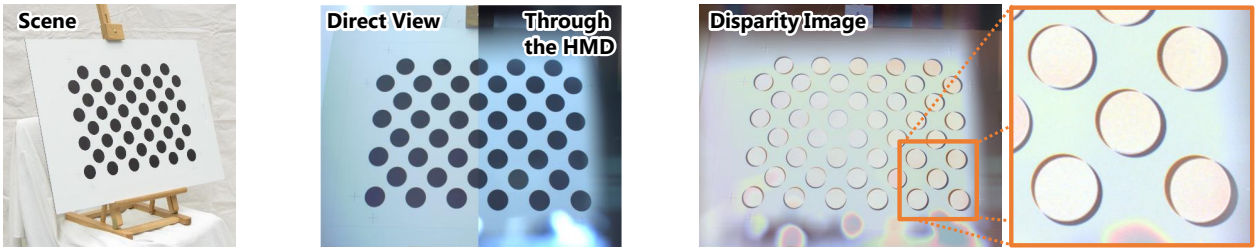
Light-Field Correction for Spatial Calibration of Optical See-Through Head-Mounted Displays

Yuta Itoh, *Student Member, IEEE*, and Gudrun Klinker, *Member, IEEE*



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

Fig. 1. An illustration of our problem: an optical distortion caused by an optical element of an OST-HMD. (a) A target in the scene. (b) A direct view by a camera and a view through the OST-HMD from the same viewpoint. (c) The absolute difference of the two images (intensity is inverted). (d) A zoomed part of (c). The distortion is radial in appearance. Note that distortion shapes can vary depending on the position of the view point and the HMD. Fig. 2 shows a corresponding HMD setup.

Abstract—A critical requirement for AR applications with Optical See-Through Head-Mounted Displays (OST-HMD) is to project 3D information correctly into the current viewpoint of the user – more particularly, according to the user's eye position. Recently-proposed interaction-free calibration methods [16, 17] automatically estimate this projection by tracking the user's eye position, thereby freeing users from tedious manual calibrations. However, the method is still prone to contain systematic calibration errors. Such errors stem from eye-/HMD-related factors and are not represented in the conventional eye-HMD model used for HMD calibration.
This paper investigates one of these factors – the fact that optical elements of OST-HMDs distort incoming world-light rays before they reach the eye, just as corrective glasses do. Any OST-HMD requires an optical element to display a virtual screen. Each such optical element has different distortions. Since users see a distorted world through the element, ignoring this distortion degenerates the projection quality. We propose a light-field correction method, based on a machine learning technique, which compensates the world-scene distortion caused by OST-HMD optics. We demonstrate that our method reduces the systematic error and significantly increases the calibration accuracy of the interaction-free calibration.

Index Terms—OST-HMD, calibration, undistortion, optical see-through display, light field, ~~IN~~DICA, SPAAM, eye tracking

1 Introduction

Optical See-Through Head-Mounted Displays (OST-HMDs) have lately gained much interest, as typified by recent consumer-level products like Google Glass, EPSON BT-200, etc. We believe the trend will mark a watershed in the democratization of Augmented Reality(AR), where people enjoy seeing AR content in the real world in real time, at any time and place. Realizing such experiences requires an OST-HMD system to keep the virtual and physical objects consistent in various aspects – temporally [37, 8, 3], visually [19, 23, 6, 5], socially, and spatially [34, 2]. Lacking any of these consistencies degrades the user experience. Spatial consistency, in particular, is one of the most fundamental requirements for AR applications with OST-HMDs [4, 14]. It is the focus of this paper.

Spatial consistency relates to presenting information geometrically aligned with real objects, i.e., virtual 3D information must be rendered on the 2D virtual screen of an OST-HMD worn by a user in perfect alignment with the 3D-2D projections of real objects onto the user's retina. To this end, the path of the 3D-2D projection rays from the 3D world through the HMD into the user's eye needs to be calculated. In other words, spatial consistency requires careful calibration of an eye-HMD system consisting of an eyeball and the virtual screen of an OST-HMD. Mission-critical AR applications such as medical AR [26, 4] especially depend on accurate calibration quality. Achieving such accurate, or even indistinguishable, registration quality is our ultimate, long-term goal.

Manual calibration methods, such as the Single Point Active Alignment Method (SPAAM) [4], estimate the 3D-2D projection by requesting users to align a series of 3D points in the physical world with 2D dots on the virtual screen. The calibration procedure is impractical for the daily use of OST-HMDs, since users have to recalibrate the system whenever they move the HMD – which happens frequently. In contrast to such manual methods, we have developed an automated calibration method, INteraction-*free* DIsplay CAlibration (~~IN~~DICA) [16]. It continuously estimates the appropriate projection matrix based on the user's current eye position, which is automatically tracked by an HMD-mounted eye camera.

Although the automated method frees users from manual calibration, our most recent calibration results still contain systematic errors due to simplistic eye-HMD modeling [17]. We have presented a sensitivity analysis of a number of calibration and registration parameters, indicating which of them are most critical for reducing systematic errors. Yet, they neglected to model an important fact of the system: optical elements of OST-HMDs distort incoming world-light rays before they reach the eye (Fig. 1), just as corrective glasses do.

In common OST-HMD designs, the light rays of an image that arrive at the user's eye are collimated, showing images at virtual infin-

- *Yuta Itoh is with the Department of Informatics at Technical University of Munich. E-mail: itoh@in.tum.de.*
- *Gudrun Klinker is with the Department of Informatics at Technical University of Munich. E-mail: klinker@in.tum.de.*

Manuscript received 18 Sept. 2014; accepted 10 Jan. 2015. Date of Publication 23 Mar. 2015; date of current version 23 Mar. 2015. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.

ity [11] or are perceived as a 3D planar screen floating mid air in practice. While this property is desirable to align the image and the user's eye easily, it requires curved optical elements which inevitably distort light rays incoming from the world [14]. Since users see a distorted world through these elements, ignoring the distortion degenerates the registration quality.

This paper proposes a method to compensate the world light-ray distortion caused by OST-HMDs' optical elements. The method estimates a 4D-to-4D mapping between the original light field and the light field distorted by the optics in offline. After computing the mapping, the method compensates distortions on the virtual screen w.r.t the eyeball center. We first validate the compensation method in the camera-based OST-HMD setup and show that the method significantly reduces the calibration error. We then further evaluate our method in an actual interaction-free OST-HMD calibration setup with a real user involved. The result shows that the compensation reduces the systematic error, and again significantly improves the overall calibration quality.

As a summary, our contribution includes the following:

- We provide a formulation of the lens distortion caused by the optical elements of HMDs that distort the transmission of light rays.
- We provide an offline calibration procedure to learn a mapping which corrects the light-field distortion. The procedure is required only once per HMD.
- We demonstrate that applying the correction method reduces the systematic error which has existed in conventional camera-based and user-based calibrations and significantly increases the calibration accuracy.

In Section 2, we present undistortion techniques for optical distortion effects in imaging systems, and discuss the relationship to our method. In Section 3, we present the conceptual foundations of our approach. Section 4 describes a technical implementation of the approach. Section 5 presents results, verifying the method in a camera-based setup and evaluating it in a user-based interaction-free OST-HMD calibration setup. The section also discusses the current status, shortcomings of the method and future steps. The major lessons learned thus far are summarized in the concluding Section 6.

2 Related Work

The following subsections present an overview of topics related to the correction of optical distortion effects in imaging systems.

2.1 Spatial Calibration of OST-HMDs

Existing calibration methods model an eye-HMD vision system, such as ours in Fig. 2, as an off-axis pinhole camera where the virtual screen S of the display is the image plane, and the eyeball center E is the camera center (Fig. 3). The model is represented by a 3-by-4 projection matrix $\mathbf{P}_{WE}$ which projects a 3D point from the world W to a user view on the screen S.

Manual methods, such as SPAAM, require at least six 3D-2D correspondences to estimate the matrix (Fig. 3 top right). On the other hand, an automated method such as ~~I~~NDICA (Fig. 3 bottom left and right) does not require such user-based alignment. Instead, it tracks the eye position and computes the projection matrix together with some precalibrated parameters.

