

# Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays

Alexander Plopski, Yuta Itoh, Student Member, IEEE, Christian Nitschke, Kiyoshi Kiyokawa, Member, IEEE, Gudrun Klinker, Member, IEEE and Haruo Takemura, Member, IEEE

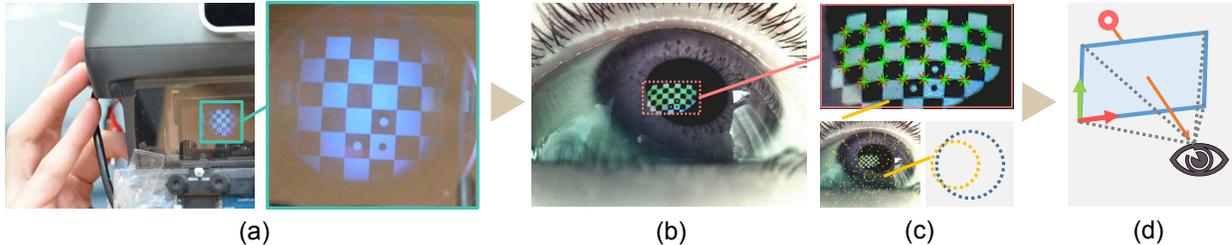


Fig. 1: Corneal Imaging OST-HMD calibration: (a) The user is shown an image on the HMD-screen. Here we used a checkerboard pattern. (b) The reflection of the image on the cornea is captured by a camera and correspondences of points shown on the HMD-screen and points in the camera image are used to estimate the position of the cornea (detected correspondences are shown in green and the reflection on the estimated cornea position in red). (c) In the upper image we show a magnification of the overlay of detected points and the reflection at the estimated cornea position. A hemisphere representing the cornea is visualized using an overlaid yellow sphere in the lower image. (d) Estimating the position of the eye center from multiple cornea centers allows us to automatically calibrate the HMD and associate points on the HMD-screen with projection rays of the scene.

**Abstract**— In recent years optical see-through head-mounted displays (OST-HMDs) have moved from conceptual research to a market of mass-produced devices with new models and applications being released continuously. It remains challenging to deploy augmented reality (AR) applications that require consistent spatial visualization. Examples include maintenance, training and medical tasks, as the view of the attached scene camera is shifted from the user's view. A calibration step can compute the relationship between the HMD-screen and the user's eye to align the digital content. However, this alignment is only viable as long as the display does not move, an assumption that rarely holds for an extended period of time. As a consequence, continuous recalibration is necessary. Manual calibration methods are tedious and rarely support practical applications. Existing automated methods do not account for user-specific parameters and are error prone. We propose the combination of a pre-calibrated display with a per-frame estimation of the user's cornea position to estimate the individual eye center and continuously recalibrate the system. With this, we also obtain the gaze direction, which allows for instantaneous uncalibrated eye gaze tracking, without the need for additional hardware and complex illumination. Contrary to existing methods, we use simple image processing and do not rely on iris tracking, which is typically noisy and can be ambiguous. Evaluation with simulated and real data shows that our approach achieves more accurate and stable eye pose estimation, which results in an improved and practical calibration with a largely improved distribution of projection error.

**Index Terms**—OST-HMD calibration, eye pose estimation, corneal imaging, optical see-through

## 1 INTRODUCTION

The quality of an augmented reality (AR) experience depends on how well virtual content is integrated into the real world—spatially, photometrically and temporally. In other words, if the world does not appear consistent, users are likely to dislike or reject the AR experience as a whole. Correct spatial alignment is of utmost importance for being able

to correctly view an augmentation [14]. If the virtual augmentation is displayed at the wrong location, the user will notice the discrepancy even if other parameters are rendered to match the scene. Therefore, it is important to ensure that the virtual content is correctly aligned.

Current state-of-the-art solutions solve this problem by capturing the world with an external camera and estimating the camera's position relative to the scene. The image captured by the camera is then augmented with virtual content, and the user is shown the augmented view from the view point of the camera. However, this view can greatly differ from the user's perspective, for example by augmenting the view of a mobile phone held by the user. Recent work has shown that displaying the augmentation from the user's perspective has potential to improve the user's perception and overall AR experience [21]. Therefore, head-mounted displays (HMDs) are likely the better choice to present AR to a user. For a long time these devices were limited to laboratories and industrial use-cases. Recent advances in display manufacturing lead to an increasing number of inexpensive, high-quality, consumer-oriented HMDs, such as video see-through adaptations of Oculus Rift (VST-HMDs) or optical see-through Google Glass and Epson Moverio BT200 (OST-HMDs). Although OST-HMDs were at the forefront of AR research [7, 33] they have been replaced by handheld devices and VST-HMDs due to their limitations: a small augmentable field of view,

• Alexander Plopski is with Osaka University.  
E-mail: plopski@ime.cmc.osaka-u.ac.jp.

• Yuta Itoh is with Technical University of Munich.  
E-mail: itoh@in.tum.de.

• Christian Nitschke is with Kyoto University.  
E-mail: christian.nitschke@i.kyoto-u.ac.jp.

• Kiyoshi Kiyokawa is with Osaka University.  
E-mail: kiyo@cmc.osaka-u.ac.jp.

• Gudrun Klinker is with Technical University of Munich.  
E-mail: klinker@in.tum.de.

• Haruo Takemura is with Osaka University.  
E-mail: takemura@cmc.osaka-u.ac.jp.

Manuscript received 18 Sept. 2014; accepted 10 Jan. 2015. Date of Publication 20 Jan. 2015; date of current version 23 Mar. 2015.  
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.

contrast problems and the requirement for constant recalibration of the system, to name a few [19].

Design issues, like the small field of view and contrast can be solved through design improvements and technical progress. Spatial alignment of the virtual content with the real world on the other hand, depends on external factors and must be solved for each user and execution through calibration. Furthermore, if the HMD moves on the user's head, the calibration is no longer ideal and the calibration process has to be repeated. Current solutions such as the Single Point Active Alignment Method (SPAAM) require extensive user input whenever the calibration is performed [2, 10, 40]. This is a very tedious process and it furthermore introduces user dependent errors. As a result the recalibration process is often skipped, which in return impacts the user's experience and acceptance of the OST-HMD.

Several researchers proposed the use of an eye-tracking camera with HMDs for eye-gaze user interaction and to study human behavior [6, 15, 32]. The eye plays an essential role in the calibration of an OST-HMD. If the eye can be tracked by an eye-tracking camera, the relation between the HMD-screen and the eye can be used to calibrate the OST-HMD. Itoh and Klinker [16] build on this idea and use an RGB camera to track the user's eye. They detect the iris contour and use the two-sphere eye-model to determine eye position and gaze-direction. The assumption of empirically measured, uniform eye-parameters allows their Interaction Free Display Calibration (INDICA) method to be applied without any manual calibration. Correctly detecting and tracking the iris is a difficult problem, as it is partially occluded by eye-lashes and eye-lids, especially when the face and gaze direction move away from the camera, and as the iris does not have a sharp edge due to a gradual transition from the transparent cornea to the white sclera. Furthermore, small errors in the detection of the iris can lead to large errors in the estimated eye pose. Research in eye-gaze tracking has shown that methods estimating the eye gaze from iris contours [27, 30, 43], as used in [16], generally achieve a low precision of about 6 degrees, which consequently results in a bad estimation of the eye position [35]. A more stable approach that is commonly used in commercial eye trackers is the pupil-center–cornea-reflection (PCCR) method [11, 13, 36, 41]. It exploits the idea that the cornea can be modeled as a spherical mirror reflecting light from the environment. Correspondence matching of specular highlights (glints) from at least two known 3D point light sources (IR-LEDs) then allows estimating the position of the cornea with known size.

We propose a similar approach (shown in Fig. 1) to compute the relationship between the OST-HMD and the eye. Different from conventional eye trackers, we do not rely on additional light sources, complex illumination algorithms and pupil tracking. Instead, we use *corneal imaging* [27, 30] to analyze the reflection from an existing, pre-calibrated OST-HMD-screen on the cornea. In an eye image, we detect the corneal reflection of the image on the screen and find correspondence pairs of screen and reflection locations to estimate the position of the physical cornea. From at least three non-coplanar cornea positions, we obtain the eye position as the center of rotation. This achieves a practical and lightweight HMD (re-)calibration and gaze tracking method, where the point of gaze (PoG) is obtained as the intersection of the gaze ray with the screen and the scene.

