Gaussian Light Field: Estimation of Viewpoint-Dependent Blur for Optical See-Through Head-Mounted Displays

Yuta Itoh, Toshiyuki Amano, Daisuke Iwai, and Gudrun Klinker

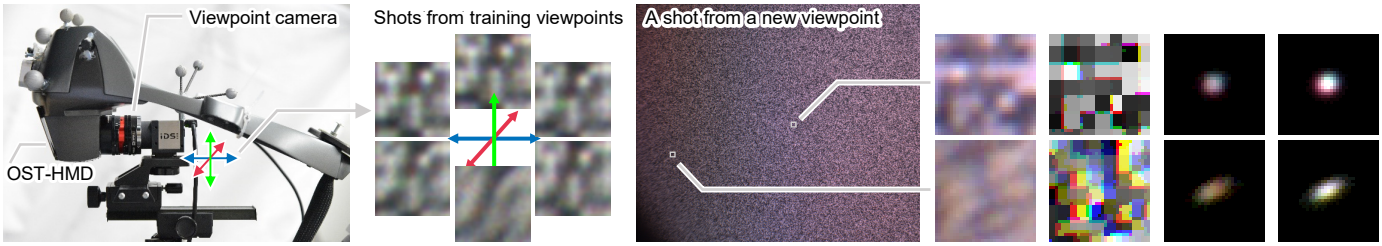

Fig. 1. An estimation result of eye position-dependent image blur of an OST-HMD by out calibration method. (a) Our experiment hardware with an OST-HMD and a viewpoint RGB camera (Sec. 4). The HMD displayed a random gray-scale noise image. (b) Images captured by the camera from different viewpoints. The patches corresponding to an identical region of the displayed image, yet blur tendencies differ among each other. (c) A captured image from a new test viewpoint. (d) Zoomed image patches. (e) Corresponding *ideal* image patches if no image blur exists (yet with optical distortion). (f) Observed Point-Spread Functions (PSFs) (g) Estimated PSFs with our Gaussian Light Field model. Note that the blur characteristics varies in not only the image space but also depending on the camera position with respect to the HMD as explained in Sec. 1.

Abstract—We propose a method to calibrate viewpoint-dependent, channel-wise image blur of near-eye displays, especially of Optical See-Through Head-Mounted Displays (OST-HMDs). Imperfections in HMD optics cause channel-wise image shift and blur that degrade the image quality of the display at a user's viewpoint. If we can estimate such characteristics perfectly, we could mitigate the effect by applying correction techniques from the computational photography in computer vision as analogous to cameras. Unfortunately, directly applying existing calibration techniques of cameras to OST-HMDs is not a straightforward task. Unlike ordinary imaging systems, image blur in OST-HMDs is viewpoint-dependent, i.e. the optical characteristic of a display dynamically changes depending on the current viewpoint of the user. This constraint makes the problem challenging since we must measure image blur of an HMD, ideally, over the entire 3D eyebox in which a user can see an image. To overcome this problem, we model the viewpoint-dependent blur as a Gaussian Light Field (GLF) that stores spatial information of the display screen as a (4D) light field with depth information and the blur as point-spread functions in the form of Gaussian kernels, respectively. We first describe both our GLF model and a calibration procedure to learn a GLF for a given OST-HMD. We then apply our calibration method to two HMDs that use different optics: a cubic prism or holographic gratings. The results show that our method achieves significantly better accuracy in Point-Spread Function (PSF) estimations with an accuracy about 2 to 7 dB in Peak SNR.

Index Terms—OST-HMD, calibration, optical see-through, chromatic aberration, point-spread functions, light field

1 Introduction

Optical aberrations of an OST-HMD could severely degrade the image quality of the display. One major effect caused by optical aberrations is the optical blur that results in undesirable, blurry images in the user's viewpoint (Fig. 1 (b)). Estimating the blur characteristics of an HMD is important since we could mitigate the effect by applying correction techniques from the computational photography as analogous to cameras (Sec. 2.3).

However, the calibration problem has two challenges that are unique to HMD systems unlike static camera systems. Firstly, existing HMDs employ various optical designs, thus their optical characteristics can be totally different for each other (Sec. 2.2). Secondly, and most importantly, the optical characteristics of HMDs are *viewpoint dependent*. The optics affect final images differently at the user's viewpoint depending on the current user's 3D eye positions in the HMD coordinates (Fig. 1 (a, b)).

The second property above makes the blur estimation problem more challenging. When an OST-HMD system, i.e. an image source and optics, and a viewpoint are fixed to each other, it is possible to measure optical characteristics in a 2D image space as analogous to cameras [22]. In practice, however, lenses (display optics) are only fixed to objects (micro display panels), not to the image planes (our eyeballs). In other words, if an eye-HMD system were a camera, it is as if we use a camera while keep moving its lens focus all the time. As a result, the dynamic change of the 3D position of eyes with respect to HMD's display panels affects how an observed image is blurred in the user's view. Therefore, we must model image blur in OST-HMD systems in a 3D spatial manner. Furthermore, the amount of blur is different for each color channel of the display.

In this work, we thereby propose a calibration method that can estimate the image blur of an OST-HMD for each color channels and for arbitrary eye positions within a given eyebox (Fig. 1 (c-g)).

- *Y. Itoh and G. Klinker are with Technical University of Munich.*
- *T. Amano is with Wakayama University.*
- *D. Iwai is with Osaka University.*

Manuscript received 29 May 2015; revised 8 Oct. 2015; accepted 18 Nov. 2015. Date of publication 8 Dec. 2015; date of current version 3 Aug. 2016. This is a preprint.

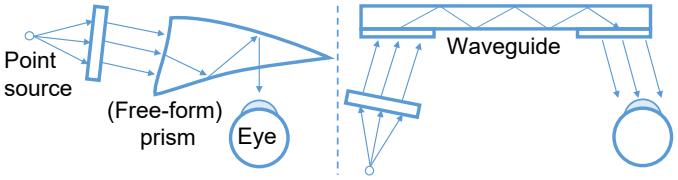

Fig. 2. Schematic diagrams of common OST-HMD designs with micro displays. Light from a point source, i.e. a pixel, is transferred to a user's eye via optical components. This could induce optical aberration at the viewpoint due to various phenomena such as refraction and diffraction. (left) With a free-form prism. (right) With a waveguide optics.

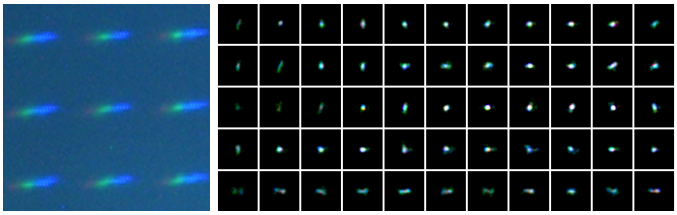

Fig. 3. Examples of impulse responses, i.e. PSFs, of an OST-HMD (Lumus DK-32). A user-perspective camera captured the display screen while drawing single pixels at various screen positions. (left) A raw image patch of the observed image that has a large chromatic aberration and image blur in the horizontal direction. (right) Various patches from different 2D positions after a noise filtering (Sec. 3.2.1).