The automated method actually has two formulations. The formulation of **Recycled ~~I~~NDICA** (Fig. 3 bottom left) reuses an old projection matrix from a prior eye position E_0 and updates this old projection matrix by taking the new eye position E into account. The formulation of **Full ~~I~~NDICA** (Fig. 3 bottom right) calculates the current projection according to precalibrated HMD parameters such as the relative pose between an eye-tracker T and a world camera W, the pose of the HMD's virtual screen S w.r.t the world camera, and the apparent size (α_x, α_y) of the virtual screen pixels [17]. This formulation has 17 degree of freedom (DoF) including the eye position.

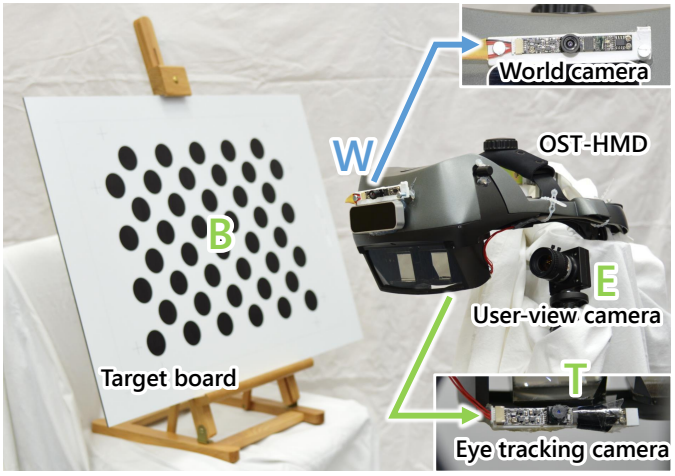


Fig. 2. Hardware setup. It shows our OST-HMD (nVisor ST60), the world camera W attached on the HMD, the eye tracking camera T fixed beneath the left optical element of the HMD, the user-view camera E, and a target board for calibration experiments.

Although ~~I~~NDICA is suitable for practical use, it is prone to contain systematic errors possibly stemming from its simplified eye-HMD system modeling [17]. The model mainly consists of two independent parts: eye-dependent and HMD-dependent parameters. The former relate to anatomical eye parameters such as the eye-ball radius. The latter relate to optical characteristics of the optical components of an HMD such as the virtual screen pose and optical distortion. This distortion effect is the prime issue that we investigate in this paper.

2.2 Undistortion for Cameras

As mentioned in the previous section, existing OST-HMD calibration methods assume the eye-HMD system to be an (off-axis) pinhole camera [1]. The model is commonly used in computer vision, where lens distortion is one of the most essential problems [32]. Parametric distortions in the 2D image space, e.g. radial and tangential distortions, affect ordinary lenses the most, and thus are commonly sufficient to perform image undistortions [36, 7]. For heavily-distorted lenses, such as fish-eye lenses or catadioptric optics, some approaches employ non-parametric distortion models [12, 29].

An important difference between conventional cameras and eye-HMD systems is that camera models may assume that the camera center w.r.t. camera's image plane is static, while HMD models must expect that the center, i.e. the user's eyeball center, is dynamic w.r.t. the image screen of the OST-HMD. Therefore, to undistort an eye-HMD system, it is necessary to estimate distortions relative to the possibly moving eyeball center.

2.3 Undistortion for HMDs

Vincent and Tjahjadi [35] propose a non-parametric approach for Head-Up Display (HUD) calibration. Their method undistorts images by first estimating a homography between ideal and distorted grid images and then computing further offsets per grid by fitting a B-spline surface to the data to compute a non-parametric undistortion model. While their method can handle complex distortions, such as those caused by HUD optics, it needs to re-learn the distortion parameters whenever the eyeball center moves.

A key observation of these undistortion methods is that they only estimate a 2D mapping between a distorted and the original image. Given a camera, a 2D mapping is only valid for one camera center known beforehand. Unlike in cameras, the eyeball center changes dynamically in an eye-HMD system. A naive way to apply these methods to the eye-HMD system is to estimate a 2D mapping once for a predefined eyeball center, and then reuse the mapping for different users [35]. Obviously, this does not assure that the learned mapping

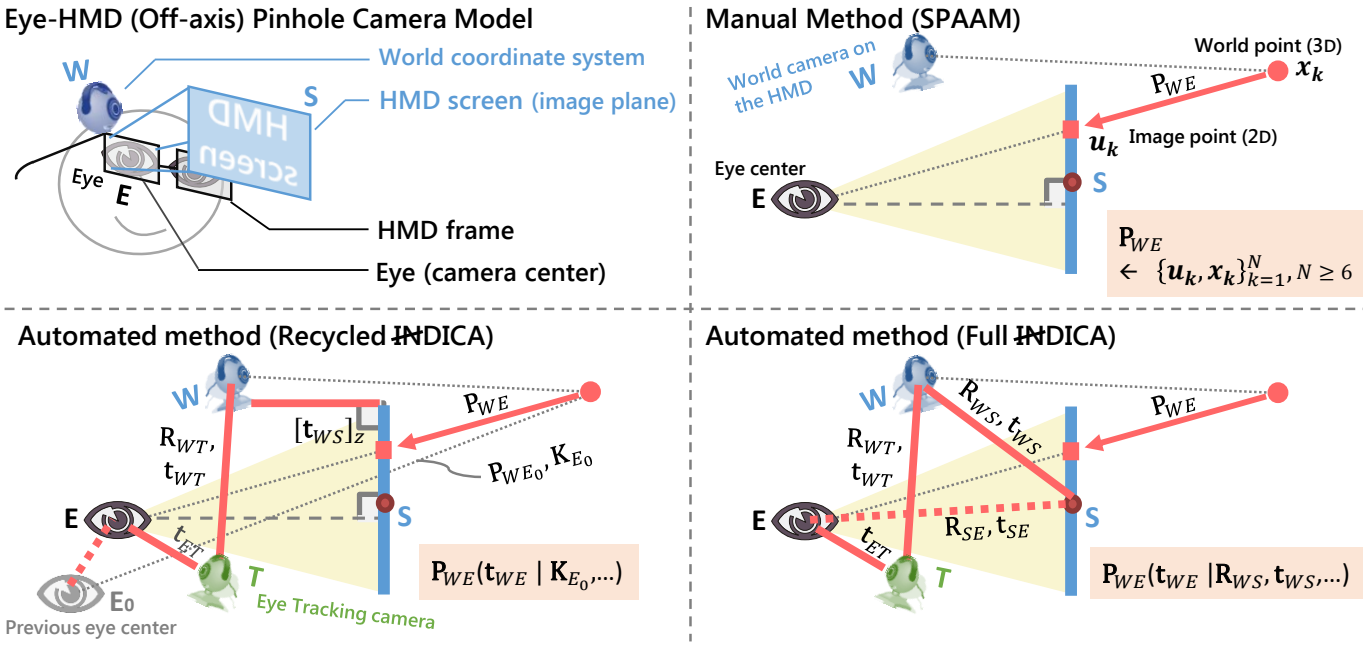


Fig. 3. Schematic overview of OST-HMD calibration methods. (Top left) The image screen is fixed to the HMD coordinate system defined on the world camera. (Top right) Manual methods, e.g., SPAAM, construct an 11-DoF projection matrix $\mathbf{P}_{WE}$ from at least six 2D-3D correspondences collected by a user. (Bottom left and right) Automated methods (Recycled INDICA and Full INDICA) [17, 16] reconstruct the projection as a function of the current eyeball position $\mathbf{t}_{WE}$ and various parameters of the current eye-HMD system. Note that none of the methods considers the distortion caused by optical elements of OST-HMDs. See Sec. 3.2 for a detailed definition of the parameters in the figure.

undistorts images properly for arbitrary eyeball centers. A second possible option would be to learn several mappings at different eyeball centers, then select a mapping of a predefined eyeball center nearest to the current position at runtime. This approach might work more accurately than the first one, yet again it does not produce a correct undistortion for every new eyeball center. The third approach would be to further learn a regression function of those 2D mappings, namely learn a super function which returns a 2D mapping given an eyeball center. This approach assumes that two 2D mappings of two eyeball centers close to each other are similar in some sense. This assumption requires a careful definition of the distance between the 2D mappings used, e.g. a distance of radial distortion coefficients, factorial distortion parameters, etc.

In this paper, we extend the last idea to the 4D domain – the light field space. Remember that we are concerned with the distortion caused by an optical element of an OST-HMD. Physically speaking, the distortion is due to the fact that the optical element distorts all incoming light rays from the scene passing through the element. Under the assumption that the optical element smoothly distorts the light rays, i.e. similar incoming light rays are distorted similarly, it is our problem to find a 4D-to-4D mapping between the original light field and the distorted light field. Once the mapping is given, we can readily create a 2D mapping for a given eyeball center.