With this paper, we make the following main contributions:

- We propose Corneal Imaging Calibration (*CIC*), a novel approach for OST-HMD (re-)calibration based on cornea position estimation using corneal imaging to obtain correspondence pairs from a calibrated HMD-screen and its content reflection on the cornea in an eye image.
- We show that the approach has major advantages over state-of-the-art methods: It is more practical as it uses simple and automatable image processing, less reliant on correct eye modeling as the error propagates less into the result, and more robust as it does not rely on iris detection and allows for a large number of correspondence pairs with sub-pixel accuracy.

- The method is suitable for high-quality eye-gaze tracking as it employs an approach similar to commercial eye-gaze trackers. Contrary to existing solutions, it is lightweight as it does not require additional hardware in form of IR light sources and complex illumination for pupil detection.

The remainder of this paper is structured as follows: Section 2 surveys approaches for OST-HMD (re-)calibration, eye tracking and corneal imaging. Section 3 explains the notations used in this paper. Section 4 provides a general overview of the OST-HMD calibration approach. Section 5 explains the proposed method, followed by the implemented experimental setup in Section 6. Section 7 discusses evaluation results compared with state-of-the-art SPAAM and automated INDICA calibration. Finally, Section 8 concludes with a summary and promising future directions.

## 2 RELATED WORK

Estimation of the eye location and the gaze direction are required for various application scenarios, such as gaze interaction. In this paper we show how corneal imaging can be used to calibrate an OST-HMD, i.e., determine the relationship between the eye and the virtual screen for robust alignment of virtual and real content, and subsequently determine the user's gaze direction and the point of regard. In this section we give an overview of the related areas.

### 2.1 OST-HMD calibration

OST-HMD calibration models the human eye as a pinhole camera whose image plane corresponds to the HMD-screen. The goal of the calibration is to determine how points in the real world project onto the HMD-screen. Given a correct calibration, the user perceives perfectly aligned virtual and real scenes.

Research on how to achieve correct calibration has been conducted over many years, from tedious first approaches with special setups [3], to the much simpler manual SPAAM calibration for mono- [40], and stereo-displays [9] and further simplifications and improvements of this technique [10, 22, 24], towards automated calibration [16, 17, 31]. A more detailed explanation of the evolution of OST-HMD calibration can be found in [16].

Display-Relative Calibration (DRC) was proposed by Owen et al. [31]. Instead of estimating all 11 projection parameters at the same time, as is done by SPAAM and its adaptations, they propose to split the calibration into an estimation of the display parameters and the eye position, which is the basis of current automated calibration approaches. The former is a static HMD-dependent parameter that can be determined in an offline process, while the latter is estimated at execution time. They describe five options for the estimation of the user's eye position, ranging from not estimating it at all, over performing a simple warping such as Easy SPAAM [10, 24], to a full 3 DOF eye position estimation with a pupilometer. The online part however, still requires user interaction.

Itoh and Klinker [16] proposed INDICA to determine the position and orientation of the eye automatically. Their method detects the user's iris by an RGB eye camera attached beneath the HMD. From the projection of the circular iris onto an ellipse in the eye-camera image, they compute the 3D iris position and gaze direction. The position of the eye's center of projection is then constrained by a geometric eye model at some distance along the negative gaze direction. At each timeframe, the method computes a new projection matrix regarding the current eye position and HMD parameters. The HMD parameters are obtained offline through a display calibration similar to DRC or from an initial calibration through SPAAM. The authors show that their approach achieves more stable eye-position estimates than that from SPAAM and improves the alignment compared to using a degraded SPAAM calibration, where the recalibration is skipped after the OST-HMD has moved. However, they also show that the current INDICA contains a systematic error that stems from their simple eye-HMD system model [17]. Additionally their method relies on iris detection for eye position estimation and suffer from the drawbacks of existing techniques, such as low accuracy and inapplicable scenarios.

## 2.2 Eye Pose Estimation

Eye pose estimation refers to recovering the position and orientation of a 3D geometric eye model relative to an eye tracking camera. This requires *image processing* to track anatomic structures (iris contour, pupil) or corneal reflections (glints, correspondences), and *geometric modeling* to reconstruct the pose of the eye model from the image information. Eye pose estimation recovers the gaze direction up to the optical axis. An additional one-time individual calibration is necessary to recover the true gaze direction or visual axis [13]. We distinguish between *passive methods* that work on any eye image and *active-light methods* that require additional controlled illumination.

Due to their reduced hardware and calibration requirements, passive methods are often applied with low-cost, non-professional and uncontrolled applications. A common strategy recovers the 3D pose of the eye model by reconstructing the circular iris from its projected elliptical contour in the image [16, 17, 23, 27, 28, 29, 35, 43]. The methods work on natural eye images, but suffer from noisy eye detection and unknown parameters. A major issue is that a single eye camera only allows reconstructing the iris up to a two-way ambiguity that needs to be resolved from further knowledge, constraints [29] and assumptions [16, 17, 35, 43].

Active-light methods are developed for accurate automatic eye tracking and require a complex hardware system with calibrated light sources and controlled illumination. The PCCR technique is largely covered in research [11, 36, 39] and applied in commercial systems. It involves a two-step approach, first estimating the position of the cornea from reflections of multiple known light sources (commonly IR LEDs), then estimating the orientation of the eye from the center or contour of the pupil, segmented using active IR illumination.

Eye pose estimation also allows for eye gaze tracking (EGT). Beside traditional scene-mounted remote trackers, advances in wearable head-mounted trackers are likely to be integrated into HMDs. Wearable eye-tracking systems have been proposed in research [6, 15, 32] and as commercial products, like the Tobii Glasses or the SMI Eye Tracking Glasses. The many applications of EGT include interactive gaze control [32], gaze-reactive behavior and state analysis [6, 34] and passive data collection. First person view applications have moved into the focus in recent years [34, 37, 38].

## 2.3 Corneal Imaging

The cornea is the transparent protective and optical outer layer of the eye. It reflects a small fraction of incident illumination that can be noticed when looking at a person's eye [4, 5]. Analyzing and exploiting such corneal reflections from eye images can be beneficial to accomplish a wide range of tasks [30].

Corneal imaging refers to analyzing the comprehensive reflected illumination from an eye image by modeling the combination of camera and cornea as a catadioptric imaging system [27]. The recovered environment map allows extraction of the visual information and application to computational tasks in vision and graphics, such as face reconstruction, relighting [26] and recognition [25]. Image capturing and quality may be improved through a super-resolution strategy [28]. Analyzing corneal reflections from a known or controlled environment allows for a variety of geometric tasks, such as cornea position estimation [11, 36, 39], cornea shape reconstruction [12] and scene pose estimation [23, 29, 35].

Our proposed method automatically detects the reflection of a pre-calibrated HMD-screen to obtain a number of correspondence pairs of 3D locations and imaged reflections. Exploiting the existing hardware of OST-HMDs, therefore, allows to combine the advantages of passive and active-light methods to achieve practical and accurate eye pose estimation. Beside HMD calibration, this allows for instantaneous EGT, where the PoG is obtained by intersecting the gaze ray with the screen and the scene.

## 3 NOTATIONS

In this section we explain the notations used throughout the remainder of the paper. We denote an object and its coordinate system by an upper case letter, such as  $S$ . A single 3D point is denoted by a bold upper

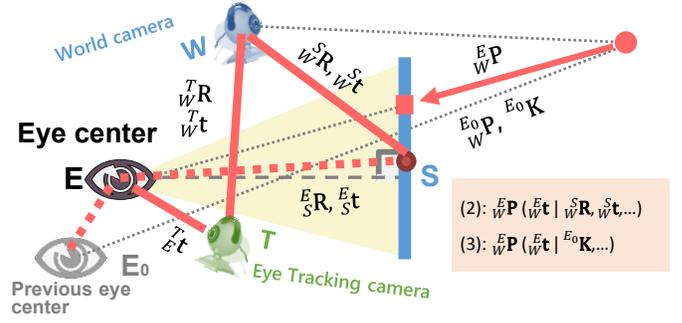


Fig. 2: Schematic overview of the automated OST-HMD calibration [16]. Recycle INDICA (3) and Full INDICA (2), reconstruct the projection as a function of the current eyeball position  $\mathbf{E}$  (and static parameters of the eye-HMD system). Section 4 elaborates the parameters in this figure.

case letter, such as  $\mathbf{P}$ , and a 2D point by a bold lower case letter, such as  $\mathbf{p}$ . If we refer to a point  $\mathbf{P}$  not in the world coordinate system, we introduce the coordinate system as an upper index to the left of  $\mathbf{P}$ , e.g.,  ${}^A \mathbf{P}$  is used to refer to  $\mathbf{P}$  in the coordinate system  $A$ .

