Display-camera calibration using eye reflections and geometry constraints

Christian Nitschke a, *, Atsushi Nakazawa a,b, Haruo Takemura a

a Cybermedia Center, Osaka University, 1-32 Machikaneyama, Toyonaka, Osaka 560-0043, Japan
b PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

A R T I C L E   I N F O

Article history:
Received 30 December 2009
Accepted 25 February 2011
Available online 4 March 2011

Keywords:
Display-camera system
Calibration
Image analysis
Visual system
Human eye model
Eye pose estimation
Corneal reflection
Catadioptric imaging

A B S T R A C T

In this paper, we describe a novel method for calibrating display-camera setups from reflections in a user’s eyes. Combining both devices creates a capable controlled illumination system that enables a range of interesting vision applications in non-professional environments, including object/face reconstruction and human–computer interaction. One major issue barring such systems from average homes is the geometric calibration to obtain the pose of the display which requires special hardware and tedious user interaction. Our proposed approach eliminates this requirement by introducing the novel idea of analyzing screen reflections in the cornea of the human eye, a mirroring device that is always available. We employ a simple shape model to recover pose and reflection characteristics of the eye. Thorough experimental evaluation shows that the basic strategy results in a large error and discusses possible reasons. Based on the findings, a non-linear optimization strategy is developed that exploits geometry constraints within the system to considerably improve the initial estimate. It further allows to automatically resolve an inherent ambiguity that arises in image-based eye pose estimation. The strategy may also be integrated to improve spherical mirror calibration. We describe several comprehensive experimental studies which show that the proposed method performs stably with respect to varying subjects, display poses, eye positions, and gaze directions. The results are feasible and should be sufficient for many applications. In addition, the findings provide general insight on the application of eye reflections for geometric reconstruction.

1. Introduction

Personal computers are turning more and more into multimedia processing machines that come with a large number of peripheral devices. One of these devices is a camera. Historically a tool for videoconferencing, advances in vision algorithms extend its potential applications. As it is mounted on a PC, it can be physically related to the device. The main output device of a PC system is a CRT or LCD monitor. Together with a camera as input device, it can be used to form a controlled system. In the past, there have been two major areas of application for such display-camera systems. One is the reconstruction of 3D object properties, such as shape and reflectance, where the display is used as a controllable planar light-source with the illuminated scene captured by the camera. The other is human–computer interaction (HCI) where the content of the display is adapted according to information about the user which is obtained from the camera.

In object reconstruction, properties of static surfaces are recovered from series of images under varying screen illumination. Photometric stereo methods have been proposed to estimate the shape of lambertian [11,21,52] objects. PC monitor illumination is not directed and focused and is therefore ideal for coping with partially lambertian [20] or non-lambertian objects [29]. Shape reconstruction methods for specular objects [58,7,17,44] identify display-camera correspondences and estimate the respective surface normals. Display illumination has been further applied to analyze the complex light interaction of transparent and translucent objects [39,43] for scene compositing [68] and relighting [56].

Vision-based user interfaces, employ computer vision to “look at people” and perform tasks such as face recognition; head, face, eye, hand, body detection and tracking; facial expression and body movement analysis; and gesture, posture and activity recognition [60,50,32]. Closely related with this work are eye gaze tracking techniques [16,26]. Often, the resulting information does not only passively affect the displayed content but also depends on information about the display itself. Such knowledge is usually obtained by calibration.

Most of the described applications require calibration to find the relation between display and camera. There are mainly two forms of calibration, geometric and radiometric. Geometric calibration
determines the internal camera parameters as well as the external display pose with respect to the camera [7,21,19,58]. Radiometric calibration establishes the relation between the light emitted by the display and the light measured by the camera [58,21,20].

In this work, we focus on geometric calibration where we seek to find the pose of the display with respect to the camera for which internal parameters are already determined. If the screen is directly visible to the camera, the calibration can easily be performed using standard techniques with correspondences obtained from patterns (e.g., checkerboard, structured light) shown on the screen [27,8,10]. This is, however, not possible in the common case where display and camera face a similar direction. In such a configuration, the calibration is achieved by analyzing reflections of a screen pattern from mirroring objects with known shape and pose [55,36]. Methods have been proposed for planar [7,21] and spherical mirrors [19,58]. However, the process is cumbersome as it involves a special mirror as well as tedious physical user interaction.

We describe a novel calibration technique that builds on the observation that the cornea of the human eye acts as a partial mirror and provides rich cues about environmental light [47]. When exploring display-camera setups, we found that the reflected content is clearly visible in eye images of a person in front of the display. This provided the motivation for introducing the idea of using corneal reflections for display reconstruction. The foundation for the proposed approach lies in a combination of Nishino and Nayar’s method for recovering the pose of an eye from its image [47] and Francken et al.’s method for screen-camera calibration using reflections on a freely moving spherical mirror [18]. We discuss a thorough experimental evaluation of this strategy, regarding individual factors, display pose, eye position, and gaze direction, and show that it results in a large error and deviation due to the unknown geometry and size of the individual eye. To compensate for this, we introduce an optimization framework based on known geometry constraints in the setup, achieving considerable improvement that should be sufficient for many applications.

Our novel method makes display-camera calibration substantially more practical and leads to several benefits compared to previous approaches:

- Since no additional hardware is necessary the method is easily distributed and can be applied in existing off-the-shelf setups.
- Without interaction and awareness the calibration can be seamlessly performed by non-expert or disabled persons and children, or, in situations where it is not desired to disclose technical details.
- Accuracy increases with the number of images used. Nevertheless, the minimum requirement is a single face image. This enables online calibration of dynamic setups and allows applications such as camera tracking.
- The method not only reconstructs the pose of the display, but also provides information about eye locations and gaze directions. This makes it ideal to realize human–computer interfaces based on eye gaze tracking.

Table 1 compares the features of the proposed method with previous approaches for display-camera calibration.

The paper makes the following contributions:

- The idea is introduced, to analyze corneal reflections of computer monitor or projection screen illumination in eye images for reconstructing the position and orientation of the screen.
- To verify the proposed strategy, thorough experimental evaluation is conducted for the straightforward combination of eye pose estimation [47] and screen-camera calibration [18], which is found to result in a large error and deviation.

- To compensate for this, an optimization framework is introduced that jointly refines eye poses, reflection rays, and display pose subject to known geometry constraints in the setup.
- A large number of comprehensive experimental studies demonstrate that stable results can be obtained under varying conditions. A quantitative and qualitative comparison with spherical mirror ground truth is provided. The gained insights are not only applicable to the subject of this work, but could also be helpful when analyzing geometric reconstruction from eye reflections in general.
- The described developments and findings enable a novel method for the geometric calibration of display-camera setups that does not require special hardware, explicit user interaction or awareness, and allows online calibration.

Earlier stages of this research have been published before. In [49] we introduce the idea of exploiting eye reflections to calibrate display-camera setups and discuss preliminary results using a single eye image with a constrained setup. In [48] we introduce the optimization framework and show initial experimental results to prove the feasibility of the method. This paper further extends on this work. It provides the most comprehensive and detailed description of the topic. Refinements have been added to eye pose estimation and the optimization framework. Furthermore, three extensive new experimental studies provide important insights regarding the proposed approach and eye reflection analysis in general.

