Vision Enhancement: Defocus Correction via Optical See-Through Head-Mounted Displays

Yuta Itoh
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany
itoh@in.tum.de

Gudrun Klinker
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany
klinker@in.tum.de

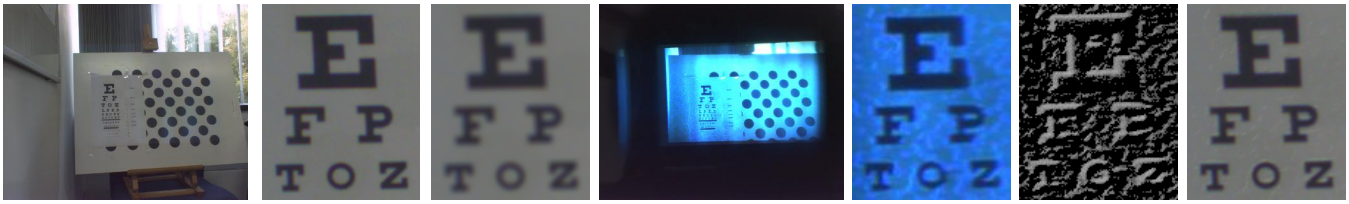

Figure 1. An overview of the vision enhancement for defocus correction via OST-HMDs. (a) An ideal user-perspective image, (b) A zoomed part of (a) capturing an eye chart, (c) A degraded user view by applying a synthetic optical blur to (b), (d) A user-perspective image with a black background and an HMD screen on which a world camera image is rendered, (e) A zoomed part of (d), (f) A compensation image computed from (e), and (g) An enhanced image created by combining (c) on (f). Note that the brightness and contrast of (f) are modified +50/-50% for the presentation purpose.

ABSTRACT

Vision is our primary, essential sense to perceive the real world. Human beings have been keen to enhance the limit of the eye function by inventing various vision devices such as corrective glasses, sunglasses, telescopes, and night vision goggles. Recently, Optical See-Through Head-Mounted Displays (OST-HMD) have penetrated in the commercial market. While the traditional devices have improved our vision by altering or replacing it, OST-HMDs can augment and mediate it. We believe that future OST-HMDs will dramatically improve our vision capability, combined with wearable sensing systems including image sensors.

For taking a step toward this future, this paper investigates Vision Enhancement (VE) techniques via OST-HMDs. We aim at correcting optical defects of human eyes, especially defocus, by overlaying a compensation image on the user's actual view so that the filter cancels the aberration. Our contributions are threefold. Firstly, we formulate our method by taking the optical relationships between OST-HMD and human eye into consideration. Secondly, we demonstrate the method in proof-of-concept experiments. Lastly and most importantly, we provide a thorough analysis of the results including limitations of the current system, potential research issues necessary for realizing practical VE systems, and possible solutions for the issues for future research.

Author Keywords

Augmented Reality; Optical See-through; Head-mounted Displays; Near-eye Displays; Vision Enhancement; Image Deblurring

ACM Classification Keywords

H.5.1.b Information Interfaces and Representation (e.g., HCI): Multimedia Information Systems – Artificial, augmented, and virtual realities; I.4.9 Image Processing and Computer Vision: Applications

INTRODUCTION

Vision is our primary means to perceive the physical world. We can achieve various tasks through vision, combined with higher-order brain functions. We have developed numerous Vision Enhancement (VE) devices for supporting and/or boosting the capability of our vision in various aspects, e.g., dynamic range (sunglasses, night vision goggles), focal length (corrective glasses), field of view (telescopes, microscopes, low-vision glasses [23, 24]), spectrum range (thermal goggles), and exposure (stroboscopic/shutter glasses [16]).

VE devices can be categorized into two different types based on their principles: *direct* and *indirect*. Direct devices like corrective glasses and sunglasses consist of optical elements that directly make use of phenomena of optical physics such as refraction and transmittance. The capabilities of direct devices are limited by those of their optical elements and the human eyes. On the other hand, indirect devices such as night-/thermal-vision goggles use external vision sensors to obtain *super-vision* which is hard or impossible to obtain by direct devices, and users see post-processed images on a display.

Indirect devices can substitute direct devices given an appropriate sensor which reproduces the same vision as the optical components of the direct devices provide. For example, if we

This manuscript is a pre-print version of Augmented Human 2015 submission.

capture an image by a camera with a huge zoom lens, and display it on a VST-HMD, we get a virtual telescope [19]. Furthermore, indirect devices can benefit from the power of computational photography by post-processing raw sensor data as professional astrometric telescope systems do.

However, as a wearable vision system, indirect devices have an essential limitation: they dispose and intercept the user's *direct* view by occluding the real world from the user's eye. Direct devices do not have this limitation, yet they are inflexible in applying different VE effects since each optical effect requires different optical components.

Recently, Optical See-Through HMDs (OST-HMDs) have penetrated the market as typified by Google Glass, EPSON BT-200, Microsoft HoloLens, and so on. The displays integrate digital images into the user's view while keeping the real scene visible through (semi-)transparent optical combiners. The OST-HMDs are potentially capable of incorporating the benefits of both types of VE devices. Along with the recent developments of mobile sensing devices, we believe that the future OST-HMD system will substitute many direct VE devices.

This ultimate goal is, unfortunately, far from what we currently have. The limitations of the current display hardware and computer vision/graphics technologies impede the realization of practical VE devices. Yet, our community has developed essential technologies such that, if combined altogether, they could establish a more practical system. However, few works tackle VE problems along the context of OST-HMD systems.

This paper conducts a fundamental study of VE in defocus correction via OST-HMDs to improve human vision. Our idea is to add a visual stimulus to a user's natural vision such that the user regains visual acuity. The stimulus is given as a compensation image displayed on an OST-HMD. The section below summarizes the contributions of our paper.

Contributions

Our main contributions include the following: (1) We provide a theoretical formulation of VE for the defocus correction via OST-HMDs. (2) We demonstrate that our VE formulation can improve the defocus effect through conceptual experiments. (3) We provide a thorough analysis of the current VE setup including limitations and possible research directions toward the realization of a practical VE system.

RELATED WORK

The VE concept has close relations to the following areas.

Projector-Camera Systems

An OST-HMD is a projector which displays an image mid air; an eye is a camera which captures the image. Thus, they form a Projector-Camera System (PCS). PCSs have two topics related to VE: improving the appearance of the projected image, and modifying that of the surface on which the projected image appears. The former includes: a defocus-blur correction [20], a temporal super-resolution system [26], and a color-correction system [7]. The latter includes: contrast and resolution enhancement of ePapers [8], a dynamic-range enhancement technique [6], and a color-enhancement system for visually impaired people [1].

The OST-HMD system differs from conventional PCSs in several ways: the image plane floats in mid air rather than being projected on physical surfaces; the spatial relationship between the eye and the display plane is dynamic; and the user-perspective image, i.e., what a user exactly sees, is not easy to obtain. The floating image plane makes a calibration of a system nuisance, and the difficulty of the user-view acquisition even devastates the image rendering process due to missing feedback from the eye.

