DehazeGlasses: Optical Dehazing with an Occlusion Capable See-Through Display

Yuichi Hiroi
Tokyo Institute of Technology
yuichi.hiroi@ar.c.titech.ac.jp

Takumi Kaminokado
Tokyo Institute of Technology
kaminokado@ar.c.titech.ac.jp

Atsushi Mori
Tokyo Institute of Technology
atsushi.mori@ar.c.titech.ac.jp

Yuta Itoh
Tokyo Institute of Technology / RIKEN AIP
yuta.itoh@c.titech.ac.jp

Figure 1: Overview of the DehazeGlasses system. (a) The front view of our system. (b, top) the Digital micromirror device (DMD) showing an occlusion mask with a white background; (b, bottom) an input image for the optical see-through head-mounted display (OST-HMD) screen to show an additive image; (c) a hazy scene captured from the viewpoint without the haze removal and (d) with the haze removal.

ABSTRACT

We present DehazeGlasses, a see-through visual haze removal system that optically dehazes the user's field of vision. Human vision suffers from a degraded view due to aspects of the scene environment, such as haze. Such degradation may interfere with our behavior or judgement in daily tasks. We focus on hazy scenes as one common degradation source, which whitens the view due to certain atmospheric conditions. Unlike typical computer vision systems that process recorded images, we aim to realize a see-through glasses system that can optically manipulate our field of view to dehaze the perceived scene. Our system selectively modulates the intensity of the light entering the eyes via occlusion-capable optical see-through head-mounted displays (OST-HMD). We built a proof-of-concept system to evaluate the feasibility of our haze removal method by combining a digital micromirror device (DMD) and an OST-HMD, and tested it with a user-perspective viewpoint camera. A quantitative evaluation with 80 scenes from a haze removal dataset shows that our system realizes a dehazed view that is significantly closer to the ground truth scene compared to the native view under a perceptual image similarity metric. This evaluation shows that our system achieves perceptually natural haze removal while maintaining the see-through view of actual scenes.

CCS CONCEPTS

• **Computing methodologies** → **Mixed / augmented reality**; • **Human-centered computing** → **Mixed / augmented reality**.

KEYWORDS

Haze Removal, Vision Augmentation, Augmented Reality, Head-Mounted Displays, Occlusion-Capable HMD

1 INTRODUCTION

Human vision often suffers from haze, small particles floating in the air. These floating particles, such as smoke, dust, fumes, and mist, scatter light in the atmosphere and attenuate the direct scene irradiance in correspondence with the scene depth. Furthermore, the particles reflect the ambient light and veil the scene uniformly as an opaque white layer called airlight [16]. Haze thus leads to low contrast, blurring, color distortion, and other visual degradation, which significantly affects human perception and behavior. The visual degradation may even cause severe accidents in some situations, such as driving in an environment with snow or fog, or evacuation and rescue activities in the event of fire or other disasters.

Removing haze to provide a clear view from a user's viewpoint helps the user to make correct decisions and actions. In the computer vision community, haze removal also attracts interest as a

preprocessing step to aid high-level visual tasks, such as object recognition, person re-identification, and visual tracking.

The early dehazing methods were based on the atmospheric scattering model [3, 25], which is a simple approximation of the haze effect, and formulated the haze-free image assuming some priors [7]. The recent rise of deep learning pursues the robust performance of haze removal without any priors [1, 26].

In this work, we consider *re-importing* these machine visions of haze removal into human vision by presenting a first-person system. Our future goal is to enhance high-level human recognition in hazy situations and to decrease the potential for misjudgment.

The most straightforward approach is *video see-through* (VST) systems, where a system feeds a video stream from a scene camera to the user's view through a monitor or an opaque head-mounted display. While such a VST system can easily apply existing dehaze methods, the approach may not be suitable for visual assistance in a wearable system [30]. A VST system loses rich light information from the real world consisting of the real light field, which has a higher resolution and dynamic range than existing displays. More importantly, VST may have a risk of blacking out the user's view under accidental shutdown or power depletion. We conclude these disadvantages make VST systems unsuitable for dehaze applications that may be used in critical tasks [29, 30].

Optical see-through (OST) approaches, where a system minimally augments the user's first-person view, have advantages over the VST approaches. First, the OST approach is less intrusive to the user's view compared to VST systems. It can retain the real-world light, and the modification of the view is minimally done by an OST-head-mounted display (HMD). Second, the OST approach is safer. The malfunction of an OST-HMD system does not block the user's view, as it merely results in the user seeing the actual view without augmentation.

Lin et al. recently adapted a dehaze method for a commercial OST display without occlusion capability [21]. Their system calculates the difference between the real scene and a dehazed image, then computes an auxiliary image that enhances the visual image by applying the just-noticeable difference decomposition [24]. Since the OST display can only add light to the scene, their method essentially loses the contrast of the scene, and the result appears as a brightness-enhanced view instead of truly removing the airlight layer.

To realize a see-through dehaze system that optically removes haze, we built an occlusion-capable see-through display named DehazeGlasses (Fig. 1). Our system combines an OST-HMD with a digital micromirror display (DMD), which can switch the direction of light rays for each pixel by using micro-electro-mechanical mirrors to occlude the incoming rays. By adopting a dehazing method for the occlusion-capable OST-HMD system, we can optically subtract the scene light to dehaze the hazy view.

To the best of our knowledge, this is the first attempt to realize an occlusion-capable see-through dehazing system to enhance the first-person view of the user.

Our main contributions include the following:

- Providing a computation method to realize haze removal using an occlusion-capable OST-HMD.
- Implementing a proof-of-concept OST AR display that can optically dehaze the first-person view of the user.
- Conducting quantitative and qualitative analyses of the system with a user-perspective camera, which show that the system achieves perceptually natural haze removal while maintaining the see-through view of actual scenes.
- Providing a thorough analysis of the current setup, including limitations and possible research directions.

2 RELATED WORK

In this section we introduce key research areas relevant to our DehazeGlasses system, namely haze removal algorithms and occlusion-capable OST-HMDs. Following this, we also elaborate other vision augmentation research works.

2.1 Haze Removal

Existing dehazing methods rely on the physical scattering model [3, 25]. This model employs the atmospheric transmission map and the atmospheric illumination. Most of the methods based on this model use prior-based techniques [7]. Schechner et al. took the airlight polarization into account on top of the scattering model [31, 32]. Recent advancements in deep neural networks have resulted in generating more natural dehaze images [1, 26].

All of these methods require image processing on the input image in order to reconstruct a dehazed image. Haze adds gray colors on top of the clear image, so the processing inevitably requires the subtraction of these colors from the input image. Since our goal is to dehaze the user's view via a see-through system, the system must be capable of *occlusion*—selectively attenuating the scene light in the user's view.

2.2 Occlusion-capable OST Displays

Occlusion-capable OST displays can selectively attenuate the incoming light to provide perceptually realistic experiences and support correct depth cues.

Recent commercial OST-HMD systems (e.g., Microsoft HoloLens and Magic Leap) use a neutral density filter placed on the outside of the display to reduce ambient light uniformly; however, these are not suitable for the haze removal display, which requires spatial control of the occlusion layer.

