Corneal Imaging System: Environment from Eyes[*]

Ko Nishino   Shree K. Nayar

Department of Computer Science   Department of Computer Science

Drexel University   Columbia University

3141 Chestnut Street,   1214 Amsterdam Avenue, MC 0401

Philadelphia, PA 19104, USA   New York, NY 10027, USA

kon@cs.drexel.edu   nayar@cs.columbia.edu

Abstract

This paper provides a comprehensive analysis of exactly what visual information about the world is embedded within a single image of an eye. It turns out that the cornea of an eye and a camera viewing the eye form a catadioptric imaging system. We refer to this as a corneal imaging system. Unlike a typical catadioptric system, a corneal one is flexible in that the reflector (cornea) is not rigidly attached to the camera. Using a geometric model of the cornea based on anatomical studies, its 3D location and orientation can be estimated from a single image of the eye. Once this is done, a wide-angle view of the environment of the person can be obtained from the image. In addition, we can compute the projection of the environment onto the retina with its center aligned with the gaze direction. This foveated retinal image reveals what the person is looking at. We present a detailed analysis of the characteristics of the corneal imaging system including field of view, resolution and locus of viewpoints. When both eyes of a person are captured in an image, we have a stereo corneal imaging system. We analyze the epipolar geometry of this stereo system and show how it can be used to compute 3D structure. The framework we present in this paper for interpreting eye images is passive and non-invasive. It has direct implications for several fields including visual recognition, human-machine interfaces, computer graphics and human affect studies.

1   What Do Eyes Reveal?

Our eyes are crucial to providing us with an enormous amount of information about our physical world. What is less often explored is the fact that the eye also conveys equally rich information to an external observer. Figure 1 shows two images of eyes. One can see a building in the example on the left and a person's face in the example on the right. At the same time, we can see that the mapping of the environment within the appearance of the eye is complex and depends on how the eye is imaged. In [33], highlights in an eye were used to locate three known point sources. The sources were then used to apply photometric stereo and relight the face. Extensive previous work has also been done on using eyes to estimate gaze [41, 29, 40, 31, 26, 11] and to identify people from their iris textures [14, 7][1]. However, virtually no work has been done on using eyes to interpret the world surrounding the person to whom the eye belongs.

In this paper, we provide the first comprehensive analysis of exactly what information is embedded within a single image of an eye. We believe this work is timely. Until recently, still and video cameras were relatively low in resolution and hence an eye in an image simply did not have a sufficient number of pixels to represent useful information. Recently, however, CCD and CMOS image detectors have made quantum leaps in terms of resolution. At this stage, one can legitimately ask the following question: What does an image of an eye reveal about the world and the person and how can this information be extracted?

Our key observation is that the combination of the cornea of an eye and the camera capturing the appearance of the eye can be viewed as a catadioptric (mirror + lens) imaging system. We refer to this as the *corneal imaging system*. Since the reflecting element (the cornea) is not rigidly attached to the camera, the corneal imaging system is inherently an uncalibrated one. We use a geometric

---

[1]The texture of the iris is known to be a powerful biometric for human identification [20, 14, 7]. It is important to note that sensors used for scanning the iris use special lighting to ensure that the reflections from the cornea (appearance of the external world) are minimized. In unstructured settings, however, the corneal reflections tend to dominate the appearance of the eye and it is exactly this effect we seek to exploit. Clearly, even in an unstructured setting, the texture of the iris will contribute to the appearance of the eye. In our work, we do not attempt to eliminate the contribution of the iris; this problem is significant by itself and will be addressed separately in the future.
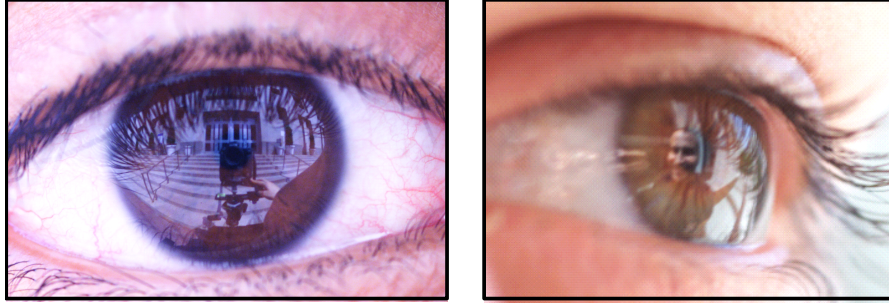
Figure 1: What does the appearance of an eye tell us about the world surrounding the person and what the person is looking at?

model of the cornea based on anatomical studies to estimate its 3D location and orientation[2]. This is equivalent to calibrating the corneal imaging system. Once this is done, we show that we can compute a precise wide-angle view of the world surrounding the person. More importantly, we can obtain an estimate of the projection of the environment onto the retina of the person's eye. We refer to this as the retinal image. From this retinal image of the surrounding world, we are able to determine what the person is looking at (focus of attention) without implanting an image detector in his/her eye.

We present a detailed analysis of the characteristics of the corneal imaging system. We show that, irrespective of the pose of the cornea, the field of view of the corneal system is greater than the field of view observed by the person to whom the eye belongs. We find that the spatial resolution of the corneal system is similar in its variation to that of the retina of the eye; it is highest close to the direction of gaze. It also turns out that this imaging system is a non-central one; it does not have a single viewpoint but rather a locus of viewpoints. We derive the viewpoint locus of the corneal system and show how it varies with the pose of the cornea. When both eyes of a person are captured in the same image, we have a catadioptric stereo system. We derive the epipolar geometry for such a system and show how it can be used to recover the 3D structure of objects in front of the person.

We believe our framework for extracting visual information from eyes has direct implications

---

[2]The computed corneal orientation may be used as an estimate of the gaze direction. However, it is important to note that gaze detection is not the focus of our work – it is just a side product.
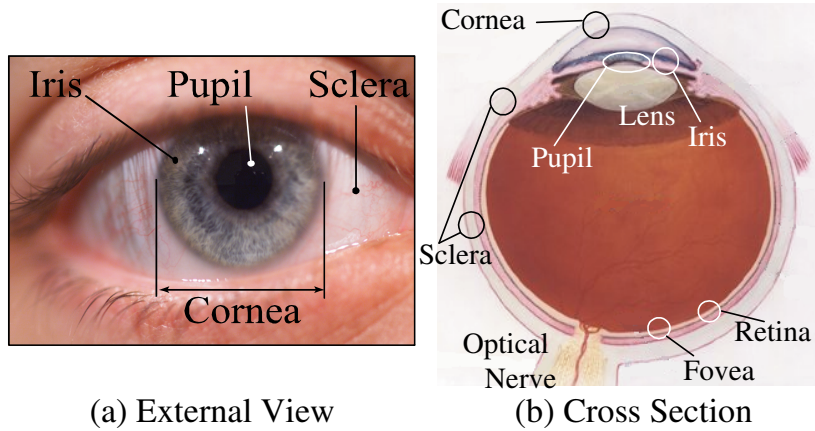
(a) External View   (b) Cross Section

Figure 2: (a) An external view of the human eye. The sclera and the cornea (behind which the pupil and the iris reside) are the most visually distinct components of the eye. (b) A horizontal cross-section of the right human eyeball. Despite its fairly complex anatomy, the physical dimensions of the eye do not vary much across people.

for various fields. In visual recognition, the recovered wide-angle view of the environment can be used to determine the location and circumstance of the person when the image was captured. Such information can be very useful in security applications. The computed environment map also represents the illumination distribution surrounding the person. This illumination information can be used in various computer graphics applications including relighting faces and objects. The computed retinal image can reveal the intent of the person. This information can be used to effectively communicate with a computer or a robot [4], leading to more advanced human-machine interfaces. The image of a person's face can tell us their reaction, while the appearance of their eyes in the same image can reveal what they are reacting to. This information is of great value in human affect studies [32, 13].

2   Physical Model of the Eye

As can be seen in Figure 2(a), the most distinct visual features of the eye are the cornea and the sclera. Figure 2(b) shows a schematic of a horizontal cross-section of the right eyeball. The cornea

consists of a lamellar structure of submicroscopic collagen fibrils arranged in a manner that makes it transparent [20, 8, 37]. The external surface of the cornea is very smooth. In addition, it has a thin film of tear fluid on it. As a result, the surface of the cornea behaves like a mirror. This is why the combination of the cornea and a camera observing it form a catadioptric system, which we refer to as the *corneal imaging system*.

To interpret an image captured by this system, we need a geometric model of its reflecting element, i.e. the cornea. In the fields of physiology and anatomy, extensive measurements of the shape and dimensions of the cornea have been conducted [20, 34]. It has been found that a normal adult cornea is very close to an ellipsoid, as shown in Figure 3. In Cartesian coordinates (x, y, z), an ellipsoid can be written as [2]

$$pz^2 - 2Rz + r^2 = 0\,, \tag{1}$$

where $r = \sqrt{x^2 + y^2}$, $p = 1 - e^2$ where e is the eccentricity and R is the radius of curvature at the apex of the ellipsoid. Now, a point $\mathbf{S}$ on the corneal surface can be expressed as

$$\mathbf{S}(t, \theta) = (S_x, S_y, S_z) = (\sqrt{-pt^2 + 2Rt}\ cos\theta, \sqrt{-pt^2 + 2Rt}\ sin\theta, t) \tag{2}$$

where $0 \leq \theta < 2\pi$ (see Figure 3). It turns out that the parameters of the ellipsoid do not vary significantly from one person to the next. On average, the eccentricity e is 0.5 and the radius of curvature R at the apex is 7.8 mm [20].

