2008－CVIM－164 (19)
2008／9／5

眼球の表面反射を用いたディスプレイカメラキャリブレーション

ニチュケ クリスティアン† 中澤 篤志† 竹村 治雄†

† 大阪大学 情報科学研究科 (サイバーメディアセンター)
〒560–0043 大阪府豊中市待兼山町 1–32
E-mail: †christian.nitschke@ist.osaka-u.ac.jp, ††{nakazawa,takemura}@cmc.osaka-u.ac.jp

あらまし 本論文では眼球の表面反射を利用したディスプレイ・カメラシステムの幾何的キャリブレーション法を提案する．ディスプレイ・カメラシステムは平面ディスプレイとカメラから構成されるシステムであり，ディスプレイ上のパターンにより照明条件を変化させることで，物体や顔の３次元形状計測を行うことができる．しかし通常の構成では，ディスプレイ面（法線）とカメラの視線が同方向を向いているため，それらの幾何的関係を求めるためには，鏡などの補助的デバイスを用いる必要があった．これに対し本手法では，人の顔を計測対象にした場合，ディスプレイ上のパターンが眼球（瞳）に反射することを利用し相互の幾何的関係を求める方法を提案する．これにより付加的なデバイスを用いることなく，連続的に顔を撮影するだけで人の顔の３次元形状復元が可能になった．

キーワード ディスプレイ・カメラシステム，３次元形状計測，眼球

Display-Camera Calibration from Eye Reflections

Christian NITSCHKE†, Atsushi NAKAZAWA†, and Haruo TAKEMURA†

† Graduate School of Information, Science and Technology (Cybermedia Center), Osaka University
1–32 Machikaneyama, Toyonaka, Osaka, 560–0043 Japan
E-mail: †christian.nitschke@ist.osaka-u.ac.jp, ††{nakazawa,takemura}@cmc.osaka-u.ac.jp

Abstract We present a technique for calibrating display-camera systems from corneal reflections in the user's eyes. Display-camera systems enable a range of vision applications that need controlled illumination including 3D object reconstruction, facial modeling or human computer interaction in everyday environments. An important issue is calibrating the pose of the display with respect to the camera. Such a calibration may be achieved using a planar mirror with attached calibration pattern or a spherical mirror of known size. However, all approaches require additional hardware and user interaction. We propose an automatic way to recover display properties from patterns that are reflected in the cornea, a mirroring device that naturally coexists in any display-camera system. By applying this strategy we also obtain a continuous estimation of eye pose which may be used to calibrate eye tracking systems and generally enhance human-computer interaction.

Key words Display Camera System, Calibration, Eye, Eye Tracking, Cornea Reflection, Model

1. Introduction

These days personal computers are turning more and more into multimedia processing machines. Several peripheral devices extend the I/O capabilities and can be considered an integral part of the setup. One of these devices is a camera, mostly used for video conferencing. However, it enables a wide range of vision applications in everyday environments, such as vision-based user interfaces tracking the human body [1] or interaction devices [2].

Every computer system includes a TFT or CRT monitor. Together with a camera, a display forms a controlled illumination system, enabling many further applications. Funk and Yang [3] apply a photometric stereo technique [4] using display illumination to reconstruct lambertian surfaces of static objects. Schindler [5] proposes a similar method which works in real-time on a notebook computer. Clark [6] gives a theoretical proof that uniformly colored planar screen patches can be used to imitate point light sources, a common assumption for photometric stereo techniques. An interesting property of displays is that they radiate the light of a single pixel into a hemisphere of directions rather than into a single direction like projectors do. This makes them especially useful for reconstructing specular and transparent objects [7]. In that spirit, Francken et al. [8] achieve impressive results extracting surface normals from specular objects. Zongker et al. [9] introduce environment matting to compute the light path through transparent objects.

Many of the described systems for controlled illumination need to know the pose of the display relative to the camera. The geometric calibration can be achieved by displaying a checkerboard calibration pattern and analyzing the correspondences in the camera image. Things get more complicated for the common case where display and camera are not facing each other. Here, a mirroring device of known shape and pose is used to reflect the calibration pattern back towards the camera. Funk and Yang [3] use a planar mirror and compute its pose from an additional pattern attached to the mirror. The process is cumbersome as it involves a specially prepared mirror, many parameters and physical user interaction. Francken et al. [10] propose a calibration technique using a spherical mirror of known size. The position of the mirror can be uniquely determined from the camera image. Further extracting the reflected corners from the display allows to compute the corresponding light rays towards the screen. The screen position is found by intersecting rays from a set of images while moving the mirror. However, these technique still requires special hardware and user interaction.

Moving towards application in non-professional everyday environments, we propose a calibration technique that does not need any additional hardware and user interaction. Our approach analyzes corneal reflections of display patterns to estimate size and pose of the monitor from a set of images of a moving user. To reconstruct pose and reflection characteristics of the eye, we employ a simple geometric model. We implemented the model and perform some basic experiments using only off the shelf components to verify the general feasibility of our approach.

Beside the described applications for display-camera systems, our technique further enables applications relying on the relation between the user's eyes and the display. The proposed technique can be applied to calibrate eye tracking systems [11] in a non-intrusive way. This enables applications relying on statistical analysis or machine learning. The display-eye-camera system generally facilitates new user interfaces for interactive applications [12]. Considering interaction with 3D content, the display can be seen as a "window" to the scene which is rendered relative to the user's pose and gaze. Further 3D cues such as depth-of-field blur may be used to increase 3D depth perception on normal displays without additional hardware [13].