Spatial light modulators (SLM), such as liquid-crystal displays (LCD) or DMDs, are commonly used to selectively block out or attenuate the transmission of the real scene in a spatially varying manner. The physical scene is focused onto the transmissive [9, 15] or reflective SLM [2], and then the display modulates the amplitude and phase of incoming rays. Maimone et al. [23] also combined an out-of-focus LCD and a transparent plate with a dot array to provide an occlusive backlight as the silhouette of the virtual object. While we follow the optics design using the single monochromatic occlusive layer (i.e., DMD) inserted in front of the user view, recently proposed OST displays via per-pixel spatial color filtering [14] realized the color-stylized occlusion by using liquid crystal on silicon (LCoS) displays and polarizers.

Since the occlusion layer is generally out of focus from the user's accommodation, the occlusive mask always appears blurred. To provide a hard-edged occlusive mask, i.e., to realize a varifocal

occlusion-capable display, both software and hardware approaches are proposed. Itoh et al. [12] proposed to compensate for the blur by overlaying a modified image on a degraded occlusion area. On the hardware side, recent approaches dynamically adjust the focal plane of the occlusion layer following the user's accommodation by physically sliding the LCD layer [5], or by using multiple focus-tunable optics such as liquid lenses and optimizing the focal length of each lens [27].

Light-field occlusion is another direction to realize the variable focal length. Maimone and Fuchs [22] proposed a light-field occlusion display with stacked LCD layers. Yamaguchi and Takagi also developed an occlusion-capable display based on integral imaging [37]. Although both displays suffer from a narrow field of view and limited spatial resolution, they realize the compact form factor and depth-dependent occlusion by combining both additive and occlusive LCD layers.

2.3 Vision Augmentation

Our system belongs to a group of vision augmentation systems that can compensate for vision impairment or further enhance human vision in a programmable manner. Sutton et al. [33] provide a detailed review of various vision augmentation systems. Thus, we provide here concise outlines of existing systems relevant to the scope of our issue.

We first overview vision augmentation systems to improve the visibility, which utilizes SLMs to manipulate the contrast or brightness of input from the physical world. Tamburo et al. developed a DMD-based smart automotive headlight to improve visibility in a driving scene, especially in risky weather conditions such as rain, fog and snowstorms [34]. Hara et al. demonstrated an LCD-based system that can remove the glare of water droplets on a transparent surface by tuning the transmittance where droplets are positioned [6]. These systems can improve the visibility by modulating the incoming rays in the driving scene or computational photography. Similarly, we aim to directly modulate the incoming light into eye and improve the visibility from the first-person view.

There are some vision augmentation systems for vision impairment using the first-person view by modulating the edge, color, and contrast of the environment light. Hwang and Peli proposed vision augmentation systems that overlay edge information onto the real scene via OST-HMDs to support visual impairment [10]. Similarly, Langlotz et al. proposed a computational glass for color-vision deficiency [19] by providing a pixel-precise color-modulated image with an OST-HMD [18].

Hiroi et al. [8] et al. proposed smart brightness-adaptation glasses that control the dynamic range of the user's see-through view. Their system consists of an LCD and an OST-HMD to realize a High-Dynamic-Range (HDR) view for human eyes. Their system dynamically and selectively modulates the scene brightness with the occlusion-capable OST-HMD by detecting the over-/under-exposed light from the scene camera.

Although their system shares a similar hardware concept to ours, our system focuses more on realizing dehaze in the see-through view, which is only possible by realizing a sharp optical-occlusion mask and the correct spatial aligment of the view, the screen, and the occlusion layer.

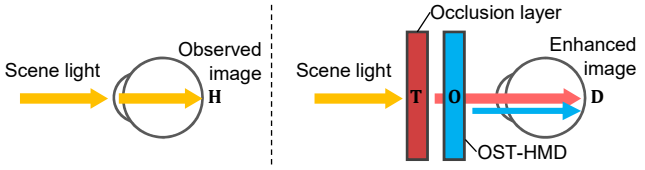

Figure 2: A schematic diagram showing a generic hardware formation of our see-through setup defined in Sec. 3.1. This generic model lacks some implementation details (Sec. 4), but just sufficient to formulate our problem.

3 METHOD

In the following, we first formulate the see-through dehaze problem and provide the mathematical notations used through the rest of the paper in Sec. 3.1. Second, we describe our base computation approach in Sec. 3.2. Finally, in Sec. 3.3, we further provide a visually less-intrusive method that aims to reduce the amount of illumination from the OST-HMD image layer.

3.1 Problem formulation and Notations

Our goal is to model a see-through haze removal filter on a glasses system equipped with an occlusion layer (e.g., with a DMD) and the additive image layer in an OST-HMD. Since the occlusion layer can only control the transparency of the light passing through it, the layer can only rescale the brightness of the scene image instead of directly subtracting its colors. This limitation prevents us from simply using existing haze removal pipelines. Instead, we have to design a dehaze approach with attenuation and addition operations only.

For simplicity, we start with a basic setup where all image layers are already spatially aligned with each other and have the same image size. In practice, we need to calibrate each image layer to associate each corresponding pixel, as elaborated in Sec. 4.4.

Our common notations and variables are as follows:

- $\mathbf{H} = \{\mathbf{h}_{ij}\} = \{h_{ijc}\} \in \mathbb{R}_+^{W\times H\times 3}$ is the observed (hazy) color scene image.
- $\mathbf{D} = \{\mathbf{d}_{ij}\} = \{d_{ijc}\} \in \mathbb{R}_+^{W\times H\times 3}$ is the dehazed color image from the viewpoint.
- $\mathbf{T} = \{t_{ij}\} \in \mathbb{R}^{W\times H}$ $(0 \le t_{ij} \le 1)$ is the $W \times H$ transmittance matrix for the occlusion layer.
- $\mathbf{O} = \{\mathbf{o}_{ij}\} = \{o_{ijc}\} \in \mathbb{R}_+^{W\times H\times 3}$ is the color image displayed on the OST-HMD.

Here, i, j are image pixel indices, W is the image width, H is the image height, and $c \in \{r, g, b\}$ is the color channel index. In the following, we may put the index c as a subscript to image matrices and color vectors to represent their channel-wise slices and elements, respectively.

Based on these notations, we formulate our problem as shown in Fig. 2. At each color channel c, the occlusion layer attenuates the incoming light of a hazey image $\mathbf{H}_c$ by $\mathbf{T}$, and the OST-HMD adds an image $\mathbf{O}_c$. We want the resulting image $\mathbf{D}_c$ to be obtained by:

$$\mathbf{D}_c = \mathbf{T} \odot \mathbf{H}_c + \mathbf{O}_c \tag{1}$$

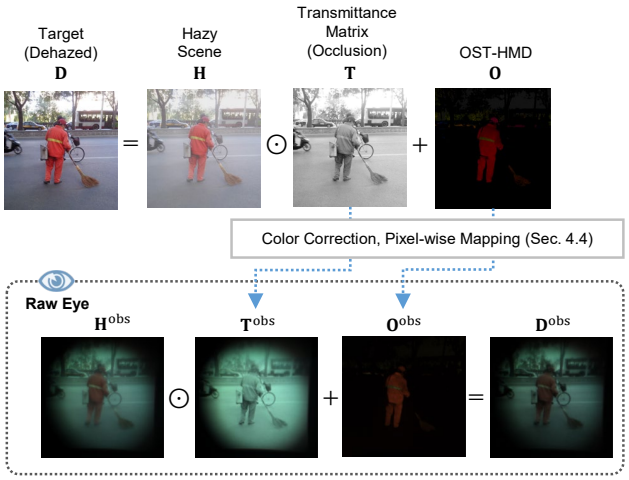

Figure 3: Overview of the computation of the transmittance matrix T and the displayed image O. (top) Computation in the captured image layer. (bottom) Observed scenes and the display images from the user view. The notations of observed images are described in Sec. 5.2. Since the observed OST-HMD image $\mathbf{O}^{\text{obs}}$ is not clearly visible in the original image, the brightness of $\mathbf{O}^{\text{obs}}$ in this figure is multiplied by 1.6.

where $\odot$ is the Hadamard (element-wise) product of two matrices of the same size. Note that the attenuation matrix $\mathbf{T}$ is shared over the color channels. Figure 3 (top) shows an overview of the computation model of $\mathbf{T}$ and $\mathbf{O}$ with an example scene.

