Simultaneous Direct and Augmented View Distortion Calibration of Optical See-Through Head-Mounted Displays

Yuta Itoh*
Technical University of Munich

Gudrun Klinker†
Technical University of Munich

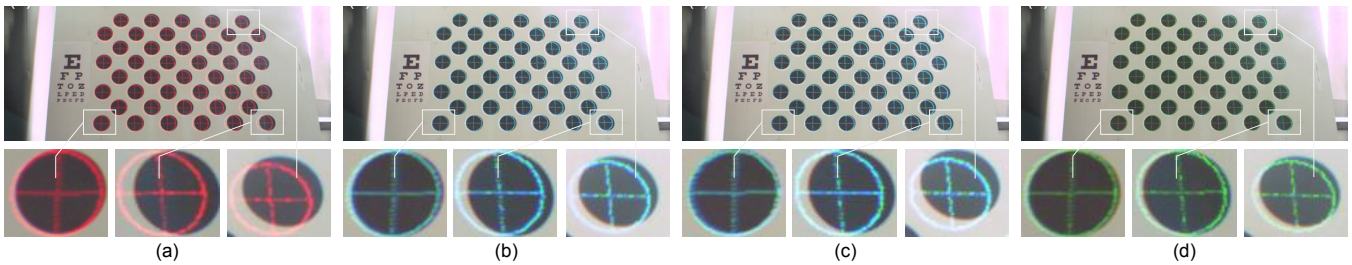


Figure 1: Demonstration of our calibration method (with a professional OST-HMD setup). A user-perspective camera captures the world and screen images from a viewpoint of test data. (a) Conventional automated calibration (Full ~~IN~~DICA [6]). (b) With image distortion corrected. (c) With scene distortion corrected. (d) Our proposed method correcting both of the two distortions simultaneously.

ABSTRACT

In Augmented Reality (AR) with an Optical See-Through Head-Mounted Display (OST-HMD), the spatial calibration between a user's eye and the display screen is a crucial issue in realizing seamless AR experiences. A successful calibration hinges upon proper modeling of the display system which is conceptually broken down into an eye part and an HMD part. This paper breaks the HMD part down even further to investigate optical aberration issues. The display optics causes two different optical aberrations that degrade the calibration quality: the distortion of incoming light from the physical world, and that of light from the image source of the HMD. While methods exist for correcting either of the two distortions independently, there is, to our knowledge, no method which corrects for both simultaneously.

This paper proposes a calibration method that corrects both of the two distortions simultaneously for an arbitrary eye position given an OST-HMD system. We expand a light-field (LF) correction approach [8] originally designed for the former distortion. Our method is camera-based and has an offline learning and an online correction step. We verify our method in exemplary calibrations of two different OST-HMDs: a professional and a consumer OST-HMD. The results show that our method significantly improves the calibration quality compared to a conventional method with the accuracy comparable to 20/50 visual acuity. The results also indicate that only by correcting both the distortions simultaneously can improve the quality.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

1 INTRODUCTION

Maintaining the consistency between virtual objects and the real world (e.g. visually, temporally, and spatially) is the fundamental requirement for realizing an immersive, indistinguishable AR experience with OST-HMDs. The spatial consistency requires an accurate calibration between an OST-HMD and a user's eye, i.e. the registration between a physical object in the direct view and a corresponding virtual content in the augmented view. The successful calibration hinges upon a proper modeling of the eye-HMD system – a virtual projector-camera system with an image screen of the HMD as the projector screen and a user's eye as the camera.

Previous works [13, 3, 4, 6, 7] break the system down into an eye and HMD part, and model them separately. The eye part involves modeling an eye optics and locating the eyeball w.r.t. the display coordinate system [6, 7, 14]. The HMD part involves modeling the image screen. The previous works estimate model parameters by measuring the eye and/or the HMD screen by cameras. This work focuses on the latter with this objective, camera-based approach.

For calibrating the HMD part, an issue often over-looked is two optical aberrations caused by the optical media of OST-HMDs: Direct-View Distortion (DVD) and Augmented-View Distortion (AVD) (See Fig. 2 too). DVD is the distortion of incoming light rays from the real world through an OST-HMD to a user's eye and AVD is the image distortion of a perceived image. A common OST-HMD design employs an optical combiner to guide the light from a light source of an OST-HMD to a user's eye [16]. As the result, the user perceives the light ray as if it appears as a virtual image floating mid air or at infinity in the user's view (Fig. 2 left). Since the combiner is an optical medium, it inevitably refracts light rays passing through itself [16] including those from a physical object in the world, i.e. DVD, and image light rays from the OST-HMD, i.e. AVD (Fig. 2 right).

The intricacy of these distortions is that the amount of each distortion depends on where and at what angle a light ray hits the combiner and passes through its medium, i.e., each user viewpoint suffers from different amounts of both distortions. Furthermore, due to the imperfection of the display optics, the image light rays may suffer from additional aberration out of the design during its passage through various optical media from the light source to the combiner.

Importantly, DVD and AVD share the same distortion characteristic in part (Fig. 2). Consider a light ray from an eye towards the optical combiner, the ray first reaches to the half-mirror of the combiner while refracted by the medium. The ray is then split into two paths: one towards the world and the other the light source while receiving additional aberrations separately. D/AVD thus consist of a shared distortion part and a individual part. This paper and existing works, however, do not explicitly separate these mixture model.