The boundary between the cornea and the sclera is called the *limbus*. The sclera is not as highly reflective as the cornea. As a result, the limbus defines the outer limit of the reflector of our imaging system. From a physiological perspective, the cornea "dissolves" into the sclera. However, in the case of an adult eye, the limbus has been found to be close to circular with radius r_L of approximately 5.5 mm. Therefore, in equation (2), the parameter t ranges from 0 to t_b where t_b is determined using $-pt_b^2 + 2Rt_b + r_L^2 = 0$ and found to be 2.18 mm.

In the remainder of this paper, we will use the above geometric model of the cornea, which well approximates a normal adult cornea. It is important to note that slight changes in the parameters of the model will not have a significant impact on our results. However, if the cornea significantly deviates from a normal adult cornea, for instance due to diseases such as keratoconus, it will be
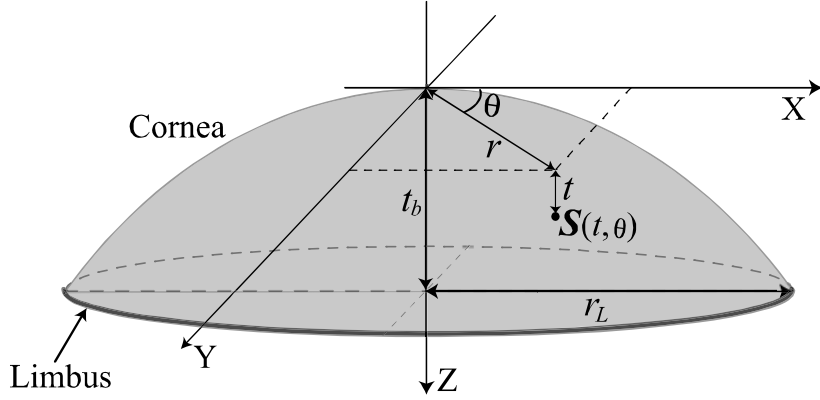
Figure 3: The cornea is modeled as an ellipsoid whose outer limit corresponds to the limbus. For a normal adult cornea, the approximate eccentricity and the radius of curvature at the apex are known. The Cartesian coordinates of corneal surface points can be parameterized with the azimuth angle θ and the height t along the Z axis.

necessary to measure its shape. This can be done by using structured light [15] and the measured shape can be directly used in our method.

3   Finding the Pose of an Eye

To explore the visual information captured by the corneal imaging system, we first need to calibrate the system. This corresponds to estimating the 3D location and orientation of the cornea from the image of an eye. We accomplish this by first locating the limbus in the image.

3.1   Limbus Localization

The limbus in an image is the projection of a circle in 3D space. Even in the extreme case when the gaze direction is perpendicular to the optical axis of the camera, the depth of this circle is only 11 mm (the diameter of the limbus). Hence, we can safely assume the camera projection model to be weak-perspective; orthographic projection followed by scaling. Under this assumption, the limbus is imaged as an ellipse.

Let (u, v) denote the horizontal and vertical image coordinates, respectively. An ellipse in an

image can be described using five parameters which we denote by a vector $\mathbf{e}$. These parameters are the center (c_u, c_v), the major and minor radii r_{max} and r_{min} and the rotation angle ϕ of the ellipse in the image plane. We localize the limbus in an image $I(u, v)$ by searching for the ellipse parameters $\mathbf{e}$ that maximize the response to an integro-differential operator applied to the image after it is smoothed by a Gaussian g_σ:

$$\max_{\mathbf{e}=(c_x, c_y, r_x, r_y, \phi)} \left| g_\sigma(r_{max}) * \frac{\partial}{\partial r_{max}} \oint_{\mathbf{e}} I(u, v) ds + g_\sigma(r_{min}) * \frac{\partial}{\partial r_{min}} \oint_{\mathbf{e}} I(u, v) ds \right| . \qquad (3)$$

Note that the filter responses are integrated along the arc ds of the ellipse. Daugman [7] proposed the use of a similar integro-differential operator to detect the limbus and the pupil as circles in an image of a forward looking eye. In contrast, our algorithm detects the limbus as an ellipse, which is necessary when the gaze direction of the eye is unknown and arbitrary.

We provide an initial estimate of $\mathbf{e}$ to the algorithm by drawing an ellipse in the image. This is accomplished with a simple user interface for positioning, resizing and rotating a predefined ellipse by dragging the center or the arc of it. We also let the user specify the arc range of this initial ellipse to be used for evaluating equation (3). In many cases, due to the large intensity gradient change between the cornea and the sclera, this arc range setting can be very crude. However, for cases when the eyelids significantly occlude the limbus or when the cornea itself occludes one side of the limbus, it becomes important to discard such occlusions from being evaluated in equation (3).

Same as [7], starting with the initial ellipse parameters, we adopt a discrete iterative implementation for obtaining the optimal ellipse parameters that maximize equation (3). We use simplex method for the iterative search. Alternatively, one can use other gradient based optimization algorithms. Furthermore, instead of manually specifying the arc range, robust estimators can be used to discard non-limbus edges.

Figure 4 shows results of the limbus localization algorithm applied to different images of eyes. Note that the limbus is accurately located for arbitrary gaze directions, despite the complex texture of the iris and partial occlusions by the eyelids.

Figure 4: Detected limbuses (red ellipses) in images of eyes. The ellipse detector successfully locates the limbus for different unknown gaze directions, despite the complex texture of the iris and partial occlusions by the eyelids.

3.2   3D Location of the Cornea

Once the limbus has been detected, we can find the 3D location of the cornea in the camera's coordinate frame. We assume the internal parameters of the camera are calibrated *a priori*[3].

As shown in Figure 5, under weak perspective projection, the limbus in 3D space is first orthographically projected onto a plane parallel to the image plane. This plane lies at the average depth of the limbus from the camera. Since the limbus is a circle in 3D space, this average depth plane always passes through the center of the limbus. As a result, the major axis of the estimated ellipse in the image corresponds to the intersection of the limbus and this average depth plane. The length of this line segment is the diameter of the limbus. Therefore, the distance to the average depth

---

[3]All eye images in this paper were captured with a Kodak DCS 760 camera with 6M pixels. Close-up views are obtained using a Micro Nikkor 105mm lens. We implemented the rotation based calibration method described in [28] to find the internal parameters of the camera.
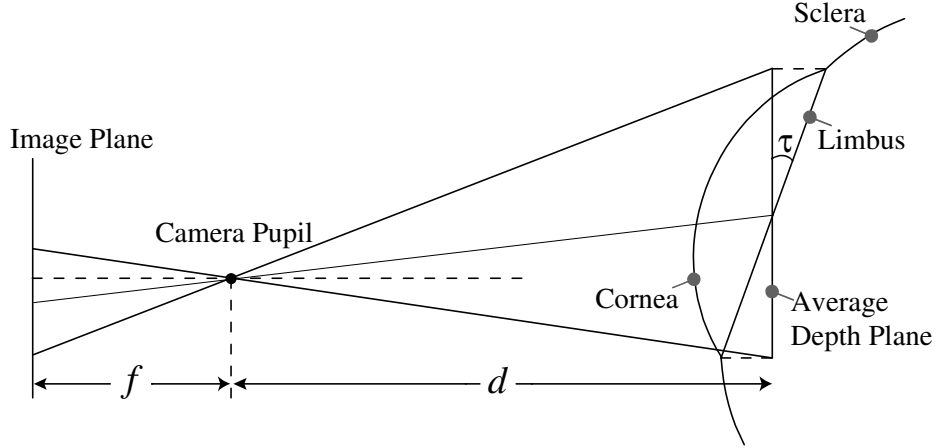
Figure 5: Weak perspective projection of the eye onto the image plane. The 3D location and orientation of the cornea in camera space can be computed from the imaged limbus.

plane, d, can be computed as

$$d = r_L \frac{f}{r_{max}}, \qquad (4)$$

where $r_L = 5.5$mm, f is the focal length in pixels and r_{max} is known from the limbus detection in Section 3.1. Note that the depth d, the ellipse center (c_u, c_v) in the image and the focal length f determine the 3D coordinates of the center of the limbus.

Figure 6 shows results of an experiment we conducted to verify the accuracy of depth (eye distance) estimation. We captured 10 images of a person's right eye at 5 different distances ranging from 75cm to 160cm. For each of the 10 images, the person was asked to change his gaze direction. Regardless of the gaze direction, the depth d was estimated with good accuracy in all cases with an RMS error of 1.9%.