Our problem is now to find $\mathbf{T}$ and $\mathbf{O}$ given $\mathbf{H}$ and $\mathbf{D}$. There are, however, two points we need to pay attention to in practice. First, to capture the $\mathbf{H}$, we need an extra scene camera aligned with the user's view [8]. Since the camera and the OST-HMD have different color profiles, their color spaces need to be calibrated [11, 19]. Second, $\mathbf{D}$ is unknown and needs to be estimated from a hazy scene image $\mathbf{H}$. The dehaze algorithm itself is not the focus of this paper, so we simply use one of the most promising dehaze algorithms in the community, FFA-Net [26], and treat the obtained image as a true dehazed image $\mathbf{D}$.

3.2 Basic computation for display images

The trivial solution for Eq. 1 is to set $t_{ij} = 0$ and $o_{ijc} = d_{ijc}$, which is to say that the system completely cuts the see-through view and use the OST-HMD as a VST system. Using the complete VST view leads to issues compared to the OST view [30], as mentioned in Sec. 1. We need to reduce the displayed colors on OST-HMD to be as minimal as possible while preserving the maximum use of the scene light.

We thus need to solve the following optimization problem for each pixel:

$$\text{minimize } ||\mathbf{o}_{ij}||, \quad \text{s.t. } ||\mathbf{d}_{ij} - (t_{ij}\mathbf{h}_{ij} + \mathbf{o}_{ij})|| = 0. \tag{2}$$

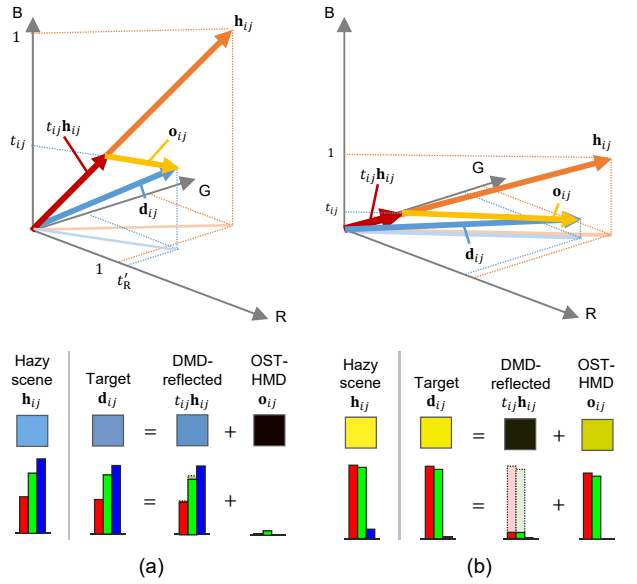

Figure 4: Visualization of the complete occlusion issue explained in Sec. 3.3 (top). The figures visualize the Eq. 1 for cases when the intensity ratio t'_c is (a) not very saturated and (b) highly saturated. The pixel value in the scene image $\mathbf{h}_{ij}$ is colored in orange, the target image $\mathbf{d}_{ij}$ in blue, the scene image reflected by the DMD $t_{ij}\mathbf{h}_{ij}$ in red, and the image displayed by the OST-HMD $\mathbf{o}_{ij}$ in yellow (bottom). RGB colors and components in each image pixel corresponding to models shown above. Note that t'_B is selected as the transmittance ratio t_{ij} in both saturation cases.

Given $\mathbf{o}_{ij} = \mathbf{d}_{ij} - t_{ij}\mathbf{h}_{ij}$ from Eq. 1 and $\mathbf{o}_{ij} \geq 0$ by definition, the above problem is reformulated as:

$$\underset{t_{ij}}{\text{argmin}} ||\mathbf{d}_{ij} - t_{ij}\mathbf{h}_{ij}||, \quad \text{s.t.}(\mathbf{d}_{ij} - t_{ij}\mathbf{h}_{ij}) \geq 0. \tag{3}$$

Given all variables are non-negative, we solve the above in the following way. We start from $t_{ij} = 0$, which meets the constraint. Gradually increasing t_{ij} from 0 only decreases the cost function, so we can safely continue increasing it until one of the color channels gives $d_{ijc} - t_{ij}h_{ijc} = 0$. Since other color channels hold $d_{ijc} - t_{ij}h_{ijc} > 0$, and we can not increase t_{ij} anymore due to the constraint, we find the solution:

$$t_{ij} = \min_{c \in (r,g,b)} t'_c, \quad t'_c = \frac{d_{ijc}}{h_{ijc}} \tag{4}$$

3.3 Alleviation of the complete occlusion issue by the occlusion layer

In the basic computation, the pixel value t_{ij} in the transmittance matrix is derived from the division of the target by the input from Eq. 4. Thus, the t_{ij} tends to be close to 0 when a color channel of the target d_{ijc} has a much smaller value than the other color channels, according to Eq. 4.

The lower value of the t_{ij} particularly affects the color compensation in pixels with high saturation, as shown in Fig. 4. When the

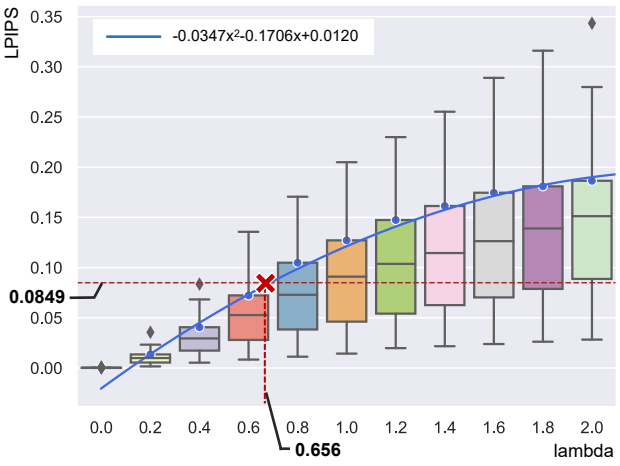

Figure 5: The determination of the penalty coefficient λ. The curve of quadratic regression is colored in blue. The red crosspoint is the intersection of the desired λ (=0.656) and the perceptually indistinguishable threshold of LPIPS (=0.0849).

pixel value in the scene image $\mathbf{h}_{ij}$ has low saturation (Fig. 4a), the transmittance ratio t_{ij} becomes higher, and the light reflected by the occlusion layer still keeps the actual scene light. Otherwise, the values of the other color channels are suppressed by the lower transmittance t_{ij} (Fig. 4b). In such pixels, the occlusion layer almost completely occludes the input, and the OST-HMD needs to add almost the same value as the target. As a result, our system works like a VST system at that pixel.