*e-mail: itoh@in.tum.de
†e-mail: klinker@in.tum.de

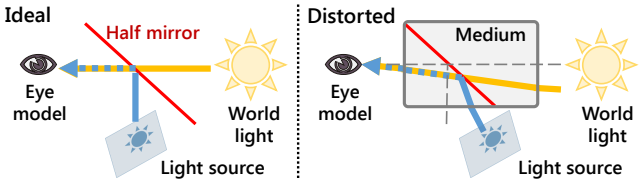


Figure 2: Schematic visualization of optical aberrations in an OST-HMD system. (Left) Ideal case. (Right) A practical case where the A/DVD present in the system. If we back-trace a ray from an eye, one notices that both distortions partially share a distortion path.

While methods exist for correcting either of the distortions independently, there is, to our knowledge, no method handling both distortions simultaneously for an arbitrary eye position w.r.t. an HMD screen.

Adapting existing AVD correction methods to DVD is not straightforward. They model the image screen of an OST-HMD as a 3D plane/surface for modeling AVD. Such modeling does not transfer to DVD directly. Our recent work proposes a camera-based calibration which corrects DVD for arbitrary eye positions based on a 4D light-field model [8]. This method treats DVD as a mapping between original and distorted light-fields; learns the mapping via non-parametric regression from a training data set offline; and computes DVD for a given new eye-position online. Under the light-field model, the AVD can be treated in a similar manner.

We propose a camera-based calibration method that corrects both distortions simultaneously for arbitrary eye positions given an OST-HMD system. Our method adapts the light-field approach to AVD and has an offline/online step. The offline step learns a cascaded mapping which consists of two light-field mappings corresponding to each distortion. The online step applies the cascaded mapping to given 3D world points and returns 2D image points. The 2D points will appear on a distorted image plane and will match the 3D points, which are also distorted, from the user's current viewpoint.

The evaluation with two OST-HMDs (a professional and a consumer OST-HMD) show that our model significantly improves the calibration quality compared to a conventional HMD model and the previous DVD-only model. The results also indicate that only correcting both distortions simultaneously can improve the quality.

We discuss limitations of the current approach mostly due to the limited capabilities of current OST-HMDs, and conclude by noting some open questions toward practical OST-HMD calibrations.

Contributions: 1. Providing a calibration method which corrects the D/AVD of OST-HMDs simultaneously for arbitrary eye positions. **2.** Demonstrating that the method improves the calibration quality of two OST-HMDs. The qualities are comparable up to a human eye of 20/50 visual acuity. **3.** Showing, with a reasoning, that only correcting both distortions can improve the final quality.

2 Related Work

A key of successful calibration is how to model the eye-HMD system. Although eye model is equivalently important, this section focuses on the HMD model which holds both DVD and AVD.

2.1 Direct-View Distortion of OST-HMD

Our previous work [8] tackles the DVD by employing the light-field correction approach. Our approach aims at capturing the optical characteristics of OST-HMDs as the shift of optical rays.

2.2 Augmented-View Distortion of OST-HMDs

A user perceives an image of an OST-HMD as a virtual screen floating in mid air. For correct registration, we need to know how this image appears in a user perspective view. We categorize existing approaches in three types: parametric and semi-/non-parametric.

Parametric Approach Parametric approaches model the screen images as a certain class of functions. A common approach is to treat the image produced by an OST-HMD as a 3D plane floating in mid air [17, 3, 2, 9, 1]. Under this model, an OST-HMD system is treated as an off-axis pinhole camera.

This model, however, is incapable of describing real OST-HMD optics. Owen et al. [13] demonstrate that the plane model does not coincide with a 3D geometry of the display measured via triangulation, and they propose a curved 3D surface model. Their surface model respects the spherical distortion caused by a curved mirror in their OST-HMD, which falls into the first-order radial distortion model. Robinett and Rolland [15] use the same distortion model. Hua et al. [5] apply a similar model to their head-mounted projective display. Lee and Hua [11] extend the surface model to 6th order radial distortions and tangential distortion.

These approaches have a common drawback in the change of the user's eye position w.r.t the HMD screen. Since the optics of an OST-HMD may distort the light of an image pixel differently at different viewpoints, the above models, learned at a single viewpoint, can cause registration errors when the eye position changes.

Semi-Parametric Approach Wientapper et al. [18] propose a semi-parametric model for Head-Up Displays (HUDs). HUDs are essentially the same as OST-HMDs except that their images are reflected on the wind shields of vehicles. Their model combines the 3D-plane and a view-dependent polynomial model. The latter employs a higher-order polynomial function of 5 parameters: a 2D image point and a 3D eye position. We call their model semi-parametric since their polynomial model is essentially non-parametric, which is based on local data points and is represented by a linear sum of polynomial kernels.

Non-Parametric Approach Klemm et al. [10] upgrade the 3D plane model by triangulating every pixel of an OST-HMD screen via the photogrammetry with structured image patterns. Their non-parametric, point-cloud approach requires an additional user adaptation since a few millimeters of error in the viewpoint position causes non-negligible registration errors.