Optics (hardware) problems can be tackled from the software side by measuring their optical characteristics a priori. For example, Heide et al. [11] showcases an aberration removal method for images photographed by simplistic convex lenses. Such lenses are lightweight, yet usually yield huge chromatic aberrations. Their method removes aberrations by locally estimating the Point Spread Functions (PSF) of lenses, and compensates their image blur through post-processing with estimated PSF priors. To transfer that valuable knowledge for imaging systems to OST-HMDs, it is crucial to establish a calibration approach that can handle optical characteristics of various OST-HMDs.

On the other hand, while their method can model aberration as local PSFs in a camera's image space, it cannot handle the view-dependent feature of OST-HMDs. Contrarily, our previous Light-Field calibration approach [17], [19] can model this 3D property for display undistortion, yet does not provide a way to model the aberration on the image plane of an OST-HMD. We find that the two approaches are neatly complementary each other.

Our proposed calibration method thus fuses the above two approaches from different domains. We demonstrate that our proposed model, the Gaussian Light Field (GLF), can handle both the viewpoint dependency and image blur in OST-HMDs simultaneously.

Note that distortion estimation is not the focus of this paper even though ignoring such distortions readily break alignment between AR contents and corresponding physical objects. Nevertheless, our calibration procedure also corrects information that can be used to the distortion estimation.

Contributions: Our main contributions include:

- Proposing a parametric, view-dependent image-blur model for OST-HMDs that stores spatial information as a light field and aberration information as point spread functions (PSF) in the HMD coordinate system (Sec. 3)
- Providing a calibration procedure to learn the model by a non-parametric regression (Sec. 4).
- Demonstrating that the learned model can predict aberrations better than a baseline method in two OST-HMDs that use completely different optics (Sec. 5).

2 Related Work

2.1 Spatial Calibration for OST-HMDs

In Augmented Reality (AR) applications, one of the long-awaited goals is to make AR experiences indistinguishable from the real world (True AR [32]). If we are exposed to such an experience, we would not be able to judge whether given stimuli are real or synthesized.

To create desired visual stimuli in the user's field of view by an OST-HMD, we must figure out diverse characteristics of a given display system such as visual (e.g. color correction [16]), temporal (e.g. tracking and rendering latency [3], [39]), and spatial characteristics (e.g. static registration between AR and real objects [2], [21]). Among them, the spatial characteristic relates to how to model the image field of a display with respect to user's eye positions for aligned rendering [2], [28]. Since eye positions are dynamic, some work integrate eye tracking into OST-HMDs so that a system can update a projection matrix for the rendering automatically [18], [30].

In previous work, we addressed on minimizing the spatial calibration error of OST-HMDs in a divide-and-conquer approach following the work by Owen et al. [28]. We first tackled the main error source due to manual HMD calibration procedures by proposing an automated calibration method based on eye tracking [18], [30].

We then further broke down the eye-HMD system, and focused on optical distortions in display optics that are another major error source after the human eye model. In recent work, we successfully eliminated spherical distortions of the direct view and image view caused by optical combiners of OST-HMDs [17], [19].

As yet another step to understand the HMD model deeper, this work now focuses on the image blur property.

2.2 OST-HMD Optics

In a common OST-HMD design with a micro display panel, an OST-HMD transfers light from its panels to user's eyes via various optical components such as prisms and waveguide plates (Fig. 2). Although optics designers aim to design HMD optics so that they can generate as sharp images as possible in predefined eyeboxes, removing various optical aberrations while meeting other design factors is still a challenging hardware issue. If we design optics that minimize aberrations, they often become bulky similar to lenses for expensive full-frame SLR cameras. This is undesirable since the total weight of an HMD is a key requirement in practice. To cope with this, Various optical designs exit, thus HMDs could hold completely different optical aberrations depending on their designs (Fig. 3 and Sec. 5).

An OST-HMD typically forms an image as the following. Let us assume that we turn on a pixel in a micro display panel and the panel is monochrome, i.e. its light consists of a single wavelength. The pixel then emits diffused light, and a magnifying optics collimates the light. The light travels through a waveguide optics to a user's eye; consequently, the eye forms the pixel's image on its retina.

The waveguides can be implemented by different optical components such as half mirrors (i.e. directly reflect the collimated light to eyes) as used in Vuzix STAR 1200XLD and in a custom HMD [13], free-form prisms [35] (Fig. 2 right), Holographic Optical

Elements (HOEs, e.g., diffraction gratings and holographic beam splitters) as used in Microsoft HoloLens and in custom HMDs [12], [26], and combination of them [6]. Thin optical plates are commonly used with HOEs to internally reflect the light (Fig. 2 left).

2.3 Potential Applications with OST-HMDs' PSFs

The prime use of the PSF information is to compute a pre-compensated image to be displayed on an OST-HMD so that an observed image becomes closer to a desired image. A key perspective we have is that an HMD-eye system can be seen as a Projector-Camera system. Under this view, knowing PSFs allows us to predict how a projected image will be degraded at a viewpoint. There are three relevant applications: projector pre-compensation [5], [8], computational displays [14], [23], and computational vision correction [1], [29]. Ideally speaking, we can maintain the image quality of an OST-HMD with estimated PSFs even if the optics is degraded.

PSFs would also help designing computational displays which can accommodate or help to optimize vision correction computation. Recently, Oshima et al. propopsed a proof-of-concept system that applies a vision correction method to an OST-HMD [27]. They demonstrate a focus correction method to improve perceptual image quality of an OST-HMD using PSFs of the display. We also demonstrated that an OST-HMD system could correct a viewer's defocused view by displaying a compensation image that is designed to cancel the defocus effect to enhance the eye sight [20]. To compute such a compensation image, the optical aberration information of an OST-HMD must be known.

In the next section, we introduce basics of optical aberrations, and introduce related calibration work on OST-HMDs.

2.4 Optical Aberration Calibration

There are five common aberrations known as the Seidel aberrations [37]: spherical, coma, astigmatism, field curvature, and distortion. Although common camera calibration methods consider the distortion often modeled as a pincushion and a radial distortion, they usually do not consider other aberrations. Our previous work [19] also considers the distortion that changes based on an eye position in a given eyebox space, but does not focus on image blur issues.

Some works investigate optical aberration issues in OST-HMDs. Ha and Rolland proposed an assessment procedure for OST-HMDs which measures the optics performance of a display by measuring the modulation transfer function [10]. Lee and Hua reported that optical distortion could affect user's depth perception in OST-HMDs since users get slight shift in their view [24]. Pohl et al. propose a chromatic distortion calibration for virtual-reality HMDs for a fixed viewpoint [31].

Spatial AR applications with projector-camera systems, such as projection mapping, have a close relationship with OST-HMDs. A mismatch between an object distance and the image focus distance from a projector often causes defocus blurs [38].

Bermano et al. proposed a radiometric compensation method for augmenting animatronics head where very precise pixel registration was required, while they pointed out that the aberration was negligible [4]. Another research, which focuses on a robust, error-tolerant radiometric compensation, also did not take the aberration into account, but regarded the drift of a projected pixel for long-term installation as a more important factor [9].

Modern projectors generally have better optics than current HMDs, because they do not have a strong size factor that limits installing sophisticated, yet bulky optical systems. Mobile projectors, however, suffer from aberration issues since they require tiny and simple optics with a space constraint. For example, Maiero et al. investigated the legibility of projected contents considering the chromatic aberration of their mobile projection system [25].