To keep the actual view even for highly saturated pixels, we introduce the alleviation term, s_{ij}, corresponding to the saturation at each pixel. This can be denoted as:

$$s_{ij} = \frac{M - t_{ij}}{M} \tag{5}$$

$$M = \max_{c \in (r,g,b)} t'_c \tag{6}$$

Then we update the t_{ij} using the s_{ij}:

$$t_{ij} \leftarrow t_{ij} + \lambda s_{ij} \tag{7}$$

where λ is the penalty coefficient.

Determination of the penalty coefficient λ. By applying the alleviation method, we compromise the replication quality of the dehaze image while suppressing the total brightness of the image shown on the OST-HMD. Since the method includes the hyper-parameter λ, we seek an objective way to choose a proper λ with reasonable justification. Note that $\lambda = 0$ is equivalent to the base method.

In general, a larger λ degrades the resulting dehaze view while reducing the total value of the OST-HMD pixels. If the degraded view is not be perceptually able to be recognized by human eyes at a glance, we may accept that result and use the corresponding λ.

One possible way to determine a λ that does not degrade the dehaze view is to increase λ for a given hazy scene until the perceptual difference between the resulting dehaze view and the view given by the base method (i.e., when $\lambda = 0$) is unnoticeable. However, this approach is impractical, as it increases the computation time excessively. Thus, instead we apply the same principle on an image dataset.

We use a public hazy image dataset (RESIDE [20][1] Synthetic Objective Testing Set). This dataset is well populated in the community and used for developing various dehaze methods [4, 26]. The dataset contains 1000 image pairs of ground truth images and their corresponding hazy images, which are synthesized by their depth images and the scattering model [3].

Figure 5 shows our process to determine λ. Using the hazy images and the approach we mentioned above, we found the best λ by measuring the perceptual similarity between the base dehazed image computed with $\lambda = 0$ and that with a different λ. We searched λ from 0 to 2.0 with a step of 0.2, and computed the third quartile (75%) of the perceptual similarity of 100 hazy images randomly chosen from the dataset. We then fit the curve to the 75% values, and found the λ that gives a perceptual similarity value that humans cannot distinguish.

For the perceptual similarity, we used the Learned Perceptual Image Patch Similarity (LPIPS) metric version 0.1 [38]. LPIPS is a metric based on deep learning and outperforms other major image metrics in terms of perception. Since LPIPS does not provide the value at which humans consider two images to be the same, we evaluated the LPIPS and decided the value for our application as follows. From the LPIPS's dataset, we took the just noticeable difference (JND) image dataset with 4800 pairs of original and visually-distorted small image patches (64x64). Each pair has a label (0-3) showing that the number of human observers that saw differences when three human observers saw each pair in a short amount of time (250 msec). On the JND dataset, we computed LPIPS values over image pairs of which the all observers could not see the difference. As the result, we get the threshold LPIPS value 0.0849 as the median of the LPIPS values over these image pairs.

By using the LPIPS and its perceptual threshold value, we obtained the threshold λ as 0.656 (Fig. 5), which we use in the evaluation in Sec. 5.

4 TECHNICAL SETUP

Here we describe our hardware/software setups and the calibration steps required for the experiment.

4.1 System Overview

Figure 6 shows a schematic diagram of our optical dehazing system. We combine a DMD-based occlusion mask with a consumer OST-HMD to selectively modulate the incoming light in both an additive and subtractive manner.

The optical design of the occlusion mask consists of a DMD, a polarized beam splitter (PBS), and a quarter waveplate. The DMD controls the reflection intensity of the incoming light by rotating several hundred thousand micro-mirrors individually, which achieves higher contrast and provides brighter images than the transmissive LCD.

To occlude the incoming light while keeping the see-through view from the eye, we implemented the optical design for the DMD similar to [14]. The PBS is made of a birefringent material and

[1] https://sites.google.com/view/reside-dehaze-datasets/

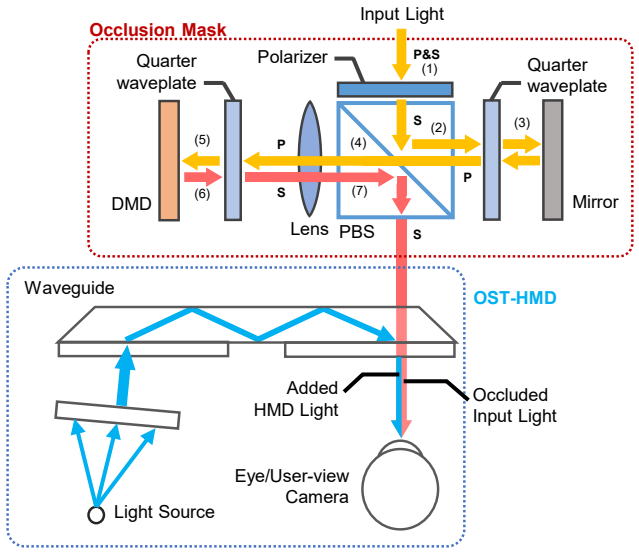

Figure 6: Schematic diagrams of our optical designs and prototype systems.

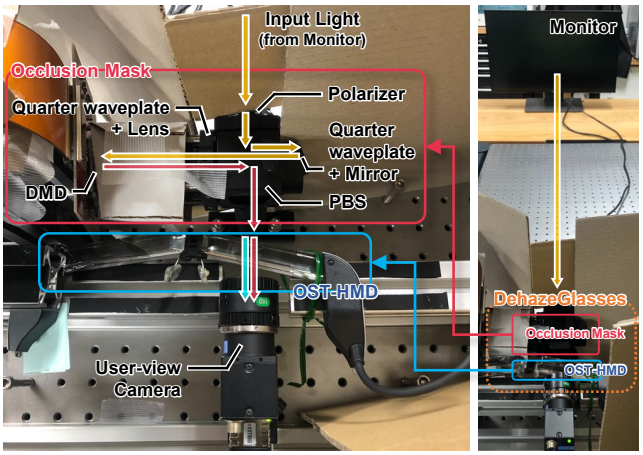

Figure 7: (left) Hardware setup with schematic visualization of light paths corresponding to Figure 6 (right). The relationship between the monitor and the entirety of our system. Input light from the hazy scene is coming from the monitor placed in front of our system.

splits the light depending on the polarization states: it transmits p-polarized light and reflects s-polarized light at an angle of 90°.

By leveraging this property of PBS, our setup realizes the following optical paths: (1) the incoming light is first s-polarized by a linear polarizer, and then (2) reflected to the right by the PBS. Then (3) the s-polarized light is converted into p-polarized light by being reflected by a mirror and passing through a quarter waveplate twice. (4) The p-polarized light passes through the PBS from right to left and also passes through the lens. After that, (5) another quarter waveplate rotates the p-polarized light into circularly polarized light, and then (6) the DMD modulates the amplitude of the light and the waveplate again rotates the light into s-polarized light. Finally, (7) the s-polarized light is again reflected by the PBS and exits in the same direction as the input light. As a result, the occlusion mask optically aligns with the view direction of the user's eye.

Note that the distance between the DMD and the lens determines where the occlusion mask is virtually located. We set this distance to equal the focal length of the lens, which means the occlusion mask is located at an infinite distance from the eye. Using the DMD for this optical design has one notable issue. The tilt angles of the micromirrors we used can only take ±12° with the tilt axis 45° diagonal to each mirror pixel. This led us to slightly tilt the DMD so that the mirrors face perpendicular to the incoming light at the path (5).