Computational Photography for Aberration Corrections

Several computational displays account for visual aberrations of human eyes [22, 13]. Such systems consist of multilayer/light-field displays and render pre-filtered images designed to cancel estimated aberrations. These displays require the user's eye position and the eye aberration profile, in practice, as a point spread function (PSF).

Low-Vision Devices for Visual Impairments

Apart from the pure computer vision/graphics fields, researchers in clinical fields have developed various low-vision devices to help visually impaired people [23, 24]. These *direct* devices are composed of several optical lenses to create tailored vision to compensating the patients' degraded vision.

Recently, Huang and Peli [14] developed a VE system which provides a patient with an edge-enhanced image for contrast enhancement. They create a user-perspective image by using a planar OST-HMD screen model and a display pose measured manually.

Vision Enhancement in Augmented Reality

In Augmented Reality (AR), there are works on overlaying images on users' views to modify/enhance their vision. Such work includes: an AR-Xray system which integrates occluded scenes into a user's view via a VST-HMD system [3], and an AR microscope that provides depth-dependent image augmentations so that viewers can grasp the focal depth of microscope imagery effortlessly [11].

METHOD

VE systems have three key computation steps: To transform sensor images so that the displayed image will coincide with the user's perspective (Step A), to estimate a user's vision, e.g., a PSF and the color sensitivity of a user's eye (Step B), and to preprocess the image to be displayed (or to modulate light from the display) so that the combined stimulus of the image and user's vision creates the desired optical effect (Step C). We refer these steps in the discussion section to associate topics to the steps. We first formulate our method based on ideal assumptions. We then relax it with practical assumptions. Through out this paper, we assume that cameras are the pinhole cameras and images are undistorted. $\mathbf{I}$ represents a 2D image. We attach $\star$ to variables for representing that their are true values.

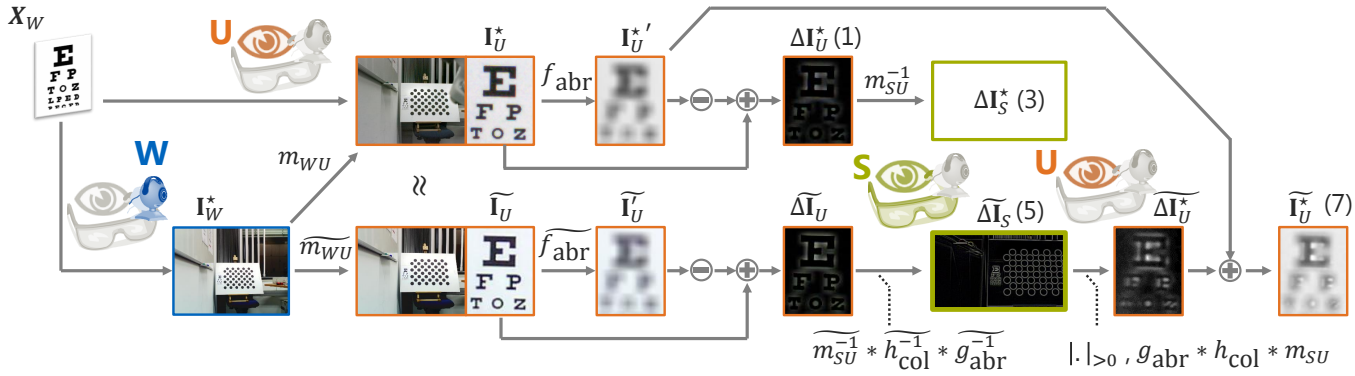

Figure 2. Schematic diagram of a general formulation of the vision enhancement for defocus correction. See the method section for more details.

Formulation

Step A is analogous to the rendering process in graphics engines: we need to bake a 2D image out of a 3D scene from the view point of a user-perspective. The intricacy of our setups is that neither we can align a camera to a user's eye position perfectly nor the camera sensor can replicate the eye sensor exactly.

We start with a User-Perspective (UP) camera U, which represents the human eye, placed in the world (Fig. 2). The camera sees a scene structure $\mathrm{X_W}$ in the world coordinate system. Typically, $\mathrm{X_W}$ is modeled as a set of 3D surfaces with material information. As the result, the camera records a user-perspective image $\mathbf{I}^\star_{\mathrm{U}}$. $\mathbf{I}^\star_{\mathrm{U}}$ is a ground-truth image that an emmetropic eye would see, whereas an ametropic eye suffers from an optical aberration $f_{\mathrm{abr}}(\cdot)$ and sees a degraded view $\mathbf{I}^{\star\prime}_{\mathrm{U}} := f_{\mathrm{abr}}(\mathbf{I}^\star_{\mathrm{U}})$. Typical aberrations include myopia, hyperopia, and astigma. If we can present a visual stimulus:

$$\Delta\mathbf{I}^\star_{\mathrm{U}} := \mathbf{I}^\star_{\mathrm{U}} - \mathbf{I}^{\star\prime}_{\mathrm{U}} \tag{1}$$

to the user's view, the user would regain the regular view as $\mathbf{I}^\star_{\mathrm{U}} = \Delta\mathbf{I}^\star_{\mathrm{U}} + \mathbf{I}^{\star\prime}_{\mathrm{U}}$. We call $\Delta\mathbf{I}^\star_{\mathrm{U}}$ as a compensation image.

We estimate $\Delta\mathbf{I}^\star_{\mathrm{U}}$ from an image taken by a world camera W attached to an OST-HMD. Without loss of generality, we treat W as the origin of the world coordinate system. Similar to the UP camera, W also sees $\mathrm{X_W}$ and captures a world-view image $\mathbf{I}^\star_{\mathrm{W}}$. Note that $\mathbf{I}^\star_{\mathrm{W}}$ is different from $\mathbf{I}^\star_{\mathrm{U}}$ since the two cameras have different extrinsic and intrinsic parameters. Therefore, we need to warp $\mathbf{I}^\star_{\mathrm{W}}$ to $\mathbf{I}^\star_{\mathrm{U}}$ by using these parameters. Let $\mathbf{P}_{\mathrm{WU}}$ and $\mathbf{P}_{\mathrm{WW}}$ be world-to-image projection matrices of U and W respectively ($\mathbf{P}_{\mathrm{WW}}$ only has the intrinsic part since W is at the origin). Given $\mathbf{P}_{\mathrm{WU}}$, $\mathbf{P}_{\mathrm{WW}}$ and the 3D structure of $\mathrm{X_W}$, we define an image warping function $m(\cdot \mid \cdot, \cdot, \cdot)$ as

$$\mathbf{I}^\star_{\mathrm{U}} = m(\mathbf{I}^\star_{\mathrm{W}} \mid \mathrm{X_W}, \mathbf{P}_{\mathrm{WW}}, \mathbf{P}_{\mathrm{WU}}) =: m_{\mathrm{WU}}(\mathbf{I}^\star_{\mathrm{W}}) \tag{2}$$

where $m_{AB}(\cdot) := m(\cdot \mid \mathrm{X_W}, \mathbf{P}_{\mathrm{W}A}, \mathbf{P}_{\mathrm{W}B})$ for given coordinate systems A and B. If the 3D structure is complex, $m_{\mathrm{WU}}(\cdot)$ can be computed by the epipolar geometry with known depth. If the structure is a plane, like our experiment setups, $m_{\mathrm{WU}}(\cdot)$ becomes an image transformation via a homography matrix.