The remainder of this paper is organized as follows: Section 2 gives an overview of existing methods in the areas of geometric display-camera calibration and eye reflection analysis. Section 3 introduces the geometric eye model and explains how to recover the pose of the eye from an image. Section 4 explains how to compute the position of a light source under varying eye poses. Section 5 describes the geometric display calibration from multiple light sources represented as markers in screen patterns. Section 6 introduces the optimization framework. Section 7 details the experimental studies and analyzes the results. Finally, Section 8 discusses the contributions of this work, shows relevant application areas, and explains limitations and Extensions.

2. Related work

We will now provide an overview of existing methods within the two fields related to this work. First, we discuss approaches for mirror-based geometric display calibration in case the screen is not directly visible to the camera. In the second part, we briefly survey areas in computer vision that exploit reflections in the cornea of the human eye. These two fields have one property in common, the combination of a mirror and a lens that forms a catadioptric imaging system [19,47]. While a camera has a single viewpoint, catadioptric systems can have a single or multiple

| Table 1 Feature matrix of geometric display-camera calibration methods. |
|------------------|-------------------|-------------------|-------------------|-------------------|
| Checkerboard pattern Planar mirror Spherical mirror This method |
| [27,8,10] | Attached pattern [21] Plain mirror [7] [18,19,58] |
| Catadioptric Hardware | | | |
| Interaction | + | | |
| Awareness | + | + | + |
| Online | + | + | |
| Eye Poses | + | + | + |
| Accuracy | + | + | + | + |
viewpoints depending on the shape of the mirror and its pose relative to the lens of the camera. Here, we cope with both kinds of systems, single viewpoint for planar and multiple viewpoints for convex mirror methods. Overviews of optical properties are found in [4,22] and [57,38] respectively.

2.1. Geometric display calibration

Funk and Yang [21] use a planar mirror and compute its pose from an additional pattern attached to it. The process is cumbersome as it involves a special prepared mirror, several known parameters and physical user interaction. Bonfort et al. [7] simplify the planar pose estimation by not requiring any marker attachments. Their first method uses a circular hard-drive platter with known interior and exterior radii. Its pose can be obtained from a single image of the circular boundaries which project to concentric ellipses. The second method [7,55] uses at least three poses of an arbitrary planar mirror: At first, a virtual camera pose is calibrated with respect to the screen reflected in each mirror plane. Then, the planes themselves are recovered from the virtual camera poses.

Tarini et al. [58], and in more detail Francken et al. [18], propose calibration techniques using a spherical mirror of known size. The position of the mirror is uniquely determined from a camera image. Extracting the corners from the reflected screen allows to compute the corresponding light rays using inverse raytracing. The real corners are then obtained by intersecting rays from different sphere positions. More recently, Francken et al. [19] propose a refined approach using a series of Gray code patterns to recover a large number of correspondences from only a small number of images. They are able to increase accuracy while reducing the number of sphere positions. Applying a convex mirror also has the advantage that rays are reflected from a wider field-of-view. This makes it ideal for the calibration of large-sized displays, e.g., TV screens in home entertainment setups together with a game console and attached camera. However, all spherical mirror techniques still require special hardware and user interaction.

2.2. Corneal reflection analysis

Since the cornea of the human eye exhibits mirror-like properties, specular reflections from eye images have been exploited in several works from different areas. In confidential data security, Backes et al. [3] present an eavesdropping technique to recover screen content from reflections in the user’s eyes at faraway locations using a telescope mounted on a camera. They improve on this in [2] by accounting for motion and defocus blur using non-blind image deconvolution. While this work is interesting, it does not apply any information about geometric and photometric properties of the eye.

In biomedicine, a detailed knowledge of corneal surface geometry is required for many optometric applications. It is usually measured by a non-intrusive optical technique known as videokeratography [41,6]. Therefore, Halstead et al. [24] describe a popular algorithm that reconstructs a 3D surface model of the human cornea from specular reflections by illumination with a pattern of concentric rings using a special device.

A lot of applications do not require the individual’s corneal geometry when analyzing reflections. In fact, it is often approximated as a spherical cap or paraboloid with static parameters since it is nearly constant among different persons [54,35]. In computer graphics and vision, corneal reflections are exploited because they capture the illumination distribution surrounding a person. Tsumura et al. [59] are the first to recover information about lighting from specular highlights in the eye. They estimate the directions toward three known light sources and reconstruct a 3D facial model for relighting using photometric stereo. Johnson and Farid [33] analyze corneal reflections to detect digital forgeries in the case where an image is composed from individuals photographed under different lighting conditions. Nishino and Nayar [47] provide the first comprehensive analysis of the visual information that is embedded within an image of the human eye. They derive a geometric model of the cornea and analyze the properties of the resulting multi-viewpoint catadioptric (corneal) imaging system. They describe several applications, such as spherical panoramas, retinal images, structure from binocular stereo, facial reconstruction and relighting [46]. Our method for eye pose estimation from a single eye image is based on the one described by Nishino and Nayar [47].

Another application of eye images is security systems using biometrics to automatically identify or verify a person. A method enabling non-intrusive and large-scale surveillance is face recognition [67,1]. Practical solutions have to perform reliably under the large variation of facial poses and illumination found in real environments. Regarding illumination, Nishino et al. [45] propose an appearance-based approach which exploits lighting conditions estimated from corneal reflections. Explicit 3D model-based face recognition can be realized based on photometric stereo from corneal reflections [47,59].

An important direction in eye analysis with applications in various disciplines [15] is eye gaze tracking to measure where a person is looking [16,25,61,26,66]. Common approaches analyze corneal reflections, retinal reflections, or imaged eye features. There is a large variety in systems with respect to (1) deployment (head-mounted/remote, with/without free head motion), (2) hardware (passive/active light, single/multiple cameras, other sensors) and (3) result (implicit image-screen or image-image mapping/explicit 3D eye pose and gaze, point of regard (PoR)).

Focusing on remote approaches, Yoo and Chung [65] exploit the fact that the cross-ratio is an invariant of projective space. They introduce a method where four infrared (IR) LEDs are placed at the corners of the screen to create corneal reflections and a fifth one near the lens to illuminate the pupil. The gazed PoR on the screen can then be estimated from the pupil center within the reflected screen boundary. The initial method is analyzed and improved on by Coutinho and Morimoto [12,13], Ko et al. [37], and Kang et al. [34]. The cross-ratios method relates to our proposed method as it exploits corneal reflections from the screen corners, however, using additional LEDs rather than screen illumination. The only method to our knowledge that exploits screen illumination is proposed by Iqbal and Lee [30]. It uses a high-speed camera to find the corneal reflection of a whole CRT monitor from its periodic flicker pattern. The reflection patch locates an eye in a face image, and its centroid is applied as a glint to estimate the PoR on the screen. There are several differences to our proposed technique: The method requires special hardware in the form of a CRT monitor and a high-speed camera. It needs individual calibration for the mapping from imaged pupil-glint vectors to screen coordinates that does not allow for head movement. Moreover, both described gaze tracking methods are purely image-based and do not obtain any explicit geometric information about eye, camera and display.