Corneal reflections have been analyzed in several other contexts to obtain information about the environment. Nishino and Nayar [14] introduce the eye-camera system as a catadioptric imaging system consisting of the corneal reflector and the camera. They explain how to extract an environment matte from an image of the eye and perform face relighting. Johnson and Farid [15] also analyze the direction of a light source in an image of the human face. Their objective is the detection of digital forgeries. In the context of confidential data security, Backes et al. [16] present an interesting eavesdropping approach. They analyze the reflections of content from an LCD monitor in the user's eyes from far-away locations by using a camera mounted with a telescope.

The remainder of this paper is organized as follows: Sec. 2. introduces a geometric eye model and explains how to recover its pose from an image of the eye. Sec. 4. specifies the corneal reflection model and shows how to compute the direction towards a light source from a corresponding corneal reflection in the image. Sec. 3. introduces a display model and describes its geometric calibration from multiple images. Sec. 5. discusses experimental results on model applicability and accuracy. Finally Sec. 6. shows interesting future directions and problems to be solved.

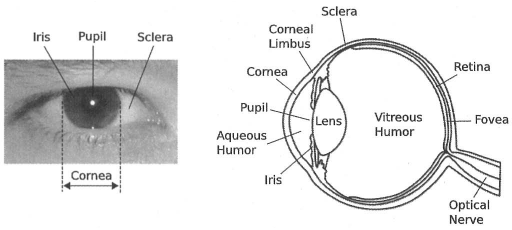


図 1 (a) An outside view and (b) a cross section of the human eye.

2. Eye-Camera System

2.1 Geometric Eye Model

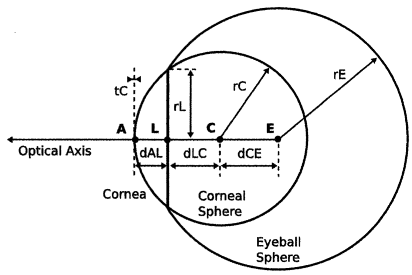
Fig. 1(a) shows an outside view of the a human eye. The most distinctive components are the pigment-textured iris with the pupil in its center. The iris is surrounded by the white sclera. The cornea is more difficult to recognize. It covers the iris and dissolves into the sclera at the corneal limbus. It consists of submicroscopic collagen fibrils that are arranged in a special lamellar structure which makes it transparent. The external surface of the cornea is very smooth and coated with a thin film of tear fluid which explains its mirror-like reflection characteristics. The cross section of the eye in Fig. 1(b) reveals that its majority is located behind the visible components.

The human eyeball is not a plain sphere. It can be subdivided into two main segments: the anterior and the posterior segment. Both can be approximated as two overlapping spheres of different size. The smaller anterior segment covers about one sixth of the eye and contains the components in front of the vitreous humour including cornea, iris, pupil and lens. The posterior segment covers the remaining five sixths.

For the proposed approach, we apply a 3D geometric model of the human eye (Fig. 2, top) which is used (1) to estimate the eye pose from an image and (2) to calculate light interaction at the eye. There are two main difficulties related to such a model: (1) the eyeball has a complex shape and (2) its anatomic parameters vary among humans. However, for most applications its sufficient to apply the before-mentioned observation and approximate the eye as two overlapping spheres. Nevertheless, parameter values vary amongst related work (Fig. 2, bottom).

2.2 3D Eye Pose Estimation

The 3D pose of an eye describes the location and orientation of the eye in a common coordinate frame. We further refer to the camera coordinate frame, where the origin $\mathbf{O} = (0, 0, 0)$ is placed at the camera pupil (Fig. 3). There exist several methods to estimate the eye pose from an image of the eye, mainly related to eye tracking. The methods can



	[12]	[14]	[19]	[20]	[21]	[22]
tC			0.55			0.45
rL	6.47*	5.50	6.15* / 6.54*	6.19*	5.80	5.59
dAL	3.53* / 4.15	2.18	3.00* / 3.55	3.05	2.50	2.80 / 3.24*
dLC	4.17				5.30	
dCE					4.70	
rC	7.70	7.80	7.80	7.80	7.80	7.26
rE			12.50	12.50*	11.50	

図 2 The geometric eye model with an overview of parameter values used in related works. Entries marked with * were calculated by applying the model. In case two values are stated for $d_{\mathbf{AL}}$, the first one does not include the thickness of the cornea.

be categorized by whether (1) they allow any camera placement [14], [17], [18] or (2) assume a rigid setup with a head-mounted camera [19], [20] and (a) whether they do passive image processing [14], [17], [18] or (b) use active controlled illumination [19], [20]. We adopt the method described by Nishino [14] to fit our eye model. The method can be used for the left and right eye in the same way. We assume the internal parameters of the camera to be known. Further mentioned points are assumed to be in camera coordinates.

The cornea itself is a paraboloid that is cut from the corneal sphere by the limbus plane. Assuming weak perspective projection, the 3D limbus circle projects to an ellipse in the image which can be described using five parameters. These are the center $\mathbf{l} = (u_\mathbf{l}, v_\mathbf{l})$, the major and minor radii $r_{\max}$ and $r_{\min}$ and the rotation angle ϕ. We estimate these parameters by fitting an ellipse to approximate the iris boundary.