4.2 Scene Camera

In addition to the above setup, we also need a scene camera that captures the hazy view **H**. Ideally, the camera should be aligned with the viewpoint of the user's eye via a half-mirror [8]. In our proof-of-concept system, we did not include this design factor. Instead, we focused on evaluating the full potential of our see-through dehaze system by placing a user-perspective camera at the viewpoint, which serves **H**.

4.3 Hardware and Software setup

Figure 7 shows our hardware setups. For the DMD and its controller, we used DLP LightCrafter 6500 from Texas Instruments (1920 × 1080 pixels). The DMD is connected to a laptop by an HDMI interface via the controller. For the OST-HMD, we used Epson Moverio BT-30E (23° field of view diagonally, 1920 × 1080 pixels). Other components were from Thorlabs: LPVISE100-A for linear polarizers, WPQ10ME-546 for the quarter waveplates, PBS251 for the PBS, BB1-E02 for the mirror, and AC254-045-A for the lens. We installed Point Grey FL3-U3-88S2C-C (4096 × 2160 pixels) as the user-perspective camera and set its gamma value to 1.0. A monitor displaying the hazy images (3840 × 2160 pixels) was placed 1.6 m away from our system. We set the camera focus and the position of the display according to the focal plane of the OST-HMD.

4.4 System Calibration

To realize dehaze functionality that is equivalent to the formulation we denoted in Sec. 3.1, spatial and color calibration is essential in our system. Therefore, we created some look-up tables (LUTs) that reflect our system properties in advance and correct input images based on them.

4.4.1 Pixel-wise mapping between displays. In the spatial correction, we needed to warp images to display on the DMD and OST-HMD so that they exactly overlap an image on the monitor. To achieve this, we used a gray code pattern and obtained pixel correspondences between the user-perspective camera and each display. We then computed approximate projection functions with neural net fitting in MATLAB. After that, we obtained the pixel-wise mappings from the DMD and OST-HMD to the monitor as LUTs by sampling from composite functions of them.

4.4.2 Color and Gamma Correction on each display. In the color correction, we need to correct input images to the DMD, OST-HMD,

and monitor, so that the camera response to each color channel becomes linear and color balance is unified. For the OST-HMD and monitor, we displayed RGB colors one-by-one in 256 levels on each display and measured them with the camera. For the DMD, we input grayscale images in 256 levels to the DMD and measured them with the camera while displaying a white background on the monitor. We then computed approximate power functions and created LUTs by sampling from composite functions of them.

5 EXPERIMENTS

We evaluated our method through an objective evaluation with quantitative measures.

5.1 Experiment setup

To conduct a fair quantitative evaluation, we introduced two conditions in the setup.

First, we placed a user-perspective camera behind the glasses, as mentioned in Sec. 4.2. We used the camera to capture hazy images $\mathbf{H}$ and dehazed images $\mathbf{D}$. Second, we used the hazy images instead of using the real scenes. Since it is hard to collect a ground truth dehazed view of different real scenes, we substituted the real view with a monitor screen and displayed various hazy scenes. For the scenes, we again used the dataset introduced in Sec. 3.3. The images displayed on the monitor were scaled so that their short side is 2000 pixels, which is close to the diameter of the visible area of the monitor from the user-perspective camera. Given this setup, we conducted our evaluation as shown below.

5.2 Experiment procedure

Our goal is to evaluate if the observed dehazed image is indistinguishable from the true dehazed view. For clarity of the evaluation, we define several notations:

- $\mathbf{D}$ is a ground truth dehaze image from the dataset, which will be displayed on the monitor.
- $\mathbf{D}^{\text{obs}}$ is an observed ground truth dehaze image captured by the user-perspective camera while the monitor is displaying $\mathbf{D}$.
- $\widehat{\mathbf{D}}^{\text{obs}}$ is an observed dehaze image captured by the user-perspective camera while the system displaying $\mathbf{T}$ on the DMD and $\mathbf{O}$ on the OST-HMD and the monitor displaying $\mathbf{H}$.
- $\mathbf{H}^{\text{obs}}$, $\mathbf{T}^{\text{obs}}$ and $\mathbf{O}^{\text{obs}}$ are observed images on each of the displays captured by the user-perspective camera.

Note that there are two options for $\mathbf{D}$. The first is the ground truth dehazed scene images provided in the dataset. The second option is the dehazed scene images we obtain by applying the dehazing method (FFA-Net) to the hazy scene images. In the experiment, we consider both options and provide their results for purposes of comparison.

Finally, our goal is to measure the performance of $\widehat{\mathbf{D}}^{\text{obs}}$ against the $\mathbf{D}^{\text{obs}}$ for given sample scenes. Since the scene can be arbitrary, we used the image dataset in Sec. 3.3 and captured 80 different scenes. We obtained a set of $\{(\mathbf{D}_k^{\text{obs}}, \widehat{\mathbf{D}}_k^{\text{obs}})\}_k$ for $\mathbf{D}$ from the ground truth images and the dehazed images via FFA-Net.

To compare $\mathbf{D}_k^{\text{obs}}$ and $\widehat{\mathbf{D}}_k^{\text{obs}}$, we used both the L2 matrix norm and the LPIPS metric. As mentioned in Sec. 3.3, we considered the LPIPS metrics to be a more suitable perceptual metric than the L2 matrix norm to compare image pairs.

5.3 Experiment Results

5.3.1 Dehaze capability. Figure 8 shows the result images observed from our haze removal systems with different computation methods, target images, and combinations of the displays. Throughout the experiments, we observed that the choice of target images, the dehaze images from the dataset or estimated from FFA-Net, did not affect the experiment results, and the conditions always had similar results.

We further quantitatively analyzed the overall image quality of both the basic and alleviation methods. Figure 9 visualizes the distribution of the LPIPS and L2 matrix norm taken under different screen conditions. Applying either the base method or the alleviation method both clearly show that our DehazeGlasses reduced the similarity metric between the final view and the target dehazed view, and so we conclude our system successfully dehazed the user's see-through view. Our system also realized lower LPIPS mean values than the noticeable threshold (red line) that our perceptual judgment analysis provided in Sec. 3.3. In other words, this suggests that our system can dehaze the scene view to be perceptually the same as the original dehazed view.

Interestingly, only applying DMD also improved the see-through view, whereas the change adding HMD (HMD+DMD) did not further improve the view. This suggests that attenuating scene light correctly has a bigger impact on dehaze application than adding colors on top.

The impact of the full setup over the DMD-only. We then investigated in what conditions the full setup outperforms over the DMD-only setup. As explanatory examples, we evaluated two contrasting scenes where (a) the full setup outperforms over the DMD-only setup and (b) a condition opposite to the condition (a). To find such scenes, we searched scenes where the difference of the LPIPS values between the full setup and the DMD-only setup are either the largest or the smallest. We show the two scenes in Fig. 10. This result suggests that the full setup reproduces the color better when the scene contains high-saturation pixels, such as the swimming float of the yellow duck in sample scenes in Fig. 10 top. On the contrary, it was hard to perceive the difference of the scenes where the DMD-only setup is superior to the full setup (Fig. 10 bottom). In other words, the full setup is comparable to the DMD-only setup in scenes with low-saturation. The full setup is, however, more effective in scenes with high-saturation.