2.4 Light-field Representation

A light field or Lumigraph is a 4D function representing the light rays passing through a 3D space (Fig. 4 bottom) [10, 22]. The representation has been used for rendering photorealistic visual effects such as reflection and refraction in computer graphics [13], and applied to model light-field displays [18, 15] and light-field cameras [25] in computational photography.

2.5 Non-parametric Regression

We use non-parametric regression to estimate the mapping between light fields. In machine learning, regression is one of the most fundamental methods. Given training data $\{(x_k, y_k)\}_k$ with k samples, a regression method finds a function $y = f(x)$ which explains the dataset best in a statistical sense. If candidates of f is limited within a function class $g(x \mid \theta)$ with parameters θ, then the problem is called parametric regression. Image undistortion based on a radial distortion model is an example of this problem. On the other hand, if f is estimated locally based on the data itself, it is called non-parametric regression. For example, the B-spline regresses a function f by tuning the amplitude of each basis function which is uniformly distributed in the data domain. The so-called kernel regression method is similar to B-splines. Yet, it is more concise in the sense that the method regresses f by radial basis functions located at each data point [30].

3 Method

This section explains the spatial calibration of the eye-HMD system and the distortion estimation of the optical elements of the display.

3.1 Notation

Bold lower/upper-case letters denote vectors/matrices such as a translation vector $\mathbf{t}$ and a rotation matrix $\mathbf{R}$. $(\cdot)^{\mathrm{T}}$ denotes transpose vectors and matrices. If a matrix is explicitly written with its elements, zero elements are left blank for clarity. Lower-case letters represent scalars. Upper-case letters denote coordinate systems such as the world coordinate system W. Given a coordinate system A, a 3D point in A is denoted by using vectors with the coordinate symbol as the lower index: $\mathbf{x}_A$. Given coordinate systems A and B, the relative transformation from A to B is described by $(\mathbf{R}_{AB}, \mathbf{t}_{AB})$ where $\mathbf{R}_{AB}$ and $\mathbf{t}_{AB}$ stand for rotation and translation respectively. Furthermore, explicit transformation of a 3D point $\mathbf{x}_A$ in A to $\mathbf{x}_B$ in B can be written as $\mathbf{x}_B = \mathbf{R}_{AB}\mathbf{x}_A + \mathbf{t}_{AB}$.

3.2 Interaction-Free Calibration in a Nutshell

As introduced in Sec. 2.1 (with Fig. 3), the eye-HMD system is commonly modeled as an off-axis pinhole camera and represented by a 3-by-4 projection matrix from the world to user's eye (user-view) as

$$\mathbf{P}_{WE}(\mathbf{t}_{WE}) := \mathbf{K}_E \begin{bmatrix} \mathbf{R}_{WE} & \mathbf{t}_{WE} \end{bmatrix} \tag{1}$$

where $(\mathbf{R}_{WE}, \mathbf{t}_{WE})$ is a transformation from the world coordinate system W (attached to the HMD) to user's eye coordinates E, and $\mathbf{K}_E$ is an

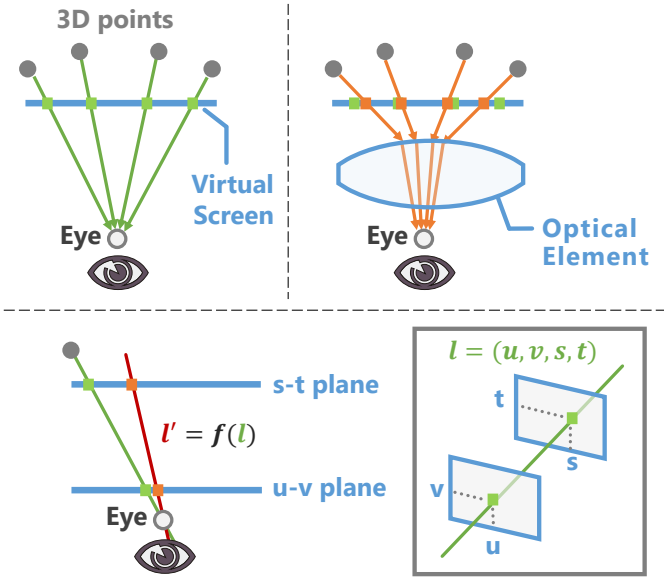


Fig. 4. Schematic drawing of the real-world distortion effect caused by the optical element of an OST-HMD. (Top left) Light rays from a user's eye to the world. The rays intersect with the image plane of the virtual screen. (Top right) The optical element of the display distorts the rays. It thus also shifts the pixel positions of the intersections. (Bottom) Modeling of the distortion as a 4D-to-4D mapping between light fields. We use the 4D Lumigraph parameterization: (u,v,s,t). Note that a distorted light ray l' is modeled to pass through the eye center and a shifted intersection pixel position.

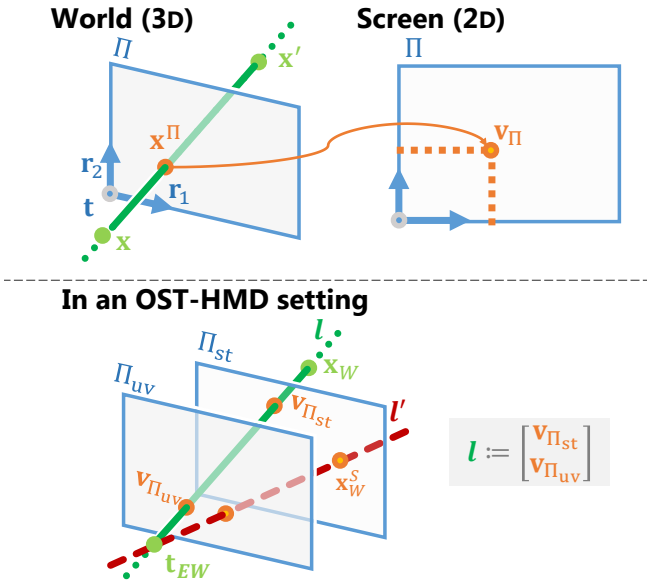


Fig. 5. Schematic diagram of the definitions of the light field w.r.t the HMD coordinate system. (Top) A 3D point $\mathbf{x}^\Pi$ is the intersection between a 3D plane Π and a 3D line passing through 3D points $\mathbf{x}$ and $\mathbf{x}'$. In the plane coordinate system, $\mathbf{x}^\Pi$ can be defined by a 2D point v_Π. (Bottom) Light rays in our OST-HMD setup in more detail, using the notation of Sec. 3.3.1.

intrinsic 3D-2D projection matrix. $\mathbf{K}_E$ has two representations [16]:

$$\mathbf{K}_E = \underbrace{\mathbf{K}_{E_0} \begin{bmatrix} 1+z_{EE_0}/z_{SE} & & -x_{EE_0}/z_{SE} \\ & 1+z_{EE_0}/z_{SE} & -y_{EE_0}/z_{SE} \\ & & 1 \end{bmatrix}}_{\text{Recycled } \cancel{\text{IN}}\text{DICA}}, \tag{2}$$

$$= \underbrace{\begin{bmatrix} \alpha_x & & c_x \\ & \alpha_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} z_{SE} & & -x_{SE} \\ & z_{SE} & -y_{SE} \\ & & 1 \end{bmatrix}}_{\text{Full } \cancel{\text{IN}}\text{DICA}}, \tag{3}$$

where S is the virtual screen coordinate system, $\mathbf{t}_{SE} = [x_{SE}, y_{SE}, z_{SE}]^\mathrm{T}$ and $\mathbf{t}_{EE_0} = [x_{EE_0}, y_{EE_0}, z_{EE_0}]^\mathrm{T}$. The scaling factor $\mathbf{a} := [\alpha_x, \alpha_y]^\mathrm{T}$ converts 3D points on the screen to pixel points. $c_x := (w-1)/2$ and $c_y := (h-1)/2$ define the image center with the pixel width w and height h. $\mathbf{K}_{E_0}$ is the intrinsic matrix of another virtual camera dependent on a previous eye position E_0.