The 3D location of the limbus center $\mathbf{L}$ can be estimated from the detected ellipse. The distance d between image plane and average depth plane is obtained as

$$d = r_L \frac{f}{r_{\max}} \tag{1}$$

where f is the focal length of the camera in pixels and r_L is the limbus radius specified by the eye model. Knowing the distance to the image plane we estimate the coordinates of the limbus center as

$$\mathbf{L} = \left(d\frac{u_\mathbf{l} - u_0}{f}, d\frac{v_\mathbf{l} - v_0}{f}, d \right) \tag{2}$$

where (u_0, v_0) are the coordinates of the principal point.

The gaze direction corresponds to the visual axis of the eye $-\mathbf{z}_{eye}$. It is described by the rotation ${}^{eye}_{cam}\mathbf{R}(\phi, \tau)$ of the eye coordinate frame centered in $\mathbf{L}$ from the camera coordinate frame. The angle ϕ around $\mathbf{y}_{eye}$ is already known as the rotation of the limbus in the image plane. The angle τ around $\mathbf{x}_{eye}$ corresponds to the tilt of the 3D limbus circle with respect to the image plane. It is obtained from the shape of the imaged limbus as

$$\tau = \arccos\left(\frac{r_{\min}}{r_{\max}}\right) \tag{3}$$

In this method, τ is always estimated as a positive angle describing an eye looking up. Nishino [14] refers to an automatic method resolving this ambiguity.

At this point, we are able to reconstruct the center of the corneal sphere $\mathbf{C}$ located at some distance $d_{\mathbf{LC}}$ from the limbus center $\mathbf{L}$ along $\mathbf{z}_{eye}$ as

$$\mathbf{C} = \mathbf{L} + d_{\mathbf{LC}} \cdot {}^{eye}_{cam}\mathbf{R}(\phi, \tau) \frac{\mathbf{L}}{\|\mathbf{L}\|} \tag{4}$$

Center $\mathbf{C}$ and radius r_C describe the corneal sphere which enables us to model the light reflection properties of the corneal surface. Knowing the gaze direction, we can further construct the eyeball sphere with radius r_E around the center $\mathbf{E}$. However, this is not necessary in our case.

3. Light Transport at the Eye

Light that is perceived by the human visual system passes through several transparent components of the eye until finally reaching the fovea. Along the direction of incoming light, these components are cornea, aqueous humor, lens and vitreous humor. At every transition, the light undergoes reflection and refraction. The light reflected at a particular transition is called the n-th Purkinje image [12]. Here, we only cope with the most prominent first Purkinje image P1 reflected from the outer surface of the cornea. We employ the geometric eye model from Sec. 2. to develop a corneal reflection model. This can be used to calculate the direction of incoming light reflecting at the corneal surface into the pupil of the camera.

3.1 Corneal Reflection Model

We capture a camera image of a human face $I(u, v)$, showing the reflection of a light source in the cornea of an eye (Fig. 4). We then determine pixel $\mathbf{s} = (u_\mathbf{s}, v_\mathbf{s})$ as the center of this reflection in the image. Knowing the internal parameters we reconstruct a 3D point $\mathbf{S}'$ on the projection ray as

$$\mathbf{S}' = \left(\frac{u_\mathbf{s} - u_0}{f}, \frac{v_\mathbf{s} - v_0}{f}, 1 \right) \tag{5}$$

We formulate point $\mathbf{S}$ on the corneal surface as

$$\mathbf{S} = t_1 \mathbf{r}_1 \tag{6}$$

where $\mathbf{r}_1 = \mathbf{S}'/\|\mathbf{S}'\|$ is the normalized direction vector. To recover $\mathbf{S}$ we compute the intersection with the corneal sphere for the scale parameter t_1 by solving the quadric equation

$$\|\mathbf{S} - \mathbf{C}\|^2 = r_C^2 \tag{7}$$

Expanding and rearranging leads to

$$t_1{}^2 \mathbf{r}_1^2 - 2t_1 \left(\mathbf{r}_1^2 \mathbf{C}\right) + \mathbf{C}^2 - r_C^2 \tag{8}$$

from where we construct the simplified quadric formula

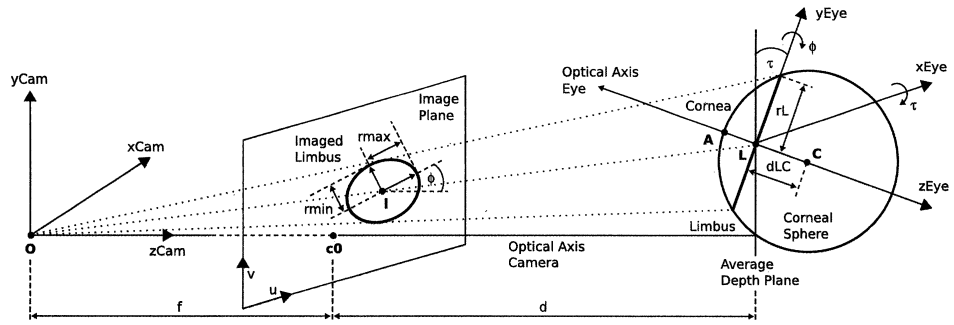


図 3 The relation between camera and eye coordinate frames for 3D eye pose estimation from an image.

$$t_1 = \left(\mathbf{r}_1^2\mathbf{C}\right) \pm \sqrt{\left(\mathbf{r}_1^2\mathbf{C}\right)^2 - \mathbf{C}^2 + r_C^2} \tag{9}$$

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

At point $\mathbf{S}$ we compute the specular reflection

$$\mathbf{r}_2 = 2\mathbf{n}_\mathbf{S} + \mathbf{r}_1 \tag{10}$$

where $\mathbf{r}_2 = \mathbf{P} - \mathbf{S}/\|\mathbf{P} - \mathbf{S}\|$ is the normalized direction vector towards the position of a light source $\mathbf{P}$ which can be expressed as a ray from $\mathbf{S}$ with

$$\mathbf{P} = \mathbf{S} + t_2\mathbf{r}_2 \tag{11}$$

at an unknown distance t_2.