5.3.2 Alleviation result. The motivation for introducing the alleviation method was to have the OST-HMD display images as minimally as possible. To verify this, we drew a curve by calculating the ratio of the pixels in an OST-HMD image $\mathbf{O}$ that are lower than a given threshold. By moving the threshold from 0 to 255, we obtain a curve starting from (0,0) and ending at (255, 1.0). The area under the curve (AUC) represents how bright a given image is. When an image is completely black, then the AUC takes its maximum value. We compute this response curve for all input OST-HMD images used in the experiment and draw the mean curves for different conditions (Fig. 11).

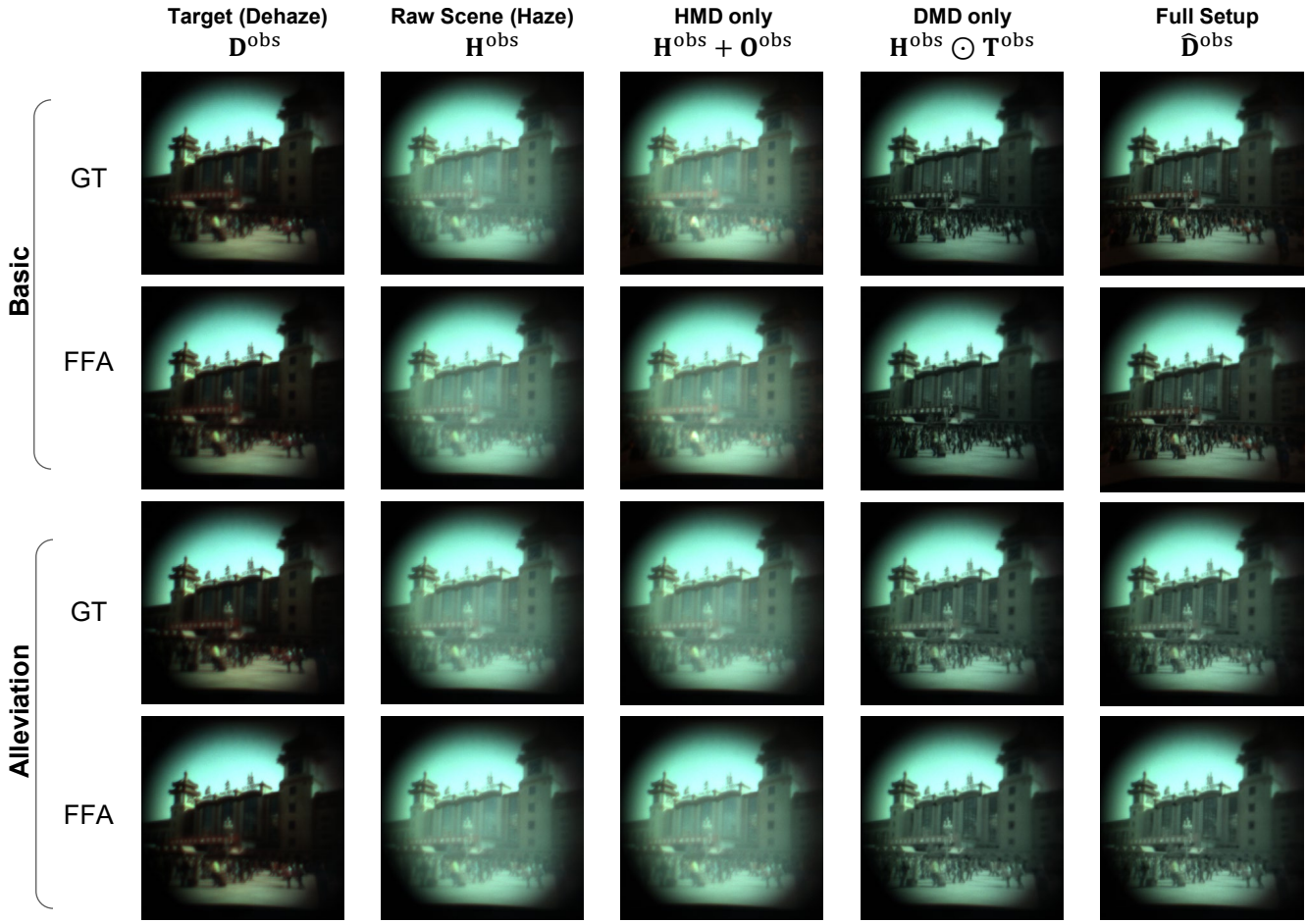

Figure 8: Results of haze removal by different combinations of the displays. Each row represents the experimental conditions: the basic computation (Sec. 3.2) or the alleviation of the complete occlusion (Sec. 3.3), and the image regarded as the target in the computation—either the ground truth dehaze image from the dataset (GT) or the estimated dehaze image using FFA-Net [26] (FFA). Each column represents the observed target scene $\mathrm{O}^{\mathrm{obs}}$, the hazy scene $\mathrm{H}^{\mathrm{obs}}$, and the observed image with different combinations of the displays. We selected the image where the difference of LPIPS between the hazy image (Raw) and the observed image (Scene+HMD+DMD) is the median from the dataset.

The results show that the AUC of the alleviation methods is clearly larger than the base method, which confirms that the alleviation method was successful in reducing the required image brightness of the OST-HMD.

6 DISCUSSIONS AND LIMITATIONS

While our proof-of-concept system showed promising results, there is still a large gap between the current system and a practical system with real-world applications. In this section, we describe the issues of the current implementation and suggest future research directions.

6.1 Aligning image screens

Our system defines several image screens (DMD, HMD, and the scene camera). As we mentioned in the calibration section (Sec. 4.4), the ideal system requires real-time dynamic calibration that calibrates the system with respect to the current position of the user's eye. Such automated calibration has been a topic of discussion in the OST-HMD community [13], but have not yet been explored deeper in occlusion-capable OST-HMD scenarios with scene cameras.

6.2 Color occlusion capability

The current system uses a DMD for occlusion. Since a DMD is essentially a dynamic mirror, each micromirror of the DMD merely controls the transmittance of the user's see-through view regardless of the colors. In other words, our occlusion system cannot attenuate a specific color band in the scene view only. If a system can attenuate each color channel separately, the system has less of a need to display compensation colors on the OST display (Sec. 3.3). We think it desirable that the OST layer renders the least intrusive images