3.3 3D Orientation of the Cornea

From the image parameters of the limbus we can also compute the 3D orientation of the cornea. The 3D orientation is represented using two angles (ϕ, τ). ϕ is the rotation of the limbus in the image plane which we have already estimated. Consider the plane in 3D on which the limbus lies. τ is the angle by which this plane is tilted with respect to the image plane (see Figure 5) and can
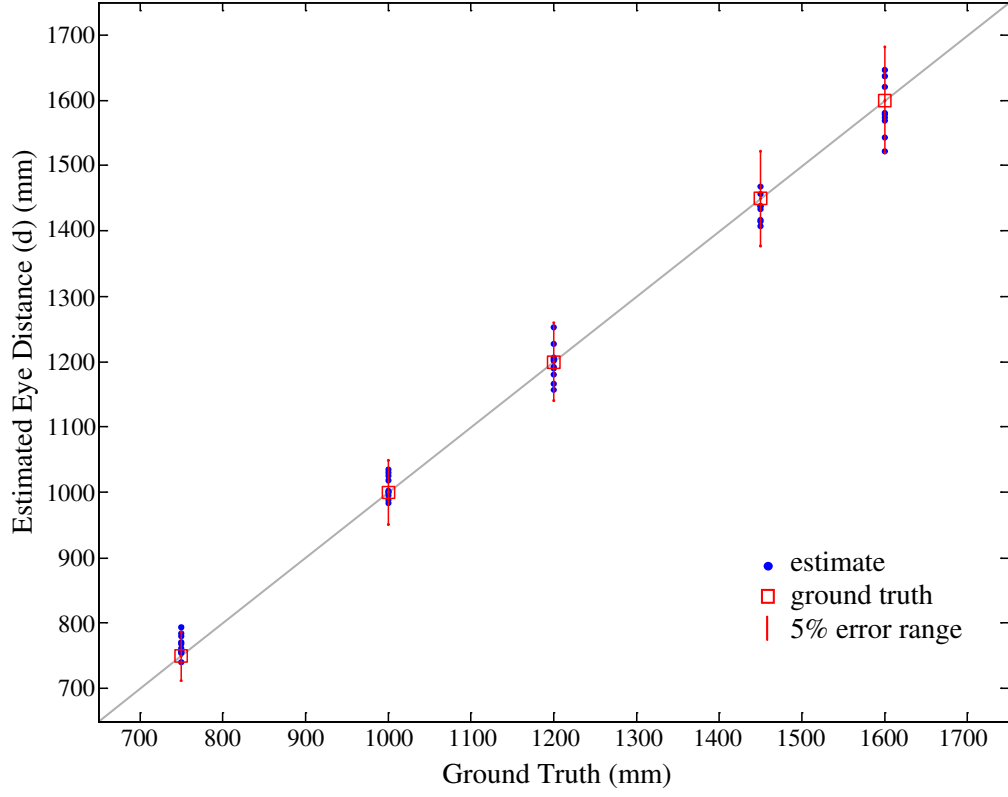
9

Figure 6: Estimation results of the distance to the eye (d of equation (4)). Images of eyes taken at 5 different distances and 10 different gaze directions for each distance were used. At each of 5 distances, 10 images were captured for different gaze directions. Regardless of the gaze direction, the depth d is estimated with good accuracy (always less than 5% error and RMS error 1.9%).

be determined from the major and minor radii of the detected ellipse:

$$\tau = \arccos \frac{r_{min}}{r_{max}} \ . \tag{5}$$

Note, however, that there is an inherent two-way ambiguity in the estimate of τ; for instance, an eye looking downward and upward by the same amount will produce the same ellipse in the image. Recently, Wang et al. [35] showed that anthropometric properties of the eye ball can be used to automatically break this ambiguity. In our experiments, we manually break the ambiguity.

The computed orientation of the cornea (the angles ϕ and τ) represents the direction of the optical axis of the eyeball. Despite the fact that the actual gaze direction is slightly offset[4], the

---

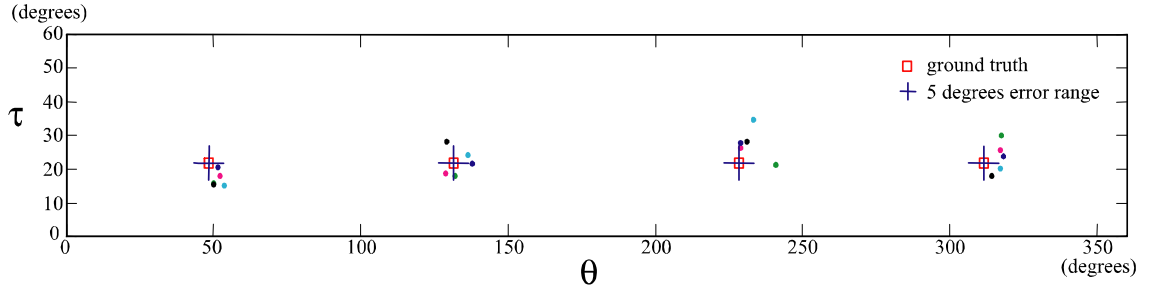[4]The actual gaze direction is slightly nasal and superior to the optical axis of the eyeball [20]. This means the

10

Figure 7: Estimation results of the orientation of the cornea (ϕ, τ). Images of 5 people looking at markers placed at 4 different known locations on a plane were used. In most cases, both angles are estimated within 5 degrees error and the RMS errors was $3.9°$ for ϕ and $4.5°$ for τ.

optical axis is commonly used as an approximation of the gaze direction. We will therefore refer to the computed 3D corneal orientation as the gaze direction.

To verify the accuracy of orientation estimation, we used images of 5 people looking at markers placed at 4 different known locations on a plane. Figure 7 shows the estimation results of the corneal orientation (ϕ, τ). Estimates of both ϕ and τ are mostly within the $5°$ error range and the RMS errors were $3.9°$ and $4.5°$, respectively. These errors are small given that the limbus itself is not a sharp discontinuity.

4 The Corneal Imaging System

In the previous section, we showed how to calibrate the corneal catadioptric imaging system. We are now in a position to investigate in detail the imaging characteristics of this system. These include the viewpoint locus, field of view, resolution and epipolar geometry (in the case of two eyes) of the system.

---

optical axis does not intersect the center of the fovea.

4.1 Viewpoint Locus

Previous work on catadioptric systems [1] has shown that only the class of conic mirrors can be used with a perspective camera to configure an imaging system with a single viewpoint. In each of these cases, the entrance pupil of the camera must be located at a specific position with respect to the mirror. Although the mirror (cornea) in our case is a conic (ellipsoid), the camera observing it is not rigidly attached to it. Therefore, in general, the corneal imaging system does not have a single viewpoint but rather a *locus of viewpoints*. Such non-single viewpoint systems have been explored in other contexts [6, 30].

Consider the imaging geometry shown in Figure 8. The reference frame is located at the apex of the cornea. Let the pupil of the camera be at $\mathbf{P}$. An incident ray $\mathbf{V}^i(t, \theta)$ can be related to the surface normal $\mathbf{N}(t, \theta)$ at the reflecting point $\mathbf{S}(t, \theta)$ and the reflected ray $\mathbf{V}^r(t, \theta)$ via the law of specular reflection

$$\mathbf{V}^i = \mathbf{V}^r - 2\mathbf{N}(\mathbf{N} \cdot \mathbf{V}^r) , \tag{6}$$

$$\mathbf{V}^r = \frac{\mathbf{P} - \mathbf{S}}{|\mathbf{P} - \mathbf{S}|} . \tag{7}$$

As seen in Figure 8, the viewpoint corresponding to each corneal surface point $\mathbf{S}(t, \theta)$ must lie along the incident ray $\mathbf{V}^i(t, \theta)$. Therefore, we can parametrize points on the viewpoint locus as

$$\mathbf{V}(t, \theta, r) = \mathbf{S}(t, \theta) + r\mathbf{V}^i(t, \theta) , \tag{8}$$

where r is the distance of each viewpoint from the corneal surface. In principle, the viewpoints can be assumed to lie anywhere along the incident rays (r can be arbitrary). However, a natural representation of the viewpoint locus is the *caustic* of the imaging system, to which all the incident rays are tangent [6]. In this case, the parameter r is constrained by the caustic and can be determined by solving [5]

$$\det J(\mathbf{V}(t, \theta, r)) = 0 , \tag{9}$$

where J is the Jacobian. Let the solution to the above equation be $r_c(t, \theta)$[5]. Then, the viewpoint

---

[5]The solution is very lengthy and there is no simple expression for it. We used Mathematica to solve equation (9).
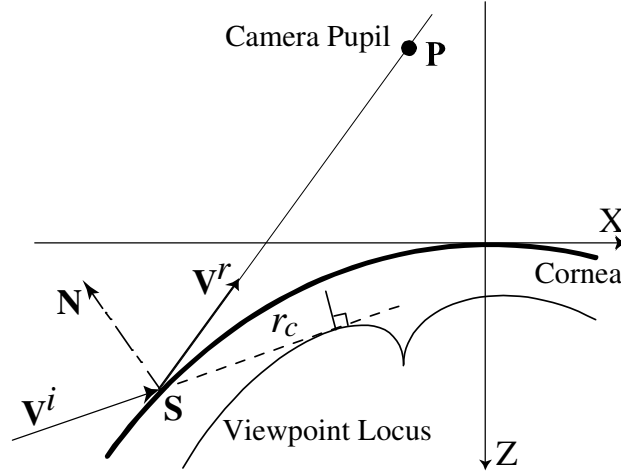
Figure 8: The locus of viewpoints of the corneal imaging system is the envelope of tangents produced by the incident rays that are reflected by the cornea into the pupil of the camera. One can see that the shape of this locus will depend on the relative orientation between the cornea and the camera.

locus is $\{\mathbf{V}(t, \theta, r_c(t, \theta)) : 0 \leq t \leq t_b, 0 \leq \theta < 2\pi\}$. Note that this locus completely describes the projection geometry of the corneal system; each point on the locus is a viewpoint and the corresponding incident ray is its viewing direction.

Figure 9 shows the viewpoint loci of the corneal imaging system for four different eye-camera configurations. As can be seen, each computed locus is a smooth surface with a cusp. Note that the position of the cusp depends on the eye-camera configuration. The cusp is an important attribute of the viewpoint locus as it can be used as the (approximate) effective viewpoint of the system [30]. Also note that the viewpoint locus is always inside the cornea; its size is always smaller than the cornea itself.

4.2   Field of View

Since the cornea is an ellipsoid (convex), the field of view of the corneal imaging system is bounded by the incident light rays that are reflected by the limbus (outer limit of the cornea).