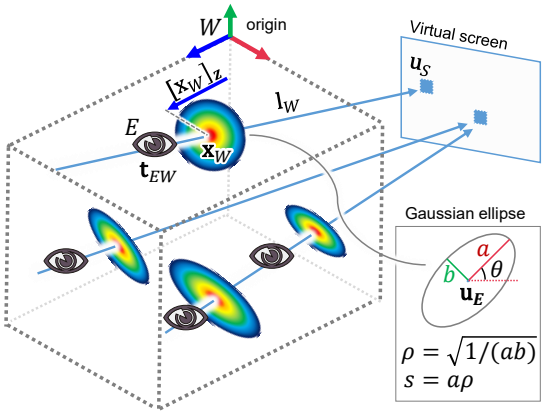

Fig. 4. A schematic diagram of our Gaussian Light Field model and its PSF parameters (Sec. 3).

3 Method

We first start from a simplistic setup with a schematic illustration in Fig. 4. A monochrome user-perspective camera E placed behind a monochrome OST-HMD W captures a gray-scale image on the HMD screen S.

The camera captures a screen point $\mathbf{u}_S$ as $\mathbf{u}_E$ in the camera image. Assuming that the camera follows a pinhole model, and given the pose of the camera with respect to the HMD coordinate system W as $(\mathtt{R}_{EW}, \mathbf{t}_{EW})$, we can conceptually define a mapping m_{ES} from the 2D screen space to the 2D image plane space as $\mathbf{u}_E = m_{ES}(\mathbf{u}_S|\mathtt{R}_{EW}, \mathbf{t}_{EW}) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. In practice, we build this mapping as look-up tables for each viewpoint at which we place the camera (Sec. 3.1).

Let $\mathtt{K}_E$ be the intrinsic matrix of the camera, then we know that the image pixel $\mathbf{u}_E$ is on a light ray $\mathbf{l}_E$, a 3D line, passing through the camera center $\mathbf{t}_{EW}$ and a backprojected point $\mathtt{K}_E^{-1}\widetilde{\mathbf{u}_E}$, where $\widetilde{\mathbf{u}_E} \in \mathbb{R}^3$ is a homogeneous vector of the camera pixel.

By defining two 3D planes in a coordinate system, then a 3D line in the system can be parametrized as a 4D vector storing intersection points of the line and the 3D planes. Given two planes in the HMD coordinate system W, and by converting the 3D points $\{\mathbf{t}_{EW}, \mathtt{K}_E^{-1}\widetilde{\mathbf{u}_E}\}$ via the camera pose $(\mathtt{R}_{EW}, \mathbf{t}_{EW})$, we get a light ray $\mathbf{l}_W \in \mathbb{R}^4$ in the HMD coordinate system corresponding to the pixels $\mathbf{u}_S$ and $\mathbf{u}_E$.

The camera E now observes a displayed image. At a camera pixel $\mathbf{u}_E$, we get an image patch $\mathtt{I}_E \in \mathbb{R}^{N\times N}$ with size N. In reality, displays suffer from image blur that deteriorate ideal image patches. Similar to a work in [11], we model the image blur as PSFs at each pixel.

A PSF can be modeled as a Gaussian kernel, a 2D filter which blurs a given image. Assuming that we have a discretized PSF $\mathtt{P} \in \mathbb{R}^{N\times N}$ corresponding to a camera image pixel $\mathbf{u}_E$, then the PSF creates a degraded image by convoluting the ideal image as

$$\mathtt{I}'_E(x,y) = \mathtt{P} \otimes \mathtt{I}_E = \textstyle\sum_{k,j} \mathtt{P}(k,j) * \mathtt{I}_E(x-k, x-j). \qquad (1)$$

The PSF $\mathtt{P}$ is correlated to both the light ray $\mathbf{l}_W$ and a 3D point $\mathbf{x}_W := \mathtt{R}_{EW}\mathtt{K}_E^{-1}(f_E * \widetilde{\mathbf{u}_E}) + \mathbf{t}_{EW}$ at which the ray intersect with the

camera's image plane (f_E is a focal length of the camera). Therefore, we get a relationship as

$$\mathtt{P} \Leftrightarrow \mathbf{l}_W,\ [\mathbf{x}_W]_z, \tag{2}$$

where $[\mathbf{x}_W]_z$ is the z element of the point. We only use the z element not the x and y because the information in the x and y elements are redundant – the light ray $\mathbf{l}_W$ contains the equivalent information implicitly.

Now our goal is to learn the relationship described by Eq. 2 over different viewpoints so that we can predict the image blur of the display at a new viewpoint. To do so, we design a data collection procedure and apply a machine learning technique to build a GLF model.

Note that our previous work in [19] essentially learns a relationship between the light ray $\mathbf{l}_W$ in the HMD coordinate system and the screen pixel point $\mathbf{u}_S$. In practice, this step is combined to associate estimated PSFs of a viewpoint to the display screen image. Our calibration procedure also collects the data required for the method in [19]. The only difference is that our procedure is done for each color channel of the display while their method works on a gray-scale image converted from an RGB.

3.1 Data Collection Procedure

We collect the dataset described by Eq. 2 for different viewpoints by using a user-perspective camera. We make assumptions that our camera is calibrated and has image resolution so that a pixel of the display can be captured by ideally more than 4 pixels of the camera. This is for a successful 2D-2D mapping estimation by structured-light patterns as introduced below. We also assume that the camera yields low chromatic blur.

We keep our explanation for a monochromatic camera setup for the simplicity. One can simply extend to a color setup by repeating the procedure for each color channel. In the experiment, we consecutively recorded each color channel every time we moved the camera to another viewpoint.

We first place a camera behind an OST-HMD, then we display the following images to be captured by the camera:

- Structured-light patterns (gray-code + sinusoidal patterns)
- Impulse images (single white dots with black background)
- Black image

The structured-light patterns give us the 2D-to-2D mapping $\mathbf{u}_E = m_{ES}(\mathbf{u}_S|\mathtt{R}_{EW}, \mathbf{t}_{EW})$ between the display image and the camera image pixels. The impulse images and the black image are used to estimate PSFs as elaborated in Sec. 3.2.

Assuming that we have tracked both the camera and the HMD, we also get the camera pose $(\mathtt{R}_{EW}, \mathbf{t}_{EW})$. We describe our tracking setup in Sec. 4.2. In addition to those data, we also display a random noise pattern for evaluations used in Sec. 5.1.

3.2 PSF Estimation

A PSF is an impulse response of an imaging system, i.e. a PSF represents how a point light source could be observed at a specific point on an imaging sensor. If we turn on a single pixel on the HMD screen with white color (an impulse image), then the light from the pixel propagates the space and casts a PSF on the camera's image plane. Given a 2D-to-2D mapping obtained by the previous section, we can extract a corresponding Region Of Interest (ROI) for a given impulse pixel (Fig. 5 top row).

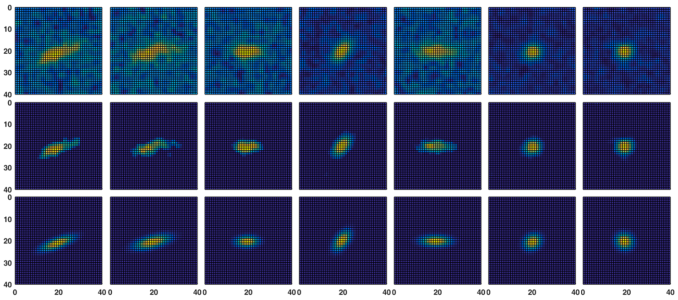

Fig. 5. Examples of PSF estimation. (top row) Observed ROIs of impulse images in a single color channel. (middle row) ROIs after the background removal. (bottom) Estimated final PSFs.