Although we can now compute $\Delta\mathbf{I}^\star_{\mathrm{U}}$ from $\mathbf{I}^\star_{\mathrm{W}}$, we can not directly provide this stimulus to the user's eye: we have to do so via the OST-HMD screen. This introduces the third camera: a screen camera S. Its position is the same as the UP camera, and its image plane (thus orientation) is defined by the image screen S of the OST-HMD. Note that S can be defined as an off-axis pinhole camera under the assumption that the screen, which is floating in mid air, is planar. If S is a real camera, it has projection matrix $\mathbf{P}_{\mathrm{WS}}$ and records an image $\mathbf{I}^\star_{\mathrm{S}}$. When properly transformed, $\mathbf{I}^\star_{\mathrm{S}}$ exactly matches to a part of $\mathbf{I}^\star_{\mathrm{U}}$ since their cameras share the same view point. We now obtain the ideal screen image to be displayed on the OST-HMD:

$$\Delta\mathbf{I}^\star_{\mathrm{S}} := m^{-1}_{\mathrm{SU}}(\Delta\mathbf{I}^\star_{\mathrm{U}}) = m^{-1}_{\mathrm{SU}}(m_{\mathrm{WU}}(\mathbf{I}^\star_{\mathrm{W}}) - f_{\mathrm{abr}}(m_{\mathrm{WU}}(\mathbf{I}^\star_{\mathrm{W}}))) \tag{3}$$

where $m^{-1}_{\mathrm{SU}}(\cdot)$, the inverse mapping of another warping function $m_{\mathrm{SU}}(\cdot)$, warps $\mathbf{I}^\star_{\mathrm{S}}$ to $\mathbf{I}^\star_{\mathrm{U}}$. Finally, displaying $\Delta\mathbf{I}^\star_{\mathrm{S}}$ on the screen gives the stimulus:

$$\mathbf{I}^{\star\prime}_{\mathrm{U}} + m_{\mathrm{SU}}(\Delta\mathbf{I}^\star_{\mathrm{S}}) = \mathbf{I}^{\star\prime}_{\mathrm{U}} + \Delta\mathbf{I}^\star_{\mathrm{U}} = \mathbf{I}^\star_{\mathrm{U}}. \tag{4}$$

In practice, we only have the estimates of the functions in the above formulation: $\widetilde{m_{\mathrm{WS}}}(\cdot)$, $\widetilde{m^{-1}_{\mathrm{SU}}}(\cdot)$, $\widetilde{f_{\mathrm{abr}}}(\cdot)$, and $\widetilde{m_{\mathrm{WS}}}(\cdot)$. Thus the screen image that we actually display becomes

$$\widetilde{\Delta\mathbf{I}_{\mathrm{S}}} := \widetilde{m^{-1}_{\mathrm{SU}}}(\widetilde{\Delta\mathbf{I}_{\mathrm{U}}}), \quad \widetilde{\Delta\mathbf{I}_{\mathrm{U}}} := \widetilde{\mathbf{I}_{\mathrm{U}}} - \widetilde{\mathbf{I}'_{\mathrm{U}}}, \tag{5}$$

$$\widetilde{\mathbf{I}_{\mathrm{U}}} := \widetilde{m_{\mathrm{WU}}}(\mathbf{I}^\star_{\mathrm{W}}), \quad \widetilde{\mathbf{I}'_{\mathrm{U}}} := \widetilde{f_{\mathrm{abr}}}(\widetilde{\mathbf{I}_{\mathrm{U}}}). \tag{6}$$

By displaying $\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}$ on the HMD screen, the user perceives

$$\widetilde{\mathbf{I}^\star_{\mathrm{U}}} := \mathbf{I}^{\star\prime}_{\mathrm{U}} + \widetilde{\Delta\mathbf{I}^\star_{\mathrm{U}}} \approx \mathbf{I}^\star_{\mathrm{U}}, \tag{7}$$

where $\widetilde{\Delta\mathbf{I}^\star_{\mathrm{U}}} := m_{\mathrm{SU}}(\widetilde{\Delta\mathbf{I}_{\mathrm{S}}})$. Here we applied the true warping function $m_{\mathrm{SU}}(\cdot)$ since this is a physical process.

Note that ordinary displays cannot handle *negative* values. We can only display $|\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}|_{>0}$ instead of $\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}$ where $|\cdot|_{>0}$ is a function which sets negative values of the image pixels to zero. We refer the VE with the former as Enhanced + (plus) and the latter Enhanced +- (plus-minus).

We have, so far, ignored some effects related to the OST-HMD optics: a color distortion by the virtual screen and an optical aberration effect against the displayed image. We further integrate these effects into the above formulation.

Color Distortion of Virtual Screen Image

Eye-HMD system has a color distortion stems from several conversions between analog and digital image signals. The

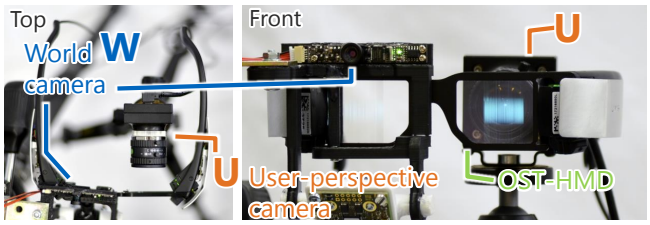

Figure 3. An OST-HMD setup used in our experiments. (Left) A top view of the setup. (Right) Front view. A blue color is displayed on the OST-HMD: Lumus DK32. The UP camera is placed behind the HMD, and the world camera is mounted on the display frame. Both cameras see a printed Snellen Chart (image by Jeff Dahl) as a reference object .

world camera receives light from the world and sends converted digital values to the display. The display emits new light based on the values. Finally, the Up camera receives the display light, and outputs new digital values. In general, this final color differs from the color from the world camera.