A point on the limbus and its corresponding incident ray direction can be written as $\mathbf{S}(t_b, \theta)$ and
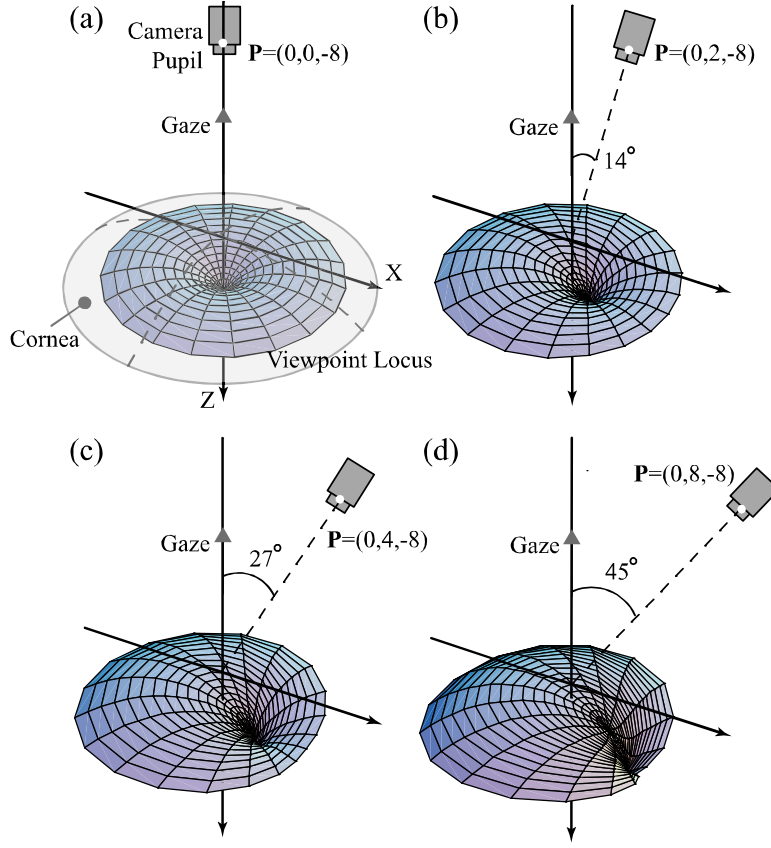
Figure 9: The viewpoint loci of the corneal imaging system for different relative orientations between the eye and the camera. The camera observes the eye from (a) $0°$ (camera pupil $\mathbf{P} = (0, 0, -8)$), (b) $14°$ ($\mathbf{P} = (0, 2, -8)$), (c) $27°$ ($\mathbf{P} = (0, 4, -8)$) and (d) $45°$ ($\mathbf{P} = (0, 8, -8)$). The viewpoint locus is always smaller than the cornea and has a cusp whose location depends on the relative orientation between the eye and the camera.

$\mathbf{V}^i(t_b, \theta)$, respectively, where t_b was defined in Section 2. Here, $\mathbf{V}^i(t_b, \theta)$ is a unit vector computed using equation (6). Let us define the field of view of the corneal system on a unit sphere. As we traverse the circular limbus, $\mathbf{V}^i(t_b, \theta)$ forms a closed loop on the unit sphere. Hence, the solid angle subtended by the corneal FOV can be computed as the area on the sphere bounded by this

14

closed loop:

$$FOV(\mathbf{P}) = \int_0^{2\pi} \int_0^{\arccos V_z^i(t_b,\theta)} sin\varphi d\varphi d\theta$$
$$= \int_0^{2\pi} (-V_z^i(t_b,\theta) + 1)d\theta. \tag{10}$$

Here, V_z^i is the Z coordinate of $\mathbf{V}^i$ and θ and φ are azimuth and polar angles defined in the coordinate frame of the unit sphere. From equation (6) we know that the above FOV depends on the camera position $\mathbf{P}$. For gaze directions that are extreme with respect to the camera, the camera may not see the entire extent of the cornea. We are neglecting such self-occlusions here as we have found that they occur only when the angle between the gaze direction and the camera exceeds 40° for a reasonable eye-camera distance. Also note that we are not taking into account other occlusions such as those due to other facial features including the eyelids and the nose.

Figure 10 depicts how the field of view varies with different eye-camera configurations. Notice how the field of view gradually decreases as the viewing angle of the camera increases (P_x increases). However, the field of view remains large for all viewing angles and for this particular setting ($P_z = 8$) it is always larger than a hemisphere ($2\pi \approx 6.28$). We found that for all the cases we encountered in the later examples, the corneal FOV was always larger or almost equal to a hemispherical field of view.

Figure 11 depicts the corneal FOVs for the four different eye-camera configurations as shaded regions of unit spheres. The monocular field of view of human perception is roughly 120° [37]. In Figure 11, the red contour on each sphere represents the boundary of this human FOV. It is interesting to note that, for all the eye-camera configurations shown, the corneal FOV is always greater than, and includes, the human FOV. That is, the corneal system generally produces an image that includes the visual information seen by the person to whom the eye belongs.

4.3 Resolution

We now study the spatial resolution of the corneal imaging system. As shown in Figure 12, we use the same corneal coordinate system as before with the camera pupil located at $\mathbf{P} = (P_x, 0, P_z)$. We can assume that the camera pupil lies on the XZ plane without loss of generality. However, the
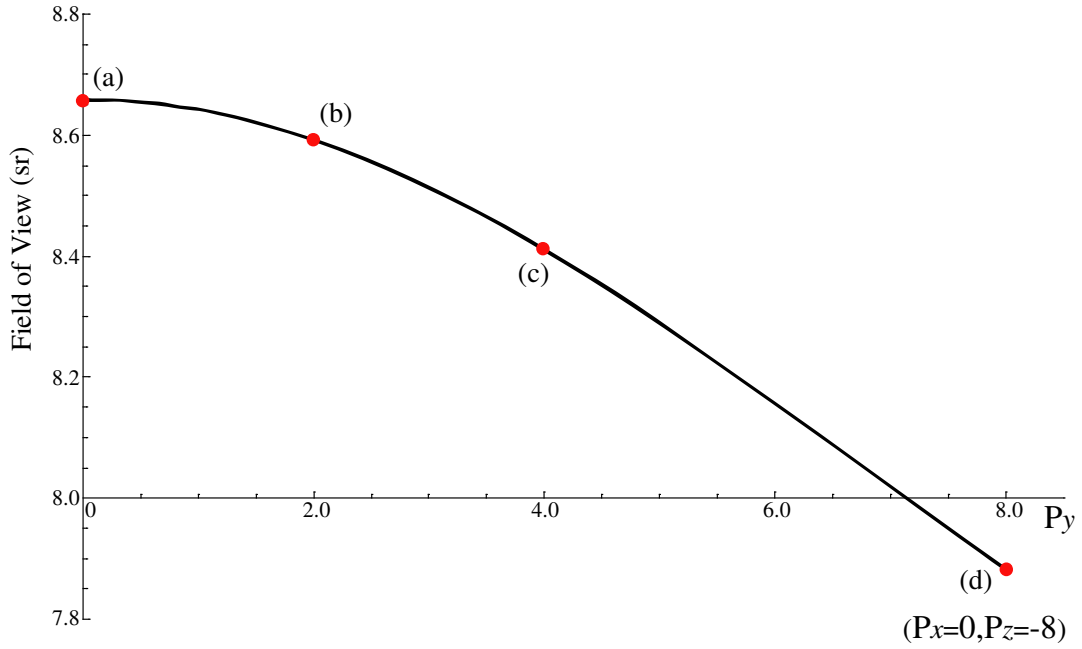
Figure 10: The field of view of the corneal imaging system as a function of the location of the camera pupil ($\mathbf{P} = (0, \mathbf{P}_y, -8)$). The red dots (a)–(d) correspond to the four configurations used in Figure 9. Note how the field of view gradually decreases as the viewing angle of the camera increases. The field of view of the corneal imaging system is generally large. In this example setting, it always exceeds a hemispherical field of view.

image plane has another degree of freedom; tilting around the Z axis as depicted in Figure 12. The angle between the pixel where the light ray hits and the center of projection $\alpha(t, \theta)$ as well as the angle between the pixel and the Z axis $\beta(t, \theta)$ can be computed for each corneal surface point.

Now, let us consider an infinitesimal area δA on the image plane. This small area is a projection of a scene area by the reflection on the cornea. The resolution of the imaging system at this particular pixel (the center of area δA) can be defined as the ratio of the infinitesimal area δA on the image plane to the solid angle $\delta \Omega$ subtended by the corresponding scene area δB from the viewpoint locus.