Recall that the refraction of an optical medium causes the AVD. The amount of refraction depends on which path the ray takes through in the medium – a light field (LF) of an image changes the shape based on the user's eye position.

We adapt our original light-field model used for the DVD – we model the image screen as a function of light rays defined by 4D rays and learn the function via non-parametric kernel regression.

3 Method

We first explain the DVD and the AVD correction separately. We then introduce a unified approach. The notations are same as [8].

3.1 Direct-View Distortion Correction

Following our previous work [8], let ${}^{D}l_{W}$ a light ray in the world coordinate system as a 4D Lumigraph: ${}^{D}l_{W} := l(\mathbf{t}_{EW}, \mathbf{x}^{s}_{W}, \mathrm{R}_{SW}, \mathbf{t}_{SW}) := l(\mathbf{x}^{s}_{W}) \in \mathbb{R}^4$. See Sec. 3.3 of [8] for the exact definition.

The DVD causes a distorted ray ${}^{D}l'_{W}$. Given a set of $({}^{D}l_{W}, {}^{D}l'_{W})$ measured from various viewpoints within the eyebox of the HMD, our LF correction approach in [8] gives a function ${}^{D}f(\cdot)$ so that ${}^{D}f({}^{D}l_{W})$ is close to ${}^{D}l'_{W}$ via a non-parametric kernel regression.

3.2 Augmented-View Distortion Correction

We measure the LF of the OST-HMD screen in a similar way as in the previous section. Instead of letting a camera see 3D world points through the medium of the HMD, we let the camera capture the image of the image screen such that the camera can identify which image pixel is corresponding to that of the camera sensor.

Let $\mathbf{u}^{s}_{E} \in \mathbb{R}^2$ be an image pixel of a camera which corresponds to $\mathbf{u}_{S} \in \mathbb{R}^2$, an image pixel of the virtual screen of the HMD. Let

Figure 3: Hardware setup. (Left) nVisor ST60. (Right) Moverio BT-100. The camera on the left image is displaced for the photography; it was set closer to the screen during actual data collections.

$\mathtt{K}_E \in \mathbb{R}^{3\times3}$ be the intrinsic matrix of the user-perspective camera located at $\mathbf{t}_{EW}$. We compute a point $\mathtt{K}_E^{-1}\widetilde{\mathbf{u}}_E^s$, where $\widetilde{\cdot}$ denotes the homogeneous vector. Given a 6DoF pose between the eye and the world coordinate systems as $(\mathtt{R}_{EW}, \mathbf{t}_{EW})$, an eye sees a ray ${}^A l'_W$ as

$$ {}^A l'_W := l(\mathbf{t}_{EW}, \mathtt{R}_{EW}\mathtt{K}_E^{-1}\widetilde{\mathbf{u}}_E^s + \mathbf{t}_{EW}, \mathtt{R}_{SW}, \mathbf{t}_{SW}) = l(\mathtt{R}_{EW}\mathtt{K}_E^{-1}\widetilde{\mathbf{u}}_E^s + \mathbf{t}_{EW}). \quad (1) $$

${}^A l'_W \in \mathbb{R}^4$ is a *distorted* ray since $\mathbf{u}_E^s$ contains the AVD already. We define an original ray ${}^A l_W$ *virtually*: a ray that would appear as $\mathbf{u}_S$ if there were not for AVD and if the 3D plane model were correct. Let α a scale parameter with a unit of [meter/pixel], then ${}^A l_W$ becomes,

$$ {}^A l_W := l(\mathbf{t}_{EW}, \alpha\mathtt{R}_{SW}\widetilde{\mathbf{u}}_S + \mathbf{t}_{SW}, \mathtt{R}_{SW}, \mathbf{t}_{SW}) = l(\alpha\mathtt{R}_{SW}\widetilde{\mathbf{u}}_S + \mathbf{t}_{SW}) \in \mathbb{R}^4. \quad (2) $$

Finally, we learn a function ${}^A f^{-1}(\cdot)$ so that ${}^A f^{-1}({}^A l'_W)$ is close to ${}^A l_W$. We now introduce a way to correct D/AVD simultaneously.

3.3 Unified Distortion Correction

For aligned visualization, a world point and a corresponding image point must eventually travel along the same light path from the combiner to the eye. However, ${}^D l_W$ reaches the user's eye as ${}^D f({}^D l_W)$ after having undergone the DVD; the corresponding virtual ray ${}^A l_W$ as ${}^A f({}^A l_W)$ after AVD. The virtue of these LFs is that they are defined in a common coordinate system: they share the same u-v and the s-t plane of the 4D lumigraph. We can thus directly bypass both distortion effects.

Our goal is to achieve ${}^A f({}^A l_W) =^D f({}^D l_W)$. Since ${}^D l_W$ is defined in the same space as ${}^A l_W$, we may apply the inverse of the AVD: ${}^A l_W =^A f^{-1}({}^D f({}^D l_W))$. Thus, if $\mathbf{u}_S$ corresponds to ${}^A l_W$, then $\mathbf{u}_S$ and the corresponding 3D point in the world align in the user's view. By definition, $\mathbf{u}_S$ is the s-t elements of ${}^A l_W$: $\mathbf{u}_S = [[{}^A l_W]_s [{}^A l_W]_t]^{\mathrm{T}}$.