3.2.1 PSF sampling

In practice, ROI images also contain background noises. They could be from the environment light and a backlight of an OST-HMD that remain visible even if the pixel intensity is set to minimum. During the calibration, we also measure the background noises by displaying a black screen. Using this base image, we filter the original PSF ROI image by subtracting the maximum noise value of a corresponding ROI of the base image (Fig. 5 middle row). Some OST-HMDs that employ Organic LEDs do not suffer from the problem since they do not emit light when setting their color to black.

Note that we call these images after the background noise filtering as *Observed PSFs* because they represent the impulse response of each corresponding pixels, which is exactly the definition of PSFs. We use this observed PSFs in the experiment (Sec. 5.1).

After the filtering, we fit a unimodal Gaussian function to the filtered ROI image (Fig. 5 bottom row). We do so by applying the expectation maximization algorithm [7]. To apply the method, we converted the intensity image into a probabilistic density map by treating the intensity value as the frequency count. After applying the algorithm, we obtain a PSF as a 2-by-2 covariance matrix $\mathtt{C}$, which represents an ellipse by $\forall\mathbf{u}$ so that $\mathbf{u}^{\mathrm{T}}\mathtt{C}\mathbf{u} = 1$.

Following the strategy in [33], we parameterize the ellipse by its orientation θ, scale s, and axis ratio $\rho \geq 1$ as (see also Fig. 4),

$$\mathtt{C} = s\mathtt{U}\Lambda\mathtt{U}^{\mathrm{T}}, \mathtt{U} := \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \Lambda := \begin{bmatrix} \rho & 0 \\ 0 & \rho^{-1} \end{bmatrix} \tag{3}$$

These parameters can be obtained via the eigenvalue decomposition. Now the learning problem falls into estimating the three parameters θ, s, and ρ from the observations.

3.2.2 GLF modeling

As a summary, we obtain the following input and output pairs,

$$\mathbf{x} =, [\mathbf{l}_W^{\mathrm{T}}[\mathbf{x}_W]_z]^{\mathrm{T}}, \mathbf{y} = [\theta, s, \rho]^{\mathrm{T}}, \tag{4}$$

through the calibration procedure. We call a set of this pair as a Gaussian Light Field (GLF) since the dataset consists of 4D light field vectors (without radiance information) and PSFs represented by Gaussian kernels.

The advantage of the GLF model is that it naturally models both image distortions and image blur of an eye-HMD system while considering their view-point dependent nature.

Now our goal is to learn a function $f(\cdot)|\mathbb{R}^5 \rightarrow \mathbb{R}^3$ from a training dataset $\{(\mathbf{x}_k, \mathbf{y}_k)\}$.

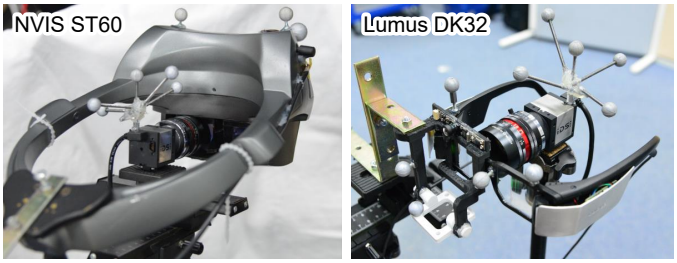

Fig. 6. Our calibration setups for data collection. (left) NVIS nVisor ST60. (right) Lumus DK-32. Marker spheres are rigidly attached to the displays and the camera (Sec. 4.1).

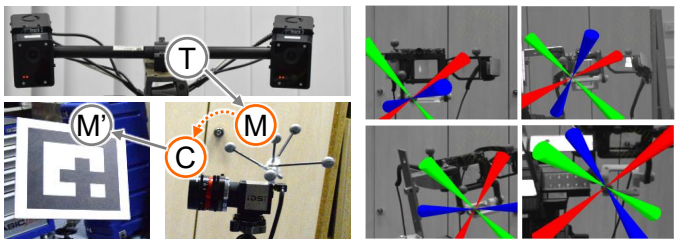

Fig. 7. Our hand-eye calibration setup. (left) A schematic diagram of the coordinate systems. (right) Visualizations of an estimated pose. With the estimated hand-eye pose, we augmented a virtual 3D cross-hair on a marker-sphere in different viewpoints correctly.

3.2.3 *Kernel regression*

To learn the GLF function f, we apply a kernel regression method on the dataset. Similar to [19], we chose a Gaussian kernel for the regression. We minimize the training error by the L2 norm with an L2 regularizer for the kernel parameters.

Since the change of PSFs are local in the HMD coordinate system, we apply a local learning strategy for one of the two HMD setup (Sec. 5.3). Given a training input set $\{\mathbf{x}_k\}$, we divide the 5D space into grids based on $\mathbf{l}_w$, thus in the 4D space. For each grid, we learn a kernel function f with training datasets whose input vectors are inside the grid. When testing a test data, we search a corresponding local function, then apply the function to the data.

3.2.4 *Baseline PSF*

As a more naive estimation, we also compute $\hat{s}$, the median of the scale s over training samples, and construct a diagonal Gaussian filter as $\mathrm{diag}([\hat{s},\hat{s}])$. We use this median Gaussian filter as a baseline PSF against our GLF-based PSF. This means that we assume the PSF is always constant over the entire eyebox of the display.

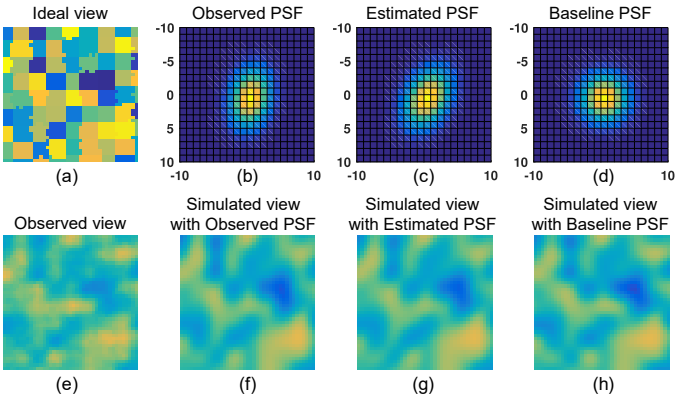

Fig. 8. Visualization of a training result. See Sec. 5.1 for details.

4 Technical Setup

We evaluate our calibration method with two OST-HMDs that employ different optics designs. Ahead to the experiment section (Sec. 5), this section describe the HMD setups and the tracking setup used for the calibration.

The experiments were done offline with Matlab R2016a on a commodity Windows 7 laptop, and computation time took some tens of seconds for a learning and few seconds for a testing.

4.1 OST-HMD Setups

The first OST-HMD is a nVisor ST60 from NVIS (Fig. 1 and Fig. 6 left), with a 1280×1024 resolution and 60° diagonal field of view (FoV). The HMD employs LCOS (Liquid Crystal on Silicon) technology, a microdisplay technology, with a one-panel design; a single LCOS panel projects each RGB color in succession with high frequency. The optics of the display is similar to those with a free-form prism (Fig. 2 left), but with a cubic prism. A semi-transparent, cubic prism in front of the user's eye reflects these light rays to the eye. The HMD has two DVI inputs for a right-eye display and a left-eye display. Our setup uses only the left-eye display.