4. Display-Eye-Camera System

Our controlled illumination system consists of three components: (1) a raster display device which acts as a light source, (2) the cornea of the human eye which reflects the light from the display and (3) a camera which captures the reflected light from the cornea. We assume a simple lightpath where the light is reflected only once at the outer surface of the cornea.

4.1 Geometric Display Model

A display can be modeled as a plane containing the pixels $\mathbf{p} = (i, j)$ where $i \in [1, rx]$ and $j \in [1, ry]$ with rx and ry denoting the resolution of the display. We want to describe pixel $\mathbf{p}$ as point $\mathbf{P} = (x, y, z)$ in camera coordinates (Fig. 4).

Knowing at least three coplanar points $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ where two are linear independent, we can reconstruct the display plane and its pose. The plane equation is defined as

$$\mathbf{n}\left(\mathbf{P} - \mathbf{P}_1\right) = 0, \quad \mathbf{n} = \left(\mathbf{P}_2 - \mathbf{P}_1\right) \times \left(\mathbf{P}_3 - \mathbf{P}_1\right) \tag{12}$$

where vector $\mathbf{n}$ is perpendicular to the plane. Knowing pixel $\mathbf{p} = (i, j)$ we obtain its corresponding point $\mathbf{P}$ as

$$\mathbf{P}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{P}_1 + \begin{pmatrix} (i - i_1)\, s_x \\ (j - j_1)\, s_y \\ z \end{pmatrix} \tag{13}$$

where s_x and s_y are two scaling parameters depending on the size of the display pixels. They can be computed as

$$s_x = \frac{x'' - x'}{i'' - i'} \quad s_y = \frac{y'' - y'}{j'' - j'} \tag{14}$$

with $i' < i''$, $j' < j''$ and corresponding coordinates x', x'', y' and y'' obtained from two linear independent points on the plane. We can now completely describe $\mathbf{P}$ by solving the plane equation for z as

$$z = \frac{-n_x\left(x - x_1\right) - n_y\left(y - y_1\right)}{n_z} + z_1 \tag{15}$$

The three basis vectors describing the orientation of the display with respect to the camera coordinate frame are given as

$$\begin{aligned} \mathbf{i} &= \frac{\mathbf{P}\left(\mathbf{p}\left(1,0\right)\right) - \mathbf{P}\left(\mathbf{p}\left(0,0\right)\right)}{\left\|\mathbf{P}\left(\mathbf{p}\left(1,0\right)\right) - \mathbf{P}\left(\mathbf{p}\left(0,0\right)\right)\right\|} \\ \mathbf{j} &= \frac{\mathbf{P}\left(\mathbf{p}\left(0,1\right)\right) - \mathbf{P}\left(\mathbf{p}\left(0,0\right)\right)}{\left\|\mathbf{P}\left(\mathbf{p}\left(0,1\right)\right) - \mathbf{P}\left(\mathbf{p}\left(0,0\right)\right)\right\|} \\ \mathbf{k} &= \mathbf{i} \times \mathbf{j} \end{aligned} \tag{16}$$

4.2 Radiometric Display Model

We use a display lighting model similar to Funk [3] which takes into account the attenuation effect of viewing angle and distance. While the viewing angle might be neglected for CRT monitors, it has a high influence on LCD monitors.

A particular pixel $\mathbf{p} = (i, j)$ emits light over a hemisphere of directions with radiance $R\left(i, j, \theta, \phi\right)$, that is described as

$$R\left(i, j, \theta, \phi\right) = R'\left(i, j\right) \cdot f\left(\theta, \phi\right) \tag{17}$$

where $R'\left(i, j\right)$ is the unattenuated radiance and $f\left(\theta, \phi\right)$ is the attenuation function. The corresponding irradiance at some point $\mathbf{S}$ in space or on a surface depends also on its distance to the display and can be modeled as

$$I_\mathbf{S} = \frac{R\left(i, j, \theta_\mathbf{S}, \phi_\mathbf{S}\right)}{\left(\mathbf{S} - \mathbf{P}\right)^2} \tag{18}$$

where the angles $\theta_\mathbf{S}$ and $\phi_\mathbf{S}$ describe the location of $\mathbf{S}$ with respect to the orientation of the screen.

4.3 Geometric Display Calibration

As explained for the geometric display model, we need to know the 3D positions of some points $\mathbf{P}_k$ with $k \geqslant 3$ on the display plane. Therefore we render a binary lighting pattern as described in Sec. 5. where the corresponding positions are represented as white circles with radius r centered in $\mathbf{p}_k$. To simplify explanations, we continue for a single point $\mathbf{P}$ without regarding its index k.