First, the area perpendicular to the chief ray that projects to surface area δA on the image plane
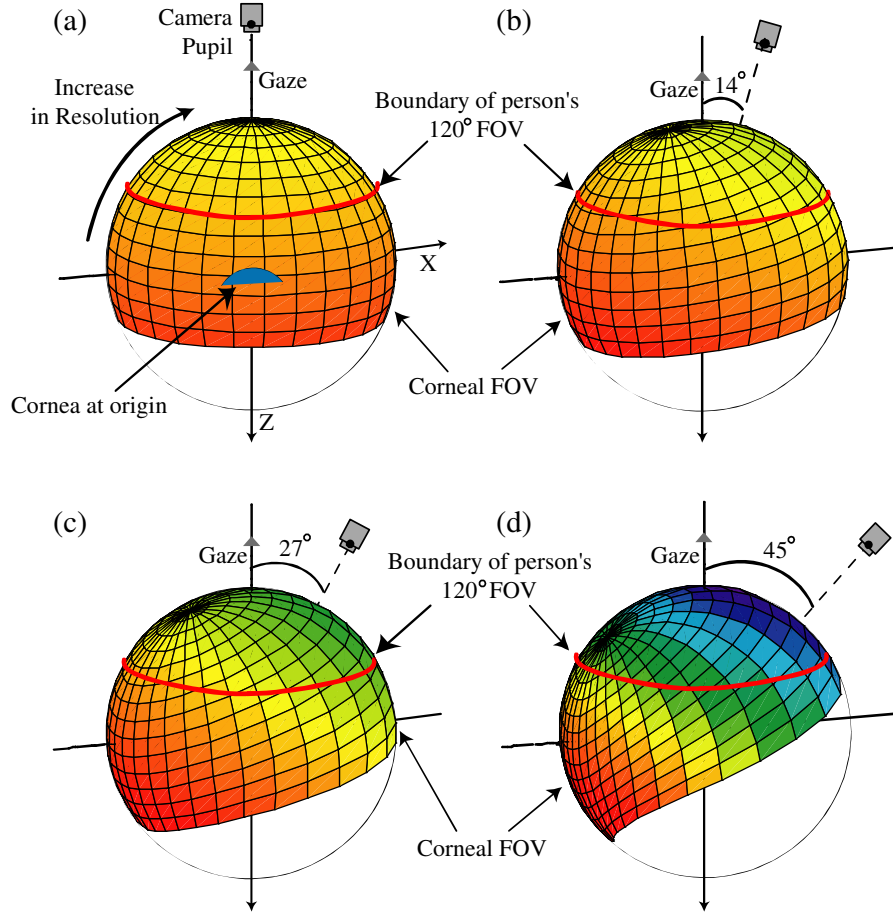
16

Figure 11: The field of view and the spatial resolution of the corneal imaging system for the four different eye-camera orientations ((a)–(d)) used in Figure 9. The shaded regions on the spheres represent the corneal FOV. The colors within the shaded regions represent the spatial resolution (resolution increases from red to blue). The red contour shows the boundary of the FOV of the human eye itself (120°). Note that the corneal FOV is always greater than this human FOV and the resolution is highest around the gaze direction and decreases towards the periphery.

is

$$\delta \tilde{A} = \delta A \cos \alpha(t, \theta) .$$

(11)

The solid angle $\delta \omega$ subtended by this area from the camera pupil is

$$\delta \omega = \left(\frac{f}{\cos \alpha} \right)^2 \delta \tilde{A} .$$
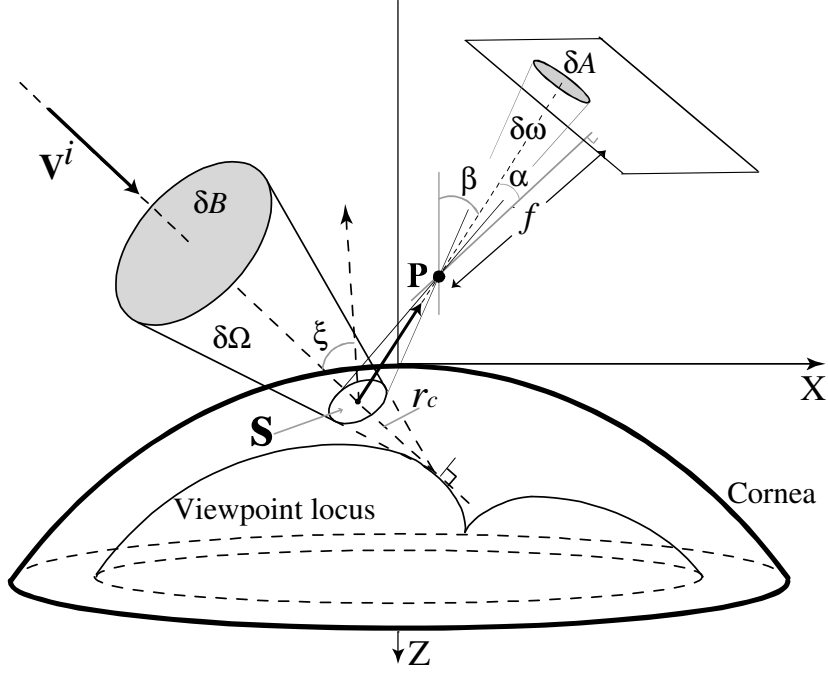
(12)

Figure 12: The resolution of the corneal imaging system can be defined as the ratio of the solid angles $\delta\omega$ and $\delta\Omega$. This resolution varies over the field of view of the corneal system.

Then, the surface area of the cornea that corresponds to this area can be computed as

$$\delta\tilde{B} = \delta\omega \left(\frac{S_z(t,\theta) - P_z}{\cos\beta} \right)^2 \cos\xi(t,\theta) . \tag{13}$$

where S_z and P_z are the Z coordinates of the corneal surface point and camera pupil. Finally, the solid angle $\delta\Omega$ from the viewpoint locus can be computed as

$$\delta\Omega = \frac{1}{r_c^2(t,\theta)} \frac{\delta\tilde{B}}{\cos\xi(t,\theta)} . \tag{14}$$

Therefore, the resolution we gain at a pixel whose corresponding incident light ray hits corneal surface point $\mathbf{S}(t,\theta)$ can be computed as

$$\frac{\delta A}{\delta\Omega} = \frac{(f\frac{\cos\beta(t,\theta)}{\cos\alpha(t,\theta)})^2 r_c^2(t,\theta)}{\cos\alpha(t,\theta)(S_z(t,\theta) - P_z)^2} . \tag{15}$$

Note equation (15) coincides the resolution equation derived in [30] when $\alpha(t,\theta) = \beta(t,\theta)$, i.e. when the optical axis of the camera and the cornea are aligned.
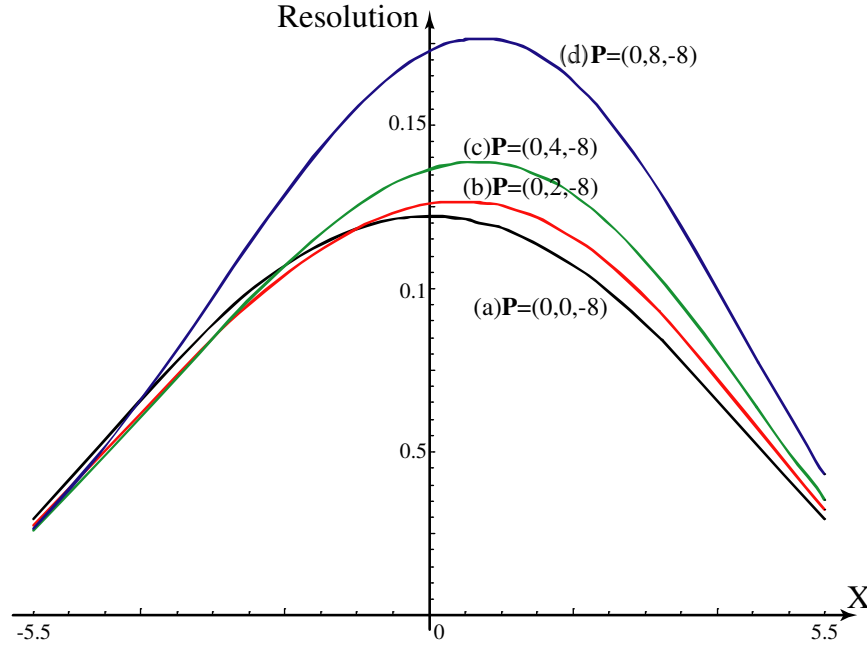
18

Figure 13: Spatial resolution characteristic of the corneal imaging system on the corneal surface. The resolution variation along the cross section of the cornea sliced by the XZ plane is depicted. The resolution is highest around the gaze direction and the maximum resolution increases as the viewing angle of the camera increases ((a) to (d)).

Figure 13 shows the spatial resolution characteristic of the corneal imaging system along the XZ cross section of the cornea. The resolution variations are depicted for the four different eye-camera configurations used in Figure 9. Note that the resolution is highest around the gaze direction (optical axis of the cornea which is the Z axis) and the maximum resolution increases as the viewing angle of the camera increases.

In Figure 11, the colors shown within each FOV represent the spatial resolution of the system. Here, resolution increases from red to blue. Notice how the resolution changes inside the red contour which corresponds to the human FOV. The highest resolution is always close to the gaze direction and decreases towards the periphery of the FOV. It is well known that the resolution of the human eye also falls quickly [6] towards the periphery [20]. The above result on resolution shows

[6]The spatial acuity of the retina decreases radially. This reduction is rapid up to $3°$ from the center and is then more gradual up to $30°$. Then, it once again decreases quickly to the periphery [20].
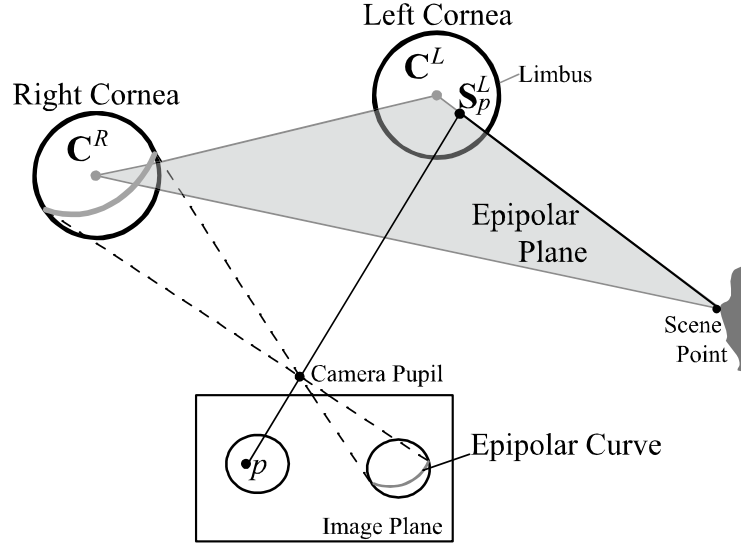
19

Figure 14: The epipolar geometry of the corneal stereo system. The epipolar plane intersects the eye at a curve which projects onto the image plane as a curve (the epipolar curve).

that, irrespective of the eye-camera configuration, the corneal system always produces an image that has roughly the same resolution variation as the image formed on the retina of the eye.