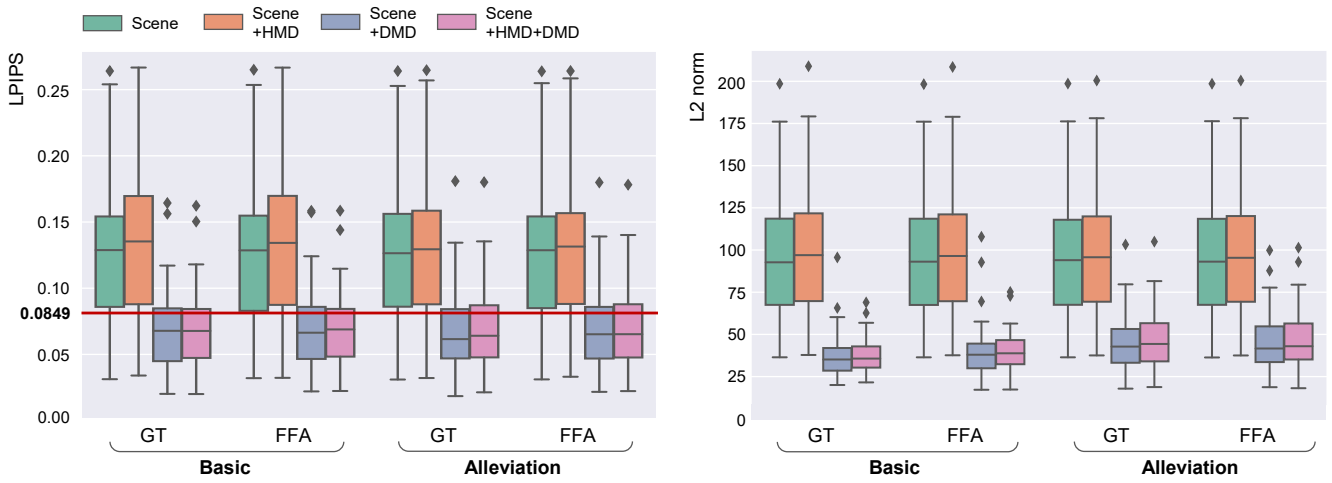

Figure 9: The overview of the dehaze results, as described in Sec. 5.3.1. (left) With the LPIPS metric. (right) With the L2 matrix norm. Over the plots, our dehaze system (HMD+DMD) clearly reduced the similarity metric between the final view and the target dehazed view, which indicates that our method successfully realized see-through view dehazing. The red line in the LPIPS plots refers to the line where our perceptual judgment analysis provided a noticeable threshold.

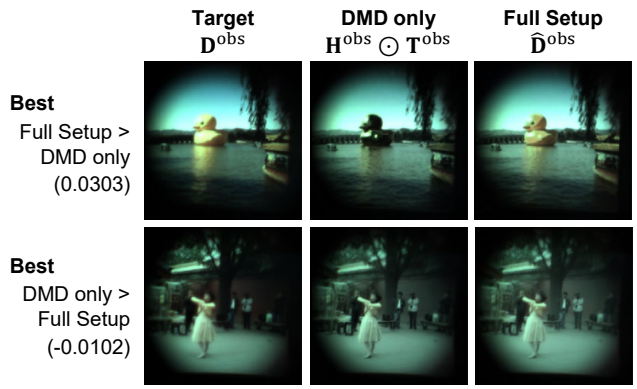

Figure 10: The results that show the scene in which the full setup is superior to the DMD-only setup (top) and the DMD-only setup is superior to the full setup (bottom), in the basic-GT condition. The values under the row legends are the differences of the LPIPS metrics between the DMD only and the full setup.

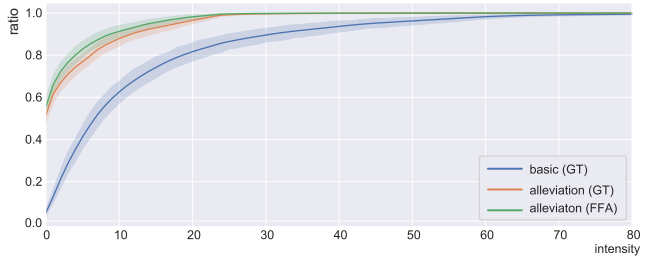

Figure 11: The intensity response curves with quartile areas for comparison between the base method and the alleviation method. The alleviation method gives curves with larger AUCs than those of the basic method, which shows that the alleviation method suppressed an excessive use of the OST-HMD. Note that we only show the results of the threshold 0 to 80 since the ratio can be regarded as 1.0 when the threshold is more than 80. We do not show the base (FFA) since it gave a results almost identical to the base (GT).

possible, following the discussion on the disadvantages of using a VST system in Sec. 1.

Existing color-occlusion displays use, for example, transmissive color LCD [28, 35] and reflective LCoS device [2, 14].

6.3 Depth of field

In practice, most of the hazy scenes that the users see are located far enough from their view to be treated as a 2D image captured from an infinite distance. Since our system renders a 2D image layer and an occlusion layer, our dehaze approach works fine when the layers are optically focused at a far distance.

This 2D scene assumption is, however, invalid when the hazy scene is close to the user's view. For example, a garden scene with fog or smoke could be hazy and contain scene objects close enough to the user to cause a deep depth of field in the user's view. In such scenarios, relying on the 2D image model causes image blur or occlusion blur [12] due to the mismatch of the screens' depth and the user's focus depth.

A possible solution to this issue is to make the system varifocal, namely making it so that the image plane of the OST display and the occlusion layer are focus-tunable. Varifocal OST-HMDs have garnered interest from the display community [17]. Researchers have also investigated ways to realize varifocal occlusion-capable displays [5, 27].

6.4 Color Correction

Calibrating the color property of each screen and the scene camera is an important issue in practice. In the evaluation, we calibrated the system in a global manner where we applied the same color pre-distortion for each pixel of a screen. Since each pixel may have slightly different color changes, the implementation of a local color calibration method that calibrates colors pixel-wise would improve system performance [18].

6.5 Scene Camera

As briefly mentioned in Sec. 4.2, the scene camera is a vital part of the system to provide the basis of the user's current first person view. There are other issues with introducing the scene camera omitted in this paper, such as view alignment, latency, and resolution.

6.6 Hardware minimization

The current version of DehazeGlasses is a proof-of-concept system built on an optical bench. For practical use, the system needs to be wearable. Although the development of a wearable OST-HMD system without occlusion is still a challenge, the community has developed some wearable designs of occlusion-capable displays [36].

7 CONCLUSION

In this work, we proposed DehazeGlasses, a see-through view dehazing system using an occlusion-capable OST-HMD. We first formulated the computation for displaying images in order to realize the dehaze functionality for our system, and explored the parameter that provides a perceptually reasonable dehazed scene while maintaining the actual view. Then, we implemented the proof-of-concept system of DehazeGlasses that consists of a DMD for the occlusion layer and OST-HMD for image overlay. We verified that the results with a dehazed image dataset show that our proof-of-concept system can clearly improve hazy views, and the perceived views from our system are difficult to distinguish from the target dehazed view. Finally, we provided the limitation of the current implementation and provided future research directions.

Our system leverages machine vision of haze removal to enhance human vision, and provides its implementation by occlusion capable OST-HMD. While we focused on the optical dehazing, our contributions potentially can be applied to other vision augmentation application that tries to bring the human vision closer to the target machine vision by using occlusion capable OST-HMDs. We hope that our system will inspire others to create new vision augmentation systems with occlusion capable OST-HMDs.

ACKNOWLEDGMENTS

This work was supported by JST PRESTO Grant Number JPMJPR17J2, Japan.

REFERENCES

[1] Cai Bolun, Xu Xiangmin, Jia Kui, Qing Chunmei, and Tao Dacheng. 2016. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Transactions on Image Processing* 25, 11 (2016), 5187–5198.

[2] Ozan Cakmakci, Yonggang Ha, and Jannick P Rolland. 2004. A compact optical see-through head-worn display with occlusion support. In *3rd IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 16–25.