Recently, Schnieders et al. [53] apply our proposed idea for reconstructing the pose of the display from corneal reflections [48] to eye gaze tracking. Instead of analyzing reflections of a special marker pattern, they directly use the curved edges from the corneal reflection of a screen showing a bright content to recover the corresponding edges in 3D and, thus, the display plane. While this method applies a single face image with two eyes and assumes the user to look at the screen, the goal of our work is to provide a general method without constraining gaze direction, and to present a comprehensive performance analysis for scene reconstruction from eye reflections under multiple eye poses from different images regarding a large number of factors.

The methods introduced in this section analyze corneal reflections from an image of a human eye which also shows the diffuse
3. Eye model and pose

Fig. 1a shows an outer view of the human eye. The most distinctive components are the colored iris and the pupil in its center. The iris is surrounded by the white sclera. The cornea is more difficult to recognize. It is the outer layer of the eye that plays the main role in focusing images on the retina. It covers the iris and dissolves into the sclera at the corneal limbus. The cornea consists of submicroscopic collagen fibrils that are arranged in a special lamellar structure which makes it transparent. Its external surface is very smooth and coated with a thin film of tear fluid that explains its mirror-like reflection characteristics [35]. A comparison of the outer view in Fig. 1a with a cross-section of the eye in Fig. 1b reveals that the main part of the eyeball is located behind skin and components that are visible from the outside.

3.1. Geometric eye model

The human eyeball is not a plain sphere; it can be subdivided into two main segments: the anterior and the posterior. The two are be approximated by two overlapping spheres of different size as shown in Fig. 1c. The smaller anterior segment covers about one-sixth of the eye and contains the components in front of the vitreous humor, including the cornea, iris, pupil and lens. The posterior segment covers the remaining five-sixths.

For the proposed approach, this geometric model is applied (1) to estimate the pose of the eye from an image and (2) to calculate light interaction at surface of the cornea. There are two main difficulties related to such a model: (a) The actual shape of the eyeball is more complex than the model and (b) its anatomical parameters vary individually. However, for this application it is sufficient to apply the aforementioned observation and approximate the eye as two overlapping spheres. Also, the common variation of parameter values among different people is sufficiently small.

3.2. 3D eye pose estimation

The 3D pose of an eye describes the location and orientation of the eye model in the camera coordinate frame where the origin \( \mathbf{0} = (0, 0, 0)^T \) is placed at the camera pupil (Fig. 2). There exist several methods to estimate the pose of the eye from an image. The methods can be categorized by whether (1) they allow arbitrary camera placement or (2) assume a rigid head-mounted setup, and whether (a) they apply passive image processing or (b) use active controlled illumination. We employ a modified version of the method described by Nishino and Nayar [47], combining (1) and (a), to fit our eye model. It can be used for both eyes in the same way. We assume the internal parameters of the camera to be known.

The cornea is modeled as a spherical cap that is cut off from the corneal sphere by the limbus plane. The corneal limbus is the surface shape discontinuity at the intersection between corneal and eyeball sphere where the cornea dissolves into the sclera. For an adult, the radius of the limbus \( r_L \) averages approximately 5.5 mm [47]. The iris is located right behind the limbus and extends below the sclera. It has a radius \( r_I \) of about 6 mm [54]. The visible part of the iris marks the limbus plane.

We assume weak perspective projection since the depth of the tilted limbus is much smaller than the distance between eye and camera. Thus, in the image the almost circular limbus projects to an ellipse that is described by five parameters: the center \( \mathbf{l} = (u, v)^T \), the major and minor radii \( r_{\text{max}} \) and \( r_{\text{min}} \), and the rotation angle \( \phi \). We estimate the values to approximate the limbus by fitting an ellipse to the imaged iris.

Therefore, the image is transformed into a binary edge image by smoothing with a Gaussian filter and extracting edges with an adaptively-thresholded Canny edge detector. We apply a simple interactive strategy: An initial guess for the ellipse parameters is obtained from a user selecting four points on the boundary of the iris. An accurate boundary is then estimated by iteratively minimizing the following error function:

\[
\text{eval} = \sum_{(u, v) \in I} D(u, v) = \begin{cases} D(u, v) & \text{for } D(u, v) \leq D_{\text{max}}, \\ D_{\text{max}} & \text{otherwise,} \end{cases}
\]

where \( E \) is the set of pixels representing the ellipse boundary, and \( D(u, v) \) is a particular pixel value in the distance-transformed binary edge image. For each value, we apply a constant upper bound \( D_{\text{max}} \).
to reduce the effect of non-edge boundary points on the estimation. This is necessary to robustly handle occlusions by the eyelids that especially occur at increasing gaze angles. Note, that the initial guess for the ellipse parameters may also be obtained automatically. There exists a multitude of approaches for eye detection and tracking; and a suitable one has to be selected based on the particular system's constraints. A recent overview can be found in [26].

We are now able to estimate the 3D position of the limbus center \( L \) from the center of the detected ellipse \( l \) and the distance to the camera \( d \) as in

\[
L = \left( d \frac{u_l - u_0}{f}, d \frac{v_l - v_0}{f}, d \right)^T, \quad d = f \frac{r_L}{r_{\text{max}}},
\]

is the focal length in pixels and \( c_0 = (u_0, v_0)^T \) the principal point. Fig. 3 shows estimation results for an increasing distance between eyes and display-camera system.

The orientation of the eye is described by its optical axis which is the line connecting limbus center \( L \), corneal apex \( A \), center of corneal sphere \( C \) and center of eyeball \( E \). The direction of the optical axis \( g \) can be used to approximate the gaze direction.\(^1\) It is obtained as in

\[
g = \begin{bmatrix}
\sin \tau \sin \phi \\
-\sin \tau \cos \phi \\
-\cos \tau
\end{bmatrix}.
\]

where \( \phi \in [0, \pi) \) is already known as the rotation angle of the limbus ellipse. Angle \( \tau \in [0, \pi/2] \) corresponds to the tilt of the limbus plane with respect to the image plane (Fig. 4). It is estimated from the shape of the ellipse up to a sign ambiguity as in

\[
\tau = \pm \arccos \left( \frac{r_{\text{min}}}{r_{\text{max}}} \right).
\]

Further knowledge is necessary to automatically resolve this ambiguity. Wang et al. [63], for example, apply anthropometric properties of the eyeball. In our case, the sign of \( \tau \) does not need to be specified in advance. The ambiguities for a set of eye images can be jointly resolved by the display-camera calibration algorithm (Section 6.3).