Equation 3 (Full ~~IN~~DICA) does not rely on knowledge about a previous eye position $\mathbf{t}_{WE_0}$. Instead, it requires the virtual screen pose $(\mathbf{R}_{WS}, \mathbf{t}_{WS})$ and the scaling vector $\mathbf{a}$ [pixel/meter]. On the other hand, Eq. 2 (Recycled ~~IN~~DICA) does not rely on these parameters, except for $[\mathbf{t}_{WS}]_z$, because it reuses the old intrinsic matrix $\mathbf{K}_{E_0}$.

Both setups also require $(\mathbf{R}_{WE}, \mathbf{t}_{WE})$ – the transformation between the world and the eye. $\mathbf{R}_{WE}$ is equivalent to $\mathbf{R}_{WS}$ in Full ~~IN~~DICA, while $\mathbf{K}_{E_0}$ is obtained by decomposing the old projection matrix in Recycled ~~IN~~DICA. Let T be the coordinates of an eye tracker rigidly mounted on the OST-HMD, then $\mathbf{t}_{WE} = \mathbf{R}_{WS}\mathbf{R}_{WT}^\mathrm{T}(\mathbf{t}_{WT} - \mathbf{t}_{ET})$ (Eq. 6 in [16]).

3.3 Distortion Estimation for OST-HMD Optics

An optical element of an OST-HMD distorts light rays incoming from the world to an eye, i.e. each light ray is mapped to a distorted light ray (Fig. 4). Our goal is to obtain this mapping $f : \mathbb{R}^4 \to \mathbb{R}^4$ between an original light field and a distorted light field after distortions by the optical element. We use the 4D Lumigraph parameterization by assigning point pairs on two planes denoted as *u-v plane* and *s-t plane* (Fig. 4 bottom).

3.3.1 Light Field Computation in OST-HMDs

In this section, we first formulate a light ray passing through a plane in a coordinate system (Fig. 5 top). We then apply the formulation to our OST-HMD calibration setup, and define original and distorted light rays (Fig. 5 bottom).

Given a position $\mathbf{t}$ and an orientation $\mathbf{R} := \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}^\mathrm{T}$, $\mathbf{t}$ and the first two row vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ span a 3D plane as $\Pi(\mathbf{R}, \mathbf{t}) := \{a\mathbf{r}_1 + b\mathbf{r}_2 + \mathbf{t} \mid a, b \in \mathbb{R}\}$. A light ray passing through two 3D points $\mathbf{x}$ and $\mathbf{x}'$ intersects with the 3D plane Π at $\mathbf{x}^\Pi$ as follows (Fig. 5 Left):

$$\mathbf{x}^\Pi(\mathbf{x}, \mathbf{x}') := \mathbf{x}' + \frac{(\mathbf{t} - \mathbf{x}')^\mathrm{T}\mathbf{r}_3}{(\mathbf{x} - \mathbf{x}')^\mathrm{T}\mathbf{r}_3}(\mathbf{x} - \mathbf{x}') \in \mathbb{R}^3. \tag{4}$$

Note that $\mathbf{x}^\Pi$ is commutative, i.e. $\mathbf{x}^\Pi(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\Pi(\mathbf{x}', \mathbf{x})$. The 3D intersection point is represented by a real-scale 2D vector in the plane's coordinate system as:

$$v_\Pi(\mathbf{x}, \mathbf{x}') := \begin{bmatrix} \mathbf{r}_1^\mathrm{T} \\ \mathbf{r}_2^\mathrm{T} \end{bmatrix} (\mathbf{x}^\Pi(\mathbf{x}, \mathbf{x}') - \mathbf{t}) \in \mathbb{R}^2. \tag{5}$$

Now consider our interaction-free OST-HMD calibration setup. We treat the HMD coordinate system as the world W. Let the virtual screen orientation and position be $\mathbf{R}_{SW} = \mathbf{R}_{WS}^\mathrm{T}$ and $\mathbf{t}_{SW} = -\mathbf{R}_{WS}^\mathrm{T}\mathbf{t}_{WS}$ respectively, and let $\mathbf{t}_{SW}^0 := \begin{bmatrix} [\mathbf{t}_{SW}]_x & [\mathbf{t}_{SW}]_y & 0 \end{bmatrix}^\mathrm{T}$. Then, we define the s-t and u-v plane as $\Pi_{st} := \Pi(\mathbf{R}_{SW}, \mathbf{t}_{SW})$ and $\Pi_{uv} := \Pi(\mathbf{R}_{SW}, \mathbf{t}_{SW}^0)$ respectively. Given a point $\mathbf{x}_W$ in W, we define a light ray l passing through $\mathbf{x}_W$ and the eyeball center $\mathbf{t}_{EW} = -\mathbf{R}_{WS}^\mathrm{T}\mathbf{t}_{WE}$ as

$$l_k := l(\mathbf{t}_{EW}, \mathbf{x}_W, \mathbf{R}_{SW}, \mathbf{t}_{SW}) := \begin{bmatrix} v_{\Pi_{st}}(\mathbf{t}_{EW}, \mathbf{x}_W) \\ v_{\Pi_{uv}}(\mathbf{t}_{EW}, \mathbf{x}_W) \end{bmatrix} \in \mathbb{R}^4. \tag{6}$$

Equation 6 represents light rays from the eyeball center when there is no distortion induced by the optical element. If we have such a

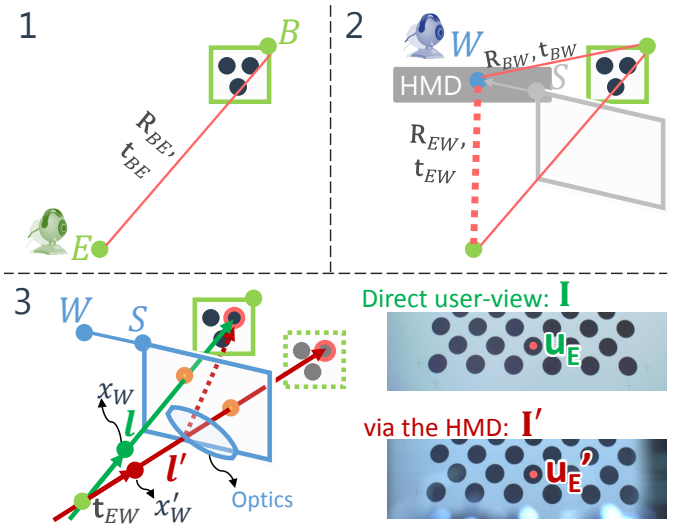


Fig. 6. Light-field collection overview. It describes each step of the collection procedure in Sec. 4.2.

distortion, then $\mathbf{x}_w$ matches, from a view point, to a 3D screen point $\mathbf{x}_w^s$ which is slightly shifted from $\mathbf{x}^{\Pi_{st}}(\mathbf{x},\mathbf{x}') \neq \mathbf{x}_w^s$. We define the distorted light ray as:

$$l'_k := l\left(\mathbf{t}_{EW}, \mathbf{x}_w^s, \mathbf{R}_{SW}, \mathbf{t}_{SW}\right) \in \mathbb{R}^4. \tag{7}$$

Finally, given a set of the light-ray pairs $\left\{\left(l_k, l'_k\right)\right\}_k$, our objective is now to learn the regression f which returns a distorted light ray given an original light ray so that the output explains the data set well.

3.3.2 Non-parametric Regression for the Distorted Light Field

The kernel regression yields a regression function $f : \mathbb{R}^N \to \mathbb{R}$ given a set of input and output pairs $\{(\mathbf{x}_k, y_k)\}_k$ [30]. We use the Gaussian kernel model, which approximates the true function $f(\mathbf{x})$ by the linear sum of Gaussian radial basis functions ϕ as:

$$f(\mathbf{x}) \approx \hat{f}(\mathbf{x} \mid \alpha) := \sum_{k=1}^{n_b} \alpha_k \phi(\mathbf{x}, \mathbf{x}_k), \tag{8}$$

$$\phi(\mathbf{x}, \mathbf{x}_k) := \exp\left(\frac{-(\mathbf{x}-\mathbf{x}_k)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_k)}{2\sigma^2}\right), \tag{9}$$

where σ is the kernel width, n_b is the number of basis functions, and $\alpha = [\alpha_1, \ldots, \alpha_{n_b}]^{\mathrm{T}}$ is the coefficient vector. The regularized least-square estimator, i.e. the maximum likelihood estimator, of the model is given by

$$\hat{f}(\mathbf{x} \mid \hat{\alpha}), \ \hat{\alpha} := (\mathbf{K} + \lambda \mathbf{I}_{n_b})^{-1} \mathbf{y}, \tag{10}$$

where $\mathbf{K}$ is defined as $[\mathbf{K}]_{ij} = \phi(\mathbf{x}_i, \mathbf{x}_j)$, λ is the model regularization parameter, $\mathbf{I}_{n_b}$ is an n_b-by-n_b identity matrix, and $[\mathbf{y}]_k = y_k$.