We denote a vector between two points with a bold lower case letter, such as  $\mathbf{v}$ , and a directional ray with an additional right arrow, such as  $\vec{v}$ . Given a ray  $\vec{v}$  we describe its unit vector as  $\hat{\mathbf{v}} = \vec{v} / \|\vec{v}\|$ . Lower case letters, for example  $d$ , are used to represent scalar values.

We represent a matrix in sans serif font, such as  $\mathbf{P}$ . In particular, we always refer to a rotation matrix as  $\mathbf{R}$ . If  $\mathbf{T}$  is the transformation from coordinate system  $A$  into coordinate system  $B$  we denote it as  ${}^B_A \mathbf{T}$ .  ${}^B_A \mathbf{T}$  is described by  $({}^B_A \mathbf{R}, {}^B_A \mathbf{t})$  where  ${}^B_A \mathbf{R}$  and  ${}^B_A \mathbf{t}$  stand for rotation and translation respectively. Furthermore, explicit transformation of  ${}^A \mathbf{P}$  to  ${}^B \mathbf{P}$  can be written as  ${}^B \mathbf{P} = {}^B_A \mathbf{R} {}^A \mathbf{P} + {}^B_A \mathbf{t}$ . We refer to the transpose of a matrix or a vector as  $(\cdot)^T$ .

## 4 OVERVIEW OF THE AUTOMATED HMD CALIBRATION

Figure 2 summarizes the automated OST-HMD calibration methods in [16]. The eye-HMD system is commonly modeled as an off-axis pinhole camera with a 3-by-4 projection matrix

$${}^E_W \mathbf{P} := {}^E \mathbf{K} \begin{bmatrix} {}^E_W \mathbf{R} & {}^E_W \mathbf{t} \end{bmatrix}, \quad (1)$$

which is a function of the dynamic eye position  ${}^E_W \mathbf{t}$  estimated online.  ${}^E \mathbf{K}$  has two representations:

$${}^E \mathbf{K} = \begin{bmatrix} \alpha_x & c_x \\ & \alpha_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} z_{SE} & -x_{SE} \\ & z_{SE} & -y_{SE} \\ & & 1 \end{bmatrix} \quad (2)$$

$$= {}^E_0 \mathbf{K} \begin{bmatrix} 1 + z_{EE_0}/z_{SE} & & -x_{EE_0}/z_{SE} \\ & 1 + z_{EE_0}/z_{SE} & -y_{EE_0}/z_{SE} \\ & & 1 \end{bmatrix}, \quad (3)$$

where  $S$  denotes the virtual screen coordinates,  ${}^E_S \mathbf{t} = [x_{SE}, y_{SE}, z_{SE}]^T$ , and  ${}^E_E \mathbf{t} = [x_{EE_0}, y_{EE_0}, z_{EE_0}]^T$ .  $\mathbf{a} := [\alpha_x, \alpha_y]^T$  is a scaling factor that 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  $h$ .  ${}^E_0 \mathbf{K}$  is the intrinsic matrix of another virtual camera defined by the old eye position  $E_0$ .

Full INDICA (2) does not rely on the old eye position  ${}^E_0 \mathbf{t}$ . Instead, it requires the HMD-screen pose  $({}^S_W \mathbf{R}, {}^S_W \mathbf{t})$  and the scaling vector  $\mathbf{a}$ . On the other hand, Recycle INDICA (3) does not rely on these parameters, except for  $[{}^S_W \mathbf{t}]_z$ , and it reuses the old intrinsic matrix  ${}^E_0 \mathbf{K}$ . Let  $T$  be the coordinate system of an eye tracker rigidly mounted on the OST-HMD, then  ${}^E_W \mathbf{t} = {}^S_W \mathbf{R} ({}^T_W \mathbf{R})^T ({}^T_W \mathbf{t} - {}^T_E \mathbf{t})$  ((6) in [16]). Thus,  ${}^T_E \mathbf{t}$  (the eye position w.r.t the eye tracker) is a key parameter of the automated OST-HMD calibration.

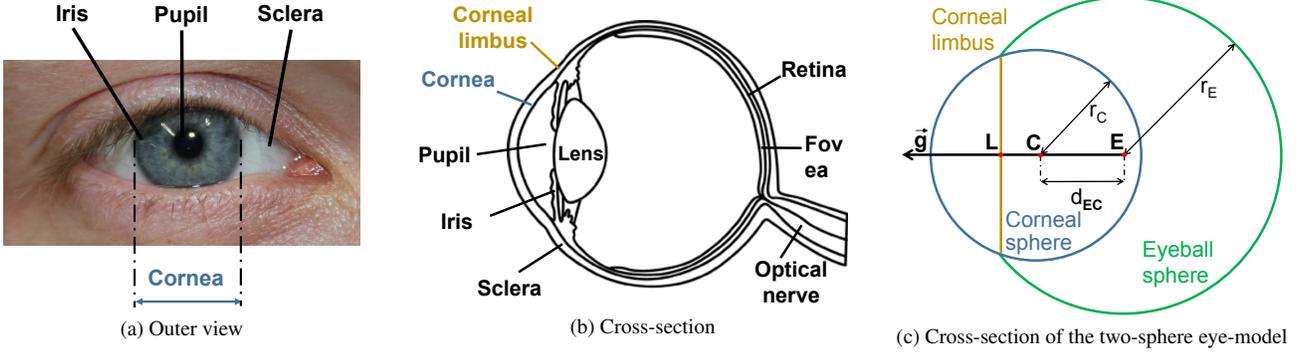


Fig. 3: The parts of the eye and the corresponding geometric eye model used in our method.

## 5 METHOD

### 5.1 Eye Model

The outer view of the eye exposes the textured iris with the pupil in the center. The iris is surrounded by the white sclera, an opaque tissue with primarily protective function (Fig. 3a). A cross-section of the eye (Fig. 3b) reveals that the eyeball is not a simple sphere. Its outer layer can be divided into two approximately spherical surfaces of different radii and centers of curvature (Fig. 3c).

The centers of the spheres lie on the optical axis, which corresponds to the gaze ray  $\vec{g}$ . Figure 3c shows the geometric eye model with the following empirically determined and established parameter values. The corneal sphere is the smaller of the two with a radius  $r_C = 7.8$  mm, and contains, among other parts of the eye, the cornea, the iris and the pupil. The larger sphere is described by the sclera and the center of rotation of the eye. The centers  $C$  and  $E$  of the corneal and eyeball spheres, respectively, are located  $d_{EC} = 5.7$  mm apart along the gaze direction  $\vec{g}$  that originates in  $E$ . The contour of the iris coincides with the intersection of the spheres, which describes the circular limbus with a radius of 5.5 mm.  $L$ , the center of the limbus, lies 5.53 mm away from the corneal sphere center  $C$  along  $\vec{g}$ .

The model is not perfect, as the user-dependent parameters are assumed to be static, and modeling the cornea and eyeball as a spherical surface is not anatomically exact. Nonetheless, we use it as a simple but effective approximation that has been successfully applied in previous studies [16, 17, 23, 27, 28, 29, 39].

About 1% [27] of the incoming light is reflected of the surface of the cornea and can be observed by an onlooker or captured by a camera. In this sense, the eye displays properties similar to a spherical mirror. Thus, the cornea-camera system can be described as a non-axial catadioptric imaging system [27]. Sphere position estimation from known 2D-3D correspondences [1, 36] can, therefore, be applied to determine the position of the cornea relative to the camera for every frame.

### 5.2 Cornea Position Estimation

Our approach determines the position of the eye center from multiple positions of the cornea center. Therefore, for each frame we first need to compute  $C$  from correspondence pairs of points in the camera image and their displayed position on the HMD-screen. This computation consists of two steps. First, we use all correct correspondences to determine the ray  $\vec{r}_{TC}$  from  $T$ , the position of the tracking camera  $T$ , towards  $C$ , represented by the unit vector  $\hat{r}_{TC}$ . In a second step, we estimate the distance  $d_{TC}$  along  $\hat{r}_{TC}$  for each correct correspondence pair separately and combine these into a stable solution.