Finally, we compute the corneal center \( C \) located at distance \( d_{LC} \) from \( L \) (Fig. 2) as in

\[
C = L - d_{LC} g,
\]

\[
d_{LC} = r_C - d_{AL} = \sqrt{r_C^2 - r_L^2} \approx 5.5 \text{ mm}.
\]

Center \( C \) and radius \( r_C \), averaging 7.8 mm [35], describe the corneal sphere which enables to model the light reflection properties of the corneal surface. Knowing the gaze direction, we might also con-
struct the eyeball sphere with radius $r_E$ around the center $E$. However, this is not necessary in our case.

4. Corneal reflection and light source position

Light that reaches the eye undergoes reflection and refraction at several transparent components, namely the cornea, aqueous humor, lens and vitreous humor, until finally reaching the fovea. The reflection at a particular transition is called the nth Purkinje image [16]. We only deal with the most prominent reflection at the outer surface of the cornea (first Purkinje image) since later reflections along the lightpath cannot be detected without special hardware [42]. We employ the geometric eye model from Section 3 to develop a corneal reflection model that is used to calculate the direction towards an unknown light source. The corresponding light source position is estimated by intersecting the reflection rays obtained under several eye positions.

4.1. Corneal reflection model

We capture an image of a human face showing the reflection of a light source at the cornea of an eye (Fig. 5). We then determine the reflection center $s = (u_s, v_s)^T$ with subpixel accuracy. Knowing the internal parameters, we construct the backprojection ray through point $S$ as in

$$S' = \left(\frac{u_s - s_0}{f}, \frac{v_s - s_0}{f}, 1\right)^T.$$  

(6)

The point of reflection $S$ on the corneal surface is then formulated as in

$$S = t_1 r_1,$$  

(7)

where $r_1 = S'/\|S'\|$ is the normalized direction vector. To recover $S$ we calculate the intersection with the corneal sphere by solving the quadratic equation

$$\|S - C\|^2 = r_C^2.$$  

(8)

Expanding and rearranging leads to

$$t_1^2 r_1^2 - 2t_1 (r_1^T C) + C^2 - r_C^2,$$  

(9)

from which we construct the simplified quadratic formula

$$t_1 = (r_1^T C) \pm \sqrt{(r_1^T C)^2 - C^2 + r_C^2}.$$  

(10)

The first intersection at the front side of the cornea is described by the smaller value of $t_1$.

We calculate the specular reflection at $S$ as

$$r_2 = 2(-r_1 \cdot n_S)n_S + r_1,$$  

(11)

where $r_2$ and $n_S$ are the outgoing and normal vector respectively. The position of light source $P$ lies on the reflection ray extending from $S$ with $P = S + t_2 r_2$ at an unknown distance $t_2$.

4.2. Light source position estimation

We capture a set of images under varying eye poses. We then find $N > 2$ eye poses and their corresponding reflection rays towards $P$. The intersection of the rays gives us the position of $P$. For $N = 2$, an approximate intersection can be obtained by triangulation. This enables estimation from only a single image containing both eyes. For $N > 2$, we have to minimize the distances between $P$ and the set of rays. At frame $l$, the distance $d_l$ between $P$ and the nearest point on the ray $P_l = S_l + t_l r_2$ is defined as in

$$d_l(P, P_l) = \frac{\|r_2 \times (S_l - P)\|}{\|r_2\|}.$$  

(12)

Knowing $\|r_2\| = 1$ and rearranging leads to

$$d_l(P, P_l) = \|r_2\|, P - r_2 \times S_l||,$$  

(13)

where $\|\cdot\|$ denotes the matrix notation of the cross product [27]. We combine the equations and formulate the problem as a least-squares minimization of $\|AP - b\|$. Point $P$ is then obtained by computing the pseudo-inverse as in

$$P = (A^T A)^{-1} A^T b, A_{3N \times 3} = \left[ \begin{array}{c} r_{21} \times S_l \\ \vdots \\ r_{2N} \times S_n \end{array} \right], b_{3N \times 1} = \left[ \begin{array}{c} r_{21} \times S_1 \\ \vdots \\ r_{2N} \times S_n \end{array} \right].$$  

(14)

5. Geometric display calibration

The controlled illumination system consists of three components: (1) a raster display device that acts as a light source, (2) the cornea of a human eye that reflects the light from the screen and (3) a camera that captures the light reflected from the cornea. Fig. 6 shows an overview of the geometric calibration algorithm.

5.1. Display-camera transformation

A display is modeled as a screen plane containing the pixels $p = (i, j)^T$ that we want to describe as points $P = (x, y, z)^T$ in camera coordinates (Fig. 5). The transformation is expressed in homogeneous coordinates as in

$$\begin{bmatrix} P_1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} P \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_0 - p_0^0 \\ p - p_0 \\ p - p_0^0 \\ p_0 \end{bmatrix},$$  

(15)

Unfortunately, matrix $T$ depends on the three unknown correspondence pairs at $p_0^0 = (0, 0)^T$, $p = (1, 0)^T$, and $p^0 = (0, 1)^T$. However, knowing $M \geq 3$ arbitrary correspondence pairs where the points $(P_k, k = 1, \ldots, M)$ are coplanar on the screen plane, we can formulate an equation system $A_t = b$ as in

$$\begin{bmatrix} i_{11} & j_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ i_{12} & j_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ i_{13} & j_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ i_{14} & j_{14} & 0 & 0 & 0 & 0 & 0 & 0 \\ i_{21} & j_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ i_{22} & j_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ i_{23} & j_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} [i_1] \\ [j_1] \\ [i_2] \\ [j_2] \\ [i_3] \\ [j_3] \\ [i_4] \\ [j_4] \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ x_M \\ y_M \\ z_M \end{bmatrix}.$$  

(16)
We solve for the vector $\mathbf{e} = (A^TA)^{-1}A^T\mathbf{b}$ containing the nine unknown matrix elements by again computing the pseudo-inverse.

### 5.2. Marker pattern architecture

To obtain the $M$ correspondence pairs, we display a static pattern composed of $M$ light sources as circular markers centered at $\mathbf{p}_k$ on a black background. We acquire face images of a moving person showing the corresponding corneal reflections in both eyes (Fig. 9). After detecting a particular iris, we extract the reflections by segmenting four regions containing the highest intensity values inside the iris with an adaptive threshold for background subtraction. Knowing the pixels corresponding to a particular region $R_k$, where $I(u,v)$ is the intensity value at pixel $(u,v)^T$ we calculate its intensity centroid $s_k$ with subpixel accuracy as in

$$s_k = \frac{\sum_{(u,v) \in R_k} I(u,v) \cdot (u,v)^T}{\sum_{(u,v) \in R_k} I(u,v)}.$$  

We then estimate the 3D marker positions $\mathbf{P}_k$ as explained in Section 4. Display and camera are facing the user. The spatial relation between the markers (indicated by $k$) is not affected by projection and reflection as the cornea has a convex shape. Increasing $M$ is beneficial since it statistically increases the accuracy of the estimated screen plane. The maximum number $M$ is limited by the number of reflections that can be recovered from an image, which depends on the particular system configuration and may be adjusted in a pre-processing step.