Since our dataset $L := \left\{\left(l_k, l'_k\right)\right\}_k$ has multivariate, 4D output, we learn four regression functions for each output dimension of the distorted light ray l'. For the ease of notation, we use $\mathbf{f}$ for representing the bundle of the four functions so that we can write $l' = \mathbf{f}(l)$. Note that, by switching the input and output, we can also learn an *undistortion* mapping $l = \mathbf{f}^{-1}(l')$.

In general, the performance of the kernel regression depends on the parameters of the kernel function and of the regularizer, i.e. σ and λ [30, 33]. We use a standard cross-validation technique [31] to automatically choose those parameters. Another pragmatic step for stable estimation is to normalize the training data so that they have zero mean and identity covariance matrices. If we apply this normalization technique to the training data, we also need to un-/normalize the out-/input by the mean and variance of the training data used for the regression.

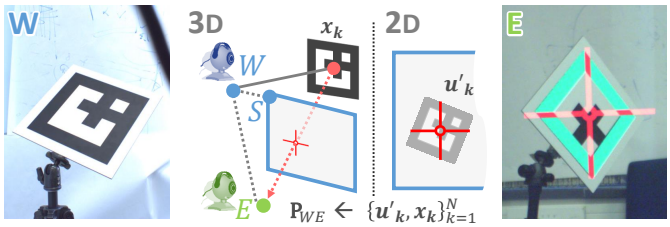


Fig. 7. Camera-based SPAAM setup. From left to right: a 3D reference marker seen by the world camera, a schematic illustration of SPAAM; a 2D crosshair is matched to the marker in the user-view camera E, and a calibration result where a 2D virtual green frame is overlaid on the board using the estimated calibration result.

3.3.3 Rendering with a Distorted Light Field

Now, we are ready to correct the optical distortion in AR visualization. In the original Full ~~N~~DICA setup [16, 17], we would project a 3D point $\mathbf{x}_w$ on the known display image plane by a projection matrix $\mathbf{P}_{WE}(\mathbf{t}_{WE})$. Instead, we now first convert $\mathbf{x}_w$ to a light ray $l\left(\mathbf{t}_{EW}, \mathbf{x}_w, \mathbf{R}_{SW}, \mathbf{t}_{SW}\right)$, and find a distorted ray $l' = \mathbf{f}(l)$. Then, we compute a distorted 2D pixel point $\mathbf{u}'$ as:

$$\mathbf{u}' = \begin{bmatrix} \alpha_x \left[l'\right]_s + c_x \\ \alpha_y \left[l'\right]_t + c_y \end{bmatrix}, \tag{11}$$

where $[\cdot]_s$ and $[\cdot]_t$ denote functions that return s and t elements of an input light ray respectively.

Note that we can define another 2D pixel point $\mathbf{u}$ from l, which represents the same pixel point that the conventional projection matrix gives. Thus, if we collect all pairs of $(\mathbf{u}, \mathbf{u}')$ corresponding to light rays that pass through each image pixel and the eye center, it generates a look-up table representing a *2D* distortion map – a common representation of lens distortions in computer vision.

4 Technical Setup

We explain our hardware setup, as well as an offline procedure for collecting original and distorted light fields for an HMD.

4.1 Hardware Setup

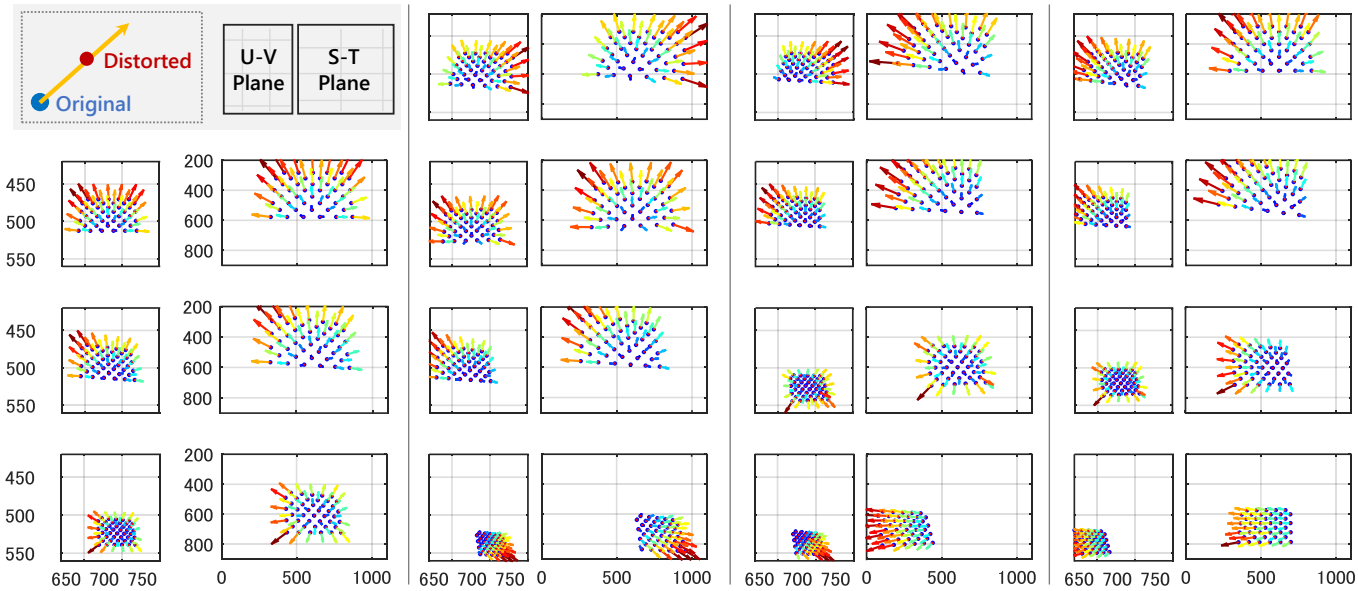
We have built an OST-HMD system equipped with an eye tracker as described below and in Fig. 2. We use an nVisor ST60 from NVIS – an OST-HMD with 1280×1024 resolution. The left-eye display is used for the current setup. An outward looking camera, a Delock USB 2.0 Camera with a 64-degree lens, serves as the world camera W. For the eye tracker T, another Delock Camera with a 55-degree lens is used. These cameras provide 1600×1200-pixel video and are attached to the HMD. The eye tracker is positioned at the bottom of the left display lens of the HMD. The default focal length of its fixed-focus lens is manually adjusted and fixed to a suitable length.

We calibrated the eye-HMD system as described in [17] to obtain offline parameters (Sec. 3.2): pose between the HMD and the eye-tracking camera $(\mathbf{R}_{WT}, \mathbf{t}_{WT})$, pose between the HMD and its virtual screen $(\mathbf{R}_{WS}, \mathbf{t}_{WS})$, and the scaling vector $\mathbf{a}$ [pixel/meter].

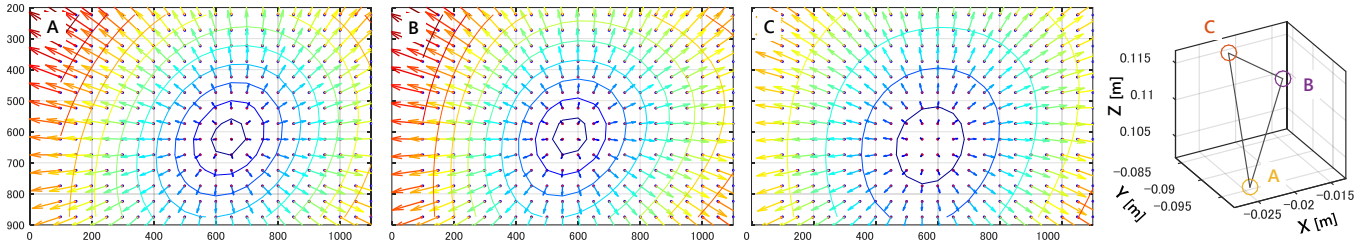
For a camera-based SPAAM experiment and for the light field estimation, we replace the human eye E by a camera. We use the UI-1240ML-C-HQ camera of iDS's together with an 8mm C-mount lens. The camera provides 1280×1024 images.