Let $h_{\mathrm{col}}(\cdot)$ be such a digital color distortion applied to the final image perceived by the UP camera. Then, the UP camera sees an image as $h_{\mathrm{col}}(\widetilde{\mathbf{I}_{\mathrm{U}}})$ at Eq. 6. We estimate the inverse of the distortion as $\widetilde{h_{\mathrm{col}}^{-1}}(\cdot)$, and redefine the displayed image $\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}$ (Eq. 5) as

$$\widetilde{\Delta\mathbf{I}_{\mathrm{S}}} = \widetilde{m_{\mathrm{SU}}^{-1}}(\widetilde{h_{\mathrm{col}}^{-1}}(\widetilde{\Delta\mathbf{I}_{\mathrm{U}}})) = \widetilde{m_{\mathrm{SU}}^{-1}}(\widetilde{h_{\mathrm{col}}^{-1}}(\widetilde{\mathbf{I}_{\mathrm{U}}} - \widetilde{\mathbf{I}'_{\mathrm{U}}})). \quad (8)$$

The UP camera finally perceives an enhanced view as $\widetilde{\mathbf{I}^{\star}_{\mathrm{U}}} = \mathbf{I}^{\star\prime}_{\mathrm{U}} + h_{\mathrm{col}}(m_{\mathrm{SU}}(\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}))$ instead of Eq. 7.

Optical Aberration of Virtual Screen Image

An HMD screen causes another aberration effect. In general, the image screen created by an OST-HMD does not necessarily appear at the same distance as an object on which a user is currently focusing. This misfocus causes another aberration on the screen image. Let $g_{\mathrm{abr}}(\cdot)$ such an aberration which is different from $f_{\mathrm{abr}}(\cdot)$, following the same derivation in the previous section, we redefine $\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}$ as

$$\widetilde{\Delta\mathbf{I}_{\mathrm{S}}} := \widetilde{m_{\mathrm{SU}}^{-1}}(\widetilde{g_{\mathrm{abr}}^{-1}}(\widetilde{\Delta\mathbf{I}_{\mathrm{U}}})) = \widetilde{m_{\mathrm{SU}}^{-1}}(\widetilde{g_{\mathrm{abr}}^{-1}}(\widetilde{\mathbf{I}_{\mathrm{U}}} - \widetilde{\mathbf{I}'_{\mathrm{U}}})), \quad (9)$$

where $\widetilde{g_{\mathrm{abr}}^{-1}}(\cdot)$ is an estimate of the inverse of $g_{\mathrm{abr}}(\cdot)$. The UP camera finally perceives $\widetilde{\mathbf{I}^{\star}_{\mathrm{U}}} = \mathbf{I}^{\star\prime}_{\mathrm{U}} + g_{\mathrm{abr}}(m_{\mathrm{SU}}(\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}))$.

Generalized Formulation

We merge the color and aberration correction functions. Assuming that the color distortion $h_{\mathrm{col}}(\cdot)$ is pixel-wise, we place the $h_{\mathrm{col}}(\cdot)$ inside the aberration function $g_{\mathrm{abr}}(\cdot)$ as $g_{\mathrm{abr}}(h_{\mathrm{col}}(\cdot))$. Then we obtain the image to be displayed on the screen as follows:

$$\widetilde{\Delta\mathbf{I}_{\mathrm{S}}} = \widetilde{m_{\mathrm{SU}}^{-1}}(\widetilde{h_{\mathrm{col}}^{-1}}(\widetilde{g_{\mathrm{abr}}^{-1}}(\widetilde{\mathbf{I}_{\mathrm{U}}} - \widetilde{\mathbf{I}'_{\mathrm{U}}}))) \quad (10)$$

$$= \widetilde{m_{\mathrm{SU}}^{-1}}(\widetilde{h_{\mathrm{col}}^{-1}}(\widetilde{g_{\mathrm{abr}}^{-1}}(\widetilde{\mathbf{I}_{\mathrm{U}}} - \widetilde{f_{\mathrm{abr}}}(\widetilde{\mathbf{I}_{\mathrm{U}}})))). \quad (11)$$

The perceived view finally becomes:

$$\widetilde{\mathbf{I}^{\star}_{\mathrm{U}}} = \mathbf{I}^{\star\prime}_{\mathrm{U}} + g_{\mathrm{abr}}(h_{\mathrm{col}}(m_{\mathrm{SU}}(\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}))). \quad (12)$$

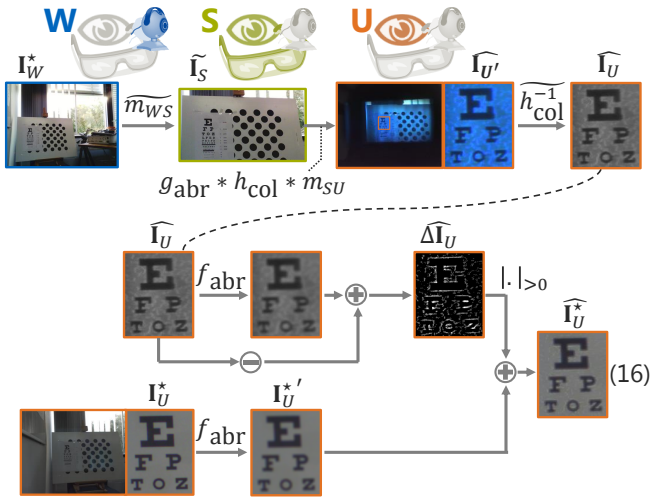

Figure 4. Schematic diagram of vision enhancement flow in the conceptual setup (Experiment 1).

The above formulation does not define how to compute each function. Because their definitions are depending on various assumptions on the scene, the display, and the eye. In the experiment section, we introduce our assumptions and show a simplified, concrete formulation based on the above general formulation.

EXPERIMENTS

We conduct two proof-of-concept experiments. Experiment 1 demonstrates the formulation under a controlled environment with various assumptions to investigate the potential of the VE. Experiment 2 relaxes the assumptions to show the limited performance of the VE with the current technology. In both the experiments, we use a UP camera instead of a real user for obtaining objective measurements.

A thorough discussion including the limitations of the setups and possible solutions for realizing a practical system is presented in the discussion section.

Hardware Setup

We have built an OST-HMD system equipped with an outward looking camera as described below and in Fig. 3. We use Lumus DK-32, an OST-HMD with 1280×720 resolution. The left eye display is used for the current setup. For the world camera, we use a USB 2.0 camera from Delock. The camera has 1600×1200 image resolution with 64° field of view. For the UP camera placed behind the OST-HMD, we use UI-1240ML-C-HQ from iDS. The camera holds 1280×1024 image resolution together with an 8mm C-mount lens. As a scene object $\mathrm{X_W}$, we use an acuity chart set on a planar calibration board. The board is placed about 1.5 m away from the display.

Experiment 1

This experiment is to assesses a potential performance of the VE by simulating a setup where a perfect OST-HMD is available. We first explain the simulation formulation, then the actual data collection procedure, and the simulation results.