The described strategy of representing each point $\mathbf{p}_k$ as a distinct marker has the advantage that reflections are easily extracted from eye images. The disadvantage, however, is that the low radiance restricts its use to setups with low environmental light. In order to increase robustness to ambient light, a bright, uniformly colored full screen pattern may be used [18]. After detecting the reflected display patch (e.g., by using a tailored color-thresholding approach [62]), the display plane can be reconstructed either using its corners or edges [53]. The presence of corneal reflections from other light sources can make it difficult to distinguish these from the marker reflections. In that case, reflected features may be matched over all images and then reconstructed. Additional geometric constraints can then be used to remove the outliers. For example, a simple distance thresholding would be effective for the common case where the display is the nearest light source to the eyes.

### 6. Optimization

Using the proposed algorithm to estimate light source positions leads to a larger error than using a spherical mirror of the same size, with the estimated positions usually being about halfway between the eyes and their true position accompanied by a large ray intersection error. This has several possible reasons. The two main sources of error are the following:

1. The individual shape and parameters of the eye are unknown. Distortion effects increase with gaze angle when reflections move away from the corneal apex towards the boundary. Moreover, the unknown radii of iris $r_i$ and corneal limbus $r_l$ influence the eye pose estimation.
2. It is difficult to exactly detect the imaged iris which gradually dissolves into the sclera. Due to iris landmarks and blood vessels, this transition is not smooth [31]. Noisy measurements directly affect the orientation of the eye pose estimation which itself is crucial for overall accuracy.

#### 6.1. Error function from known geometry constraints

The error for the reconstructed display can be significantly decreased with a small modification of the estimated eye poses and imaged reflections. These serve well as initial guesses and can be further adjusted by an optimization that minimizes a convex error function $e$ defined as the weighted sum of three error terms as in

$$e = \frac{w_1 e_1 + w_2 e_2 + w_3 e_3}{w_1 + w_2 + w_3}.$$  

The intersection error $e_1$ is defined as the average distance of the reflected light rays to their estimated intersection points as in

$$e_1 = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \| \mathbf{P}_i - \mathbf{P}_j \|.$$  

where $\mathbf{P}_i$ is the point on ray $j$ having minimal distance to the corresponding intersection point $\mathbf{P}_i$. The size error $e_2$ is defined as the average absolute error of the distances between all $M$ estimated marker positions as in

$$e_2 = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{i<j}^M \| \mathbf{P}_i - \mathbf{P}_j \| - GT_{i,j,k} \|,$$

where $\mathbf{P}_i$ and $\mathbf{P}_j$ are two estimated marker positions and $GT_{i,j,k}$ the ground-truth distance from the known display size. Finally, the plane error $e_3$ is defined as the average absolute distance of the $M$ estimated marker positions to their approximated best fit plane containing the centroid as in

$$e_3 = \frac{1}{M} \sum_{i=1}^M \| \mathbf{p}_i \cdot \mathbf{n} + \mathbf{p} \|.  \quad (21)$$
This plane is given by $P_i \cdot n + p = 0$ in Hessian normal form and obtained from orthogonal regression. In order to calculate it, we stack the estimated light source positions into a matrix $A_{M \times 4}$ as in

$$A = \begin{bmatrix} P_1 & 1 \\ \vdots & \vdots \\ P_M & 1 \end{bmatrix}$$

and compute its singular value decomposition $A = UDV^T$. The plane unit vector $n$ and distance from origin $p$ are then obtained as in

$$\begin{bmatrix} n \\ p \end{bmatrix} = \frac{1}{\sqrt{v_1^2 + v_2^2 + v_3^2}} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix},$$

where $v$ is the singular vector corresponding to the smallest singular value.

### 6.2. Optimization strategy

The proposed optimization strategy comprises three subsequent steps:

1. First, we jointly optimize the estimated positions for all corneal sphere centers $\{C_j\}_{j=1, \ldots , N}$ with weights for the error terms set as $w_1 = 0$, $w_2 = 1$, and $w_3 = 1$.
2. Next, we add the imaged reflection centroids $\{s_{ij}\}_{i=1, \ldots , M, \ j=1, \ldots , N}$ and jointly optimize them together with the modified corneal sphere centers from the last step. The weights remain unchanged.
3. Finally, we set weight $w_1 = 2$ and repeat the last step.

Each step is performed using Powell’s direction set method with the number of brackets set to nine. We advance to the subsequent step after changes fall below some threshold $\epsilon$.

Evaluating the proposed technique, we found the adjustment by the optimization step to be small for projections of the corneal spheres containing the imaged irides. Pupillary distances between left and right eye corresponding to the same face image were matching the measured ground truth. This leads to the conclusion that a spherical corneal curvature model of constant size is a valid assumption for the proposed approach. We further obtain convergence to the same result when adding small perturbations to the initial estimation, thus making it possible to apply the technique with low performance hardware. The technique can be useful for robust eye gaze tracking and also performs well at correcting calibration results obtained from a spherical mirror.

### 6.3. Resolving the sign ambiguity in eye orientation

Regarding 3D eye pose estimation in Section 3.2, there remains a sign ambiguity for the limbus tilt angle $\tau$ that could not be determined automatically from the ellipse parameters. However, after obtaining an initial guess for all eye poses and reflection centroids, it is possible to jointly estimate the missing signs $\{S_{ij}\}_{i\in[-1,1]}$ by applying the geometry constraints introduced within this section. This is done by computing the value of size error $e$ under all $2^N$ sign combinations and selecting the one with minimal error.
Table 2
Personal statistics for the test subjects. The data sets are ordered by age. There is no significant correlation between parameters. Myopia (near-sightedness) occurs in six subjects. Five of them wear glasses that were taken off for the experiments. A single subject wears contact lenses that were kept on.

<table>
<thead>
<tr>
<th>Gender</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>21</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>29</td>
<td>35</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Height</td>
<td>166</td>
<td>168</td>
<td>183</td>
<td>161</td>
<td>167</td>
<td>171</td>
<td>174</td>
<td>176</td>
<td>166</td>
<td>179</td>
<td>154</td>
</tr>
<tr>
<td>Pupillary distance</td>
<td>6.3</td>
<td>6.5</td>
<td>6.4</td>
<td>6.5</td>
<td>5.8</td>
<td>6.6</td>
<td>6.6</td>
<td>5.9</td>
<td>6.7</td>
<td>7.1</td>
<td>6.3</td>
</tr>
<tr>
<td>Myopia</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>Contact lens</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
</tbody>
</table>

Fig. 9. Example images of moving subjects, acquired in the absence of environmental light. Intensity scaling is applied for better visibility.

Fig. 10. (a)–(d) Cropped facial images with iris ellipse fitting results. Intensity scaling is applied for better visibility. (e)–(h) A pair of corresponding shape and reflection images with fitting results for each spherical mirror.

7. Experiments

In the following section, we explain several comprehensive experimental series that were conducted in order to thoroughly analyze the performance of the proposed method as well as to obtain further knowledge for reconstructing information from eye reflections in general.