4.2 Light Field Collection

This section describes our offline calibration procedure for collecting training data of original and distorted light fields. For learning the regression function $l' = \mathbf{f}(l)$, we collect a set of original and distorted light ray pairs: $L_i = \left\{\left(l_{ik}, l'_{ik}\right)\right\}_k$ for a number of viewpoints i. Measurements from different viewpoints are necessary so that the regression can cover various eye positions in applications. Our collection procedure requires the following (Fig. 2): a user-view camera E, an



(a) Measured light fields. We visualize u-v and s-t planes of each light field $\{L_i\}$ measured from different viewpoints. Colored arrows originate from original image points and pass through corresponding distorted points. Their color-coded length is proportional to the distance between the point pairs. Distortions are mainly radial but their shape changes for each viewpoint.



(b) Testing the learned regression with three artificial eyeball positions different from those in the training dataset. Three color plots show the original and distorted light rays on the s-t plane. The 3D plot on the right visualizes the three eye positions $\mathbf{t}_{EW}$ used in this example. The positions are within a 1.5 cm^3 space. Different eyeball positions result in different distortions.

Fig. 8. Light-field mapping computation. (a) Measured light fields. (b) Estimated distortion maps.

OST-HMD with a world camera W, and a fiducial target board B fixed in a scene. We assume that the cameras and the OST-HMD's virtual screen are already calibrated. The procedure is as follows (Fig. 6):

1. Place the user-view camera E and the 3D target B in the scene, and let the camera capture a direct-view image $\mathbf{I}$. Then from $\mathbf{I}$ and the camera's intrinsic matrix $\mathbf{K}_E$, estimate the pose of the target as $(\mathbf{R}_{BE}, \mathbf{t}_{BE})$.

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

3. From $\mathbf{I}$ and $\mathbf{I}'$, extract corresponding 2D points $\mathbf{u}_E$ and $\mathbf{u}'_E$. Then compute their 3D position in W as

$$\mathbf{x}_W := \mathbf{R}_{EW}\mathbf{K}_E^{-1}\widetilde{\mathbf{u}_E} + \mathbf{t}_{EW},\ \mathbf{x}'_W := \mathbf{R}_{EW}\mathbf{K}_E^{-1}\widetilde{\mathbf{u}_E}' + \mathbf{t}_{EW}, \qquad (12)$$

where $\widetilde{\cdot}$ represents homogeneous vectors. Finally, compute an original light ray $l := l\left(\mathbf{t}_{EW}, \mathbf{x}_W, \mathbf{R}_{SW}, \mathbf{t}_{SW}\right)$ and its distorted $l' = l\left(\mathbf{t}_{EW}, \mathbf{x}'_W, \mathbf{R}_{SW}, \mathbf{t}_{SW}\right)$.

As the result, we get a set of the light-ray pairs $L_i = \left\{\left(l_{ik}, l'_{ik}\right)\right\}_k$.

In our experiment, we used a calibration board with a 4-by-11 asymmetrical circle grid, and measured the distortion from 19 different view points, $\mathbf{t}_{EW}$. This yielded total 836 ($= 4 \times 11 \times 19$) light ray pairs. We have not analyzed how many viewpoints are sufficient to estimate the mapping correctly.

5 Experiment

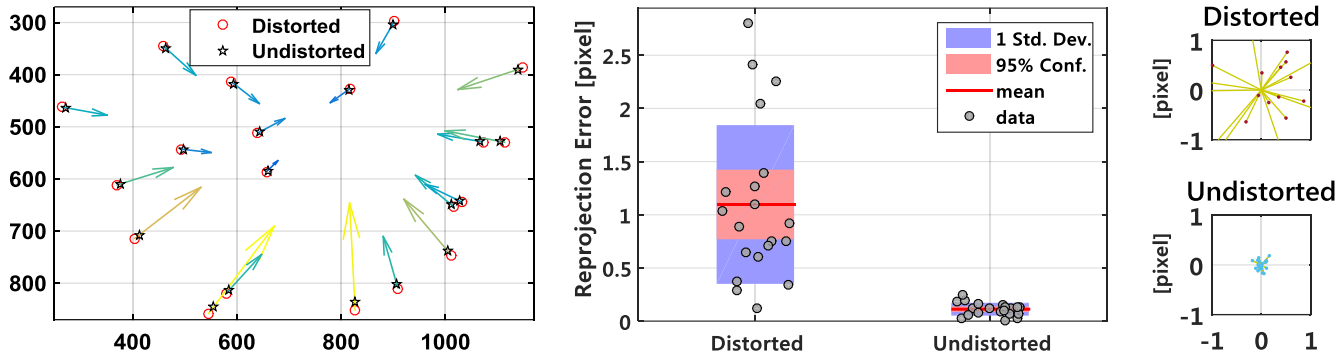
We conducted two calibration experiments: a camera-based OST-HMD calibration experiment and a user-based calibration. The camera-based calibration purely assesses the validity of our distortion correction method, and the user-based calibration further demonstrates its performance in a realistic OST-HMD calibration with a real user. Before going into the calibration experiments, we first elaborate the result of the light-field distortion learning.

5.1 Distortion Model Learning

After we collected a training data set $\{L_i\}_i$ as explained in Sec. 4.2, we learned the light-field mapping function $\mathbf{f}: \mathbb{R}^4 \to \mathbb{R}^4$ through the kernel regression method (Sec. 3.3.2). We used a Matlab implementation of the regression[1], which includes the cross validation step. We chose n_b (=100) random light rays from the training data for the basis functions in each regression. Note that we can also compute the inverse map $\mathbf{f}^{-1}$ by switching the input and output.

Figure 8 summarizes the result. Fig. 8a visualizes the u-v and s-t planes of several L_i among the 19 sets. The figure illustrates the difference between each corresponding light-ray pair $\left(l_{ik}, l'_{ik}\right)$ by drawing direction vectors from original to distorted 2D points on the planes.

[1] http://www.ms.k.u-tokyo.ac.jp/software.html



(a) Distorted 2D points ($\{\mathbf{u}_k\}$ in Fig. 7) observed during the camera-based SPAAM and their undistorted 2D points. Colored arrows originate from the distorted points and go through the undistorted points. The arrows' lengths are proportional to the distance of each point pair.

(b) Reprojection errors of original (Distorted) and corrected (Undistorted) datasets. The corrected dataset results in a smaller reprojection error than the original. The difference of the errors was statistically significant ($p \approx 8.28\text{e-}07 \ll 0.05$, two-sample t-test).

(c) Distributions of reprojection error vectors aligned to the origins.

Fig. 9. Camera-based calibration experiment. (a) Observed and corrected 2D points. (b) Reprojection errors. (c) Error vectors.

The lengths of the vectors are proportional to their point pairs' distances, for intuitive understanding. Since the virtual screen defines the s-t plane, the s-t plane figures show actual distortions observed by each view point. The visualizations show that different distortions occur at each viewpoint. Overall, the distortions are concentric similar to radial distortions.

Figure 8b tests the obtained regression function for three different view points – eyeball positions that moved within a 1.5 cm^3 space (the rightmost column). The left three columns demonstrate that the regressed function outputs different distortions for each new eye position.

5.2 Distortion Correction for Camera-based Calibration

Recall that eye-HMD system relies on the eye model and the OST-HMD model. This work focuses on improving the HMD model by taking the distortion from optical elements into account. Therefore, we first separate the eye part, another source of systematic error, for the primary validation of the distortion compensation method. As in the work by Gilson et. al. [9], our procedure uses a user-view camera E (Fig. 7) instead of human operators. We next evaluate our distortion compensation method in a camera-based manual calibration (SPAAM). The user-view camera was the same as the one used in the training dataset acquisition.

In the camera-based SPAAM (Fig. 7), we set up the user-view camera and the OST-HMD as described in Sec. 4.2. We rendered a 2D crosshair on the display. We placed a square marker in the world such that the camera saw both the center of the marker and the crosshair at the same pixel $\mathbf{u}'_k$ on S. Then we recorded $\mathbf{u}'_k$. At the same time, the world camera computed the 3D position of the marker center $\mathbf{x}_k$ in W. We repeated this procedure $N(=20)$ times, resulting in N pairs of 2D-3D correspondences $\left\{(\mathbf{u}'_k, \mathbf{x}_k)\right\}_k$. After the data collection, we compared two conditions. The first condition was an ordinary SPAAM, where we computed a projection matrix from the raw data and computed its reprojection error. In Fig. 9a, *circles (Distorted)* denote the original distorted 2D points measured during this step.