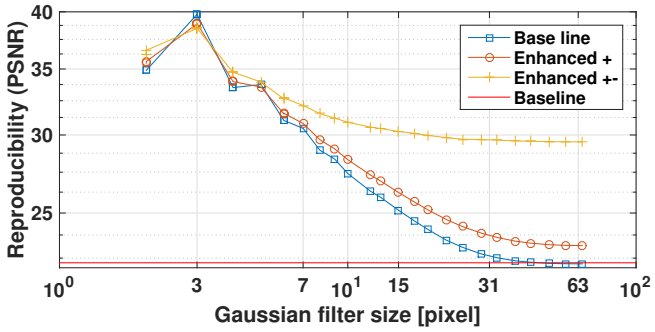

Figure 5. The result of the experiment 1. The log-scale X axis is the size of Gaussian blur filter used to compute degraded user-perspective images $\mathbf{I}_{\mathrm{U}}^{\star\prime}$. The log-scale Y axis is the PSNR between the ground truth image $\mathbf{I}_{\mathrm{U}}^{\star}$ and: $\mathbf{I}_{\mathrm{U}}^{\star\prime}$ (Degraded), the enhanced images with the positive image $\widehat{\mathbf{I}_{\mathrm{U}>0}^{\star}}$ ($\mathrm{E}^{\star}$, Enhanced +); with the complete filter $\widehat{\mathbf{I}_{\mathrm{U}}^{\star}}$ ($\mathrm{E}_{>0}^{\star}$, Enhanced +-), and the UP image $\widehat{\mathbf{I}_{\mathrm{U}}}$ created from the world view $\mathbf{I}_{\mathrm{W}}^{\star}$ (Baseline). The VE technique improves the quality of the degraded images.

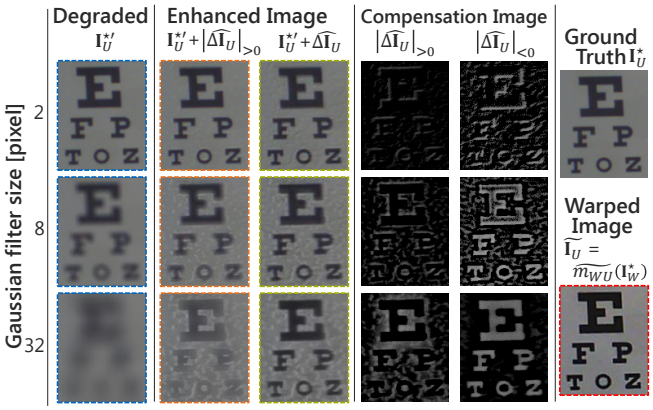

Figure 6. Sample image from the experiment 1. The enhanced images are clearer compared to the degraded images. Note that the brightness and contrast of the compensation images here are modified to +40/-40% for a presentation purpose. See also Fig. 5.

Simulation Procedure and Assumptions

Instead of directly obtaining $\widehat{\Delta\mathbf{I}_{\mathrm{U}}^{\star}}$ by displaying $\widehat{\Delta\mathbf{I}_{\mathrm{S}}}$, we partially synthesize it by introducing several assumptions (Fig. 4). Some of the assumptions naturally stem from the fact that we use an UP camera instead of an user. Some others are for simplifications due to our lack of methods to estimate some of the true functions. We emphasize it here that the upcoming sections investigate potential solutions for incorporating real users and removing the simplifications. We now explain our simplified VE procedure.

First of all, the UP camera and the world camera capture $\mathbf{I}_{\mathrm{U}}^{\star}$ and $\mathbf{I}_{\mathrm{W}}^{\star}$ respectively as same as the original formulation. We then warp $\mathbf{I}_{\mathrm{W}}^{\star}$ directly by $\widetilde{m_{\mathrm{WS}}}(\cdot)$, and display $\widetilde{\mathbf{I}_{\mathrm{S}}} := \widetilde{m_{\mathrm{WS}}}(\mathbf{I}_{\mathrm{W}}^{\star})$ on the HMD. While blocking the world light, we let the UP camera capture $\widetilde{\mathbf{I}_{\mathrm{S}}}$ as $\widehat{\mathbf{I}_{\mathrm{U}'}} := g_{\mathrm{abr}}(h_{\mathrm{col}}(m_{\mathrm{SU}}(\widetilde{\mathbf{I}_{\mathrm{S}}})))$. Finally, based on the assumption that light is additive, we process $\mathbf{I}_{\mathrm{U}}^{\star}$ and $\widehat{\mathbf{I}_{\mathrm{U}'}}$ in a software to synthesize a deblurred image.

We introduce another assumption to simplify the warping function: we limit the scene structure X_{W} to a planar 3D surface with a 2D marker. The UP camera and the world camera track the 2D marker, and we compute the relative poses among the cameras and the plane. As the result, we obtain projection matrices $\mathbf{P}_{\mathrm{WU}}$ and $\mathbf{P}_{\mathrm{WW}}$. Also, given an OST-HMD calibration method explained in the next section, we obtain the projection matrix $\mathbf{P}_{\mathrm{WS}}$ as well. As the result, $\widetilde{m_{\mathrm{WS}}}(\cdot)$ becomes an image translation by a homography matrix.

We compute a compensation image $\widehat{\Delta\mathbf{I}_{\mathrm{U}}}$ by applying the estimates of the inverse of $h_{\mathrm{col}}(\cdot)$ and $g_{\mathrm{abr}}(\cdot)$:

$$\widehat{\mathbf{I}_{\mathrm{U}}} := \widetilde{h_{\mathrm{col}}^{-1}}(\widetilde{g_{\mathrm{abr}}^{-1}}(\widehat{\mathbf{I}_{\mathrm{U}'}})), \quad \widehat{\Delta\mathbf{I}_{\mathrm{U}}} := \widehat{\mathbf{I}_{\mathrm{U}}} - f_{\mathrm{abr}}(\widehat{\mathbf{I}_{\mathrm{U}}}), \tag{13}$$

where we assume $f_{\mathrm{abr}}(\cdot)$ as a Gaussian blur with a known diameter σ [pixel]. We synthesize a degraded UP view as $\mathbf{I}_{\mathrm{U}}^{\star\prime} := f_{\mathrm{abr}}(\widehat{\mathbf{I}_{\mathrm{U}}})$, and evaluate the VE performance by changing σ. To ignore $g_{\mathrm{abr}}^{-1}(\cdot)$, We also place X_{W} almost at the same distance as the image screen when seen by the UP camera. To simplify the color correction, we only consider a gray-scale color, and we approximate the inverse of the color correction $h_{\mathrm{col}}^{-1}(\cdot)$ as a scaling factor c defined as follows:

$$\widetilde{h_{\mathrm{col}}^{-1}}(\cdot) = c^{-1} * (\cdot), \tag{14}$$

where $c := \mathrm{mean}(\widehat{\mathbf{I}_{\mathrm{U}'}})/\mathrm{mean}(\mathbf{I}_{\mathrm{U}}^{\star})$.$\mathrm{mean}(\cdot)$ is the mean color of a given image.