7.1. Experimental setup

The setup contains a 19-in. display with 1280 × 1024 (5:4) resolution, 250 cd/m² brightness, 500:1 contrast ratio and 170°/170° (H/V) viewing angles. We use a Point Grey Flea2 camera at 2448 × 2048 resolution mounted on a Fujinon HF35SA-1 lens with viewing angles 14°35'10''58''. Intrinsic camera parameters are calibrated using OpenCV functions [10]. The camera is placed at about 30 cm distance above and behind the display. Test subjects are seated with their faces positioned about 50–60 cm in front of the display (Fig. 7a). We show four markers {P_k | k = 1, ..., 4} as white circles of radius r = 0.25 in.\(^1\) (6.35 mm) at the corners of the

\(^1\) Inches are used for better comparison with display sizes.
This creates a pattern with a diagonal size $d_v$ of 18 in. (457.2 mm). We capture face images of a test subject moving in front of the display in the absence of environmental light. Regarding eye orientation, we apply random measurements with a tilt angle $\tau < 20^\circ$. As shown in Experiment 4, calibration accuracy does not vary significantly within this range; however, it decreases rapidly for larger angles.

## 7.2. Experiment 1: varying test subjects

Experimental verification was performed with 11 test subjects. Table 2 shows that there is no significant correlation between different parameters such as age, body height and pupillary distance among test subjects. We acquired data sets of 10 face images per subject that are used for reconstruction (Fig. 9). The imaged irides
have an average diameter of 160 pixels with the reflected screen occupying about $30 \times 25$ pixels [Fig. 10a–d].

Experimental results for the accuracy of reconstructed display poses can be found in Fig. 11 and Table 3. After optimization, the standard deviation between the center points of the estimated screen planes decreased considerably from 81.60 to 10.28 mm. It is important to note that no statistical significance could be obtained between test subjects with normal eyes, near-sightedness uncorrected, and corrected with contact lenses (Table 4). This means that the method can be applied to any of these conditions despite them having a particular impact on eye appearance, shape, and size that cannot be compensated. Fig. 7b–d shows results where the reconstructed screen matches the real one given in Fig. 7a. The average and standard deviation of corneal sphere position adjustment in optimization are $<0.01/0.67$, $0.02/0.39$, and $8.94/11.92$ mm for x, y, and z-coordinates respectively. This shows that the error in corneal position estimation is largest along the depth direction.

7.3. Experiment 2: varying display pose

We mounted the display on a turntable, operated using a Chuo Seiki QT-CM2 stage controller, and took data sets of 10 face images of a single person at discrete display orientations of $0^\circ$, $10^\circ$, $20^\circ$, $30^\circ$, and $40^\circ$ which act as ground truth (Fig. 14a). We further took data sets for a large and a small spherical mirror with 20 images per mirror. The large mirror with a radius of 25.4 mm is similar in size to the one used by Francken et al. [18] and acts as ground

| Table 4 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Top-right       | Top-left        | Bottom-left     | Bottom-right    | Center          |
| P-value        | 0.144           | 0.905           | 0.356           | 0.273           | 0.953           |
| Bonferroni     | 1.727           | 10.863          | 4.268           | 3.275           | 1.016           |
| Holm           | 1.007           | 0.905           | 1.778           | 1.638           | 1.016           |

| Table 5 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|

<table>
<thead>
<tr>
<th></th>
<th>(a) Error to GT</th>
<th>(b) Error to mirror L</th>
<th>(c) Residual errors after optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orientation (deg)</td>
<td>Position (mm)</td>
<td>Orientation (deg)</td>
<td>Intersection (mm)</td>
</tr>
<tr>
<td>Pre-Opt Eye</td>
<td>17.33</td>
<td>10.02</td>
<td>319.44</td>
</tr>
<tr>
<td>Mirror S</td>
<td>1.90</td>
<td>1.21</td>
<td>21.70</td>
</tr>
<tr>
<td>Mirror L</td>
<td>1.34</td>
<td>1.28</td>
<td>0.00</td>
</tr>
<tr>
<td>Opt Eye</td>
<td>0.82</td>
<td>0.37</td>
<td>11.18</td>
</tr>
<tr>
<td>Eye L</td>
<td>0.49</td>
<td>0.30</td>
<td>9.95</td>
</tr>
<tr>
<td>Mirror L</td>
<td>0.31</td>
<td>0.10</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Fig. 12. Experimental results for the display orientation experiment (2). Comparison of orientation angles estimated from eyes and spherical mirrors (a) before and (b) after optimization, with the turntable ground truth.
Fig. 13. Experimental results for the display orientation experiment (2). (a) The experimental setup with varying display pose. Each column corresponds to a particular display orientation (from left to right: 0°, 10°, 20°, 30°, 40°). (b) The displays are reconstructed from eye reflections after optimization, rendered by looking down onto the \(xz\)-plane (red, blue) of the camera frame. Each mark along the axes indicates a multiple of 20 cm. (c) Comparison between the three displays reconstructed from eyes (white), large (green), and small spherical mirrors (red) after optimization. All three planes are estimated with a low position and orientation error. (d) Same as (c), but before optimization. The screen planes from mirror reflections are estimated slightly smaller and rotated, in front of the actual position, with a screen size error \(e_2\) of about 25 mm. In contrast, the screen plane from eye reflections is estimated incorrectly with an average of 319.4 mm in front of the one obtained from the ground truth of the large mirror, with an error \(e_2\) of about 200 mm. (e)–(f) Same as (c)–(d) rendered similar to (a) by looking from the side onto the \(yz\)-plane (green, blue).

Fig. 14. Experimental results for the display orientation experiment (2) at 0°. The (a) position and (b) orientation error describe the deviation from the ground-truth estimate obtained using the large mirror. An increasing number of face images used for calibration leads to a decreasing error and convergence towards the final result.

Fig. 15. Image data for the display-eye distance experiment (3). The sequence shows the right eye of the test subject at increasing eye distances from the display-camera setup. Intensity scaling is applied for better visibility. Larger distances result in difficult conditions and image data, due to low illumination and small display size.
truth. The small mirror with a radius of 7.9 mm is similar in size to the corneal sphere. This makes it possible to independently analyze the errors from small reflector size and unknown shape difference (asphericity). The eye pose estimation algorithm was adapted for the mirror. The imaged large and small mirrors have an average diameter of about 650 and 210 pixels with the reflected screen occupying about $95 \times 75$ and $30 \times 25$ pixels respectively (Fig. 10e–h).

Accuracy is estimated using two measures: The position error $e_P$ describes the deviation in the center position of the display. The orientation error $e_O$ describes the deviation in the normal direction of the display. Let $X_{GT}$ and $n_{GT}$ denote the ground truth for the display center point and the normal direction, and let $X_{eye}$ and $n_{eye}$ denote the estimates obtained using eyes, then the error is defined as

$$e_P = \|X_{eye} - X_{GT}\|,$$

$$e_O = \arccos(n_{eye} \cdot n_{GT}).$$

Experimental results for the accuracy of reconstructed display poses are found in Fig. 12 and Table 5. After optimization, the result from the eyes correctly matches the actual display pose and is only slightly worse than the result for the small mirror. It outperforms the results obtained from both spherical mirrors before optimization which coincide with the results for the method in [18], shown in [19]. Fig. 13 offers a detailed visual comparison. The strategy further achieves a decreasing error and convergence with increasing number of face images used for calibration (Fig. 14).