Let a single frame contain  $N$  correspondence pairs  $\{P, P'\}$  of  $P'$ , a point on the image plane of  $T$ , and  $P$ , a point displayed on the HMD-screen. For the  $i$ -th pair, denoted by  $\{P_i, P'_i\}$ , light emitted from  $P_i$  reflects on the cornea at a point  $P''_i$  and projects onto  $P'_i$  (Fig. 4). Following that, according to Snell's law,  $P_i, P'_i$  and  $P''_i$  span a plane of reflection  $\pi_i$ , which contains  $T$  and  $C$ . The normal of the plane  $\pi_i$ ,  $\vec{n}_i$ , can be computed from  $\hat{r}_{TP'_i}$  and  $\hat{r}_{TP_i}$ , as  $\vec{n}_i = \hat{r}_{TP'_i} \times \hat{r}_{TP_i}$ .

Given two correspondence pairs  $\{P_i, P'_i\}$  and  $\{P_j, P'_j\}$  that describe two non-parallel planes  $\pi_i$  and  $\pi_j$ , the ray  $\vec{r}_{TCij}$  is obtained as the intersection of  $\pi_i$  and  $\pi_j$  since  $T$  and  $C$  are contained in both planes. In general, an erroneous correspondence pair  $\{P_k, P'_k\}$ , for example due to false matching or measurement errors, will describe a plane  $\pi_k$  whose normal will not be perpendicular to  $\vec{r}_{TC}$ . If  $\{P_i, P'_i\}$  and  $\{P_j, P'_j\}$  are correct correspondences, meaning they are inliers of  $\vec{r}_{TC}$ , the ray  $\vec{r}_{TCij}$  will be perpendicular to the normals of the majority of the planes described by the correspondences in the current frame. We determine inlier correspondence-pairs with

$$\left| \left( \hat{r}_{TCij} \right)^T \hat{n}_k \right| \begin{cases} \leq t & \text{if } \{P_k, P'_k\} \text{ is an inlier,} \\ > t & \text{if } \{P_k, P'_k\} \text{ is an outlier.} \end{cases} \quad (4)$$

We use an empirically estimated threshold  $t = 0.0001$  to account for noise and extract the largest subset of  $M$  inliers from the original  $N$  correspondences.

The  $M$  planes spanned by the inliers contain  $\vec{r}_{TC}$  as it is the intersection of all planes  $\pi_1, \dots, \pi_M$ . We estimate  $\hat{r}_{TC}$  as the nullspace of the matrix  $A = [\hat{n}_1 \hat{n}_2 \dots \hat{n}_M]^T$ . Singular value decomposition of  $A$  results in three matrices  $U$ ,  $D$  and  $V$  with  $A = UDV^T$ , where  $\hat{r}_{TC}$  is the last column of  $V$ .

Given the ray  $\hat{r}_{TC}$ , the distance  $d_{TCi}$  from the camera center  $T$  to the cornea center  $C$ , which is supported by the  $i$ -th inlier correspondence pair  $\{P_i, P'_i\}$ , can be computed within  $\pi_i$ . Figure 4 shows the geometric relationship within  $\pi_i$ . The ray  $\vec{u}$ , from  $T$  through  $P'_i$ , intersects the corneal sphere in the point  $P''_i$ , where  $P_i$  reflects at the corneal sphere and projects into the camera. The position of  $P''_i$  is still unknown, as its position varies for different distances  $d_{TCi}$  and with the radius  $r_C$  of the cornea. The position of  $P''_i$  can be expressed as  $P''_i = d\hat{u}$ , where  $d$  is the distance along  $\hat{u}$ , so that  $P''_i$  lies on the surface of the corneal sphere with  $\|P''_i - C\|^2 = r_C^2$ .  $\vec{u}$  reflects at  $P''_i$  as  $\vec{v}$ .

Let  $\vec{w} = P_i - P''_i$  be the ray from the point of reflection towards the point displayed on the HMD-screen.  $\vec{v}$  and  $\vec{w}$  will coincide for the correct distance  $d_{TCi}$ , with

$$\hat{v} \times \hat{w} = 0. \quad (5)$$

Reformulating (5) in  $d_{TCi}$  results in a 6-th degree polynomial with two complex, two real negative and two real positive solutions. Detailed explanation of the computation can be found in [1, 29]. Enforcing that the distance is positive and real leads to two possible solutions for  $d_{TCi}$ . While the correct estimation will remain stable for all planes, the false solutions will vary greatly. We use the median of all values as the estimate of the correct distance. The position is further refined by solving

$$\tilde{C} = \operatorname{argmin}_C \sum_{i=1 \dots M} \hat{v}_i \times \hat{w}_i. \quad (6)$$

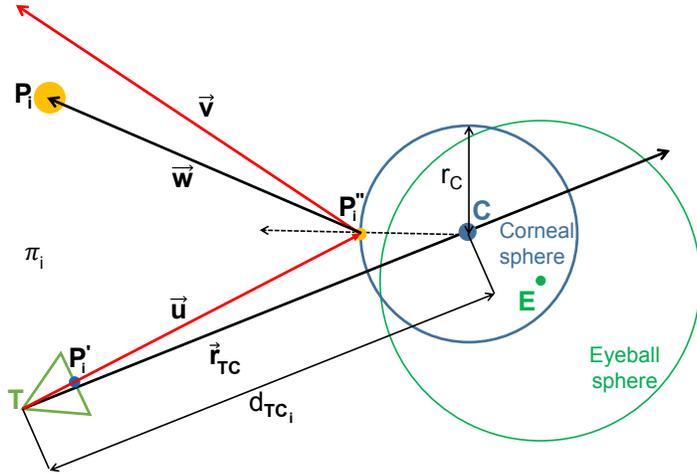


Fig. 4: After computing the ray  $\vec{r}_{TC}$ , the distance  $d_{TC_i}$  from the tracking camera to the cornea center, for a correspondence pair of  $P_i'$  on the image plane and  $P_i$  on the HMD-screen, can be determined within the plane of reflection,  $\pi_i$ . The plane  $\pi_i$  is spanned by  $T$ ,  $C$ ,  $P_i'$  and  $P_i$ . The ray  $\vec{u}$  from  $T$  through  $P_i'$  intersects the corneal sphere in  $P_i''$ , where it reflects as  $\vec{v}$ , according to Snell's law. For the correct distance  $d_{TC_i}$ ,  $\vec{v}$  will correspond to  $\vec{w}$ , the ray from  $P_i''$  to  $P_i$ , and intersect with  $P_i$ .

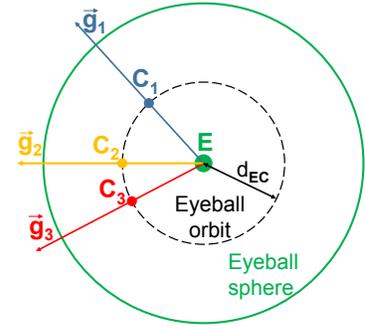


Fig. 5: Cross-section view of rotations of the eyeball. The corneal spheres centered at  $C_1$ ,  $C_2$  and  $C_3$  result from three different gaze directions. The centers lie on a spherical orbit of radius  $d_{EC}$  around  $E$  (shown in green).

### 5.3 Eye Position Estimation

When the gaze direction changes, the eye rotates around  $E$ , and  $C$  moves along a spherical orbit with radius  $d_{EC}$  around  $E$ . Figure 5 shows the cross-section of a rotating eye. Following this, we can reconstruct  $E$  from multiple cornea centers. For a known  $d_{EC}$ ,  $E$  can be estimated from three cornea centers. Three cornea centers and the radius  $d_{EC}$  describe two possible eye centers. However, we can eliminate the solution located closer to the camera, since this arrangement is anatomically implausible. We fit a sphere to all estimated cornea centers to increase the robustness of the estimation. To compute a stable eye position  $\hat{E}$  it is necessary to remove outliers, which are not within a distance  $d$  from the surface of the sphere supported by the majority of the calculated centers. To determine the inlier subset we first compute an eye center  $E_0$  for any possible combination of three cornea centers and count the number of inliers, with

$$\| \|E_0 - C_k\| - d_{EC} \| \begin{cases} \leq d & \text{if } C_k \text{ is an inlier,} \\ > d & \text{if } C_k \text{ is an outlier.} \end{cases} \quad (7)$$

In our experiments we found that a threshold  $d = 0.3$  mm provides the best results. For the remaining  $K$  inlier cornea centers we compute  $\hat{E}$ , from the initial guess  $E_0$ , with

$$\hat{E} = \operatorname{argmin}_E \sum_{i=1 \dots K} \| \|C_i - E\| - d_{EC} \|. \quad (8)$$

**Eye Gaze Tracking** Eye gaze tracking has many applications, such as gaze-based interaction, gaze or eye-movement sensing and gaze-reactive performance analysis. Although not the main goal of our approach, it allows computing gaze without the need for an additional calibration. After estimating the eye position  $E$ , the gaze direction for subsequent frames is obtained as  $\vec{g} = C - E$ .