[3] Earl. J. Cartney. 1977. Optics of the Atmosphere: Scattering by Molecules and Particles. *Physics Bulletin* 28, 11 (Nov 1977), 521–521. https://doi.org/10.1088/0031-9112/28/11/025

[4] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. 2018. Gated Context Aggregation Network for Image Dehazing and Deraining. *CoRR* abs/1811.08747 (2018). arXiv:1811.08747 http://arxiv.org/abs/1811.08747

[5] T. Hamasaki and Y. Itoh. 2019. Varifocal Occlusion for Optical See-Through Head-Mounted Displays using a Slide Occlusion Mask. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1961–1969. https://doi.org/10.1109/TVCG.2019.2899249

[6] Takenori Hara, Hideo Saito, and Takeo Kanade. 2009. Removal of Glare Caused by Water Droplets. *2009 Conference for Visual Media Production* (2009), 144–151.

[7] Kaiming He, Jian Sun, and Xiaoou Tang. 2011. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 12 (Dec 2011), 2341–2353. https://doi.org/10.1109/TPAMI.2010.168

[8] Yuichi Hiroi, Yuta Itoh, Takumi Hamasaki, and Maki Sugimoto. 2017. AdaptiVisor: Assisting Eye Adaptation via Occlusive Optical See-through Head-mounted Displays. In *8th Augmented Human International Conference (AH '17)*. ACM, Article 9, 9 pages. https://doi.org/10.1145/3041164.3041178

[9] Isela Howlett and Quinn Smithwick. 2017. Perspective correct occlusion-capable augmented reality displays using cloaking optics constraints: Perspective correct occlusion-capable AR displays. *Journal of the Society for Information Display* (01 2017). https://doi.org/10.1002/jsid.545

[10] Alex D Hwang and Eli Peli. 2014. An augmented-reality edge enhancement application for Google Glass. *Optometry and vision science: official publication of the American Academy of Optometry* 91, 8 (2014), 1021.

[11] Yuta Itoh, Maksym Dzitsiuk, Toshiyuki Amano, and Gudrun Klinker. 2015. Semi-parametric color reproduction method for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* 21, 11 (2015), 1269–1278.

[12] Yuta Itoh, Takumi Hamasaki, and Maki Sugimoto. 2017. Occlusion Leak Compensation for Optical See-Through Displays using a Single-layer Transmissive Spatial Light Modulator. *IEEE Transaction on Visualization and Computer Graphics* 23, 11 (2017), 2463–2473.

[13] Yuta Itoh and Gudrun Klinker. 2014. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *IEEE Symp. on 3D User Interfaces (3DUI)*. IEEE, 75–82.

[14] Yuta Itoh, Tobias Langlotz, Daisuke Iwai, Kiyoshi Kiyokawa, and Toshiyuki Amano. 2019. Light Attenuation Display: Subtractive See-Through Near-Eye Display via Spatial Color Filtering. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1951–1960. https://doi.org/10.1109/TVCG.2019.2899229

[15] Kiyoshi Kiyokawa, Mark Billinghurst, Bruce Campbell, and Eric Woods. 2003. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *2nd IEEE International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 133.

[16] Harald Koschmiere. 1924. Theorie der horizontalen sichtweite. *Beitrage zur Physik der freien Atmosphare* (1924), 33–53. https://ci.nii.ac.jp/naid/20001360955/

[17] George Alex Koulieris, Kaan Akşit, Michael Stengel, Rafał K. Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. *Computer Graphics Forum* 38, 2 (May 2019), 493–519. https://doi.org/10.1111/cgf.13654

[18] Tobias Langlotz, Matthew Cook, and Holger Regenbrecht. 2016. Real-Time Radiometric Compensation for Optical See-Through Head-Mounted Displays. *IEEE Trans. on Visualization and Computer Graphics* 22, 11 (2016), 2385–2394.

[19] Tobias Langlotz, Jonathan Sutton, Stefanie Zollmann, Yuta Itoh, and Holger Regenbrecht. 2018. ChromaGlasses: Computational Glasses for Compensating Colour Blindness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 390, 12 pages. https://doi.org/10.1145/3173574.3173964

[20] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. 2019. Benchmarking Single-Image Dehazing and Beyond. *IEEE Transactions on Image Processing* 28, 1 (2019), 492–505.

[21] Kai-En Lin, Kuang-Tsu Shih, and Homer Chen. 2017. Enhancing the perception of a hazy visual world using a see-through head-mounted device. In *2017 IEEE International Conference on Image Processing (ICIP)*. 4397–4401. https://doi.org/10.1109/ICIP.2017.8297113

[22] Andrew Maimone and Henry Fuchs. 2013. Computational Augmented Reality Eyeglasses. In *12th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 29–38.

[23] Andrew Maimone, Douglas Lanman, Kishore Rathinavel, Kurtis Keller, David Luebke, and Henry Fuchs. 2014. Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH Emerging Technologies*. ACM, 20.

[24] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2006. A Perceptual Framework for Contrast Processing of High Dynamic Range Images. *ACM Trans. Appl. Percept.* 3, 3 (July 2006), 286–308. https://doi.org/10.1145/1166087.1166095

[25] Srinivasa G. Narasimhan and Shree K. Nayar. 2002. Vision and the Atmosphere. *International Journal of Computer Vision* 48, 3 (01 Jul 2002), 233–254. https://doi.org/10.1023/A:1016328200723

[26] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. 2019. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. arXiv:cs.CV/1911.07559

[27] Kishore Rathinavel, Gordon Wetzstein, and Henry Fuchs. 2019. Varifocal Occlusion-Capable Optical See-through Augmented Reality Display based on Focus-tunable Optics. *IEEE Transactions on Visualization and Computer Graphics* 25, 11 (Nov 2019), 3125–3134. https://doi.org/10.1109/TVCG.2019.2933120

[28] TJ Rhodes, Gavin Miller, Qi Sun, Daichi Ito, and Li-Yi Wei. 2019. A transparent display with per-pixel color and opacity control. In *ACM SIGGRAPH 2019 Emerging Technologies*. ACM, 5.

[29] Jannick P Rolland and Henry Fuchs. 2000. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators & Virtual Environments* 9, 3 (2000), 287–309.

[30] Jannick P Rolland, Richard L Holloway, and Henry Fuchs. 1995. Comparison of optical and video see-through, head-mounted displays. In *Photonics for Industrial Applications*. SPIE, 293–307.

[31] Yoav Y. Schechner, Srinivasa G. Narasimhan, and Shree K. Nayar. 2001. Instant dehazing of images using polarization. In *CVPR (1)*. 325–332.

[32] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. 2003. Polarization-based vision through haze. *Applied optics* 42, 3 (2003), 511–525.

[33] Jonathan Sutton, Tobias Langlotz, and Yuta Itoh. 2019. Computational Glasses: Vision augmentations using computational near-eye optics and displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*.

[34] Robert Tamburo, Eriko Nurvitadhi, Abhishek Chugh, Mei Chen, Anthony Rowe, Takeo Kanade, and Srinivasa G Narasimhan. 2014. Programmable automotive headlights. In *European Conference on Computer Vision*. Springer, 750–765.

[35] Gordon Wetzstein, Wolfgang Heidrich, and David Luebke. 2010. Optical image processing using light modulation displays. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 1934–1944.

[36] Austin Wilson and Hong Hua. 2017. Design and prototype of an augmented reality display with per-pixel mutual occlusion capability. *Optics express* 25, 24 (2017), 30539–30549.

[37] Yuta Yamaguchi and Yasuhiro Takaki. 2016. See-through integral imaging display with background occlusion capability. *Applied Optics* 55, 3 (2016), A144–A149.

[38] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.