4 Technical Setup

We arranged two experimental setups with two different OST-HMDs. The first one is a professional HMD which has a higher resolution and a larger field of view than the second one – a consumer OST-HMD. Both setups use a 4×11-asymmetrical circle-grid board as the 3D world reference. We place it at about 1.5m away from the displays so that user-perspective cameras can see the board and the image of each display sharply at the same time.

4.1 Professional OST-HMD Setup

The first setup (Fig. 3 left) uses the exact same one in [8].

4.2 Consumer OST-HMD Setup

The second setup is a consumer OST-HMD (Fig. 3 right) – a Moverio BT-100 from EPSON with 960×540 resolution and 23° diagonal FoV. The left-eye display is used for the setup. The HMD employs an HTPS TFT LCD panel with a color filter for each display. Light from the panel is guided to semi-transparent, half mirrors in front of a user's eye, and the mirrors reflect the light to the eye.

This HMD has a composite video input. We use a VGA-composite converter (Ligawo PC-TV). It generates a composite video signal from an input digital image with 656×496 resolution. As a result, the HMD renders a stretched image compared to the original resolution.

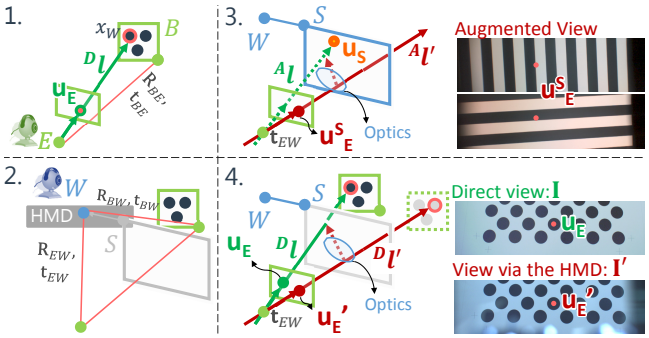


Figure 5: An overview of the data collection steps described in Sec. 4.4. Images are from the professional HMD setup.

The outward looking camera is the same Delock USB 2.0 model as the professional setup. Since this HMD has much narrower FoV than the ST60, we select a different user-perspective camera with narrower FoV. The camera is 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).

4.3 Image LF Acquisition via Structured Patterns

We first describe how to measure the LF of a display screen, i.e. $\{{}^A l'_W\}$. Given a user-perspective camera which is seeing the image screen of an OST-HMD, we need to collect 2D-to-2D correspondences $\{(\mathbf{u}_E^s, \mathbf{u}_S)\}$. We display structured patterns (gray-code/sinusoidal for pixel-/subpixel-level matching) on the screen, and match screen image points $\{\mathbf{u}_E^s\}$ to camera image pixels $\{\mathbf{u}_S\}$. We use a software[1] by Yamazaki et al. [19]. Figure 4 (c,d) are some patterns shown on the display of HMDs captured by a camera. Figure 4 (e,f) show learned mappings.

We follow the calibration procedure in Sec 3.2 of [7] to determine an initial 3D plane pose $(\mathtt{R}_{SW}, \mathbf{t}_{SW})$ and a pixel-to-meter scale α from $\{(\mathbf{u}_E^s, \mathbf{u}_S)\}$, and the intrinsic matrix $\mathtt{K}_E$ of the user-perspective camera. Finally, Eq. (1) and (2) give ${}^A l'_W$ and ${}^A l_W$ respectively.

4.4 Training Data Sampling

We describe how to collect training data of the original and the distorted LF offline. We need a set of original and distorted light ray pairs $\left\{({}^D l_W, {}^D l'_W, {}^A l'_W, {}^A l_W)\right\}$ for various viewpoints within the eye box such that the learned regressions can cover various eye positions in applications. Our collection procedure requires the following: a user-view camera E, an OST-HMD with a world camera W, and a fiducial target board B fixed in a scene. We assume that the cameras are calibrated. The procedure is as follows (see Fig. 5):

1. Place the user-perspective camera E and the 3D target B in the scene, and let E capture a direct-view image $\mathbf{I}$. Then from $\mathbf{I}$ and the intrinsic matrix $\mathtt{K}_E$ of E, estimate the pose of the target as $(\mathtt{R}_{BE}, \mathbf{t}_{BE})$.

2. Place the OST-HMD in front of the user-view camera, and let the user-view camera capture a distorted-view image $\mathbf{I}'$. Let the world camera W on the HMD capture the 3D target and estimate the pose $(\mathtt{R}_{BW}, \mathbf{t}_{BW})$. Using this pose and $(\mathtt{R}_{BE}, \mathbf{t}_{BE})$, compute the pose of the user-view camera relative to the world, $(\mathtt{R}_{EW}, \mathbf{t}_{EW})$.

3. Without touching the hardware, block the world light coming through the display, e.g. by putting a black sheet in front of the HMD, then capture the structured patterns as described in Sec. 4.3.