**Drift Detection** As the HMD calibration and the gaze estimation require a good estimate of  $E$  it is necessary to determine when the HMD has moved and re-estimate  $E$ . Therefore, let  $E_0$  be the current eye position and  $C_j$  the cornea position for the current frame. If the HMD-screen has not moved, it follows that  $d_{EC} = \|E_0 - C_j\|$ . This does not hold if the HMD has moved or the estimation of the cornea has failed. Therefore, we observe subsequent frames: If the majority of these frames supports  $E_0$  we conclude that  $C_j$  is the result of an erroneous estimation. On the other hand, if the majority suggests that

the HMD has moved we estimate a new  $E$ . We use a sliding window to determine a stable  $E$  that continuously accounts for HMD movement. The size of the sliding window depends on the desired stability, while three frames are enough to obtain a guess for a new eye center position, a larger number ensures more stable results (discussed in detail in Section 7.2.3).

## 6 IMPLEMENTATION

### 6.1 Hardware Setup

We have built an OST-HMD system equipped with an eye tracker as described below and in Fig. 6. We use an NVIS nVisor ST60 OST-HMD with  $1280 \times 1080$  resolution. The left-eye display is used for the current setup. An outward looking Delock USB 2.0 camera with a 64-degree lens serves as the world camera  $W$ , and another inward looking Delock camera with a 55-degree lens as the eye tracker  $T$ . The cameras are attached to the HMD and provide  $1600 \times 1200$  resolution video.

The position of the tracker is chosen to be at the bottom of the left-eye display of the HMD. The default focal length of its fixed-focus lens is manually adjusted and fixed to a suitable length.

### 6.2 Feature Detection in Corneal Images

An important step in the computation is a reliable assignment of 3D coordinates to 2D screen pixels. The 3D location of the displayed pattern is known from the HMD-screen calibration, e.g., [17], which assigns a 3D location to every pixel on the screen. Although it is desirable to include all points reflected in the eye, we use only the subset that can be robustly detected. Here, we use the inner corners of a displayed checkerboard pattern to compute matches, as these can be robustly detected even in images distorted by the spherical reflection, and their positions can be determined with sub-pixel precision. Note that to realize a non-intrusive online calibration, the checkerboard may be displayed at a high-enough framerate to become imperceptible, or replaced with any arbitrary screen content that allows robust detection of 2D-3D correspondences.

We use a C++ implementation of the LibCBDetect [8] library, as the original MATLAB implementation cannot be used for real-time calibration, the goal of automated calibration methods. Our naïve implementation on an Intel i7-7000 with 32 GB RAM performs the detection step in under a second on a  $1600 \times 1200$  image. An optimized

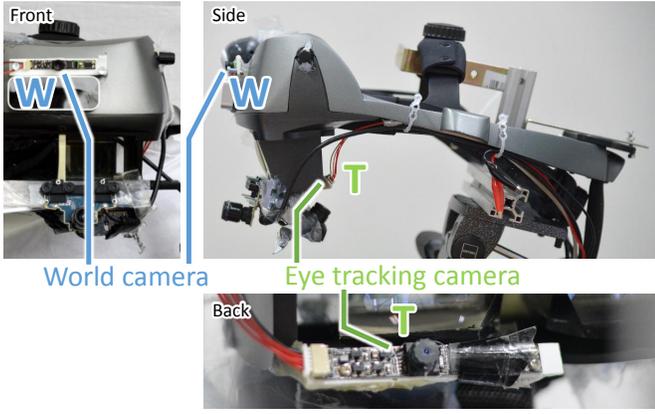


Fig. 6: Hardware setup. (Upper left) A front view of our OST-HMD (NVIS nVisor ST60). (Upper right) A side view of the HMD. (Bottom right) A view from behind the HMD optics. The world camera attached on the HMD, and the eye tracking camera fixed beneath the left optical element of the HMD.

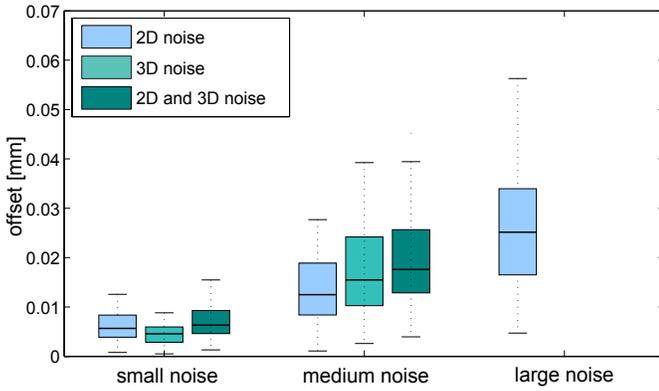


Fig. 7: Deviation of estimated eye positions from the ground truth due to data perturbed by three levels of noise: Small noise, which is likely to occur; medium noise, which will occur from small inconsistencies or calibration errors; and large noise, which is unlikely to occur.

implementation (e.g., on a GPU) and cropping of the examined area will further improve the processing speed, which we plan to do in the future. LibCBDetect detects the inner corners of a checkerboard, even if reflected on a sphere, with sub-pixel accuracy, and arranges them into a grid of at least 9 points, thus further improving the robustness to outliers. The library returns multiple disconnected grids if one or more points of the inner grid are not detected.

We align each grid separately and collect the aligned points afterwards into correspondences for each image. To determine the location of the grids, we have printed a number of dots into the squares of the checkerboard (Fig. 9a-c). Their location is static, and a detected imprinted square allows to align the origin of the coordinate systems. Here, we employ pattern matching to detect imprinted squares, as follows: As the corners of each square are known, we can re-project the enclosed area onto a squared image and compare the result with each possible template using an SAD similarity measure. The pattern with the smallest difference is chosen.

Assuming that the orientations of image plane and HMD-screen are aligned and that the captured image is a reflection, allows to align the orientations of the displayed and detected grids. Given the orientation and location, each point on the detected grid can be matched with its 3D coordinate on the screen. After correspondence matches have been computed, the estimation of the cornea position and the subsequent (re-)estimation of the eye center can be conducted in real-time.

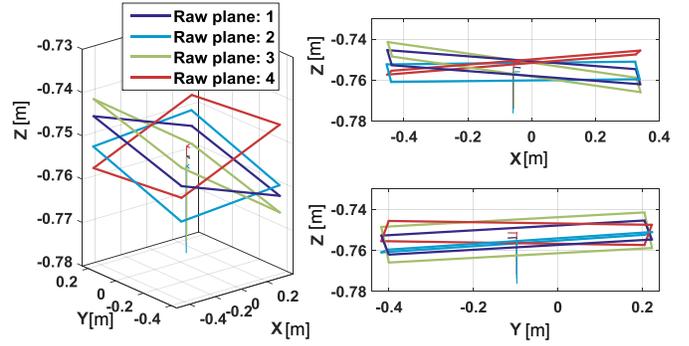


Fig. 8: Intermediate output from the off-line spatial HMD-screen calibration step for computing  $({}^S_W R, {}^S_W t)$ . (Left) 3D visualization of four raw HMD-screen planes estimated from four view points. Our algorithm averages the planes to get a single pose. Each plane is slightly slanted in a similar orientation. (Right top) X-Z-plane view. (Left) Y-Z-plane view. Note that the Z-axis, i.e., the view direction, is stretched to visualize the rotation instability.

## 7 EXPERIMENTS

### 7.1 Simulated Environment

To evaluate our method we generate simulated data according to the two-sphere eye-model. The virtual screen is positioned 700 mm behind the camera, similarly to our HMD-screen, and the virtual eye-tracking camera has the same intrinsic parameters and captures an image of the same size as the eye-tracking camera used in our experiments. First, we project 16 cornea positions from 16 different gaze directions into the camera. For each cornea position we select pixels located above the projection of the cornea center which are arranged in a grid pattern, similar to the observations in our HMD calibration experiment (described in Section 7.2). For each selected pixel  $\mathbf{p}$  we compute the back-projection ray, its reflection at the cornea, and the position of the corresponding point  $\mathbf{P}$  as the intersection of the reflection ray with the virtual screen. To evaluate how susceptible our method is to noise, we add zero-mean noise with standard deviations  $\sigma_{\mathbf{p}}$  and  $\sigma_{\mathbf{P}}$  (defined below) to  $\mathbf{p}$  and  $\mathbf{P}$ , respectively. We test the following scenarios:

1. Only the pixels have noise (representing erroneous detection of feature points in the camera image),
2. Only the 3D points have noise (simulating erroneous HMD-screen calibration),
3. Both, pixels and points present similar noise levels.

