical see-through head-mounted displays. In *Proceedings of Augmented Human International Conference*, pages 1–8. ACM, 2015. 1.2, 3, 2.2

[16] Daisuke Iwai, Shoichiro Mihara, and Kosuke Sato. Extended depth-of-field projector by fast focal sweep projection. *Transactions on Visualization and Computer Graphics*, 21(4):462–470, 2015. 2.2

[17] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. An interactive 360 light field display. In *Special Interest Group on Computer GRAPHics*, page 13. ACM, 2007. 2.1, 7

[18] R Clark Jones. On the point and line spread functions of photographic images. *JOSA*, 48(12):934–937, 1958. 1.3.1

[19] Daniel Kahneman and Jackson Beatty. Pupil diameter and load on memory. *Science*, 154(3756):1583–1585, 1966. 2, 4.1

[20] Douglas Lanman and David Luebke. Near-eye light field displays. *ACM transactions on graphics (TOG)*, 32(6):220, 2013. 2.1, 7

[21] Sheng Liu, Dewen Cheng, and Hong Hua. An optical see-through head mounted display with addressable focal planes. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 33–142. IEEE, 2008. 2.1

[22] Gordon D Love, David M Hoffman, Philip JW Hands, James Gao, Andrew K Kirby, and Martin S Banks. High-speed switchable lens enables the development of a volumetric stereoscopic display. *OSA Optics Express*, 17(18):15716–15725, 2009. 2.1

[23] Kevin J MacKenzie, David M Hoffman, and Simon J Watt. Accomodation to multiple focal plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision*, 10(8):22, 2010. 2.1

[24] Andrew Maimone and Henry Fuchs. Computational augmented reality eyeglasses. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 29–38. IEEE, 2013. 2.1

[25] Carlos Montalto, Ignacio Garcia-Dorado, Daniel Aliaga, Manuel M Oliveira, and Feng Meng. A total variation approach for customizing imagery to improve visual acuity. *ACM transactions on graphics (TOG)*, 34(3):28, 2015. 1.2, 2

[26] Carlos* Montalto, Ignacio* Garcia-Dorado, Daniel Aliaga, Manuel M. Oliveira, and Feng Meng. A total variation approach for customizing imagery to improve visual acuity. *ACM transactions on graphics (TOG)*, 34(3):28:1–28:16, 2015. 2.2, 5.2

[27] Bunyo Okumura, Masayuki Kanbara, and Naokazu Yokoya. Augmented reality based on estimation of defocusing and motion blurring from captured images. In *IEEE International Symposium on Mixed and Augmented Reality*, pages 219–225. IEEE, 2006. 3

[28] Vitor F Pamplona, Manuel M Oliveira, Daniel G Aliaga, and Ramesh Raskar. Tailored displays to compensate for visual aberrations. *ACM transactions on graphics (TOG)*, 31(4):81, 2012. 2.1, 2.2

[29] 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. 1.3.1, 3

[30] Theodore Schaefer Jr, J Brinton Ferguson, Joseph A Klein, and Edward B Rawson. Pupillary responses during mental activities. *Psychonomic Science*, 12(4):137–138, 1968. 3.2

[31] Brian T Schowengerdt and Eric 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. 2.1

[32] HM Simpson and Shirley M Hale. Pupillary changes duaring a decision-making task. *Perceptual and Motor Skills*, 29(2):495–498, 1969. 3.2

[33] DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010. 1.1

[34] Rui Wang, Brandon Pelfrey, Andrew T Duchowski, Donald H House, et al. Online gaze disparity via bioncular eye tracking on stereoscopic displays. In *3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 184–191. IEEE, 2012. 5.2

[35] John I Yellott and John W Yellott. Correcting spurious resolution in defocused images. In *Electronic Imaging*, pages 64920O–64920O. International Society for Optics and Photonics, 2007. 2.2

[36] Li Zhang and Shree Nayar. Projection defocus analysis for scene capture and image display. In *ACM transactions on graphics (TOG)*, pages 907–915. ACM, 2006. 2.2, 5.2