Now, our approximated compensation image and our enhanced image become

$$\widehat{\Delta\mathbf{I}_{\mathrm{U}}} = c^{-1} * \widehat{\mathbf{I}_{\mathrm{U}'}} - f_{\mathrm{abr}}(c^{-1} * \widehat{\mathbf{I}_{\mathrm{U}'}}). \tag{15}$$

$$\widehat{\mathbf{I}_{\mathrm{U}}^{\star}} := \mathbf{I}_{\mathrm{U}}^{\star\prime} + \widehat{\Delta\mathbf{I}_{\mathrm{U}}}, \quad \widehat{\mathbf{I}_{\mathrm{U}>0}^{\star}} := \mathbf{I}_{\mathrm{U}}^{\star\prime} + |\widehat{\Delta\mathbf{I}_{\mathrm{U}}}|_{>0}. \tag{16}$$

We compute the enhancement error as

$$\mathrm{E}^{\star} := \mathrm{PSNR}(\mathbf{I}_{\mathrm{U}}^{\star}, \widehat{\mathbf{I}_{\mathrm{U}}^{\star}}), \ \mathrm{E}_{>0}^{\star} := \mathrm{PSNR}(\mathbf{I}_{\mathrm{U}}^{\star}, \widehat{\mathbf{I}_{\mathrm{U}>0}^{\star}}), \tag{17}$$

where $\mathrm{PSNR}(\cdot,\cdot)$ represents Peak Signal-to-Noise Ratio (PSNR) of two given images. Similar images yield higher PSNR, and go to infinity if the two images are identical. Note that $\mathrm{E}_{>0}^{\star} \geqq \mathrm{E}^{\star}$ holds. Note that the above derivation of the compensation image is valid only if our generalized formulation is true (Eq. 11 and 12).

Data Acquisition for $\mathbf{I}_{\mathrm{U}}^{\star}$ and $\widehat{\mathbf{I}_{\mathrm{U}'}}$

Our data collection procedure is as follows:

1. Calibrate the world/UP cameras and the HMD screen.
2. Place the UP camera toward the board X_{W}, and capture $\mathbf{I}_{\mathrm{U}}^{\star}$.
3. Place the HMD in front of the UP camera and capture $\mathbf{I}_{\mathrm{W}}^{\star}$.
4. Compute $\mathbf{P}_{\mathrm{WW}}$ and $\mathbf{P}_{\mathrm{WS}}$ (also $\mathbf{P}_{\mathrm{WU}}$ for the experiment 2) from the Step 1 and Step 2.
5. From $\mathbf{I}_{\mathrm{U}}^{\star}$ and $\mathbf{I}_{\mathrm{W}}^{\star}$, compute the 6-DoF pose between the cameras and the parameters of the board as a 4D vector.
6. From the vector, $\mathbf{P}_{\mathrm{WW}}$, and $\mathbf{P}_{\mathrm{WS}}$, estimate $\widetilde{m_{\mathrm{WS}}}(\cdot)$ as a 2D homography mapping.
7. Display the warped image $\widetilde{\mathbf{I}_{\mathrm{S}}}$ on the display. Let the UP camera capture the displayed image as $\widehat{\mathbf{I}_{\mathrm{U}'}}$ while blocking the world light by, e.g., making the room completely dark.

At Step 1, we treat the display screen as a virtual 3D plane with its pose defined relative to the UP camera. For the OST-HMD calibration, we used method described in [15].

After this process, we obtain $\mathbf{I}^{\star}_{\mathrm{U}}$ and $\widehat{\mathbf{I}_{\mathrm{U}'}}$ for the evaluation. Note that we only compare a region of the images which includes the acuity chart (see Fig. 6).

Enhancement Result of Experiment 1

Figures 5 and 6 show the results. The red line (Baseline) in Fig. 5 shows $\mathrm{PSNR}(\mathbf{I}^{\star}_{\mathrm{U}}, \widetilde{\mathbf{I}_{\mathrm{U}}})$ as a baseline. This is equivalent to a user perspective VST-HMD setup where the display replaces the actual user view by the world camera. Note that we also adjusted the color balance of $\widetilde{\mathbf{I}_{\mathrm{U}}}$ to that of $\mathbf{I}^{\star}_{\mathrm{U}}$ by Eq. 14. The blue line with square markers (Degraded) is a plot of $\mathrm{PSNR}(\mathbf{I}^{\star}_{\mathrm{U}}, \mathbf{I}^{\star\prime}_{\mathrm{U}})$, which is what a degraded eye would see the world without VE. The orange line with circle markers (Enhanced +) is $\mathrm{E}^{\star}$ and yellow with crosses (Enhanced +-) is $\mathrm{E}^{\star}_{>0}$ from Eq. 17.

As expected, all methods suffer from the increasing amount of the Gaussian blur. However, both Enhanced + and Enhanced +- in general improve the degraded images while keeping the qualities higher than just displaying the warped world-camera view, i.e., Baseline. Enhanced +- shows a significant improvement over Enhanced +. This suggests developing an adjustable opaque layer in the display is desirable for practical VE systems. Such systems can visualize *negative* values. The negative color contributes dominantly when extreme blur is present (see the third row of Fig. 6).

Experiment 2

This experiment follows the original formulation in the method section (Fig. 2) to assess a more realistic setup that a practical VE system should follow. Data collection is done in a similar manner as the previous section.

We first prepared $\widetilde{\Delta\mathbf{I}_{\mathrm{S}}}$. For computing $\widetilde{\mathbf{I}'_{\mathrm{U}}}$, we again treated $\widetilde{f_{\mathrm{abr}}}(\cdot)$ as a Gaussian blur. The same blur is added to $\mathbf{I}^{\star}_{\mathrm{U}}$ for synthesizing $\mathbf{I}^{\star\prime}_{\mathrm{U}}$. The UP camera captured the displayed image $\widetilde{\Delta\mathbf{I}^{\star}_{\mathrm{U}}}$. When capturing, we blocked the world light from the display in the same way as the experiment 1. Then we computed $\widetilde{\mathbf{I}^{\star}_{\mathrm{U}}}$ (Eq. 12). Note that we had to tune the image gamma of $\widetilde{\Delta\mathbf{I}^{\star}_{\mathrm{U}}}$ before fusing it with $\mathbf{I}^{\star\prime}_{\mathrm{U}}$ to reduce the backlight ambient color of the display panel that made pixels bluish even if the digital color of the pixels are set to zero.

Figure 7 summarizes the result. Compared to the experiment 1, the improvement was negligibly small. One of the main cause might be the color distortion $h_{\mathrm{col}}(\cdot)$ of the display, which distorts the brightness of image and thus the compensation image.

LIMITATIONS OF THE CURRENT EXPERIMENTS

Our current setups, especially in experiment 1, use various assumptions and simplifications. We list them with the corresponding step characters from the method section:

L1 (A) $\mathrm{X_W}$ is limited to a 3D plane and 6-DoF pose between the cameras are known.

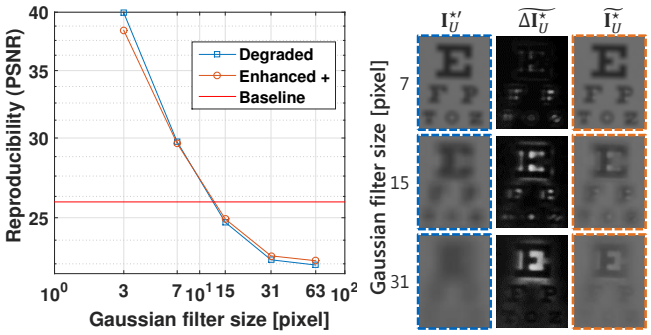

Figure 7. Result of experiment 2. (Left) Error plot. (Right) Result images, where we modified the brightness of the compensation images +70% for the presentation purpose.