4. From $\mathbf{I}$ and $\mathbf{I}'$, extract corresponding 2D points $\mathbf{u}_E$ and $\mathbf{u}'_E$. Compute their 3D position in W as $\mathbf{x}_W := \mathtt{R}_{EW}\mathtt{K}_E^{-1}\widetilde{\mathbf{u}_E} + \mathbf{t}_{EW}$ and $\mathbf{x}'_W :=$

[1] http://www.dh.aist.go.jp/~shun/research/calibration/

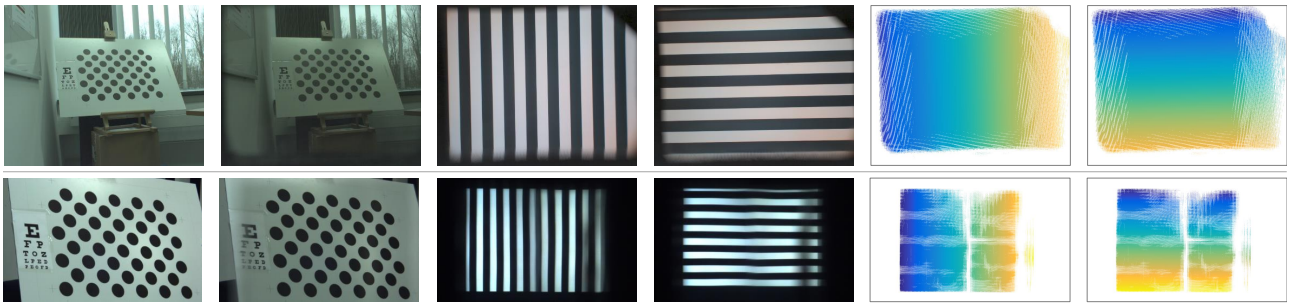


Figure 4: Examples of training data and processed images. Top row is for the ST60 and the bottom for BT-100. Each column is: (a) a scene captured by the user-perspective camera directly, (b) through the optical combiner, (c,d) a gray-code pattern displayed on the screen and captured by the camera, and (e,f) learned camera-to-hmd pixel mappings of horizontal and vertical directions. The mappings are color coded.

$\mathrm{R}_{EW}\mathrm{K}_E^{-1}\widetilde{\mathbf{u}'_E} + \mathbf{t}_{EW}$. Finally, compute an original light ray $l := l(\mathbf{x}_W)$ and its distorted $l' = l(\mathbf{x}'_W)$. Figure 4 shows collected samples.

As a result, we get a set of desired light-ray pairs with a maximum of 44 ($= 4 \times 11$) pairs. We collect such sets for a number of viewpoints. Due to the limited FoV of the image screen and/or failures in the structured-light matching, some 3D points on the board do not have corresponding 2D points on the screen (Fig. 4 (e,f)).

For the professional HMD setup, we collected training data from 17 viewpoints. We used all 17 sets for learning the DVD, and 10 for the AVD. For the consumer HMD setup, we collected training data from 8 viewpoints. We used 8 for the DVD and 7 for the AVD. Subsequently, for both setups, we took test data from new viewpoints – different from those of the training data sets.

The differences of the number of data sets used for each distortion estimation are from various reasons: a) Some data sets were old one that captured only for DVDs. b) image screens were not visible from a viewpoint, i.e., the user-perspective camera was outside of the ideal eyebox. c) A missing pose between the cameras since the user-perspective camera could not see the calibration board due to the occlusion caused by the frame of an HMD. The discussion section covers remaining issues towards establishing simpler and stabler calibration procedure.

5 Experiment

5.1 Error Measurements based on Viewing Angles

Existing calibration methods evaluate their accuracy by 2D projection errors: the distance between a measured and an estimated 2D point in a planar coordinate system (e.g. the image plane of a user-perspective camera [3, 11] or a physical board in the scene [12]).

Instead, we employ the Viewing Angle (VA) error between the forward-projected rays of the two 2D points from a user's viewpoint (Fig. 6). Let us call a ray from the original 2D point as the base ray $\mathbf{n} \in \mathbb{R}^3$, $\|\mathbf{n}\| = 1$, and let the other from the perturbed 2D point $\mathbf{n}' \in \mathbb{R}^3$, $\|\mathbf{n}'\| = 1$. We first consider a plane which is tangential to the unit sphere at $\mathbf{n}$. The plane has its x/y axis along a latitude/longitude line on $\mathbf{n}$ heading to west/north (Fig. 6 left). We then define the VA error $\theta := \arccos(\mathbf{n}^{\mathrm{T}}\mathbf{n}')$ and the direction angle ρ (Fig. 6 right).

The VA error constitutes a common measurement over OST-HMDs of different resolutions and is compatible to the human visual acuity; thus should be used instead of the reprojection error.

5.2 Experiment Procedure

For each HMD setup, our procedure was the following: collect a training and a test data set as described in Sec. 4.4, compute mappings ${}^{D}f(\cdot)$ and ${}^{A}f^{-1}(\cdot)$ via a kernel regression following [8], and test them with the test data. The number of basis functions was 50.

Experiments were done mostly with MATLAB R2014b. For the pose estimation of the calibration board, we used our open-source C++ tracking framework, Ubitrack, with the OpenCV library.

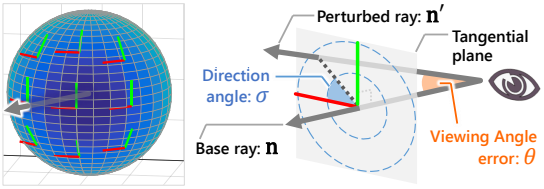


Figure 6: Illustration of the VA error. (Left) A 3D unit sphere with tangential coordinate systems. (Right) The VA error and the direction angle. The x/y axis of the tangential plane is orthogonal to the base ray and is aligned with the longi-/latitude lines of the sphere.