We capture a set of N images for a static pattern under

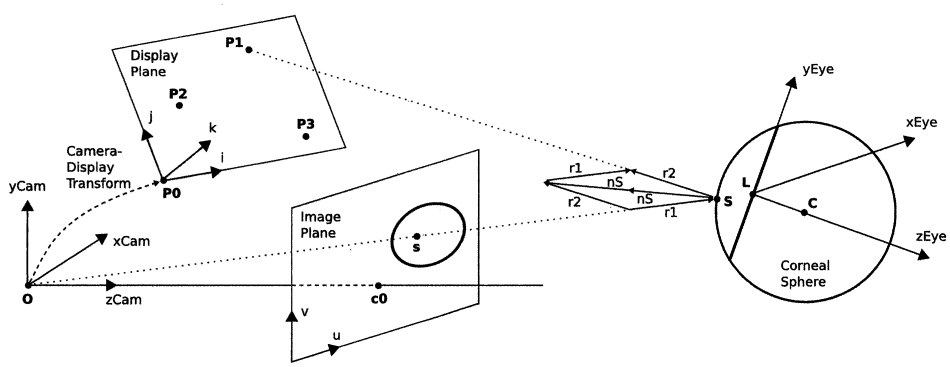


図 4 The setup for computing the direction towards a lights source. The model is applied to estimate the pose and size of a 3D display plane from a set of images showing corneal reflections of at least three planar light sources $\mathbf{P}_k$.

varying eye positions. For each image, we describe $\mathbf{P}$ as a point on a ray reflected from the cornea by

$$\mathbf{P} = \mathbf{S} + t_2\mathbf{r}_2 \tag{19}$$

as explained in Sec. 3. Computing the intersection of all N rays leads to the position of $\mathbf{P}$. However, generally the rays do not intersect due to measurement errors and system model simplifications. Thus, we define the ray intersection as a least-squares problem where the sum of squared distances d_l^2 between the rays and $\mathbf{P}$ should be minimal. A solution is obtained by solving the linear equation system

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\begin{bmatrix} \mathbf{r}_{21}^T \\ \vdots \\ \mathbf{r}_{2N}^T \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} t_{21} \\ \vdots \\ t_{2N} \end{bmatrix} \tag{20}$$

that is constructed from $N \geqslant 3$ different rays. The solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ is obtained by computing the pseudo-inverse of the coefficient matrix $\mathbf{A}$ as

$$\mathbf{A}^{-1} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T \tag{21}$$

This leads to

$$\begin{aligned} \mathbf{P} &= \frac{1}{N}\sum_{l=1}^{N}(\mathbf{S}_l + t_{2l}\mathbf{r}_{2l}) \\ d &= \frac{1}{N}\sum_{l=1}^{N}|\mathbf{P} - (\mathbf{S}_l + t_{2l}\mathbf{r}_{2l})| \end{aligned} \tag{22}$$

where $\mathbf{P}$ describes the point with nearest average distance d to the set of rays.

Knowing the points $\mathbf{P}_k$ and display resolution rx, ry, we are able to describe the display size and pose with respect to the camera coordinate frame.

5. Experiments

5.1 Setup

Our setup contains a 57 inch, 1280×720 (16:9) Epson Livingstation LS57P2 TFT LCD display and a Point Grey Flea2 camera with 1024×768 resolution. The intrinsic camera parameters are calibrated using a checkerboard pattern and OpenCV functions. We use the display to simulate a range of smaller rendered displays.

The camera is placed at varying distances d_{DC} in front of the display center with its optical axis aligned parallel to the **z**-axis of the display. Test persons are seated facing the camera with the center between both eyes at increasing distances d_{CE} in front of the camera (Fig. 5). For the experiments, we simulate a range of rendered displays and capture the corresponding face images. For each image, we find the limbi of both eyes and estimate the pose of the corneal sphere. We then extract the set of corneal reflections (CR) and compute the rays towards the corresponding corners of the rendered display. Since we know the distance between display and eyes, we can reconstruct the 3D position on the display.

5.2 Lighting Pattern Architecture

We employ a binary pattern centered on the real display showing the four corners of a particular rendered display. Each corner is rendered as a white circle around point $\mathbf{V}_k$ where $k = 1 \ldots 4$ with radius r to approximate a point light source (Fig. 6). In the ideal case, r spans only a single pixel. However, it had to be set to some higher value to achieve a measurable camera response. On the other side, r can not be set too high so as not to overexpose CR regions and invalidate the distant point light source assumption.

In the following, we show how to generate the pattern images. We assume known radius r, diagonal length d_r, resolution rx_r and ry_r of the real display as well as diagonal length d_v and aspect ratio $a_v = w_v/h_v$ of the rendered display. The indices r and v denote parameters of real and virtual rendered display. All resolutions are assumed in pixels, all lengths in the same unit.

At first, we retrieve the length of the edges of the real display as

$$w_r = \sqrt{\frac{d_r^2}{1 + 1/a_r^2}}, \quad h_r = w_r\frac{1}{a_r} \tag{23}$$

where $a_r = w_r/h_r$. In the same way, we calculate the length of the edges of the rendered display w_v and h_v. The corners $\mathbf{V}_k$ of the rendered display are now defined as

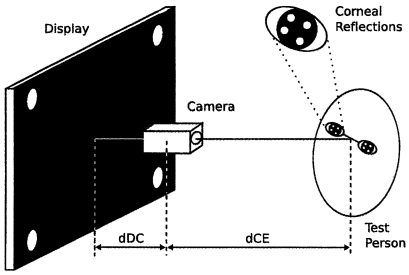


図 5 The experimental setup with display, camera and test person exhibiting corneal reflections of the pattern in both eyes.

$$\mathbf{V}_k = \left(\frac{rx_r \pm rx_v}{2}, \frac{ry_r \pm ry_v}{2} \right) \tag{24}$$

where $rx_v = a_v ry_v$ denotes the resolution of the rendered display.