L2 (A) We refined manually the 2D position between $\mathbf{I}^{\star}_{\mathrm{U}}$ and $\widehat{\mathbf{I}_{\mathrm{U}'}}$ before computing the compensation images.

L3 (A) An ideal user-perspective image $\mathbf{I}^{\star}_{\mathrm{U}}$ is available, and the eye position is known w.r.t. the HMD camera.

L4 (B) The aberration model $f_{\mathrm{abr}}(\cdot)$ is a Gaussian blur.

L5 (B,C) Images are in gray. Only the distortion of color intensity is corrected by a simple color correction (Eq. 14).

L6 (B,C) Optical elements do not distort world illumination.

L7 (C) $g^{-1}_{\mathrm{abr}}(\cdot)$ is the identity function, i.e. we ignored it.

L8 (C) The relative image resolution of the display screen is higher than that of the user-perspective image.

These issues have to be solved for a practical VE. In the next section, we discuss practical solutions for each relevant topic.

ISSUES TOWARD PRACTICAL VISION ENHANCEMENT

We analyze issues reaming in the current VE system and discuss possible solutions. We clustered the issues along the three steps of VE, and referred corresponding limitations in the headlines.

(A) Transformation of Sensor Images

UP Rendering with Arbitrary Scene Structure (L1,L2,L3):

Our VE method requires to map the world view to the user perspective view. The mapping changes by both the eye position of a user and the structure of a scene. There are hardware and software solutions to estimate a correct mapping.

A hardware solution is to align the optical paths of the world camera and the eye (the UP camera) by a half mirror [27]. This way, both cameras optically share the same viewpoint, and we may opt out the $\mathrm{X_W}$ from the warping function. However, the mirror has to be dynamically configurable to keep the virtual center of the camera according to the dynamic eye position w.r.t. the world camera.

A software solution is to compute the depth of the scene $\mathrm{X_W}$ and warp the current world view to the UP view while taking the depth of each pixel into account [30, 4]. This solution requires a 3D sensing. Since two viewpoints are different, we face with the occlusion problem that a part of the scene

visible by an UP camera is not visible by the world camera. Such occluded part of the scene thus can not be displayed.

Both the solutions require a high-speed 3D eye tracking. With an eye-tracking camera, another possible technique to provide $\mathbf{I}^{\star}_{\mathrm{U}}$ directly is the corneal imaging [18]. It analyzes reflected image on eye cornea to see what a user actually see.

OST-HMD Calibration (L2,L3):
Related to the mapping issue above, another issue is pertaining to the calibration of the OST-HMD screen. The pose of the screen image is necessary to compute the correct projection for the virtual camera S. Although common HMD calibration methods calibrate the screens as 3D planes, this is often invalid due to the complex optics of the HMD. Furthermore, the HMD optics distort the user perspective view itself as corrective glasses do. An ideal HMD calibration method must take these issues into its HMD model for the calibration.

(B) Estimation of User's Vision

Appearance Correction (L4,L5,L6):
Even if we have a correct UP rendering, we still need to estimate the color distortion, $h_{\mathrm{col}}(\cdot)$, so that a displayed color perceived by a user is consistent with the visual stimulus that the user receives from the scene directly. A solution is to calibrate the color of the display (and the world camera) beforehand by a UP camera. Some work for such OST-HMD color corrections already exists [28, 9].

Another unsolved, challenging issue is that photoreceptor cells of eyes have different sensitivity than image sensors.

Eye Aberration Estimation (L4):
In our experiment, we assumed that the eye aberration, $f_{\mathrm{abr}}(\cdot)$, is a simple Gaussian blur. Namely, we only considered a defocus basis of Zernike polynomials, a common aberration model employed in optometry [5]. Optometry researchers have worked on estimating the profile with wavefront sensors [12, 25, 17] or video-based techniques [29]. Recently a mobile system has been proposed [21].

(C) Preprocessing and Rendering of Filter Images

Image Rendering (L5,L6,L7):
In addition to $\widetilde{f_{\mathrm{abr}}}(\cdot)$, we need the second aberration $\widetilde{g^{-1}_{\mathrm{abr}}}(\cdot)$ for the virtual screen. If the image screen is at the same distance as an object at which a user is focusing, we may thus ignore this aberration or treat is as $\widetilde{f_{\mathrm{abr}}}(\cdot)$. Practically, this assumption is unlikely. If we have the aberration estimate $\widetilde{f_{\mathrm{abr}}}(\cdot)$, the display screen pose, and the scene information $\mathrm{X_W}$, then it would be possible to estimate $\widetilde{g_{\mathrm{abr}}}(\cdot)$. We then invert $\widetilde{g_{\mathrm{abr}}}(\cdot)$. A software solution is to employ an image deconvolution method such as [20] to compute $\widetilde{g^{-1}_{\mathrm{abr}}}(\cdot)$.

Another solution by hardware is a retinal HMD. Its rendering is unaffected by crystalline lenses. The display employs the Maxwellian-view optical system [31], and realizes focus-free images [2], thus we can physically ignore $\widetilde{g_{\mathrm{abr}}}(\cdot)$.

Resolution between eyes and OST-HMDs (L8):
This would be most challenging issue in terms of hardware. The maximum angular resolution of the human eye around the fovea is approximately half an arcsecond, about 8.3e-3 degrees. The display we used has 40-degrees field of view with the 1280×720 pixels, which yields about 31e-3 degrees at the finest. Thus the display resolution is still far lower than that of our eyes. As far as we know, there have been no OST-HMDs thus far with resolutions comparable to that of human eye. Again retinal displays would be a possible solution with a foveated rendering.

Other issues

Vision Enhancement against Corrective Glasses
Corrective glasses are well-established direct, analog VE devices for aberration corrections. Potential benefits of our VE system based on OST-HMDs over such devices are: personalization of enhancement depending on users' aberration types that change over age, correction of higher-order aberration that the glasses can not correct, and simultaneous enhancement of other vision problems such as color blindness.

Overcoming Physical Limit of our Visual Acuity
In other word, can we make $\widetilde{\mathbf{I}^{\star}_{\mathrm{U}}} = \mathbf{I}^{\star\prime}_{\mathrm{U}} + \widetilde{\Delta\mathbf{I}^{\star}_{\mathrm{U}}}$ better than $\mathbf{I}^{\star}_{\mathrm{U}}$?

An interesting question of the VE for the defocus correction is whether emmetropic people can benefit from the system. For example, if our display has a higher resolution than that of eye retina, then the display can create a *super-resolution* image $m_{\mathrm{SU}}(\widetilde{\Delta\mathbf{I}_{\mathrm{S}}})$ which is finer than the eye retina can perceive. There is no benefit of doing this in terms of the spatial resolution. Perhaps, a vibro-imaging system [10] can overcome this limitation by temporally modulating the display image to achieve a perceptual *super-resolution*.