5.3 Results with the Professional OST-HMD (Fig. 7)

We first explain the results of the independent distortion estimations, i.e. ${}^{D}f({}^{D}l_W)$ vs. ${}^{D}l'_W$ and ${}^{A}f^{-1}({}^{A}l'_W)$ vs. ${}^{A}l_W$, with Fig. 7a. The top row shows the boxplots of the estimation errors of the AVD and the DVD, respectively. The bottom row shows their histograms. Note that the blue lines in the plots are at one arc minute, i.e. 1/60 (0.016...) degree, which is equivalent to the critical gap size for emmetropic (standard) human visual acuity. If a calibration result crosses these lines, such an AR experience would be, as it were, *retinally* aligned, thus indistinguishable to the eyes. Both corrections improved the calibration accuracy. Especially, the DVD estimation half crossed the threshold.

We now examine the results of the actual calibrations, i.e. ${}^{A}l_W$ vs. ${}^{A}f^{-1}({}^{D}f({}^{D}l_W))$, with Fig. 7b. The boxplot at the top summarizes the calibration errors. Correcting either of the two distortions separately did improve the overall calibration quality compared to without any corrections. Correcting both of them further decreased the errors. However, the mean error was higher than the visual acuity. We conjecture that this stems from the relatively high errors in the AVD estimation compared to the DVD.

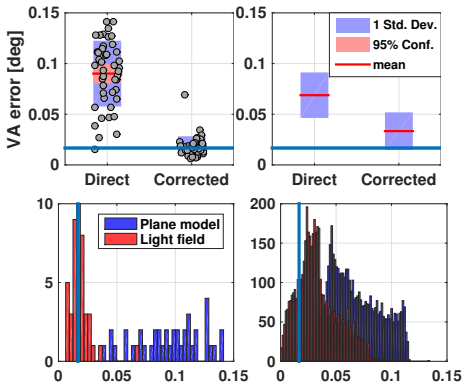
Nevertheless, the mean error reaches to the level of around 20/50 visual acuity. If those people see AR contents with this calibration without eyeglasses, it would appear in indistinguishable registration quality – if the eye model and position are perfectly estimated.

Figure 7b bottom shows each error with corresponding direction angle. The bias decreases by applying the distortion corrections.

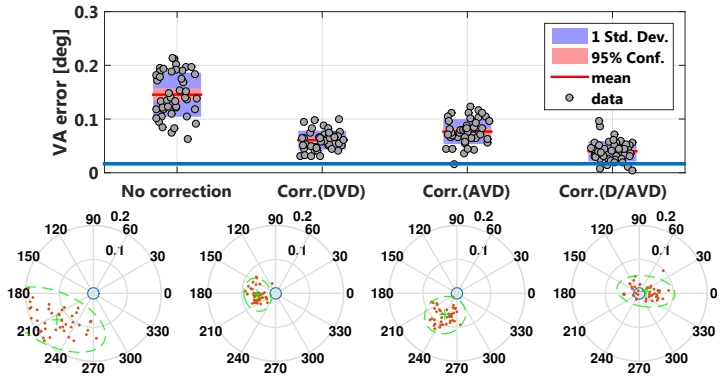
5.4 Results with the Consumer OST-HMD (Fig. 8)

Figure 8a left shows that the DVD estimation achieves the mean error to fall below the visual acuity line – as for the professional setup. In the AVD estimation (Fig. 8a right), the errors decreased overall, yet, the majority of the samples stay above the line.

Compared to the results of ST60, that of the BT100 (Fig. 8b) reveal an intriguing fact: correcting either of the D/AVD independently makes the results worse than applying no corrections – even though the DVD estimation achieves the error almost to the level of human visual acuity (Fig. 8a). Only when the two corrections are

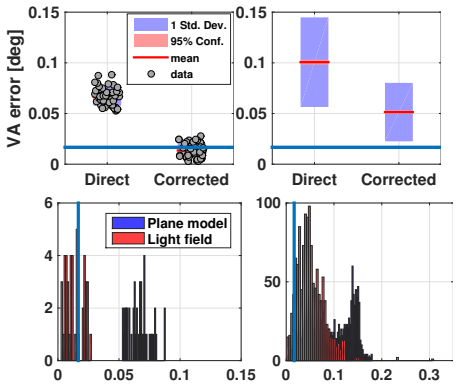


(a) Results of D/AVD estimations, i.e. ${}^{D}l'_{W}$ vs. ${}^{D}f({}^{D}l_{W})$ and ${}^{A}l_{W}$ vs. ${}^{A}f^{-1}({}^{A}l'_{W})$. (Top) Boxplots of VA errors of each estimation. (Bottom) Their histograms. In both cases two-sample t test returned p values almost zero.