Assuming correct eye model parameters  $r_C$  and  $d_{EC}$ , we evaluate combinations of the following small to large noise levels:  $\sigma_{\mathbf{p}} = \{0.2, 0.5, 1\}$  pixel and  $\sigma_{\mathbf{P}} = \{0.5, 2, 10\}$  mm. For each estimated cornea position, we calculate the re-projection error between the forward projection of  $\mathbf{P}$  and  $\mathbf{p}$  and discard estimations with a mean re-projection error  $> 2$  pixels, so that only accurate cornea estimates contribute to the eye center estimation. Figure 7 shows the eye center deviation for 100 executions at each noise level. For  $\sigma_{\mathbf{P}} = 10$  mm, the estimation does not succeed, as no corneal spheres satisfy the requirements. This shows the importance of a good pre-calibration for the HMD-screen. Our tests show that the expected accuracy for the detection of  $\mathbf{p}$  is enough to produce stable results.

Additionally, we investigate the influence of a false assumption for the cornea radius  $r_C$  by adding offsets  $\tau = \{-2, -1, 0, 1, 2\}$  mm with ideal measurements for  $\mathbf{p}$  and  $\mathbf{P}$ . We observe that the deviation consisted only of a depth-error that behaved similar to a linear function of  $\tau$ . However, if the system has extremely low noise levels, we expect that it will be possible to estimate  $r_C$  as described in [41], thus improving the overall results. Although only correspondences from two IR light sources are used in [41], our experiments show that their findings persist, even for a larger number of correspondences.

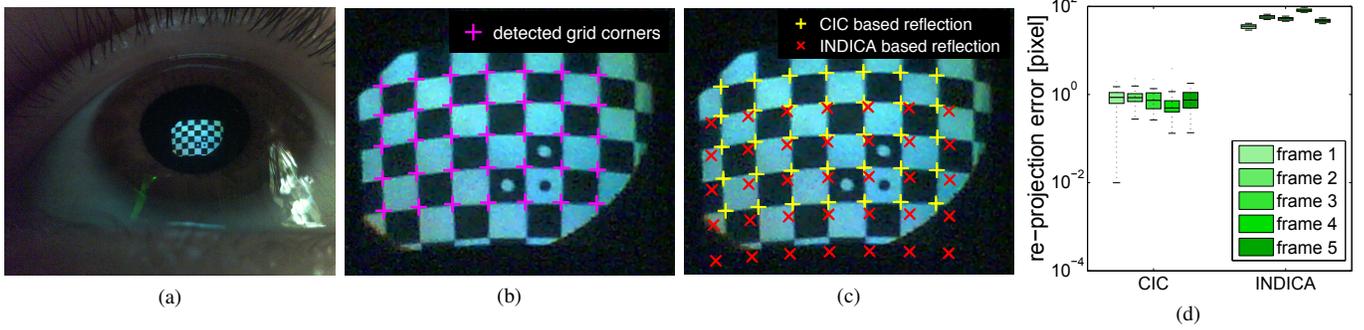


Fig. 9: Cornea position estimation evaluation from re-projection errors of the reflected checkerboard corners into the camera. (a) Image of the reflected checkerboard shown on the HMD-screen. (b) Detected checkerboard corners in the image. (c) Re-projected corners using corneal sphere positions estimated by INDICA and CIC. (d) The large re-projection error for INDICA persists throughout the sequence (Note the logarithmic scale of the error).

## 7.2 HMD Calibration

Although, results of CIC and INDICA do not depend on the expertise of the user, the results of the SPAAM calibration may degenerate, if done by a novice. Therefore, to evaluate the performance of our method in a realistic scenario we calibrate the HMD from multiple sequences, which were recorded by a single user well acquainted with the SPAAM calibration. For each sequence the HMD was placed so that the eye-tracking camera could capture a sharp image of the reflection on the user’s cornea. Each sequence is comprised of two steps: first the user looked at various points displayed on the screen of the HMD. This part is used to estimate the position of the eye center by our method. The second part was a SPAAM calibration to acquire ground truth data of the projection of 3D coordinates onto the screen. To calibrate the system with INDICA, we selected viable frames from the complete sequence. During each recording session the user was asked to keep the HMD fixed on the head. From each sequence we manually removed invalid frames, which could not be used by any method, for example because the user was blinking or the image contained extensive motion blur. While for each sequence the first part was recorded within approximately 2 minutes, the second part required more than 10 minutes.

### 7.2.1 Screen Position

To ensure that the screen pose is calibrated correctly, we captured images of the screen reflected in a spherical mirror of known size instead of the human eye. As the sphere projects onto a conic in the camera, its position can be reconstructed from the occluding contour of the imaged mirror by the algorithm described in [42]. Our method computes the position of a sphere from 2D-3D correspondences, and, thus, we expect results similar to those acquired by [42]. We use five images to verify this.

The average offset is 0.3 mm, which shows that the screen calibration is very precise, although it is not ideal. This could be a result of our assumption of a planar HMD-screen. Visualizing the screen captured in every camera frame of the calibration step, we observe slight offsets of the screen planes (Fig. 8), which may result from a slightly curved screen. These observations coincide with the results of [20, 31].

### 7.2.2 Cornea Position

The first step of an automated calibration method estimates the position of the cornea. In this section, we compare the results of the HMD-screen reflection on the cornea, estimated by CIC and INDICA. Both methods assume the same two-sphere eye-model and should obtain the correct cornea position. Therefore, the reflection of the checkerboard on the estimated cornea should project onto the detected points, for example those shown in Fig. 9b. As can be seen in Fig. 9c, the reflection on the cornea estimated by CIC almost coincides with the expected points, while the reflection for INDICA greatly differs. The large error is consistently obtained throughout the experimental evaluation

(Fig. 9d) with an average error of 54.936 pixel. This indicates that the eye position estimated by INDICA is only a rough approximation of the actual eye center.

### 7.2.3 Eye Position

The minimal solution for the calibration is error prone due to outliers and ill-posed data. Although it would be ideal to collect a large dataset for calibration, this is not possible in most application scenarios. Also, an automated calibration should require as little time as possible before the user can experience the correct augmentation whenever the application is started or the HMD has moved on the head. Therefore, it is also necessary to determine the minimum dataset size that allows a reliable calibration.

In this section we evaluate the required dataset size for the SPAAM, INDICA and CIC calibrations. With an increasing number of frames each method will converge towards a stable position. We assume that our recorded datasets are large enough for each method to successfully converge onto a position that we define as ground truth. We then select 100 random combinations of  $N$  input data (2D-3D correspondences for SPAAM, frames for INDICA and CIC) from each dataset and evaluate the deviation from the ground truth. The results are shown in Fig. 10. We use at least 6 point pairs for SPAAM, 2 frames for INDICA and 3 frames for CIC. In contrast to any of the three methods, CIC allows us to evaluate the quality of the input data, as the re-projection error  $e$  of the screen reflection on the estimated cornea position is known. Therefore, we assume that all frames with  $e > 2$  pixel are likely to be outliers and remove them from the estimation to further improve the results. Stricter thresholds will naturally speed up the convergence rate at the cost of an increasing number of discarded frames. For CIC we show the results of both, the filtered and unfiltered data. As the re-projection error for approximately 80% of the recorded frames was below the threshold, we believe that it is viable to employ filtering in an application scenario. Our observations show that the converged position of the eye-center estimation for the unfiltered dataset deviates by approximately 0.5 mm from  $\hat{\mathbf{E}}_F$ , the result of the filtered dataset. In some cases, it may not be possible to use filtering due to a generally large re-projection error. Therefore, we observe the deviation of the estimated eye position from  $\hat{\mathbf{E}}_F$  for the general case in Fig. 10c.

The position estimated by SPAAM remains unstable even with 16 samples with an error of approximately 10 mm. INDICA immediately converges to an error of less than 1 mm, which is faster than our method that requires 7 and 16 frames for the filtered and unfiltered datasets, respectively. The high quality of INDICA results from a similar gaze direction for the frames used to estimate the eye center, as the iris detection requires the contour of the iris to be visible. Our method, on the other hand, does not restrict the gaze direction and performs similarly. Furthermore, as is shown in [35], the approach taken by INDICA is very sensitive to erroneous measurements and converges to an incorrect eye center position.