The second OST-HMD is a DK32 from Lumus (Fig. 6 right), with 1280×720 resolution and 40° diagonal FoV. The display optics is based on polarized waveguide (See Fig. 2 right). light from the display panels go through a thin glass plate while being reflected, then sent to user's eyes from the grating exits. To ensure a larger eye box, the display have multiple grating exits aligned vertically. The left eye display is used for the current setup.

For a user-perspective camera, we used a UI-2280SE-C-HQ Rev.3 from iDS. It has a 2/3" sensor and provides 2448×2048 images, together with a 25mm C-mount lens (FL-CC2514-5M from RICOH).

4.2 Tracking Setup

To compute the training dataset according to Eq. 4, we need the camera pose $(\mathrm{R}_{EW}, \mathbf{t}_{EW})$ with respect to the OST-HMD coordinate system.

We employed an outside-in tracking system consisting of two ARTTRACK2 cameras. The system tracks constellations of retroreflective markers, marker *trees*, and give their rigid 6 Degree-of-Freedom (DoF) pose in a sub-millimeter accuracy at 60Hz. We attached a marker tree to each devices (Fig. 6).

Importantly, the tracking system only gives the pose of the marker tree on the camera, and does not give the camera pose $(\mathrm{R}_{EW}, \mathbf{t}_{EW})$ directly. Therefore, we had to calibrate the camera with respect to its marker tree. This is a typical problem in tracking as known as the hand-eye calibration [34]. We calibrated the system beforehand by using our calibration library Ubitrack [15].

Fig. 7 shows our hand-eye calibration setup where letters denote: Trackers (T), a Marker tree (M), an optical Marker (M'), and the Camera (C). We estimate the pose from the tree to the camera.

5 Experiment

5.1 Experiment Design

In this experiment, we evaluate our method on two OST-HMDs. For each HMD setup, we repeat a training step and a test step. In a training step, we collect a training dataset for a given HMD setup (Sec. 3.1). By using the dataset, we learn GLF (and the baseline PSF in Sec. 3.2.4) through the procedure described in Sec. 3.2. In a test

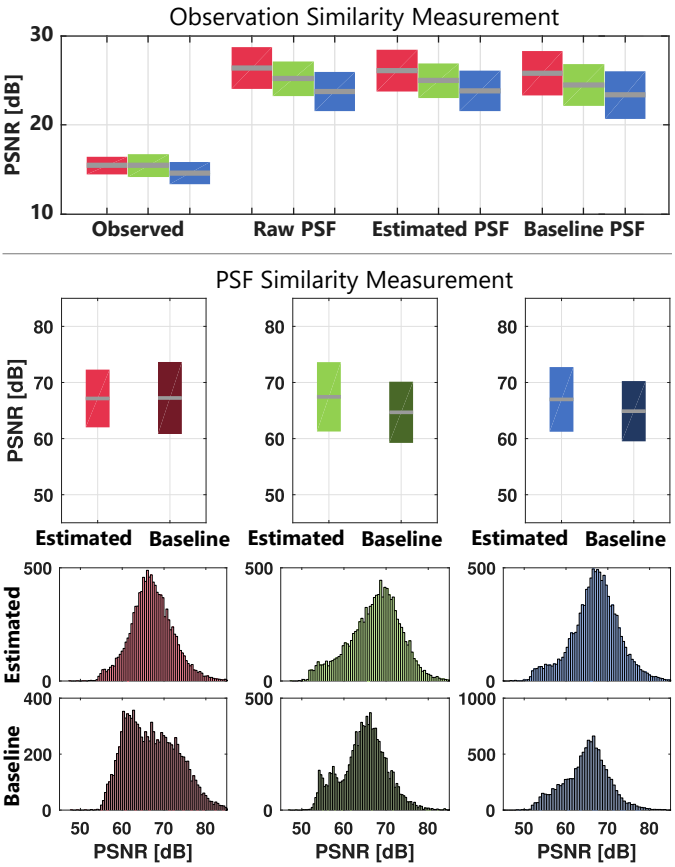

Fig. 9. Summary of the evaluation for the ST60 setup. (top) The observation similarity measurements. (bottom) The PSF similarity measurements as boxplots and their histograms. Each color in the figures corresponds to each color channel of the display. See Sec. 5.2 for the detail.

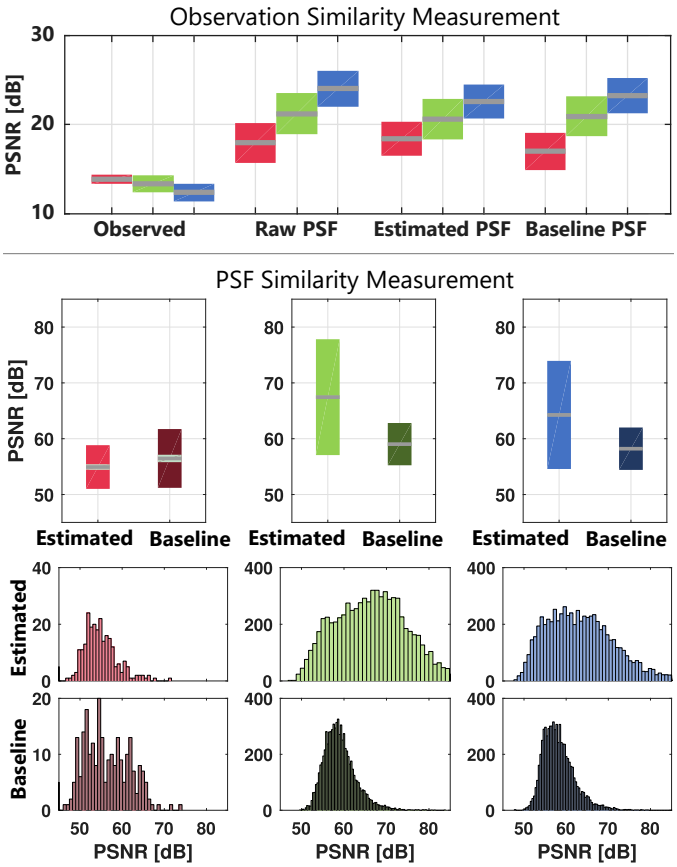

Fig. 10. Summary of the evaluation for the DK32 setup. The notations are same as Fig. 9. Note that we exclude the red channel from the evaluation due to its few sample size we could collect.

step, we collect a test dataset from a new viewpoint, and estimate their PSFs via the learned GLF. In the end, we evaluate the estimated PSFs against Observed PSFs (Sec. 3.2.1) and Baseline PSFs (Sec. 3.2.4).

To recall the definition of PSFs, we itemize them here:

- Observed PSF: Given as an 2D image of impulse response with background noise removal applied (Sec. 3.2.1)
- Estimated PSF: Estimated from our proposed GLF model.
- Baseline PSF: Estimated from a naive PSF estimation without considering viewpoint dependency (Sec. 3.2.4)

For our evaluation of the estimated PSFs, we define and use two different similarity measurements: the PSF similarity and the observation similarity.

5.1.1 *The PSF similarity*

The PSF similarity is the similarity between Observed PSFs (Fig. 8 (b)) and other PSFs, i.e. estimated and Baseline PSFs (Fig. 8 (c) and (d)), in their filter image space.

Note that we cannot compare Observed PSFs against the rest by PSF parameters – Observed PSFs are row images that are not parametrically modeled by a unimodal Gaussian function.