7.4. Experiment 3: varying display-eye distance

In this experiment we evaluated the performance with increasing eye distance from the display-camera setup. Calibration was performed at six discrete display-eye distances of 35, 65, 95, 125, 155, and 185 cm. The interval and step size are chosen with some considerations: A minimum distance of 35 cm is necessary to capture both eyes within a single face image and to allow some head movement. A distance of 65 cm is approximately the one used for Experiments 1 and 2, and, thus, enables for comparison. A maximum distance of 185 cm is rather uncommon for a 19-in. display, but is tested in order to evaluate the theoretical limitation. It also marks the upper bound for obtaining usable data for this setup since it results in very difficult conditions and image data as can be seen in Fig. 15. Due to the inverse-square law for intensity, the marker radius needed to be increased to $r = 0.5$ in. at 155 and 185 cm. At larger distances, it was not possible to extract independent centroids for each marker since the distance between imaged reflections becomes too small. Also, it was not possible to perform ellipse fitting to detect the iris boundary due to the large amount of noise.

### Table 6

<table>
<thead>
<tr>
<th>Display-eye Distance (mm)</th>
<th>(a) Error to GT</th>
<th>(b) Residual errors after optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Position (mm)</td>
<td>Orientation (deg)</td>
</tr>
<tr>
<td>Pre-Opt</td>
<td>Avg</td>
<td>Stddev</td>
</tr>
<tr>
<td>350</td>
<td>204.74</td>
<td>7.39</td>
</tr>
<tr>
<td>650</td>
<td>401.80</td>
<td>20.71</td>
</tr>
<tr>
<td>950</td>
<td>576.97</td>
<td>60.17</td>
</tr>
<tr>
<td>1250</td>
<td>759.58</td>
<td>76.25</td>
</tr>
<tr>
<td>1550</td>
<td>1156.03</td>
<td>77.58</td>
</tr>
<tr>
<td>1850</td>
<td>1361.72</td>
<td>94.43</td>
</tr>
<tr>
<td>Opt</td>
<td>350</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>679.91</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>950</td>
<td>17.22</td>
</tr>
<tr>
<td></td>
<td>1250</td>
<td>211.47</td>
</tr>
<tr>
<td></td>
<td>1550</td>
<td>227.59</td>
</tr>
<tr>
<td></td>
<td>1850</td>
<td>491.34</td>
</tr>
</tbody>
</table>

Fig. 16. Experimental results for the display-eye distance experiment (3). A plot of the (a) position and (b) orientation error in Table 6. Columns represent the average error, black bars the standard deviation. The error increases with distance. Usable results are achieved for display-eye distances up to 100 cm (camera-eye distance of 110 cm), from where the error increases rapidly.
A particular calibration result was obtained from a data set of 10 face images capturing both eyes of a single person. For statistical evaluation, we took 10 independent trials at each distance (600 images in total). Calibration was also performed for the large spherical mirror using data sets of 20 images. The optimized mirror result at 35 cm has been found to be most accurate and is used as ground truth. The camera is fixed above and approximately 10 cm behind the display. Internal camera parameters are calibrated separately at each distance since the focus of the lens had to be adjusted.

Experimental results obtained before and after optimization are shown in Fig. 16 and Table 6. Optimization can effectively reduce the overall error along the whole range that increases with distance. Results show relatively good accuracy for close ranges. At 95 cm, the average position error amounts to 17.72 mm, and the average orientation error to 7.04°. At larger distances, we observe a sharp increase, especially for the position error. Depending on the intended use, the accuracy should be sufficient for distances up to 1 m. However, regarding the size of the display, real usage scenarios will probably not involve distances larger than 65 cm.

7.5. Experiment 4: varying gaze angle

The following experiment studies the influence of eye orientation. For four test subjects, calibration was performed at six discrete limbus tilt angles $\tau$ of 0°, 6°, 12°, 18°, 24°, and 30°. Each corresponding data set comprises eight face images including both eyes, distributed evenly along the whole range of eye rotation angles $\phi$ with steps at 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°. Fig. 17 explains the experimental setup. Refer to Fig. 4 for details about the eye orientation representation. Gaze markers are attached to the wall according to the distribution of gaze directions for test subjects located at a distance of 170 cm. The display is mounted on a tripod. The camera is fixed with an adjustable arm at 10 cm in front of the display center and 100 cm in front of the center marker for $(\phi, \tau) = (0°, 0°)$ with its optical axis perpendicular to display plane and wall. The ground-truth display pose is obtained from 15 images of the large spherical mirror, distributed evenly along the camera field of view at a distance of approximately 80 cm in front of the display. Under the same conditions, data sets are acquired for moving test subjects gazing at the respective markers. Gaze markers for angles $\tau$ of 0°, 6°, and 12°, occluded from the perspective of the test subjects, are rendered with dark blue color in order to not affect the measurements.

The setup is designed with some considerations: The camera is placed at a rather uncommon location in front of the display center. With that configuration, there is no bias regarding angle $\phi$, since each gaze marker corresponding to the same angle $\tau$ and each of the four display calibration markers have equal distance to the camera. The display-camera setup is designed to minimize occlusion of the gaze markers. The maximum tilt angle $\tau$ of 30° marks the upper boundary for the marker reflections to be located on the corneal surface. The subject-wall distance is maximized in order to reduce the impact of slight head movement on the gaze direction; where 170 cm is the largest distance that allows attaching the center marker at a comfortable sitting height.

Fig. 18 shows captured eye images for the whole set of gaze directions. Iris detection produces applicable results. For an increasing tilt angle $\tau$ the angular error in gaze direction amounts to 3.62°, 3.96°, 3.79°, 3.36°, 3.01°, and 4.94° (RMSE). The reflected marker pattern moves towards the corneal boundary. For one subject, several markers at 30° were already out of the corneal boundary and have not been used for calibration. Distortion increases for the whole pattern as well as for each reflected marker patch. Both effects vary according to individual differences in corneal shape.

Experimental results obtained before and after optimization are shown in Fig. 19 and Table 7. Optimization can effectively reduce the overall error along the whole range of tilt angles $\tau$. Estimation error is lowest for 6–18, slightly larger for 24, and large for 30°, which is a result of pattern distortions and deviations in corneal curvature and shape. Interestingly, the error at 0° goes up slightly.
Consider the common case where the user is located at a distance of 60 cm in front of the center of a 19-in. display. The camera is placed above or below the display. For iris detection, the maximum gaze angle when looking at the display amounts to approximately $17/27^\circ$ (H/V). For display calibration, the maximum angular deviation for the marker reflections from the corneal apex amounts to approximately $17/14^\circ$. The values are within the applicable limits of eye and display pose estimation.