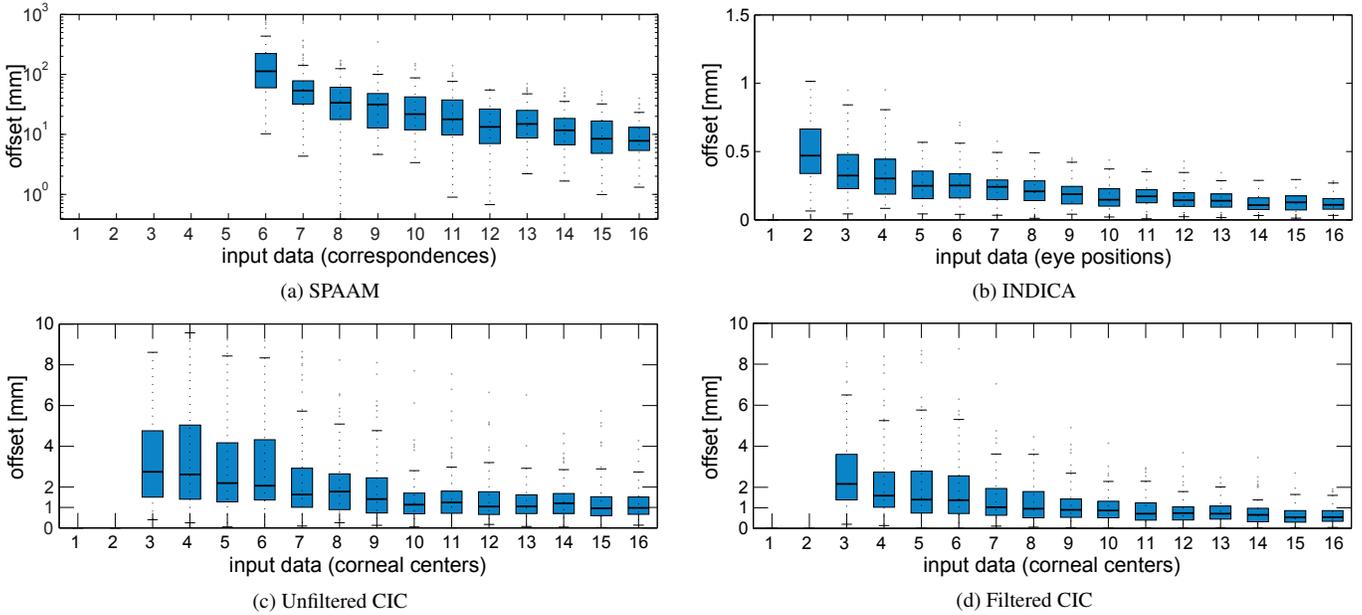


Fig. 10: The convergence rate to a stable eye position determines the time required for the calibration, i.e., how fast a usable result can be obtained. We show the convergence rate for (a) SPAAM (in logarithmic scale), (b) INDICA, (c) CIC with unfiltered data and (d) CIC with a preprocessing step, where we discard probable outliers. We remove all cornea centers, which resulted in a screen reflection re-projection error  $> 2$  pixel. Although CIC converges slower than INDICA, it converges to an expected error of  $< 1$  mm within 7 frames in (d) and 16 frames in (c). SPAAM calibration fails to converge to an acceptable value even after 16 correspondences are used.

#### 7.2.4 Projection Error

In this section we use point correspondences from the second part of the recording session to evaluate the projection error for various setups and calibration methods. We compare the SPAAM, INDICA and CIC calibration methods. For the automated calibration methods we evaluate the Recycle and Full calibration approaches, which were discussed in Section 4.

The results of the projection are shown in Fig. 11. Similar to results reported in [16, 17], the SPAAM calibration has the smallest projection error among the compared methods. This result is expected, as SPAAM incorporates inaccuracies resulting from, e.g., user errors or distortion of the screen into the projection matrix. Our method performs better than the INDICA for both, the Recycle and Full setup. The similar results achieved by INDICA and CIC can be explained by the small deviation of the estimated eye centers. The estimated positions were on average 2.03 mm apart. The eye is modeled as a pinhole camera into which we project points at a distance of 2-3 m. In this model, an offset of 2.03 mm does not introduce a substantial error. Additionally, while the shift in the position, compared to CIC, will degrade the results for parts of the dataset (e.g., points projected to the left of the correct position) it will reduce the error for other parts (e.g., points projected to the right of the correct position), thus disguising the degraded performance. Such an erroneous shift can be observed in the error distribution of the point projections. For each 3D point  $\mathbf{P}$  we compute an error vector  $\mathbf{e} = \mathbf{p} - \mathbf{p}_u$ , where  $\mathbf{p}$  is the projection of  $\mathbf{P}$  onto the screen, after the calibration, and  $\mathbf{p}_u$  is the point aligned by the user. For each calibration method we use  $N$  such projections to compute the mean vector

$$\mathbf{e}_0 = \frac{1}{N} \sum_{i=1}^N \mathbf{e}_i, \quad (9)$$

where  $\mathbf{e}_i$  is the error vector for point  $i$ . We show the error vectors for each calibration approach in Fig. 12. As expected, the SPAAM algorithm computes an ideal distribution of the errors. INDICA shows a strong error tendency, while CIC shows a much more uniform result. The remaining vector  $\mathbf{e}_0$  may result from inaccuracies in the eye model, user errors while aligning the screen with the 3D points, or sub-optimal HMD-screen calibration.

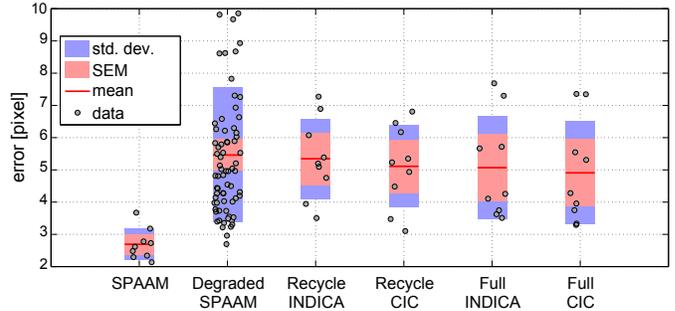


Fig. 11: Projection error for various calibration approaches. CIC performs slightly better than INDICA for both, Recycle and Full calibration approaches. SEM stands for the standard error of the mean.

## 8 CONCLUSION

In this paper, we present CIC, a novel approach for automated spatial calibration of an OST-HMD. This method employs corneal imaging instead of iris detection to determine the position of the user's eye. We employ an HMD with pre-calibrated camera and screen positions to establish correspondences of points displayed on the screen and their reflections on the cornea, as captured by an eye-tracking camera. The correspondences allow to compute the position of the user's cornea, and at least three frames with a moving cornea allow to compute the position of the user's eye center. We show that the position estimated by CIC is closer to the real position, which can improve the calibration results. CIC requires an unobstructed light-path for the HMD-screen reflecting at the user's eye into the camera. The use of contact lenses, instead of glasses, will likely not impact the solution as contact lenses do not refract the light and are attached directly to the cornea surface. However, the impact of vision-enhancing devices on the results of CIC remains to be investigated.

Nonetheless, the results are still not ideal, as CIC did not outperform the manual calibration by SPAAM. In our simulation, we determined that errors in the 3D location of the correspondences and deviation of

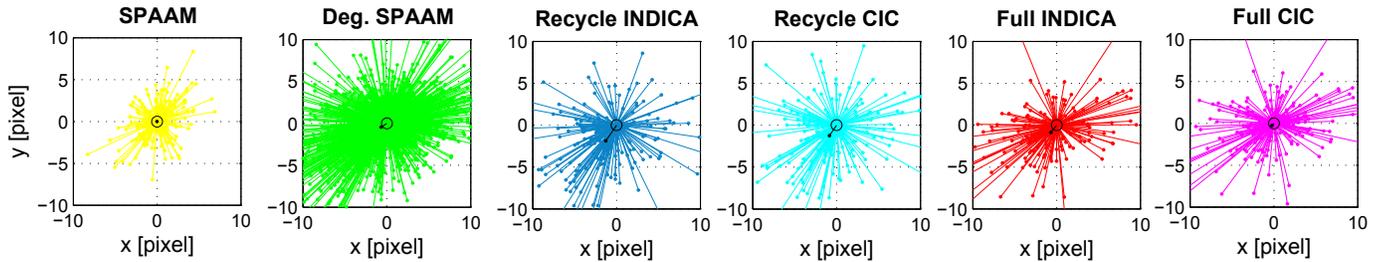


Fig. 12: Error vector distribution for each evaluated method. Each projection of a 3D point describes a vector from the estimated position to the position aligned by the user. The mean error vectors are shown in black. The position estimation by our method improves the error distribution for both, the Recycle and the Full calibration.

the cornea size are the major error sources. Another reason may be that the acquired ground truth information was affected by the optical distortion of the light paths through the HMD optics. Accounting for this non-planar deformation [18] may significantly improve the results. In the future we plan to use an improved HMD-screen calibration, e.g., [20], to acquire better 3D information. Given improved correspondences we plan to investigate if CIC may also be used to estimate user specific parameters, such as  $r_C$  and  $d_{EC}$ .