CONCLUSION AND FUTURE WORK

This paper proposes a VE concept for defocus correction of human eyes via OST-HMDs. Our main contributions are: (1) We provide a theoretical formulation of VE while incorporating with constraints of the optical relationship between the HMDs and human eyes. (2) We conduct proof-of-concept experiments, with cameras and an OST-HMD, to demonstrate that the method improves a degraded image quality. (3) More importantly, we provide a thorough analysis of the current VE setup including limitations, issues and possible solutions toward the realization of a practical VE system.

Future work directions involve: OST-HMD color correction, eye aberration estimation, deconvoluted image rendering, and extension to the full-color imagery. Besides on that, a study with actual users is also desirable. We hope this work will serve as a foundation for improving the VE techniques.

ACKNOWLEDGMENTS

The project received funding from the European Union's Seventh Framework Programmes for Research and Technological Development under PITN-GA-2012- 316919-EDUSAFE.

REFERENCES

1. Amano, T., and Kato, H. Appearance control by projector camera feedback for visually impaired. In *CVPRW*, IEEE (2010), 57–63.

2. Asai, N., Matsuda, R., Watanabe, M., Takayama, H., Yamada, S., Mase, A., Shikida, M., Sato, K., Lebedev, M., and Akedo, J. Novel high resolution optical scanner actuated by aerosol deposited pzt films. In *Micro Electro Mechanical Systems (MEMS)*, IEEE (2003), 247–250.

3. Avery, B., Sandor, C., and Thomas, B. H. Improving spatial perception for augmented reality x-ray vision. In *VR* (2009), 79–82.

4. Baričević, D., Lee, C., Turk, M., Höllerer, T., and Bowman, D. A. A hand-held ar magic lens with user-perspective rendering. In *ISMAR* (2012), 197–206.

5. Bennett, M., and Quigley, A. Creating personalized digital human models of perception for visual analytics. In *User Modeling, Adaption and Personalization*. Springer, 2011, 25–37.

6. Bimber, O., and Iwai, D. Superimposing dynamic range. In *TOG*, vol. 27 (2008), 150.

7. Bimber, O., Iwai, D., Wetzstein, G., and Grundhöfer, A. The visual computing of projector-camera systems. In *Computer Graphics Forum*, vol. 27, Wiley Online Library (2008), 2219–2245.

8. Broecker, M., Smith, R. T., and Thomas, B. H. Adaptive substrate for enhanced spatial augmented reality contrast and resolution. In *ISMAR* (2011), 251–252.

9. David Hincapie-Ramos, J., Ivanchuk, L., Sridharan, S. K., and Irani, P. Smartcolor: Real-time color correction and contrast for optical see-through head-mounted displays. In *ISMAR* (2014), 187–194.

10. Fujimori, N., and Ando, S. Super-resolution reconstruction algorithm using vibro-imaging and correlation image sensor. In *Society of Instrument and Control Engineers (SICE), Annual Conference on*, IEEE (2012), 2028–2033.

11. Giusti, A., Taddei, P., Corani, G., Gambardella, L., Magli, C., and Gianaroli, L. Artificial defocus for displaying markers in microscopy z-stacks. In *TVCG*, vol. 17 (2011), 1757–1764.

12. Grimm, B., Goelz, S., Bille, J. F., et al. Objective measurement of wave aberrations of the human eye with the use of a hartmann-shack wave-front sensor. *The Journal of the Optical Society of America A (JOSA A) 11*, 7 (1994), 1949–1957.

13. Huang, F.-C., Wetzstein, G., Barsky, B. A., and Raskar, R. Computational light field display for correcting visual aberrations. In *SIGGRAPH Posters* (2013), 40.

14. Hwang, A. D., and Peli, E. An augmented-reality edge enhancement application for google glass. *Optometry & Vision Science 91*, 8 (2014), 1021–1030.

15. Itoh, Y., and Klinker, G. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *3DUI* (2014), 75–82.

16. Koizumi, N., Sugimoto, M., Nagaya, N., Inami, M., and Furukawa, M. Stop motion goggle: augmented visual perception by subtraction method using high speed liquid crystal. In *AH* (2012), 14.

17. Lombardo, M., and Lombardo, G. Wave aberration of human eyes and new descriptors of image optical quality and visual performance. *Journal of Cataract & Refractive Surgery (JCRS) 36*, 2 (2010), 313–331.

18. Nitschke, C., Nakazawa, A., and Takemura, H. Corneal imaging revisited: An overview of corneal reflection analysis and applications. *Information and Media Technologies 8*, 2 (2013), 389–406.

19. Oskiper, T., Sizintsev, M., Branzoi, V., Samarasekera, S., and Kumar, R. Augmented reality binoculars. In *ISMAR* (Oct 2013), 219–228.

20. Oyamada, Y., and Saito, H. Defocus blur correcting projector-camera system. In *Advanced Concepts for Intelligent Vision Systems*, Springer (2008), 453–464.

21. Pamplona, V. F., Mohan, A., Oliveira, M. M., and Raskar, R. Netra: interactive display for estimating refractive errors and focal range. In *TOG*, vol. 29 (2010), 77.

22. Pamplona, V. F., Oliveira, M. M., Aliaga, D. G., and Raskar, R. Tailored displays to compensate for visual aberrations. In *TOG*, vol. 31 (2012), 81.

23. Peli, E. Vision multiplexing: an engineering approach to vision rehabilitation device development. *Optometry & Vision Science 78*, 5 (2001), 304–315.

24. Peli, E. Treating with spectacle lenses: a novel idea!? *Optometry & Vision Science 79*, 9 (2002), 569–580.

25. Platt, B. C., et al. History and principles of shack-hartmann wavefront sensing. *Journal of Refractive Surgery 17*, 5 (2001), S573–S577.

26. Sajadi, B., Gopi, M., and Majumder, A. Edge-guided resolution enhancement in projectors via optical pixel sharing. In *TOG*, vol. 31 (2012), 79.

27. Shimizu, S., and Fujiyoshi, H. Acquisition of 3d gaze information from eyeball movements using inside-out camera. In *AH* (2011), 6.

28. Sridharan, S. K., Hincapié-Ramos, J. D., Flatla, D. R., and Irani, P. Color correction for optical see-through displays using display color profiles. In *VRST* (2013), 231–240.

29. Suryakumar, R., Meyers, J. P., Irving, E. L., and Bobier, W. R. Application of video-based technology for the simultaneous measurement of accommodation and vergence. *Vision research 47*, 2 (2007), 260–268.

30. Tomioka, M., Ikeda, S., and Sato, K. Approximated user-perspective rendering in tablet-based augmented reality. In *ISMAR* (2013), 21–28.

31. Westheimer, G. The maxwellian view. *Vision research 6*, 11 (1966), 669–682.