Fig. 18. Image data for the gaze-angle experiment (4). (a) Imaged irides of the left eye of a single test subject for the set of eye orientation angles $\phi$ and limbus tilt angles $\tau$. Intensity scaling is applied for better visibility. The head was fixed, with the midpoint of the eye baseline centered in the camera image. To not only show iris deformation but also translation of its imaged position, each image patch is centered at the invisible backprojected center $C$ of the estimated corneal sphere. With increasing angle $\tau$, the reflected marker pattern moves towards the corneal boundary. Distortion increases for the whole pattern as well as for each reflected marker patch. This leads to a rapidly increasing display pose estimation error for reflections near the corneal boundary. Above $30^\circ$, some of the markers move out of corneal boundary and cannot be detected. (b) Imaged irides of the left eye of another test subject at $30^\circ$. Comparison shows larger distortion and missing reflections which relates to individual differences in corneal shape.

Fig. 19. Experimental results for the gaze-angle experiment (4). A plot of the (a) position and (b) orientation error shown in Table 7. The error is lowest for 6–18 and slightly larger for 0° and 24°. It rapidly increases with larger angles. For the common case, where a user is looking onto a 19-in. display from a distance of 60 cm, the maximal gaze angle and reflection deviation from the corneal apex are within the applicable limits of eye and display pose estimation.
Important conclusions are the following:

- The absolute error to the ground truth can be considerably decreased. Convergence is achieved with an increasing number of images.
- The error increases with distance from the display-camera setup. Nevertheless, the applicable range for this method has been found to exceed common use cases.
- The error remains stable with increasing display orientation and gaze angle for the applicable range of this method that also covers common use cases. Larger angles lead to distortions for reflections near the corneal boundary and should therefore be avoided.

### 8.2. Implications

With the proposed method, we established and verified the integration of eye reflection analysis with display-camera systems. Despite the difficult working conditions, the results are good and should be sufficient for many applications. We believe that this work has the potential to facilitate novel developments in the community and helps to generally increase usability and acceptance of applications "outside the laboratory". The unique characteristics of the method enable applications in novel situations and system configurations. An overview of potential implications is given in the following.

**Calibration-free applications.** Since calibration is achieved implicitly without requiring interaction and awareness, the method can be applied where a dedicated calibration procedure is not possible. This could be for any of several possible reasons: a lack of time, when attention is required for other tasks such as re-arranging the display in driver-assistance systems or at the workplace. Second, a lack of ability, when working with non-experts, physically/mentally disabled people, or children and infants [23]. Other reasons could include hiding technical details of the system or seamlessly integrating with art decors. Dynamic setups. The calibration does not require interaction and may be performed online. This allows applications where the relation between display and camera is changing. Examples for changing camera pose include hand-held video cameras, and pan tilt zoom cameras (PTZ) in surveillance and vision-based interfaces. Examples for changing display pose include hand-held/mobile devices, and projection displays, such as head-up displays in cars (HUD) or special displays in augmented reality [5].

The proposed method recovers the geometric relation between display, camera, and eyes. This can be beneficial for applications in...
Human–computer interaction. The method enables improved calibration-free remote eye gaze tracking where the PoR is obtained by relating gaze direction and display plane. Screen-based eye gaze tracking has many applications in different fields [16,25,26]. However, there is no restriction to planar screens: A 3D PoR on an arbitrary surface is obtained when the estimated eye pose is related with a model of the environment or an image-based environment map. This can be further combined with eye reflection analysis for scene reconstruction or eye pose refinement.

Surveillance and security. It has been shown that display content can be recovered from reflections in the eyes of a person in front of a PC, from far-away locations [3,2]. The quality of the result may be improved by undistortion which requires knowledge about display pose, eye pose and shape. Furthermore, it is possible to extend this to real-time monitoring of the interaction with mobile devices. Knowledge of eye gaze and display reflections may also be beneficial for technical improvement and to introduce context information in iris-based biometric systems [14,9].

Photometric stereo. There exist several works using display-camera systems for scene reconstruction by photometric stereo [64]. Knowledge about the distribution of environmental light sources is important and can be recovered from an image of the eye [47,59,33]. The proposed method lays the foundation to exploit this information in the context of display-camera setups, leading to performance improvement in calibration and application.

Medicine. Analyzing the relation between display content and eye poses can help to diagnose patterns related to physical and mental degrading of the visual and motor system. After having detected a particular condition, corrective actions can be provided through modification of the displayed content. Moreover, such information could also be used in order to detect and support correct 3D perception with auto-stereoscopic displays [28,40].

8.3. Limitations

The scope of this paper is to provide an in-depth analysis of the applicability of eye reflections for display-camera calibration. There are limitations when using the implementation in its current form within real conditions. Details for a fully automatic calibration procedure largely depend on the requirements of each particular setup. Necessary extensions include

- a strategy for calibrating camera parameters, e.g., directly from eye images [33],
- a technique for tracking a first guess for the eye region in a video [26],
- a scheme for discarding unusable frames that do not include an eye, have too low quality, or relate to configurations known to result in decreased accuracy,
- a pattern architecture that increases information throughput and allows robust reflection extraction in the presence of environmental light and varying iris colors [62], e.g., using coded markers, and
- an extension to suppress complex light interaction at different layers of the eye and to handle more complicated light paths [39], e.g., occurring when users wear glasses.

The described approach as well as related works apply static patterns. Systems are therefore limited to static scenarios where calibration and application are separate steps. The proposed method, however, has the potential for realizing online calibration since it does not require interaction. In order to achieve this in practice, it is necessary to detect display reflections relating to the same spatial location under a varying application content [56].

The accuracy of the system highly depends on the performance of image-based eye pose estimation. More complex geometric eye models have to be tested in order to better approximate the shape of the eye as the eyeball is slightly flattened in the vertical plane [54] and corneal topology is complex [6]. It may be beneficial to model the eye geometry as two intersecting ellipsoids and include the radii as shape parameters in the optimization process. The aspect ratio of the eyeball can be calibrated from a single iris image where the user looks directly into the camera. Strategies for calibration of each individual’s eye geometry may lead to further improvements.

Acknowledgments

This work was supported in part by the JST Precursory Research for Embryonic Science and Technology (PRESTO) program and the MIC Strategic Information and Communications R&D Promotion Programme (SCOPE), Japan. The opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and should not be interpreted as representing the views or official policies, either expressed or implied, of the sponsors.

References

Haruo Takemura received his B.S., M.S., and Ph.D. degree in Engineering Science from Osaka University, Japan in 1982, 1984, and 1987 respectively. In 1987, he joined Advanced Telecommunication Research Institute, International. In 1994, he joined Nara Institute of Science and Technology as associate professor at Graduate School of Information Science and Technology. From 1998 to 1999, he was a visiting associate professor at University of Toronto, Canada. In 2001, he became a full professor at Cybermedia Center, Osaka University. Since 2005, he also serves as vice-director and since 2007 as director of Cybermedia Center. His research interests include interactive computer graphics, human–computer interaction and mixed reality. He is a member of IEICE, IPSJ, VRSJ, HIS, IEEE, and ACM.