4.4   Epipolar Geometry of Eyes

When both eyes of a person are captured in an image, the combination of the two corneas and the camera can be viewed as a catadioptric stereo system [22, 23]. The methods we developed to calibrate a corneal imaging system can be used to determine the relative orientations between the two corneas and the camera. Here, we derive the epipolar geometry of a corneal stereo system [27].

Consider the corneal stereo system shown in Figure 14. Once we locate the two limbuses in the image (Section 3), we can compute the 3D coordinates of the cornea (center of the limbus) $\mathbf{C^R}$ and $\mathbf{C^L}$ and the orientation (θ^R, τ^R) and (θ^L, τ^L) for the right and left cornea, respectively. Now, let us denote the orientations of the major axes of the right and left limbuses in the image with unit vectors $\mathbf{A}^R = [\cos \phi^R \ \sin \phi^R \ 0]^T$ and $\mathbf{A}^L = [\cos \phi^L \ \sin \phi^L \ 0]^T$. Then the corneal surface points of the right cornea can be expressed in the camera's coordinate frame (origin at the camera pupil

20

and Z axis perpendicular to the image plane) as

$$\mathbf{S}^R(t,\theta) = \mathbf{T}(\mathbf{A}^R, \tau^R)(\mathbf{S}(t,\theta) - [0\ 0\ t_b]^T) + \mathbf{C}^R\ , \tag{16}$$

where $\mathbf{T}(\mathbf{A}, \tau)$ is the 3×3 matrix for rotation by τ around the axis $\mathbf{A}$, $t = t_b$ is the plane that the limbus lies on and the superscript T stands for transpose. $\mathbf{S}$ and $\mathbf{C}$ are column vectors. $\mathbf{S}^L(t,\theta)$ can also be derived with the same equation by substituting $\mathbf{A}^R, \tau^R$ and $\mathbf{C}^R$ accordingly. Then, the surface normals for each point on each cornea can also be computed as,

$$\mathbf{N}^R(t,\theta) = \mathbf{T}(\mathbf{A}^R, \tau^R)\mathbf{N}(t,\theta). \tag{17}$$

where, $\mathbf{N}(t,\theta)$ is the surface normal of $\mathbf{S}(t,\theta)$ computed in the corneal coordinate frame (Figure 3).

Now let us consider a 2D image point $\mathbf{p}^L$ which lies in the imaged left cornea:

$$\mathbf{p}^L = [p_x^L\ p_y^L\ -f]^T\ , \tag{18}$$

where f is the focal length. Here, the image coordinates (p_x, p_y) are normalized with respect to the width, height and the center of projection of the image. The surface point on the left cornea corresponding to this image point $\mathbf{S}_{\mathbf{p}}^L = \mathbf{S}^L(t_{\mathbf{p}^L}, \theta_{\mathbf{p}^L})$ can be computed by solving

$$\gamma\frac{-\mathbf{p}^L}{|\mathbf{p}^L|} = \mathbf{S}^L(t_{\mathbf{p}^L}, \theta_{\mathbf{p}^L}) \tag{19}$$

for $\gamma, t_{\mathbf{p}^L}$ and $\theta_{\mathbf{p}^L}$ which satisfies $0 \leq t_{\mathbf{p}^L} < t_b$ and the corresponding γ is the smallest positive of the solutions.

As we have seen in Section 4.1, the viewpoint locus is small compared to the distance between the two eyes. Therefore, as shown in Figure 14, for scenes that are not extremely close to the eyes, we can safely assume the viewpoint locus of each cornea to be a point located at the center of the limbus, i.e. $\mathbf{C}^R$ and $\mathbf{C}^L$. As a result, as shown in Figure 14, the epipolar plane corresponding to point $\mathbf{p}^L$ in the left cornea intersects the right cornea at a curve. The points on this 3D curve must satisfy the constraint

$$\left(\frac{\mathbf{C}^R - \mathbf{C}^L}{|\mathbf{C}^R - \mathbf{C}^L|} \times \frac{\mathbf{S}_{\mathbf{p}}^L - \mathbf{C}^L}{|\mathbf{S}_{\mathbf{p}}^L - \mathbf{C}^L|} \right) \cdot (\mathbf{S}^R(t,\theta) - \mathbf{C}^R) = 0\ . \tag{20}$$

21

The above equation can be solved to obtain the azimuth angles $\tilde{\theta}(t)$ that correspond to the 3D curve. We now have the 3D curve on the right cornea: $S^R(t, \tilde{\theta}(t))$. The projection of this 3D curve onto the image plane is a 2D curve

$$\left(\frac{fS_x^R(t, \tilde{\theta}(t))}{S_z^R(t, \tilde{\theta}(t))}, \frac{fS_y^R(t, \tilde{\theta}(t))}{S_z^R(t, \tilde{\theta}(t))} \right) , \tag{21}$$

where the subscripts denote the Cartesian coordinates of $\mathbf{S}$. For each point in the image of the left cornea, we have the corresponding epipolar curve within the image of the right cornea. These epipolar curves can be used to guide the process of finding correspondences between the corneal images and recover the structure of the environment.

5　The World from Eyes

We are now in a position to fully exploit the rich visual information captured by the corneal imaging system. We first show that a single image of an eye can be used to determine the environment of the person as well as what the person is looking at.

5.1　Where Are You?

Once we have calibrated the corneal imaging system as described in Section 3, we can trace each light ray that enters the camera pupil back to the scene via the corneal surface using equation (6). As a result, we are able to recover the world surrounding the person as an environment map. We will represent this map as a spherical panorama.

Figure 15 shows spherical panoramas computed from the images of eyes shown in Figure 1. These panoramas are cropped to include only the corneal FOV. As predicted in Section 4.2, the corneal FOV is large and provides us a wide-angle view of the surrounding world. This allows us to easily determine the location and circumstance of the person at the time at which their image was captured. For instance, we can clearly see that the person in the left example is in front of a building with wide stairs. We see that the person in the right example is facing another person.
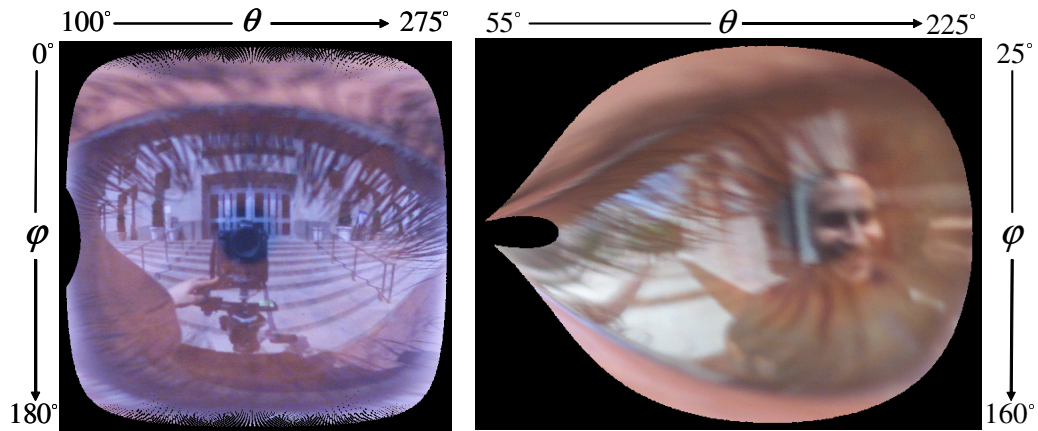
Figure 15: Spherical panoramas computed from the eye images shown in Figure 1. Each spherical panorama captures a wide-angle view of the world surrounding the person to whom the eye belongs. It reveals the location and circumstance of the person when the image was captured.

Since these panoramas have large fields of view, one can navigate through them using a viewer such as Apple's QuickTime VR^{TM}.

Note that the computed environment map includes the reflections from the iris as well. It is preferable to subtract this from the environment map for better visualization of the scene surrounding the person. However, the problem of separating reflections from the iris and the cornea is significant by itself and will be pursued in future work. Also, notice that the dynamic range and absolute resolution of the recovered environment maps are limited by those of the camera we use.

If we have multiple images of the eye taken from the same location[7] as shown in Figure 16(a), we can merge multiple spherical panoramas to obtain a even wider-angle view of the surrounding environment. However, since the estimates of corneal orientation include a few degrees of errors (see Section 3.3) simply overlaying spherical panoramas computed from different eye images can introduce misalignment artifacts. In order to precisely register these multiple spherical panoramas, we developed a simple registration algorithm.

The registration of multiple spherical panoramas can be accomplished by minimizing the dis-

---

[7]Note that, since 3D coordinates and orientation of the cornea are computed in the camera's coordinate frame, if the camera is moved, its location has to be known in order to register the computed spherical panoramas.