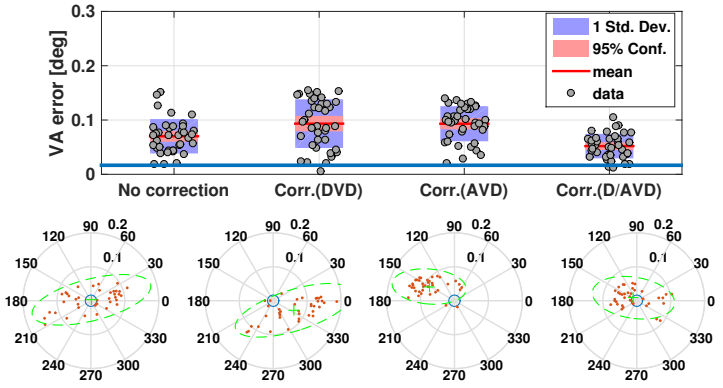


(b) Results of calibrations. (Top) The boxplot of each calibration methods in angle errors. (Bottom) Error vectors in polar coordinates. Dashed green lines visualize covariances with a confidence value 95 %. A multiple comparison test confirmed any of the two conditions are significantly different mean values.

Figure 7: Calibration results for the professional OST-HMD (NVIS nVisor ST60). Blue lines represent the standard visual acuity value: 1 arcmin.



(a) Results of D/AVD estimations. The two-sample t test returned p values almost zero for both cases.



(b) Results of calibrations. Unlike the previous setup, a multiple comparison test found significant difference except: between original and the proposed (p=0.090); and between the single correction methods (p=1.0).

Figure 8: Calibration results for the consumer OST-HMD (Moverio BT100). All notations are same as the Fig. 7.

combined, the error back down below the level of the original error.

We suppose this is because the combined part of the rays from the world and from the screen partially share an optical path inside the optical combiner (Fig. 2). The BT-100 employs a flat plastic housing for its combiner. Both light rays are possibly almost parallel when they join at the half-mirror, then they pass through the same medium while receiving the same aberration. Thus, correcting one of the two rays for A/DVD degrades the overall quality. Only the combined correction of both rays improves the accuracy.

6 Discussion

We examine the results of the two setups in combination. We start with the estimation results of the DVD (Fig. 7a/8a left) and the AVD (Fig. 7a/8a right) separately, then look at the actual calibration results with those estimated distortions (Fig. 7b and 8b). We further discuss the limitations of our method and open issues.

DVD Estimation | Fig. 7a left and 8a left Without corrections, the professional HMD setup has a higher mean error (0.09 deg.) than the consumer setup (0.067 deg.). This result is understandable since the HMD uses a thick cubic prism combiner which yields larger DVD than the flat-plate combiner of the consumer HMD. After the corrections, the DVD reduces the mean error as low as the standard human visual acuity in both setups.

AVD Estimation | Fig. 7a right and 8a right Without corrections, both setups show VA errors that are larger than in the DVD case. This can be that the AVDs were not as smooth as the DVDs: viewpoint changes could not be accommodated as well since the regression can not account for radical local changes.

Unlike the DVD case, the professional HMD has a lower mean AVD error (0.07 deg.) than the consumer (0.1 deg.). The consumer HMD has a huge non-linear distortion due to two separate half mirrors used in the optical combiner (Fig. 4 (c,d)). The center columns of the screen are especially blurry, and the pattern matching in the region failed (Fig. 4 (e,f)). On the other hand, the professional HMD achieved almost perfect matching thanks to its clear image.

Overall Calibration Results| Fig. 7b and 8b As mentioned in the experiment section, correcting the A/DVD simultaneously improves the total calibration accuracy in the professional setup (Fig. 7b) significantly. In the consumer setup (Fig. 8b), the result was not significant, yet it shows that the unified correction reverts errors that happen when only one distortion is corrected.

In both setups, the final accuracy does not reach at the human visual acuity level. Since the DVD estimation reached at the visual acuity level in both setups, it is likely that improving the AVD estimation does so the overall calibration.

Limitations and Issues of the Proposed Method The current method has a number of issues that need to be solved in order to establish a practical HMD calibration routine as easy as current camera calibration software.

Estimation quality for the AVDs: Our methods can not yet estimate AVD to the level of the human visual acuity while the DVD is corrected at the desirable level (Fig. 7a and 8a). Possible causes

are: the number of data sets, tuning parameters in the regressions, the image matching accuracy, and the pose tracking accuracy of cameras. We discuss some issues in the following paragraphs.

Number of data sets: We do not know how much training data is sufficient to achieve desirable accuracy and which viewpoints are best for the training phase. At least, if we knew a valid eye box in which any user's eye would stay, we could limit the positions of the view point to this box for collecting the data.

Hyper parameters of nonlinear regression: As Sec. (5.2) says, the nonlinear regressions need several parameters. We do not know what the best ones. We assumed that the number of basis functions might have the strongest effect. However, our informal examination did not generate strong matching differences somewhere between 50 to 200. Too few basis functions, such as 10, failed as expected.

Image matching and: As in the BT100 case, a partially blurry image of the screen makes it difficult to detect image pixel correspondences for computing the LF. This questions the use of the LF model which is based on the geometrical optics: light from a point source is treated as an ideal ray at each viewpoint. Expanding the model such that it can handle *blurred* light rays might improve the matching. This also requires a different matching approach than our current matching based on the structured patterns.