Due to the orientation of the display, radiance attenuation becomes a problem for correctly detecting a camera response. Therefore, we have to use larger r for larger rendered display sizes. For our setup, a linear increase of r is sufficient, with

$$r = r_{\min} + s\,(r_{\max} - r_{\min}) \tag{25}$$

where $s \in [0,1] = d_v - d_{v\,\min}/d_{v\,\max} - d_{v\,\min}$ is the scale.

5.3 Image Acquisition

We generate and display a series of patterns for increasing rendered display sizes d_v from 10 to 43 inch in steps of one inch. We capture a data set of corresponding face images as fast as possible. Though, the framerate is limited to 10 fps to guarantee correct display-camera synchronization. To obtain enough camera response, we capture at maximum aperture, gain of 30 dB and exposure time of 100 ms, further limiting the framerate. We assume the display is the only lightsource in our setup. In the future, iris image separation [23] and false CR detection might be used to compensate for this.

5.4 Light Source Estimation

The acquired images show both eyes of a test person with CRs (Fig. 5). After detecting a particular limbus we extract the CRs by segmenting four regions containing the highest intensity values from its bounding box. We apply an adaptive threshold for background subtraction. Knowing the pixels corresponding to a particular region R_k, we compute its intensity centroid $\mathbf{s}_k$ with subpixel accuracy as

$$\mathbf{s}_k = \frac{\sum_{(u,v)\in R_k} I\,(u,v) \cdot \begin{pmatrix} u \\ v \end{pmatrix}}{\sum_{(u,v)\in R_k} I\,(u,v)} \tag{26}$$

where $k = 1 \ldots 4$. To increase the accuracy of the estimated centroid value in the future, we have to consider the reflection from the display on a patch of the corneal surface. We then estimate pixel $\mathbf{s}_k$ from the 3D centroid of the patch. We also have to find out if the measured reflected intensity results only from corneal reflection or from an overlay with iris and lens reflections.

5.5 Results

We took data sets for two test persons at increasing camera-eye distance d_{CE} from 25 to 95 cm. At a fixed distance of 35 cm, the reflections of the rendered display span a

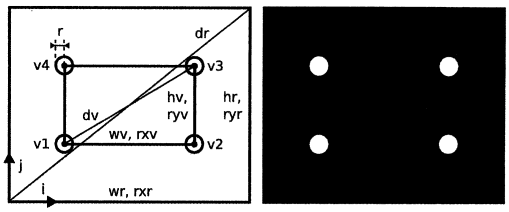


図 6 (a) The display pattern architecture with (b) a particular generated pattern.

range of 10 to 35 pixels (Fig. 7, top). Due to display attenuation, the pattern circle radius r had to be linearly increased from 0.5 to 3 inch. For higher rendered display sizes d_v, we notice a distortion of the CR area resulting from the shape of the cornea which might have an effect on centroid computation. The impact of low resolution can be noticed for increasing camera-eye distance (Fig. 7, bottom). The limbus bounding box decreases from about 100 to 20 pixels. In our setup, the visual axis of the eye and the optical axis of the camera are aligned in parallel. However, we detect a constant small elliptical distortion of the limbus where the vertical axis measures only 92% of the horizontal axis, meaning that the limbus is not purely circular. CR extraction fails for small rendered display sizes at distances between 65 and 85 cm and completely fails at distances higher than 85 cm.

We compare the estimated points to the ground truth on the real display plane. We measure the position error as the $\mathbb{R}^3$ euclidean distance between estimated and ground truth light source positions. We further measure the error in diagonal length as the signed distance between estimated and ground truth rendered display sizes. In the current experiments, we manually detected the parameters of the imaged limbus only once per data set. While capturing, we fixed the heads of the test persons and asked them to gaze at a small dark blue mark in the center of the display. However, head position and gaze did not remain static which explains the higher frequency variation in the experimental results (Fig. 8). In this setup, only the error in diagonal length is meaningful.

Fig. 8 shows experimental results for the accuracy of estimated 3D light source positions. (a) We measure highest accuracy for a corneal sphere radius of 7.8 mm which is also proposed in related work. Corneal shape and resulting surface normals have the highest impact on accuracy among all system parameters. (b) The accuracy variance increases with camera-eye distance due to a higher impact of head-movements and deviations in centroid extraction resulting from low resolution. (c) We observe no particular variation in model response between left and right eye. (d) We make an interesting observation comparing the results for different persons. While variation is small for the same person, it is high among different persons, increasing with rendered display size. We assume this to be related to anatomical differences. Even with this simple geometrical eye model under varying head positions, the error from estimated diagonal lengths and point positions does not exceed 20%. We