The second condition incorporated our distortion compensation method. First of all, before starting the above data collection, we let the user-view camera see the marker without placing the HMD. We thereby obtained the camera pose w.r.t the marker. Then, we placed the HMD in front of the camera without moving it. Since the world camera on the HMD saw the same marker (as in Sec. 3.3.1), we could compute $\mathbf{t}_{EW}$ and a 3D ray $\mathbf{x}^s_w$ by back projecting the distorted 2D point $\mathbf{u}'_k$. We thus obtained the distorted light ray l'_k.

Using the inverse mapping $\mathbf{f}^{-1}$, we estimated the original light ray as $l_k = \mathbf{f}^{-1}\left(l'_k\right)$: We computed *undistorted* 2D positions that the camera would have observed if there had been no distortion by the optical element (*Undistorted* points in Fig. 9a). Based on these corrected 2D points and the original 3D points, we estimated a projection matrix and compute its reprojection error.

Figure 9b is the comparison of the reprojection errors from the two conditions. It shows that our corrected condition (the right bar) leads to a significantly lower reprojection error compared to the original (the left bar). In SPAAM, we used the Direct Linear Transform (DLT) and Levenberg-marquadt (LM) method for estimating initial and optimized projection matrices. The DLT method does not model distortion in 2D space. The LM method we used does not include any distortion terms. And, Fig. 9c visualizes the distributions of the error vectors.

The fact that our correction method significantly reduced the reprojection error indicates that the method removed a systematic error caused by the optical element of an HMD which has not been considered in the standard HMD calibration framework.

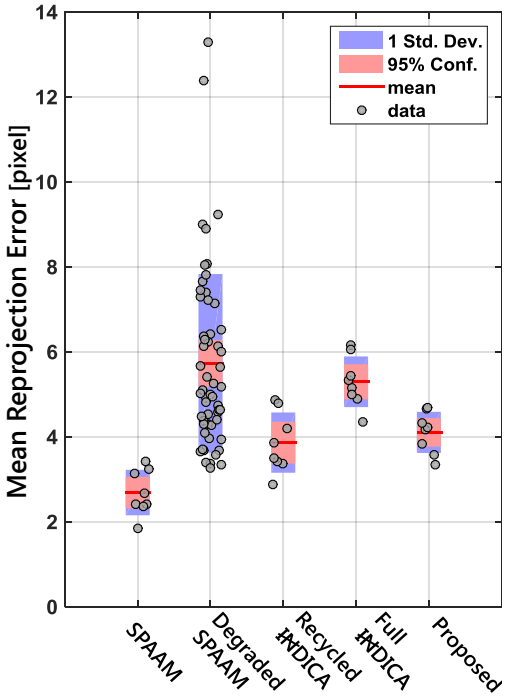
5.3 Distortion Correction for User-based Calibration

We further evaluate our method in a user-based calibration experiment where a real user conducts 2D-3D correspondence acquisitions manually.

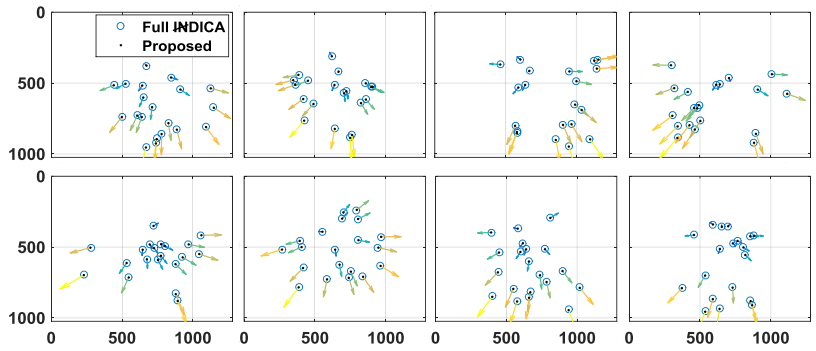
We follow the experiment design in [17]. An expert user of SPAAM has to collect sets of 2D-3D correspondences while letting the eye tracking camera record eye images to compute eye positions offline. The user has eight data collection sessions. The user is asked to take the HMD off and put it back on after each session to simulate a realistically degrading condition (Degraded SPAAM) with users staying on the initial calibration parameters even when their HMDs have moved on their head. For each session, the user collects 20 correspondences. We use the collected correspondence sets to analyze SPAAM, Degraded SPAAM, Recycled ~~IN~~DICA, Full ~~IN~~DICA and our correction method. Since our method requires the spatial parameters of the virtual screen of an OST-HMD, the method can be seen as an extension of Full ~~IN~~DICA which uses the same parameters. Figure 10 summarizes the result of the experiment.

Figure 10a shows the box plot of the average reprojection errors for each calibration session. Our proposed correction method improved the reprojection error compared to Full ~~IN~~DICA to a statistically significant level. On the other hand, the improvement was not significant compared to Recycled ~~IN~~DICA. The discussion section analyzes this observation. All ~~IN~~DICA-based methods demonstrate more stable results than the Degraded SPAAM, corroborating the findings of other work.

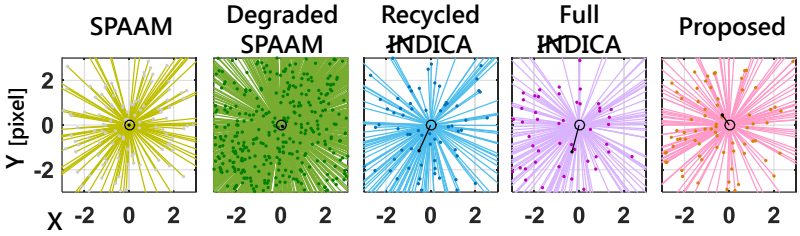
Figure 10b visualizes the effect of the distortion correction. It draws reprojected 2D points of Full ~~IN~~DICA and of the proposed



(a) Boxplot of the mean reprojection errors for each calibration session of each method. Our method improves Full ~~IN~~DICA accuracy significantly ($p \approx 6.37e\text{-}04 \ll 0.05$, two-sample t-test), while it did not defeat the Recycled ~~IN~~DICA ($p \approx 0.46$).



(b) Reprojected 2D points by the two methods from each acquisition step. The amount of distortion correction is visualized by arrows with scaling. The estimated distortions are similar to what we learned in the regression dataset.



(c) Distribution of the reprojection error vectors. Black vectors drawn from the center of each plots are the mean value of each distribution. As expected, SPAAM has a mean vector which is almost at the center. The proposed method made the mean vectors of Full ~~IN~~DICA closer to the center, thus our method reduced the systematic errors existed in our previous automated calibration method.

Fig. 10. User-based calibration experiment. (a) Box plot of the mean errors. (b) Distribution of reprojection error vectors. (c) Visualization of reprojected 2D points on the virtual screen.

method for each data acquisition session. From the error vectors between the points, estimated distortions look concentric and radial.

Figure 10c presents the error vectors of each method in separate boxes. The error vectors are defined as vectors from 2D points corrected by the user to 2D points reprojected from corresponding 3D points. In the visualization, the error vectors are shifted such that they all start at the origin. Each box also visualizes the mean of the end points of the error vectors. SPAAM shows an almost centered mean value. This is expected since the LM method estimates a projection matrix such that the mean error is minimized even if there are outliers. Since the 2D-3D *ground truth* data would contain noisy samples due to the manual alignment, the SPAAM result is likely to be overfitted.

On the other hand, the mean errors of our previous ~~IN~~DICA methods exhibit large offsets from the center. In other words, the reprojected 2D points of the methods are shifted in a particular direction – the methods contain systematic errors in their projection matrices. However, our correction method *shifts back* the mean error of Full ~~IN~~DICA closer to the center. Therefore, our method reduces the systematic errors that the previous automated method (Full ~~IN~~DICA) had.

5.4 Discussion

Throughout the two experiments, our correction method increased the calibration accuracy significantly and reduced the systematic errors which have existed in our previous interaction-free calibrations.

In the camera-based experiment, our method demonstrated that it improved the SPAAM calibration to subpixel level by precorrecting the distortion caused by the OST-HMD's optical elements. A natural question following this result was how much our method is contributing to the real, user-based OST-HMD calibration.