As image similarity, we use Peak Signal-to-Noise Ratio (PSNR). As an alternative image similarity, we also tested the structural similarity (SSIM [36]), which is motivated to correlate the quality perception of the human visual system. Since both distances resulted in a similar error tendency and displayed images are mere random noise images, we stick on the PSNR for the rest of the paper.

5.1.2 *The observation similarity*

Now we explain the observation similarity. We evaluate if the estimated PSFs result in view-point images that are similar to the actual observation. We synthesize an ideal viewpoint image by warping the reference screen image to the viewpoint by the 2D-to-2D mapping m_{ES} (Fig. 8 (a)). We then predict expected view-point images by applying the estimated PSFs (Figs. 8 (f-h)) on the ideal viewpoint image. Finally, we compare the predicted images with an actually observed image (Fig. 8 (e)). PSNR is again used for error measurement.

5.2 OST-HMD Setup with a Cubic Prism

The first evaluation is for the ST60 setup. Figure 9 summarizes the evaluation results. We collected 10 training viewpoints and 1 test viewpoint.

5.2.1 *The observation similarity results*

Figure 9 top shows boxplots of the observation similarity for each PSF types. In "Ideal Image", we compare the observed images (Fig. 8 (e)) against the ideal viewpoint images (Fig. 8 (a)), a blur-free images that the camera would see if there was no blur effects. Our results showed that PSFs with the proposed method yields significantly higher similarity measures than the baseline PSFs in all the color channel (two-sample t-test, $p \approx 0$ for the red and the green channel and $p < 0.05$ in the blue).

We further look into the PSF similarity measurement that gives another insightful analysis on the proposed method.

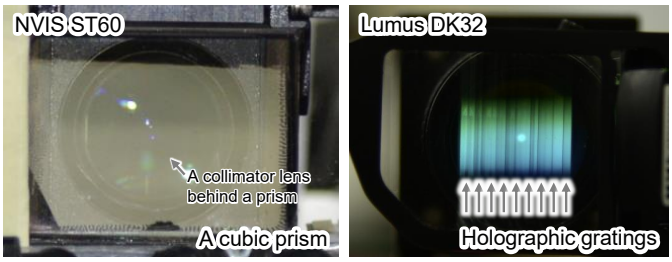

Fig. 11. Differences in the optics design of two OST-HMDs. (left) ST60 with a cubic prism. A photo of the right cubic prism of the display. A collimator lens installed on top of the prism is seen as a reflected, semitransparent image in the prism. (right) DK-32 with parallel waveguide exit plates. This HMD causes discontinuous changes in PSFs between each exit grating plates, which degenerates the GLF model based on Gaussian functions.

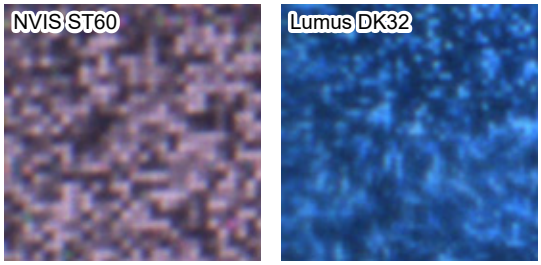

Fig. 12. Viewpoint image samples. Due to the FoV and the display quality, ST60 yields a clearer image than the DK32, which tends to include local blur.

5.2.2 *The PSF similarity results*

Figure 9 bottom shows boxplots and histograms of the PSF similarity measurements. In the green and blue channels, our method achieves significantly higher PSF similarities compared to the baseline ($p \approx 0$ for both color channels, with 2.8 and 2.1 mean dB improvements). On the other hand, the baseline method gives significantly higher mean similarity value than our method in the red channel, yet the difference is rather small (about 0.08dB).

The histograms of the measurements show additional insights that our method is superior to the baseline. In the histograms of the baseline method, several peaks exist. In the baseline method, each color channel is assigned to a single Gaussian filter, and these simplistic PSFs could not handle varying PSFs over the field of view at the given viewpoint. The peaks also suggest that the t-tests might not be appropriate for the comparisons.

On the other hand, as expected, our GLF method yields higher normality in the histograms for all the color channel. This fact supports that the proposed model successfully predicted PSFs for the new viewpoint while changing their parameters.

5.3 OST-HMD with Holographic Gratings

The second evaluation is for the DK32 setup. Fig. 10 summarizes the results. We collected 11 training viewpoints and 1 test viewpoint.

Since the display uses holographic gratings to transfer an image (Fig. 11 right), we observed that the PSFs of the display have strong locality compared to the ST60, i.e. a small change in the viewpoint position can drastically affect observed PSFs. To mitigate regression errors caused by the discontinuity of the PSFs over the input space, we decided to enforce a locality in the input space as detailed below.

We split the light field input space, $\mathbf{l}_w \in \mathbb{R}^4$, into N grids while keeping the depth feature $[\mathbf{x}_w]_z \in \mathbb{R}$ as is. The idea behind is the following. The change of a light ray could switch a corresponding

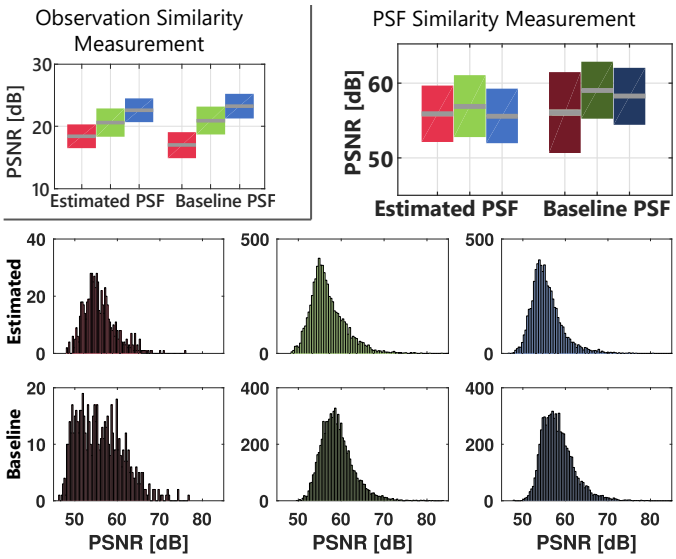

Fig. 13. Evaluation result of the DK32 setup when the local sampling was being disabled. As elaborated in Sec. 5.3.2, the regression fails to find a function which fits well to the global data at once.

waveguide plate through which the ray travels, thus a corresponding PSF could drastically change its shape. On the other hand, the change of the depth along the ray would modify the size of the PSF smoothly, thus we assume PSFs are continuous on the depth.

By heuristics of checking training errors, but not the test errors, we set $N = 25$, thus $25^4 \approx 390$k grids in the 4D space of the light field. Note that the block size N could be objectively chosen by introducing model selection scheme from the machine learning community.

Note that we disregard the red channel data in the analysis since it did not give reliable number of training samples due to the biased brightness profile of the display, which tend to create bluish color even if we display white (Fig. 12 right). Thus, we discuss only the green and blue channel data. The red channel data is, however, included in the figure for the completeness.

5.3.1 *The observation similarity results*

As shown in Fig. 10 top, our method slightly improves the mean observation similarity compared to the Baseline PSFs in the green and blue channels, and this difference was statistically significant (two-sample tests, $p \approx 0$). However, overall the PSNRs are lower than those from the ST60 setup. The reason could be due to the image quality of the display itself (Fig. 12).