23

(a)   (b)

Figure 16: (a) 5 images of an eye with different extreme gaze directions taken from the same viewpoint. (b) Spherical panoramas computed from each eye image in (a).

tortion in the overlaid spherical panoramas. Since the pixel values of each spherical panorama are computed from the 3D coordinates and orientation of the cornea, this minimization should be accomplished by varying the three parameters: d and (ϕ, τ). The canonical way to measure the distortion would be to use colors at each point on the environment map. However, since the iris texture underlies each pixel value which we do not attempt to separate out in this work, direct use of pixel values would not be a robust metric of distortion. Instead, we manually specify corresponding points inside the corneal region in each eye image pair and minimize the angular distance between those points on the environment map. Although three corresponding points will provide a closed form solution, for robustness, we specify n (≥ 3) points and accomplish iterative least square minimization.

Figure 17 shows a spherical panorama computed by merging spherical panoramas computed from the 5 images in Figure 16(a). Figure 16(b) shows individual spherical panoramas computed from each eye image. Eyelids and eyelashes in the middle eye image were masked out before registration. Figure 17 clearly covers a larger field of view compared to the left spherical panorama in Figure 15, where only the middle image in Figure 16(a) is used. Notice that we can see windows located above the entrance of the building, the blue sky above the building, a black pole which was right next to the person's right eye, the green loan field on the extreme right side, etc. These provide more information to understand the surrounding world of the eye.
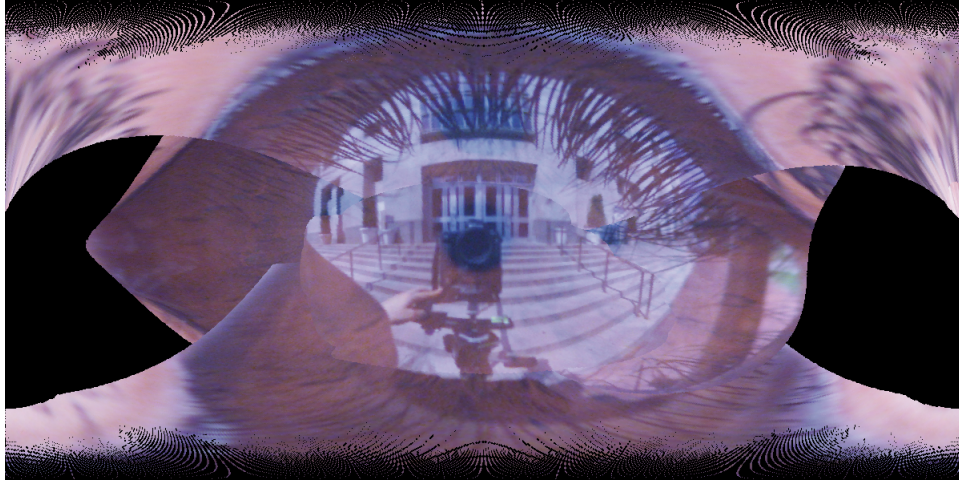
24

Figure 17: Registered and merged spherical panoramas computed from the eye images in Figure 16(a). A wider-angle view of the surrounding environment can be obtained compared to the spherical panorama computed from a single eye image (left of Figure 15).

5.2 What Are You Looking at?

Since we know the gaze direction, each eye image can be used to determine what the person is looking at from the computed environment map. The gaze direction tells us exactly which point in the environment map lies along the optical axis of the eye. A perspective projection of the region of the environment map around this point gives us an estimate of the image falling on the retina centered at the fovea. We will call this the foveated retinal image.

Figure 18 shows several spherical panoramas and foveated retinal images computed from images of eyes captured in various settings[8]. We show the foveated retinal images with a narrow field of view ($45°$) to better convey what the person is looking at. In scenario (I), we see that the person is in front of a computer monitor and is looking at the CNN logo on a webpage. The person in scenario (II) is meeting two people in front of a building and is looking at one of them who is smiling at him. The person in scenario (III) is playing pool and is aiming at the yellow ball. Note that although the eye-camera configurations in these three cases are very different, all the computed foveated retinal images have higher resolution around the center ($\approx$ gaze direction) as predicted by

---

[8]Note that these results are obtained from single images.

25

(a) Eye image (cropped)   (a) Eye image (cropped)   (a) Eye image (cropped)

(b) Spherical Panorama (cropped)   (b) Spherical Panorama (cropped)   (b) Spherical Panorama (cropped)

(c) Foveated retinal image (45° FOV)   (c) Foveated retinal image (45° FOV)   (c) Foveated retinal image (45° FOV)

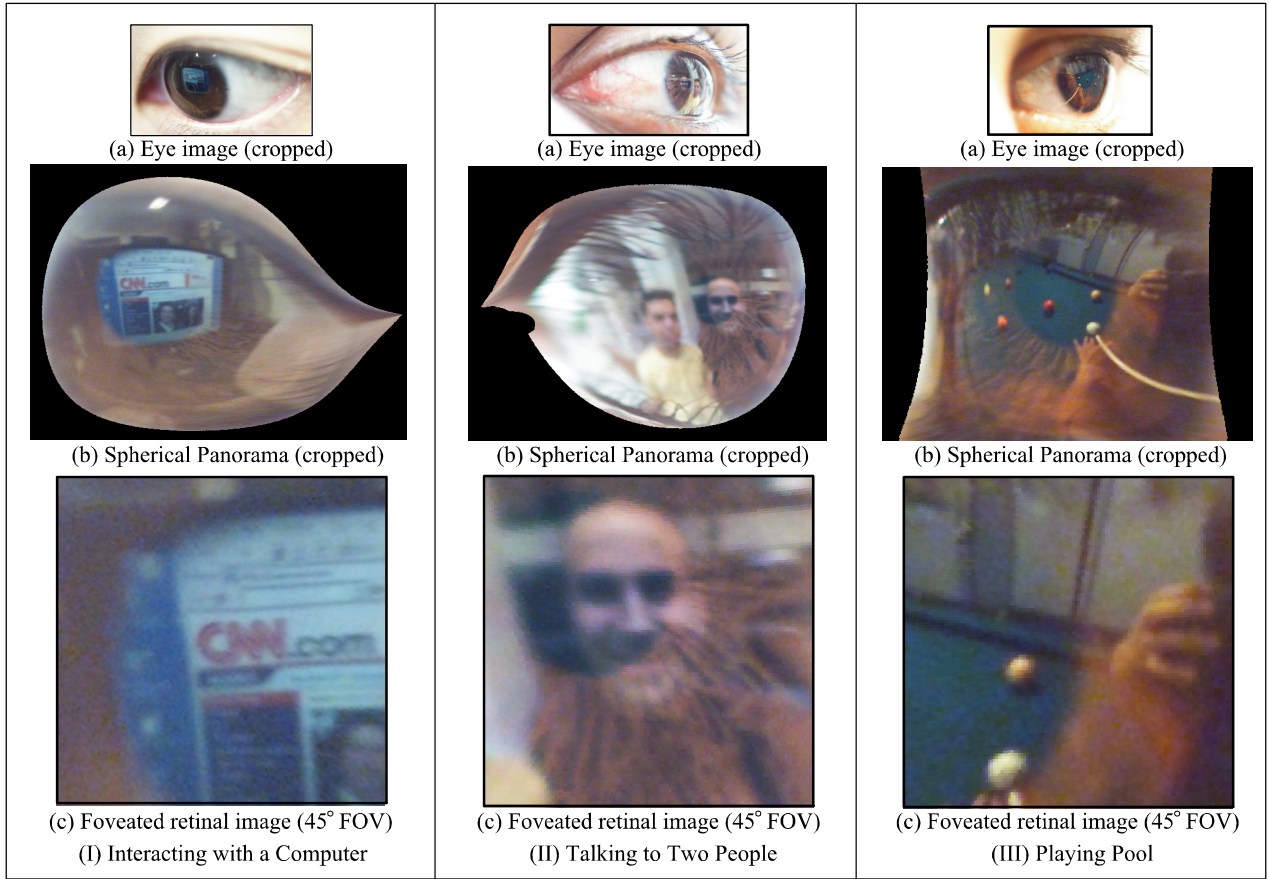(I) Interacting with a Computer   (II) Talking to Two People   (III) Playing Pool

Figure 18: Examples of spherical panoramas and foveated retinal images computed from images of eyes. Three different examples are shown. Each example includes (a) a cropped image of the eye, (b) a cropped image of the computed spherical panorama and (c) a foveated retinal image with a 45° field of view. The spherical panoramas and foveated retinal images clearly convey the world surrounding the person and what and where the person is looking. This information can be used to infer the person's circumstance and intent.

our analysis in Section 4.3. The above examples show that the computed foveated retinal images clearly convey where and what the person is looking at in their environment.

26

5.3 What is its Shape?

As described in Section 4.4, when both eyes of a person are captured in an image, we can recover the structure of the environment using the epipolar geometry of the corneal stereo system. Since our computed environment maps are inherently limited in resolution, one cannot expect to obtain detailed scene structure. However, one can indeed recover the structures of close objects that the person may be interacting with.

Figure 19(a) shows an image of the eyes of a person looking at a colored box. In Figure 19(b), the computed epipolar curves corresponding to four corners of the box in the right (with respect to the person in the image) cornea are shown on the left cornea. Note that the epipolar curves pass through the corresponding corner points. A fully automatic stereo correspondence algorithm must contend with the fact that the corneal image includes the texture of the iris. Removal of this iris texture from the captured image is a significant problem by itself and is beyond the scope of this paper. Here, we circumvented the problem by manually specifying the corresponding points on the computed epipolar curves. From these correspondences, we reconstructed a wire-frame of the box, as shown in Figure 19(c). This ability to recover the structure of what the person is looking at can be very useful in communicating with a machine such as a robot. For instance, the robot can be programmed by capturing a video of a person's eyes while he/she manipulates an object.