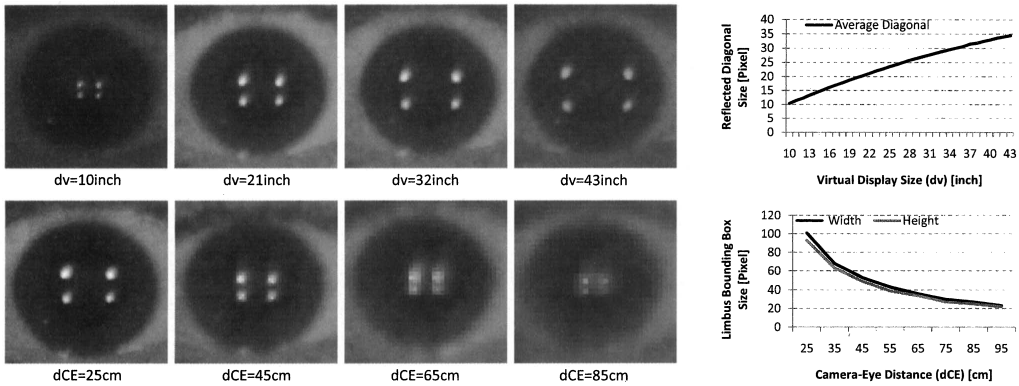


図 7 The images show the iris and corneal reflections on the right eye of a test person. The top row exhibits the effects of an increasing rendered display size d_v (for fixed camera-eye distance d_{CE} of 35 cm). The diagram shows the increasing average diagonal size measured from the images in pixels. The bottom row exhibits the effects of an increasing camera-eye distance d_{CE} (for fixed rendered display size d_v of 25 inch). The diagram shows the decreasing resolution of the limbus bounding box.

believe to obtain more accurate measurements with an improved system in the future.

6. Conclusion and Future Work

With this work we want to lay the fundamentals for the application of display-camera systems in non-professional everyday environments. We proposed a calibration technique without special hardware and user interaction. Our technique relies only on corneal reflections. We introduced a three-step approach: (1) detecting eye pose, (2) computing light rays from corneal reflections and (3) estimating display pose and size from multiple eye poses. We implemented a simple geometric eye model and performed basic experiments using only off the shelf components showing the general feasibility of our approach.

This work is our initial step towards an applicable calibration technique. Nevertheless, there are several things to be done in the future: The first issue relates to the system model. Further theoretical and practical accuracy evaluation has to be done. We are planning more complex experiments with varying eye poses. Different geometric eye models have to be tested in order to better approximate the eye shape. Strategies to automatically calibrate anatomic parameter variation have to be investigated to minimize the error along different users. We are further interested in common model parameter error variances and their effect on accuracy and will create a framework for testing with synthetic data. It is also interesting to consider a reorganization of the system: eye parameters might be calibrated under known camera and display parameters [24].

The next issue relates to the display pattern and extraction of the CRs. We have to model the effect of point light source positions and sizes encoded in the pattern. It is necessary to consider different pattern architectures (e.g. gradient, colored), pattern sequences and corresponding calibration strategies. It is also interesting to analyze the effect of complex light interaction at the eye.

As our aim is to apply the technique with off the shelf hardware, we need to consider practical problems such as image resolution, noise and ambient lighting. Moreover, accurate automatic eye tracking [17] and pose estimation has to be applied. Several users wear glasses or contact lenses. Their effect on the light path has to be investigated [7].

Finally, we are considering more advanced geometric calibration strategies involving both eyes. Applying epipolar geometry between two eyes [14] might enable a calibration from a single frame. Moreover, calibration might be extended to different parameters such as photometric properties.

文　献

[1] W. T. Freeman, D. B. Anderson, P. A. Beardsley, C. N. Dodge, M. Roth, C. D. Weissman, W. S. Yerazunis, H. Kage, K. Kyuma, Y. Miyake and K. Tanaka. Computer Vision for Interactive Computer Graphics. *Computer Graphics and Applications*, Vol. 18, No. 3, pp. 42–53, 1998.

[2] M. Greenspan and I. Fraser. Tracking a Sphere Dipole. *16th Intl. Conference on Vision Interface*, pp. 154–161, 2003.

[3] N. Funk and Y.-H. Yang. Using a Raster Display for Photometric Stereo. *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, pp. 201–207, 2007.

[4] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Shape from Shading*, MIT Press, pp. 513–531, 1989.

[5] G. Schindler. Photometric Stereo via Computer Screen Lighting for Real-time Surface Reconstruction. *Fourth International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT'08)*, 2008.

[6] J. J. Clark. Photometric Stereo with Nearby Planar Distributed Illuminants. *Third Canadian Conference on Computer and Robot Vision (CRV '06)*, p. 16, 2006.

[7] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch and M. Magnor and W. Heidrich. State of the Art in Transparent and Specular Object Reconstruction. *Eurographics - STAR*, 2008.

[8] Y. Francken, T. Cuypers, T. Mertens, J. Gielis and P. Bekaert. High Quality Mesostructure Acquisition Using Specularities. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[9] D. E. Zongker, D. M. Werner, B. Curless and D. H. Salesin. Environment Matting and Compositing. *ACM SIGGRAPH '99*, pp. 205–214, 1999.