5.3.2 *The PSF similarity results*

As shown in Fig. 10 below, our method improves the PSF estimation accuracy even further compared to the ST60 setup in both the green and the blue channel ($p \approx 0$, with 7.6 and 6.0 mean dB improvements), even though strong local blur existed in the DK32 as shown in Fig. 3. We observed that the local sampling strategy is necessary for a successful calibration for this display optics. We explain why the local sampling worked through an extra experiment.

Figure 13 is the extra evaluation result of the DK32 based on the same training and test datasets used in the above evaluation. The only difference is that we applied the global sampling instead of the local one, i.e. we did a regression over the entire training dataset only once. As one can see in the figure, the estimation failed to predict PSFs correctly. To investigate this failure of the global sampling, we

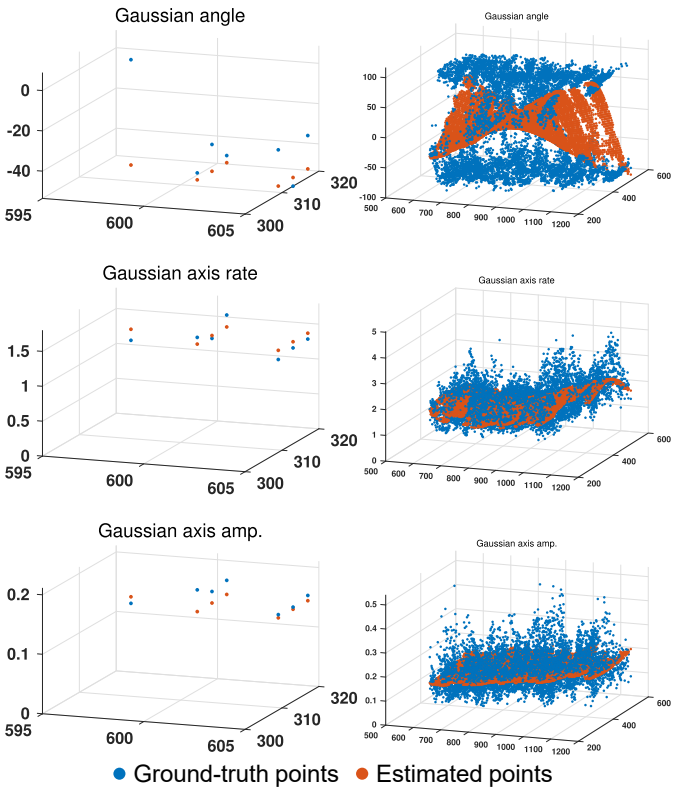

Fig. 14. Visualization of the learning results (PSFs parameters) in the viewpoint camera image pixel domain with/without the local data sampling in the DK32 setup. (left) With the local sampling (data only from one sample block). (right) Without. In the right column, due to the PSF locality of the display, our GLF estimation fails to fit *jumping* data. On the other hand, in the left column with the local sampling, our estimation nicely fits to the local, small sample sets. See also Fig. 15 to check how the sampling affects the size of local data sets.

visualize how a learned function and its training dataset distribute in the data domain. Note that we do not discuss the red channel because of the same reason explained in 5.3.

Figure 14 shows a visualization of both the learned function with the local sampling (Fig. 14 left column) and the learned function with the global sampling (Fig. 14 right column). Each row visualizes how the functions based on the two sampling approaches estimate the PSF parameters P (the axis angle, rate, and length). In each figure, the vertical axis corresponds to a PSF parameter, and the other two axes are first 2 elements of the input light ray $\mathbf{l}_W$ so that we can visualize the 5D input space ($\mathbf{l}_W$ and $[\mathbf{x}_W]_z$) in 3D with an extra output dimension. Blue dots represent measurement from a test dataset, and orange points are estimated values from each learned functions given the input data of the test dataset. For the visualization of the local sampling effect shown on the left column of Fig. 14, we took a typical regression result of a data block randomly taken from one of the N^4 local data blocks.

In the right column of Fig. 14, the learned function with the global sampling fails to fit the function to large local changes. On the other hand, the learned function with the local sampling (Fig. 14 left column), gives better estimates for this data block.

To make the global estimation work, we may tune a model parameter (the regularizer) of the kernel regression so that the regression function can express more radical changes. However, this leads to unstable estimation in practice due to the overfitting.

Figure 15 shows histograms of how many data points are used in each data block of the local-sampling for both training and test

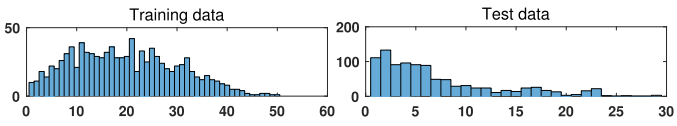

Fig. 15. Histograms of the sample size of training and test dataset used for each local kernel regression in the DK32 setup. The x axes represent sample size per local sampling blocks and the y axes are count of those blocks. If no local sampling is used, the training and the test data could be 80k and 7.1k samples respectively.

datasets of the DK32 setup. The number of samples used in one block is reduced to at most around 50 and 30 for both the training and the test data. On the other hand, the global sampling regresses a single function with 80,000 data samples and estimated 7,100 samples in the testing step.

6 Discussion and Future Work

Through the experiments, we demonstrated that our GLF method is capable to estimate the chromatic blur of OST-HMDs for a given viewpoint. Yet, there are some issues worth investigating further.

Validity of the PSF Model: In the current method, we use unimodal Gaussian kernels to characterize the PSFs. This model could be too simplistic to represent certain optical aberration types. For example, a comatic aberration is typically wedge-shaped, thus cannot be modeled by a single Gaussian kernel. If a user's viewpoint is too close to the display, an image would appear with a front bokeh, which is usually ring shaped, again can not be modeled by a Gaussian kernel. Another point to be improved is that our PSFs do not optimize the amplitude either since they are always normalized to one.
A possible option for handling these complex PSF shapes is to use a richer function model such as Gaussian mixture model. However, a more complex model might affect the computation time.

Eye Orientation and GLF: There is another issue in the definition of our GLF model. The current GLF model implicitly assumes that the viewpoint camera is frontal to the HMD screen. In other words, we treated the model as if the orientation of the viewpoint-camera does not affect observed PSFs, which is not necessarily true. If the camera (the eye) rotates, then orientation of the angle of its lens (the eye's crystal lens) changes, thus the observed PSFs could also vary. Therefore, we would need to integrate the orientation of a viewpoint in the GLF learning step. This potential effect should be investigated deeper in future work.

Locality in the Regression: Although our local sampling technique works moderately for the DK32 setup, it is worth considering to apply a more sophisticated regression algorithm to improve the estimation performance. For example, in the DK32 setup, it would be reasonable to select subsets of data (4D light rays) so that each data corresponds to a particular waveguide plate(s). In some OST-HMDs with waveguide plates, an exit piece consists of several optics that are aligned side-by-side, and their optical effects are not necessarily continuous. Thus learning independent GLFs for each of the optical components would be a reasonable extension of the current method.

Other Remarks: Although it is not the focus of this work, we have not built a system that can achieve the GLF estimation real-time with a dynamic viewpoint change. Eye positions change all the time during the use of HMDs in the daily life as analogous to eyeglasses. Thus a real-time system would be necessary for more practical continuous use such as deblurring systems [20], [27].

We could also extend the method to correct the direct-view blur, i.e. chromatic blur of user's real-world view that the user sees through semitransparent optics of OST-HMDs. We could lern this by observing reference objects in the real world through HMDs.