Camera tracking: We used a calibration board for the pose tracking between the user-perspective camera and the world camera on the HMD. We do not know if the pose accuracy was high enough or not. In this sense, a sensitivity analysis via simulations, as we did for the interaction-free calibration [7], might reveal the key factors in the entire calibration procedure.

DVD as a LF model: There is yet another issue in the current LF model for the DVD. The model treats both original and distorted light rays as those that pass through the center of the eyeball. The assumption is true for the original ray. However, a distorted ray (${}^{D}l'_{w}$) does not necessarily do the center – e.g. a convex lens shift the focal point of convergent light rays. Nevertheless, the experiments of the DVD estimation achieved desirable accuracy even though we did not consider this possible misalignment.

Computation time: We have not yet implemented a real time system to render the corrected virtual view. Our projection function is not a simple perspective projection any more. To compute the complex function, we might need a sampling approach somewhat similar to a ray tracing approach. Since we can not directly compute which image pixel corresponds to which 3D point in the scene, a naive way would be to sample a bundle of light rays passing through a given eye position, and to check where they hit the image screen.

User-based evaluation: Although this paper focuses on the HMD-dependent factors only, it is necessary to see how much our correction method contributes to the user-based interaction-free calibration [6]. It would be also interesting to consider how to measure the VA error, if we can not get the ground truth eye position. Given an accurate eye-tracking algorithm, one might make a compromise and use the position from the tracker as the ground-truth.

Hardware Approach for DVD There are OST-HMDs that employ retinal-scanning (e.g., Brother AirScouter) or pupil-division optics (e.g., Olympus MEG4.0). Theoretically, such displays do not suffer from the DVD problem: either they do not use the optical combiner or the combiner is small enough such that the world light reaches to user's eye directly.

7 Conclusion

This paper presents a calibration method which corrects optical aberrations that degrade the quality of OST-HMD calibration. Unlike existing methods, our calibration method corrects both DVD and AVD simultaneously for arbitrary eye position. Our approach expands a light-field correction approach developed for DVD to the AVD, and cascades two distortion corrections to cancel both distortions at the same time. Our method is camera-based and has an offline learning and an online correction step. The evaluation shows that the method significantly improves the calibration quality compared to conventional methods. The overall registration accuracy was comparable to 20/50 visual acuity. Furthermore, the results indicate that only by correcting both distortions simultaneously can improve the accuracy. We also analyzes limitations of the method and possible research directions.

8 Acknowledgments

The authors wish to thank Christoph Resch for a discussion on structured light measurements. The project received funding from the European Union: PITN-GA-2012- 316919-EDUSAFE.

References

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

[2] Y. Genc, F. Sauer, F. Wenzel, M. Tuceryan, and N. Navab. Optical See-Through HMD Calibration: A Stereo Method Validated with a Video See-Through System. In *ISAR*, pages 165–174, 2000.

[3] S. J. Gilson, A. W. Fitzgibbon, and A. Glennerster. Spatial calibration of an optical see-through head-mounted display. *Journal of Neuroscience Methods*, 173(1):140–146, 2008.

[4] R. L. Holloway. Registration error analysis for augmented reality. *Presence*, 6(4):413–432, 1997.

[5] H. Hua, C. Gao, and N. Ahuja. Calibration of a head-mounted projective display for augmented reality systems. In *Proc. ISMAR*, pages 176–185. IEEE, 2002.

[6] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *Proc. 3DUI, Minneapolis, USA, March 29-30*, pages 75–82. IEEE, 2014.

[7] Y. Itoh and G. Klinker. Performance and sensitivity analysis of INDICA: Interaction-free display calibration for optical see-through head-mounted displays. In *Proc. ISMAR*, pages 171–176. IEEE, 2014.

[8] Y. Itoh and G. Klinker. Light-field correction for spatial calibration of optical see-through head-mounted displays. *TVCG (Proc. VR)*, 21(4):471–480, Apr. 2015.

[9] 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.

[10] M. Klemm, H. Hoppe, and F. Seebacher. Non-parametric camera-based calibration of optical see-through glasses for augmented reality applications. In *Proc. ISMAR*, pages 273–274. IEEE, 2014.

[11] S. Lee and H. Hua. A robust camera-based method for optical distortion calibration of head-mounted displays. In *Proc. of VR*, pages 27–30. IEEE, 2013.

[12] E. McGarrity, Y. Genc, M. Tuceryan, C. B. Owen, and N. Navab. A new system for online quantitative evaluation of optical see-through augmentation. In *ISAR*, pages 157–166. IEEE, 2001.

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

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

[15] W. Robinett and J. P. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence*, 1(1):45–62, 1992.

[16] J. Rolland and H. Hua. Head-mounted display systems. *Encyclopedia of optical engineering*, pages 1–13, 2005.

[17] M. Tuceryan, Y. Genc, and N. Navab. Single-point active alignment method ("spaam") for optical see-through hmd calibration for augmented reality. *Presence*, 11(3):259–276, 2002.

[18] F. Wientapper, H. Wuest, P. Rojtberg, and D. Fellner. A camera-based calibration for automotive augmented reality head-up-displays. In *Proc. ISMAR*, pages 189–197. IEEE, 2013.

[19] S. Yamazaki, M. Mochimaru, and T. Kanade. Simultaneous self-calibration of a projector and a camera using structured light. In *Proc. CVPRW*, pages 60–67, June 2011.