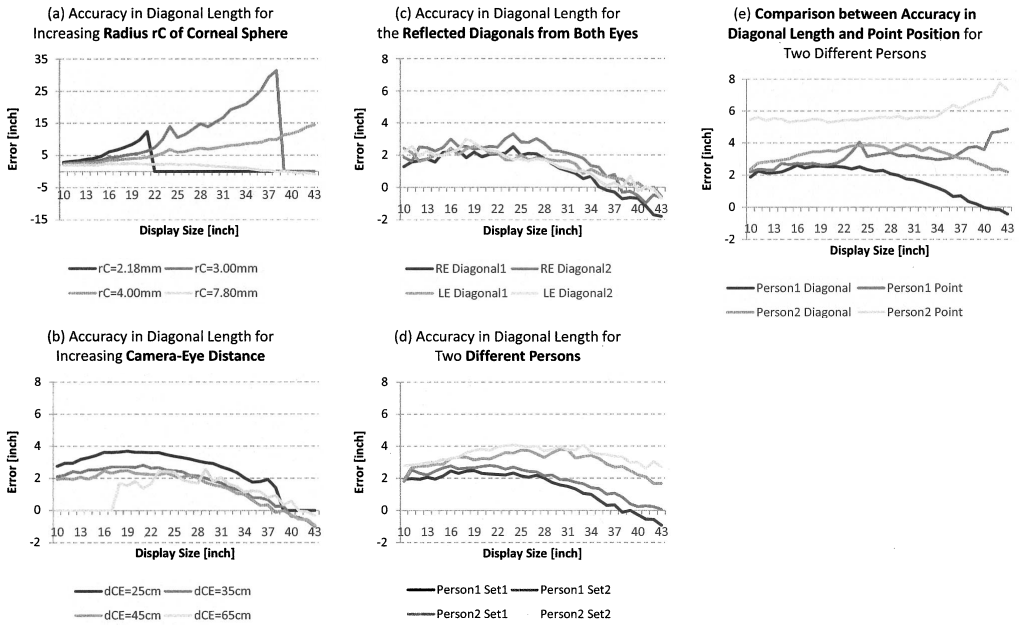


図 8 Experimental results showing the accuracy of estimated 3D light source positions. The calculation is done by extracting corneal reflections from captured face images. The position of a particular light source is estimated at known display-eye distance along the corresponding light ray. All diagrams show the accuracy for a range of display sizes (rendered patterns) from 10 to 43 inch. (a) shows the results for an increasing corneal sphere radius r_C. Small radii result in high errors and failing reflections at large display sizes. A radius of 7.8 mm (as proposed in related works) achieves the highest accuracy and will be used for the following results. (b) shows the results for an increasing camera-eye distance. Variance increases with distance because head-movements and deviations in centroid extraction have a higher impact. (c) shows the variance for the diagonals between left and right eye. The model responds similar for both eyes. (d) shows the variance between two test persons with two data sets per person. Variation is noticeably higher between persons. (e) shows a comparison between accuracy of diagonal length and point estimation.

[10] Y. Francken, C. Hermans and P. Bekaert. Screen-Camera Calibration using a Spherical Mirror. *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, pp. 11–20, 2007.

[11] A. Villanueva and R. Cabeza. Models for Gaze Tracking Systems. *Image and Video Processing*, Vol. 2007, No. 3, pp. 1–6, 2007.

[12] C. H. Morimoto and M. R. M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, Vol. 98, No. 1, pp. 4–24, 2005.

[13] T. Ishizawa, T. Okatani and K. Deguchi. Gaze-Reactive Image Display for Enhancing Depth Perception by Depth-of-Field Blur. *11th Meeting on Image Recognition and Understanding (MIRU 2008)*, pp. 113–120, 2008.

[14] K. Nishino and S. K. Nayar. Corneal Imaging System: Environment from Eyes. *International Journal of Computer Vision*, Vol. 70, No. 1, pp. 23–40, 2006.

[15] M. K. Johnson and H. Farid. Exposing Digital Forgeries Through Specular Highlights on the Eye. *Lecture Notes in Computer Science*, Vol. 4567, pp. 311–325, 2008.

[16] M. Backes, M. Dürmuth and D. Unruh. Compromising Reflections-or-How to Read LCD Monitors around the Corner. *IEEE Symposium on Security and Privacy*, pp. 158–169, 2008.

[17] R. Valenti and T. Gevers. Accurate Eye Center Location and Tracking Using Isophote Curvature. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[18] H. Wu, Q. Chen and T. Wada. Visual Direction Estimation from a Monocular Image. *IEICE Trans. on Information and Systems*, Vol. E-88D, No. 10, pp. 2277–2285, 2005.

[19] H. Hua, P. Krishnaswamy and J. P. Rolland. Video-based eyetracking methods and algorithms in head-mounted displays. *Optics Express*, Vol. 14, No. 10, pp. 4328–4350, 2006.

[20] F. Li, S. Kolakowski and Jeff Pelz. Using Structured Illumination to Enhance Video-Based Eye Tracking. *IEEE Intl. Conference on Image Processing*, pp. 373–376, 2007.

[21] A. Lefohn, B. Budge, P. Shirley, R. Caruso and E. Reinhard. An Ocularist's Approach to Human Iris Synthesis. *IEEE Computer Graphics and Applications*, Vol. 23, No. 6, pp. 70–75, 2003.

[22] C. R. Nave. Scale Model of Human Eye. *http://hyperphysics.phy-astr.gsu.edu/Hbase/vision/eyescal.html*, August 7, 2008 (last visited).

[23] H. Wang and S. Lin and X. Ye and W. Gu. Separating corneal reflections for illumination estimation. *Neurocomputing*, Vol. 71, No. 10–11, pp. 1788–1797, 2008.

[24] J. A. Sakamoto, H. H. Barrett and A. V. Goncharov. Inverse optical design of the human eye using likelihood methods and wavefront sensing. *Optical Express*, Vol. 16, No. 1, pp. 304–314, 2008.