Obtaining a dense set of correspondences of points on the HMD-screen and their reflections may provide valuable information about the surface of the cornea. Reconstructing and using this shape instead of the simple two-sphere model may further improve the stability of the results. Correspondence matching with real screen content, or fast imperceptible pattern display would allow for continuous non-intrusive re-calibration.

Approaches that employ IR light sources for eye-gaze tracking have achieved a high-quality estimation of the user's gaze with a similar approach. Their results greatly outperform the accuracy of passive trackers based on the iris contour. From these observations we will investigate if eye gaze tracking with CIC can produce similar results.

## ACKNOWLEDGMENTS

This research is funded in part by Grant-in-Aid for Scientific Research(B), #24300048 from Japan Society for the Promotion of Science (JSPS), Japan and by the European Union's Seventh Framework Programme for Research and Technological Development under PITN-GA-2012-316919-EDUSAFE.

## REFERENCES

- [1] A. Agrawal and S. Ramalingam. Single image calibration of multi-axial imaging systems. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1399–1406, 2013.
- [2] M. Axholt, M. A. Skoglund, S. D. O'Connell, M. D. Cooper, S. R. Ellis, and A. Ynnerman. Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration. In *Proc. IEEE Virtual Reality Conference (VR)*, pages 27–34, 2011.
- [3] R. T. Azuma. *Predictive Tracking for Augmented Reality*. PhD thesis, University of North Carolina, 1995.
- [4] M. Backes, T. Chen, M. Duermuth, H. Lensch, and M. Welk. Tempest in a teapot: Compromising reflections revisited. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 315–327, 2009.
- [5] M. Backes, M. Duermuth, and D. Unruh. Compromising reflections-or-how to read lcd monitors around the corner. In *Proc. IEEE Symposium on Security and Privacy (SP)*, pages 158–169, 2008.
- [6] A. Bulling, D. Roggen, and G. Tröster. Wearable eog goggles: Seamless sensing and context-awareness in everyday environments. *Journal of Ambient Intelligence and Smart Environments*, 1(2):157–171, 2009.
- [7] T. Caudell and D. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proc. Hawaii International Conference on System Sciences*, volume 2, pages 659–669, 1992.
- [8] A. Geiger, F. Moosmann, O. Car, and B. Schuster. Automatic calibration of range and camera sensors using a single shot. In *Proc. International Conference on Robotics and Automation (ICRA)*, pages 3936–3943, 2012.
- [9] 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 *Proc. IEEE/ACM International Symposium on Augmented Reality (ISAR)*, pages 165–174, 2000.
- [10] Y. Genc, M. Tuceryan, and N. Navab. Practical solutions for calibration of optical see-through devices. In *Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 169–175, 2002.
- [11] E. D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 53(6):1124–1133, 2006.
- [12] M. A. Halstead, B. A. Barsky, S. A. Klein, and R. B. Mandell. Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals. In *Proc. ACM Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 335–342, 1996.
- [13] D. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):478–500, 2010.
- [14] R. L. Holloway. Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):413–432, 1997.
- [15] Y. Ishiguro, A. Mujibiyah, T. Miyaki, and J. Rekimoto. Aided eyes: Eye activity sensing for daily life. In *Proc. Augmented Human International Conference (AH)*, pages 25:1–25:7, 2010.
- [16] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *Proc. IEEE Symposium on 3D User Interfaces (3DUI)*, 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 *Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 171–176, 2014.
- [18] Y. Itoh and G. Klinker. Light-field correction for spatial calibration of optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Virtual Reality 2015)*, 21(4):xxx–eee, April 2015.
- [19] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff. Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In *Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 177–186, 2014.
- [20] M. Klemm, H. Hoppe, and F. Seebacher. Non-parametric camera-based calibration of optical see-through glasses for augmented reality applications. In *Proc. IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, pages 273–274, 2014.
- [21] G. Liestol and A. Morrison. Views, alignment and incongruity in indirect augmented reality. In *Proc. IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, pages 23–28, 2013.
- [22] P. Maier, A. Dey, C. A. L. Waechter, C. Sandor, M. Tonniss, and G. Klinker. An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration. In *Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 267–268, 2011.
- [23] A. Nakazawa and C. Nitschke. Point of gaze estimation through corneal surface reflection in an active illumination environment. In *Proc. European Conference on Computer Vision (ECCV)*, pages 159–172, 2012.
- [24] N. Navab, S. Zokai, Y. Genc, and E. M. Coelho. An on-line evaluation system for optical see-through augmented reality. In *Proc. IEEE Virtual*

- Reality Conference (VR)*, pages 245–246, 2004.
- [25] K. Nishino, P. N. Belhumeur, and S. K. Nayar. Using eye reflections for face recognition under varying illumination. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 519–526, 2005.
  - [26] K. Nishino and S. K. Nayar. Eyes for relighting. In *Proc. ACM Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 704–711, 2004.
  - [27] K. Nishino and S. K. Nayar. Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1):23–40, 2006.
  - [28] C. Nitschke and A. Nakazawa. Super-resolution from corneal images. In *Proc. British Machine Vision Conference (BMVC)*, pages 22.1–22.12, 2012.
  - [29] C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration using eye reflections and geometry constraints. *Computer Vision and Image Understanding*, 115(6):835–853, 2011.
  - [30] C. Nitschke, A. Nakazawa, and H. Takemura. Corneal imaging revisited: An overview of corneal reflection analysis and applications. *IPSJ Transactions on Computer Vision and Applications*, 5:1–18, 2013.
  - [31] C. B. Owen, J. Zhou, A. Tang, and F. Xiao. Display-relative calibration for optical see-through head-mounted displays. In *Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 70–78, 2004.
  - [32] H. M. Park, S. H. Lee, and J. S. Choi. Wearable augmented reality system using gaze interaction. In *Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 175–176, 2008.
  - [33] J. P. Rolland, R. L. Holloway, and H. Fuchs. Comparison of optical and video see-through, head-mounted displays. In *Proc. Photonics for Industrial Applications (SPIE)*, pages 293–307, 1995.
  - [34] E. Schneider, T. Villgrattner, J. Vockeroth, K. Bartl, S. Kohlbecher, S. Bardins, H. Ulbrich, and T. Brandt. Eyesecam: An eye movement-driven head camera for the examination of natural visual exploration. *Annals of the New York Academy of Sciences*, 1164(1):461–467, 2009.
  - [35] D. Schnieders, X. Fu, and K.-Y. K. Wong. Reconstruction of display and eyes from a single image. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1442–1449, 2010.
  - [36] S.-W. Shih, Y.-T. Wu, and J. Liu. A calibration-free gaze tracking technique. In *Proc. International Conference on Pattern Recognition (ICPR)*, pages 201–204, 2000.
  - [37] A. Tsukada and T. Kanade. Automatic acquisition of a 3d eye model for a wearable first-person vision device. In *Proc. ACM Symposium on Eye Tracking Research and Applications (ETRA)*, pages 213–216, 2012.
  - [38] A. Tsukada, M. Shino, M. S. Devyver, and T. Kanade. Illumination-free gaze estimation method for first-person vision wearable device. In *Proc. IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2084–2091, 2011.
  - [39] N. Tsumura, M. N. Dang, T. Makino, and Y. Miyake. Estimating the directions to light sources using images of eye for reconstructing 3D human face. In *Proc. IS&T/SID Color Imaging Conference (CIC)*, pages 77–81, 2003.
  - [40] 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.
  - [41] A. Villanueva and R. Cabeza. A novel gaze estimation system with one calibration point. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, 38(4):1123–1138, 2008.
  - [42] K.-Y. K. Wong, D. Schnieders, and S. Li. Recovering light directions and camera poses from a single sphere. In *Proc. European Conference on Computer Vision (ECCV)*, pages 631–642, 2008.
  - [43] H. Wu, Q. Chen, and T. Wada. Visual direction estimation from a monocular image. *IEICE Transactions on Information and Systems*, E88-D(10):2277–2285, 2005.