In the user-based calibration, our method also improved the calibration accuracy against Full ~~IN~~DICA. However, the accuracy had no significant difference against Recycled ~~IN~~DICA. A reason might lie in the recycled projection matrix in Recycled ~~IN~~DICA. In the experiment, the recycled projection matrix was from a standard user-based SPAAM, which means that the user aligned *distorted* 2D points to 3D points. And the DLT and LM methods estimated a projection matrix which best fit the distorted correspondences to the extent allowed by the perspective camera model. Thus, the recycled projection matrix partially contributed to an implicit compensation of the optical distortion in Recycled ~~IN~~DICA.

We conject that this is why the Recycled ~~IN~~DICA is yielding as low a mean error as our correction method while showing higher error variance – a systematic error possibly induced by the forcibly fit projection matrix.

Even though the original SPAAM is prone to overfit the given 2D-3D correspondences, the automated methods have not performed as accurately as the SPAAM calibration, yet. Why? Why do such gaps still exist? We have several hypotheses stemming from the fact that ~~IN~~DICA models the eye-HMD vision system as a naive pinhole camera, which is not true when we scrutinize the optical models of OST-HMD optics and the anatomical model of the human eye.

Firstly, OST-HMDs have distortions in their virtual screen, whereas projection has been assumed to be planar in our current model. Our correction method considers an optical phenomenon that the optical elements distort incoming world light rays. In the same manner, the elements also distort virtual screens perceived by users into a non-planar surface [20, 21, 27]. Even the assumption that we treat the virtual screen as a collection of 3D points floating in mid air is violated when the light sources are collimated as in retinal displays. Camera-based experiments would suffice to justify and evaluate those HMD-related hypotheses.

Secondly, the visual axis of the human eye differs from the optical axis in the two-sphere eye model we employed [28] This issue requires actual users. A camera integrated in a prosthetic eye might be an alter-

native, yet we have no clue how accurately such a system can mimic the real eye.

Yet another issue is the reliability of our 2D-3D correspondence dataset, which is collected manually. Although the dataset was produced by a human expert, the 2D-3D dataset may still contain a large amount of noise: if the noise is more than a few pixels in the true projected 2D points, it would be meaningless to argue about calibration errors in subpixel range – or impossible to obtain major significance despite the potential of a new method.

What would help to justify this last hypothesis is to conduct the same experiment with many different subjects in a very controlled environment such as in [24]. Perhaps such a study can create a benchmark dataset as a by product. Similarly, our method would also require a proper user-study as a follow-up.

As a final remark, let us reconsider the distortions by the optical elements. In the experiments with the nVisor ST60, the estimated distortion was rather concentric. Other OST-HMDs may have different characteristics. For example, EPSON BT-100 has a linear distortion due to its thick planar optical element. As another example, Lumus DK32, a high-end OST-HMD with ultra-thin light-guide elements, seems to have little distortion. Thus it might not benefit from our distortion corrections as much as the ST60 does. In this way, as a follow up study, it would be interesting to apply our non-parametric distortion correction to various HMDs.

6 Conclusion and Future Work

We have proposed an OST-HMD calibration method. It corrects an optical distortion which conventional eye-HMD models have not considered – the distortion of the light rays passing through the optical elements of OST-HMDs. Our method consists of both offline and online steps. In the offline step, we learn a 4D-to-4D light field mapping which converts each original light ray to its distorted ray. The learning is done by first collecting the light rays measured with/-out an optical element, then computing the mapping via a non-parametric regression. Then, at the online step, the method compensates the distortion by using the mapping given an eyeball position from the interaction-free OST-HMD calibration method. Our experiments show that the correction method reduces the systematic error which has existed in both conventional camera-/user-based calibrations, and also significantly improves calibration accuracy.

Future work directions involve: considering the distortion of the virtual screen [20, 21, 27] which is assumed to be planer in this paper, deepening the understanding of the eye-dependent parameters [28], investigating the possibility of automated frame-wise OST-HMD calibrations, establishing and refining ways to compare different calibration methods with both subjective [24] and objective error measurements, overcoming the latency issue which is also another dominant aspects directly affects to the spatial registration quality [37], and so on.

7 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. Axholt. Pinhole camera calibration in the presence of human noise. *Linköping University Electronic Press*, 2011.

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

[3] R. Azuma and G. Bishop. A frequency-domain analysis of head-motion prediction. In *Proceedings of ACM SIGGRAPH 1995*, pages 401–408, 1995.

[4] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, Aug. 1997.

[5] O. Bimber, A. Grundhöfer, G. Wetzstein, and S. Knödel. Consistent illumination within optical see-through augmented environments. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2003*, pages 198–207, 2003.

[6] O. Cakmakci, Y. Ha, and J. P. Rolland. A compact optical see-through head-worn display with occlusion support. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2004*, pages 16–25, 2004.

[7] F. Devernay and O. D. Faugeras. Straight lines have to be straight. *Journal of Machine Vision and Applications*, 13(1):14–24, 2001.

[8] J. Didier, D. Roussel, and M. Mallem. A time delay compensation method improving registration for augmented reality. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA) 2005, April 18-22, 2005, Barcelona, Spain*, pages 3384–3389, 2005.

[9] 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.

[10] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proceedings of ACM SIGGRAPH 1996*, pages 43–54, 1996.

[11] R. R. Hainich and O. Bimber. *Displays: fundamentals and applications*. CRC press, 2011.

[12] R. I. Hartley and S. B. Kang. Parameter-free radial distortion correction with center of distortion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(8):1309–1321, 2007.

[13] W. Heidrich, H. P. A. Lensch, M. F. Cohen, and H.-P. Seidel. Light field techniques for reflections and refractions. In *Rendering Techniques*, pages 187–196, 1999.

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

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

[16] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI), Minneapolis, MN, USA, March 29-30, 2014*, pages 75–82, 2014.

[17] Y. Itoh and G. Klinker. Performance and sensitivity analysis of INDICA: Interaction-free display calibration for optical see-through head-mounted displays. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2014*, pages 171–176, 2014.

[18] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. Rendering for an interactive 360 light field display. 26(3):40, 2007.

[19] K. Kiyokawa, Y. Kurata, and H. Ohno. An optical see-through display for mutual occlusion with a real-time stereovision system. *Computers & Graphics*, 25(5):765–779, 2001.

[20] M. Klemm, H. Hoppe, and F. Seebacher. Non-parametric camera-based calibration of optical see-through glasses for augmented reality applications. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2014*, pages 273–274, 2014.

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

[22] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of ACM SIGGRAPH 1996*, pages 31–42, 1996.

[23] S. Liu, D. Cheng, and H. Hua. An optical see-through head mounted display with addressable focal planes. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2008*, pages 33–42, 2008.

[24] K. Moser, Y. Itoh, K. Oshima, E. Swan, G. Klinker, and C. Sandor. Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Virtual Reality 2015)*, 21(4):491–500, Aprial 2015.

[25] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. In *Computer Science Technical Report CSTR*, volume 2, 2005.

[26] S. Nicolau, L. Soler, D. Mutter, and J. Marescaux. Augmented reality in laparoscopic surgical oncology. *Surgical oncology*, 20(3):189–201, 2011.

[27] C. B. Owen, J. Zhou, A. Tang, and F. Xiao. Display-relative calibration for optical see-through head-mounted displays. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2004*, pages 70–78, 2004.

[28] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura. Corneal-imaging calibration for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Virtual Reality 2015)*, 21(4):481–490, Aprial 2015.

[29] M. Qiu and S. D. Ma. The nonparametric approach for camera calibra-

tion. In *Proceedings of IEEE International Conference on Computer Vision (ICCV) 1995*, pages 224–229, 1995.

[30] B. Scholkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[31] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 111–147, 1974.

[32] P. F. Sturm, S. Ramalingam, J. Tardif, S. Gasparini, and J. Barreto. Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(1-2):1–183, 2011.

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

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

[35] C. Y. Vincent and T. Tjahjadi. Multiview camera-calibration framework for nonparametric distortions removal. *IEEE Transactions on Robotics*, 21(5):1004–1009, 2005.

[36] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(11):1330–1334, 2000.

[37] F. Zheng, T. Whitte, A. Lastra, P. Lincoln, A. State, A. Maimonek, and H. Fuchs. Minimizing latency for augmented reality displays: Frames considered harmful. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2014*, pages 195–200, 2014.