Our method is also applicable to virtual reality displays to collect their chromatic blur that tends to be stronger than that of OST-HMDs.

7 Conclusion

This work presents a calibration method to estimate the optical characteristics of a given OST-HMD especially the image blur that are viewpoint-dependent. A challenging and unique problem in OST-HMDs (also in virtual reality HMDs) is that the optical characteristic of a display is dependent on a user's viewpoint. No optical systems with static optical components can easily be designed to meet all optical requirements for all possible eye positions.

Instead of approaching this issue from the hardware side, we tackled this problem from a software side and proposed a viewpoint-dependent image blur model so-called Gaussian Light Field (GLF). We demonstrated that our GLF model is capable to estimate the chromatic blur of two different OST-HMDs with different optics.

8 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] M. Alonso Jr and A. Barreto. Pre-compensation for high-order aberrations of the human eye using on-screen image deconvolution. In *Engineering in Medicine and Biology Society. Proceedings of the 25th Annual International Conference of the IEEE*, volume 1, pages 556–559. IEEE, 2003.

[2] R. Azuma and G. Bishop. Improving static and dynamic registration in an optical see-through HMD. In *Proceedings of SIGGRAPH*, pages 197–204. ACM, 1994.

[3] R. T. Azuma. *Predictive tracking for augmented reality*. PhD thesis, Citeseer, 1995.

[4] A. Bermano, P. Brüschweiler, A. Grundhöfer, D. Iwai, B. Bickel, and M. Gross. Augmenting physical avatars using projector-based illumination. *ACM TOG*, 32(6):189:1–189:10, Nov. 2013.

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

[6] D. Cheng, Y. Wang, C. Xu, W. Song, and G. Jin. Design of an ultra-thin near-eye display with geometrical waveguide and freeform optics. *Optics express*, 22(17):20705–20719, 2014.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[8] M. Grosse, G. Wetzstein, A. Grundhöfer, and O. Bimber. Coded aperture projection. *ACM TOG*, 29(3):22, 2010.

[9] A. Grundhofer and D. Iwai. Robust, error-tolerant photometric projector compensation. *Image Processing, IEEE Transactions on*, 24(12):5086–5099, Dec 2015.

[10] Y. Ha and J. Rolland. Optical assessment of head-mounted displays in visual space. *Applied optics*, 41(25):5282–5289, 2002.

[11] F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb. High-quality computational imaging through simple lenses. *ACM TOG*, 32(5):149, 2013.

[12] K. Hong, J. Yeom, C. Jang, J. Hong, and B. Lee. Full-color lens-array holographic optical element for three-dimensional optical see-through augmented reality. *Optics letters*, 39(1):127–130, 2014.

[13] H. Hoshi, N. Taniguchi, H. Morishima, T. Akiyama, S. Yamazaki, and A. Okuyama. Off-axial hmd optical system consisting of aspherical surfaces without rotational symmetry. In *Electronic Imaging: Science & Technology*, pages 234–242. Int'l Society for Optics and Photonics, 1996.

[14] F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar. Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays. *ACM TOG*, 33(4):59, 2014.

[15] M. Huber, D. Pustka, P. Keitler, F. Echtler, and G. Klinker. A system architecture for ubiquitous tracking environments. In *Proceedings of IEEE and ACM ISMAR*, pages 1–4. IEEE Computer Society, 2007.

[16] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker. Semi-parametric color reproduction method for optical see-through head-mounted displays. *IEEE TVCG*, 21(11):1269–1278, 2015.

[17] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker. Simultaneous Direct and Augmented View Distortion Calibration of Optical See-Through Head-Mounted Displays. In *Proceedings of IEEE ISMAR*, pages 43–48, 2015.

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

[19] Y. Itoh and G. Klinker. Light-field correction for spatial calibration of optical see-through head-mounted displays. *IEEE TVCG (Proceedings Virtual Reality 2015)*, 21(4):471–480, April 2015.

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

[21] A. L. Janin, D. W. Mizell, and T. P. Caudell. Calibration of head-mounted displays for augmented reality applications. In *Virtual Reality Annual International Symposium*, pages 246–255. IEEE, 1993.

[22] N. Joshi, R. Szeliski, and D. J. Kriegman. Psf estimation using sharp edge prediction. In *CVPR 2008.*, pages 1–8. IEEE, 2008.

[23] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar. Content-adaptive parallax barriers: optimizing dual-layer 3d displays using low-rank light field factorization. *ACM TOG*, 29(6):163, 2010.

[24] S. Lee and H. Hua. Effects of configuration of optical combiner on near-field depth perception in optical see-through head-mounted displays. 22(4):1432–1441, 2016.

[25] J. Maiero, E. Kruijff, A. Hinkenjann, and G. Ghinea. Picozoom: A context sensitive multimodal zooming interface. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, pages 1–6, June 2015.

[26] H. Mukawa, K. Akutsu, I. Matsumura, S. Nakano, T. Yoshida, M. Kuwahara, K. Aiki, and M. Ogawa. 8.4: distinguished paper: a full color eyewear display using holographic planar waveguides. In *SID Symposium Digest of Tech. Papers*, volume 39, pages 89–92. Wiley Online Library, 2008.

[27] K. Oshima, D. C. Rompapas, K. Moser, E. Swan, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato. Sharpview: Improved legibility of defocussed content on optical see-through head-mounted displays. In *Demo under submission for: ACM International Symposium on Mixed and Augmented Reality*, 2015.

[28] C. B. Owen, J. Zhou, A. Tang, and F. Xiao. Display-relative calibration for optical see-through head-mounted displays. In *Proceedings of IEEE ISMAR*, pages 70–78, 2004.

[29] V. F. Pamplona, M. M. Oliveira, D. G. Aliaga, and R. Raskar. Tailored displays to compensate for visual aberrations. *ACM TOG*, 31(4):81:1–81:12, July 2012.

[30] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura. Corneal-imaging calibration for optical see-through head-mounted displays. *IEEE TVCG*, 21(4):481–490, 2015.

[31] D. Pohl, G. S. Johnson, and T. Bolkart. Improved pre-warping for wide angle, head mounted displays. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, pages 259–262. ACM, 2013.

[32] C. Sandor, M. Fuchs, A. Cassinelli, H. Li, R. Newcombe, G. Yamamoto, and S. Feiner. Breaking the barriers to true augmented reality. *arXiv preprint arXiv:1512.05471*, 2015.

[33] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *Image Processing, IEEE Transactions on*, 16(2):349–366, 2007.

[34] R. Y. Tsai and R. K. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *Robotics and Automation, IEEE Transactions on*, 5(3):345–358, 1989.

[35] Q. Wang, D. Cheng, Y. Wang, H. Hua, and G. Jin. Design, tolerance, and fabrication of an optical see-through head-mounted display with free-form surface elements. *Applied optics*, 52(7):C88–C99, 2013.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[37] J. C. Wyant and K. Creath. Basic wavefront aberration theory for optical metrology. *Applied optics and optical engineering*, 11(s 29):2, 1992.

[38] L. Zhang and S. Nayar. Projection defocus analysis for scene capture and image display. *ACM TOG*, 25(3):907–915, 2006.

[39] F. Zheng, T. Whitted, A. Lastra, P. Lincoln, A. State, A. Maimone, and H. Fuchs. Minimizing latency for augmented reality displays: Frames considered harmful. In *Proceedings of IEEE ISMAR*, pages 195–200. IEEE, 2014.