6 Implications

In this paper, we presented a comprehensive framework for extracting visual information from images of eyes. Using this framework, we can compute: (a) a wide-angle view of the environment of the person; (b) a foveated retinal image which reveals what the person is looking at; and (c) the structure of what the person is looking at. Our approach is a purely image-based one that is passive and non-invasive. We believe our framework has direct implications for the following fields.

Visual Recognition: From the computed environment map, one can easily determine the location of the person as well as infer the circumstance he/she was in when the image was captured. Such information can be of great value in surveillance and identification applications.

(a) Input Image

(b) Epipolar Curves

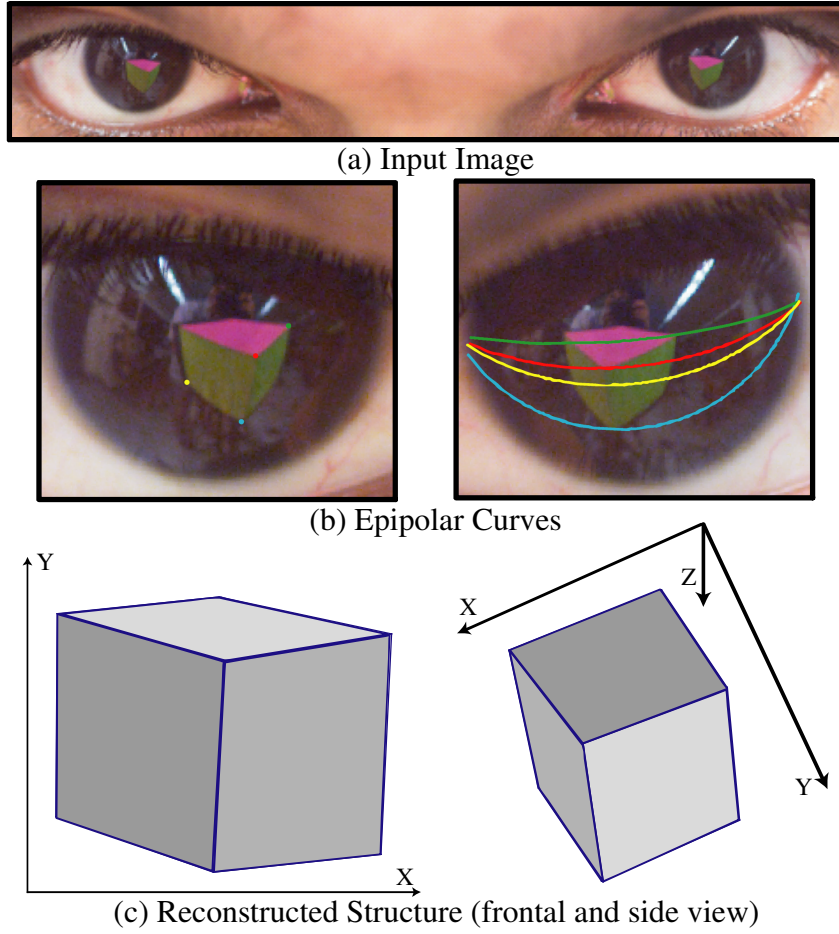(c) Reconstructed Structure (frontal and side view)

Figure 19: (a) An image of two eyes looking at a box. (b) The epipolar curves on the left cornea corresponding to four corners of the box in the right cornea. (c) A reconstructed wire-frame of the box.

Human-Machine Interfaces: Our results can extend the capability of the eye as a tool for human-machine interactions [4, 18, 16]. The eye can serve as more than just a pointer; it is an imaging system that conveys details about the person's intent. In the context of teaching robots [19, 17], the ability to reconstruct the structure of what the person is looking at is powerful.

Computer Graphics: The environment map recovered from an eye not only visualizes the scene surrounding the person but it also tells us the distribution of the lights illuminating the person [9]. Such illumination information can be very useful, for instance, to relight objects in the scene or the face of the person [21, 3, 10, 33, 24].

Human Affect Studies: By capturing the eyes as well as the facial/body expressions in an image, it is possible to record what the person is looking at and his/her reaction to it with perfect synchronization. Such data is valuable in human affect studies. Such studies give us important insights into human emotions [12, 13] and the way social networks work [32, 36].

7 Conclusion

In this paper, we have shown that eyes can serve as much more than just visual sensors for the person to whom the eye belongs. We showed that the combination of the cornea of an eye and the camera viewing the eye form a catadioptric imaging system. We can self-calibrate this corneal imaging system by locating the limbus in the image and computing the 3D coordinates and orientation of the cornea in the camera's coordinate frame. Once this self-calibration is done, we can compute a wide-angle view of the surrounding environment and an approximation of the projection on the retina of the person. Furthermore, we showed that from two eyes in an image, we can compute the structure of the object in the scene. In other words, we can exploit the visual information embedded in the appearance of eyes which convey rich information about the person's intent and circumstances. We believe these results have strong implications in various fields.

As future work, we plan to address the problem of separating iris and corneal reflections from the appearance of the cornea. Although this remains as a difficult problem due to the complex structure of the texture of the iris, we believe methods based on polarization [38, 39] would provide a robust solution.

References

[1] S. Baker and S.K. Nayar. A Theory of Single-Viewpoint Catadioptric Image Formation. *IJCV*, 35(2):1–22, Nov. 1999.

[2] T.Y. Baker. Ray tracing through non-spherical surfaces. *Proc. of The Royal Society of London*, 55:361–364, 1943.

[3] V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3D Faces. In *ACM SIG-GRAPH 99*, 1999.

[4] R.A. Bolt. Eyes at the Interface. In *ACM CHI*, pages 360–362, 1982.

[5] D.G. Burkhard and D.L. Shealy. Flux Density for Ray Propagation in Geometrical Optics. *JOSA*, 63(3):299–304, Mar. 1973.

[6] S. Cornbleet. *Microwave and Optical Ray Geometry*. John Wiley and Sons, 1984.

[7] J.G. Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. *IEEE TPAMI*, 15(11):1148–1161, Nov. 1993.

[8] H. Davson. *Physiology of the Eye*. Macmillan, 5th edition, 1990.

[9] P. Debevec. Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography. In *ACM SIGGRAPH 98*, pages 189–198, 1998.

[10] P. Debevec, T. Hawkins, C. Tchou, H-P. Duiker, and W. Sarokin. Acquiring the Reflectance Field of a Human Face. In *ACM SIGGRAPH 00*, pages 145–156, 2000.

[11] Y. Ebisawa. Improved Video-Based Eye-Gazed Detection Method. *IEEE Trans. on Instrumentation and Measurement*, 47(4):948–955, 1998.

[12] P. Ekman. Facial Expression of Emotion. *American Psychologist*, 48:384–392, 1993.

[13] P. Ekman and E. L. Rosenberg, editors. *What the Face Reveals*. Oxford University Press, New York, 1997.

[14] L. Flom and A. Safir. Iris Recognition System. US patent 4,641,349, 1987.

[15] M.A. Halstead, B.A. Barsky, S.A. Klein, and R.B. Mandell. Reconstructing Curved Surfaces from Specular Reflection Patterns Using Spline Surface Fitting of Normals. In *ACM SIGGRAPH 96*, pages 335–342, 1996.

[16] T.E. Hutchinson, K.P. White, K.C. Reichert, and L.A. Frey. Human-computer Interaction using Eye-gaze Input. *IEEE TSMC*, 19:1527–1533, Nov./Dec. 1989.

[17] K. Ikeuchi and T. Suehiro. Toward an Assembly Plan from Observation, Part 1: Task Recognition with Polyhedral Objects. *IEEE Trans. Robotics and Automation*, 10(3):368–385, Jun. 1994.

[18] R. Jacob. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *ACM CHI*, pages 11–18, 1990.

[19] S.B. Kang and K. Ikeuchi. Toward Automatic Robot Instruction from Perception –Mapping Human Grasps to Manipulator Grasps. *IEEE Trans. on Robotics and Automation*, 13(1), Feb. 1997.

[20] P.L. Kaufman and A. Alm, editors. *Adler's Physiology of the Eye: Clinical Application*. Mosby, 10th edition, 2003.

[21] S.R. Marschner and D.P. Greenberg. Inverse Lighting for Photography. In *IS&T/SID Color Imaging Conference*, pages 262–265, 1997.

[22] S.K. Nayar. Sphereo: Recovering depth using a single camera and two specular spheres. In *SPIE: Optics, Illumination and Image Sensing for Machine Vision II*, 1988.

[23] S.A. Nene and S.K. Nayar. Stereo Using Mirrors. In *IEEE ICCV 98*, pages 1087–1094, Jan. 1998.

[24] K. Nishino and S.K. Nayar. Eyes for Relighting. *ACM Trans. on Graphics (Proceedings of SIGGRAPH 2004)*, 23(3):704–711, 2004.

[25] K. Nishino and S.K. Nayar. The World in Eyes. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 444–451, 2004.

[26] T. Ohno, N. Mukawa, and A. Yoshikawa. FreeGaze: a gaze tracking system for everyday gaze interaction. In *Proc. of the Symposium on ETRA*, pages 125–132, 2002.

[27] T. Pajdla, T. Svoboda, and V. Hlaváč. Epipolar geometry of central panoramic cameras. In *Panoramic Vision : Sensors, Theory, and Applications*. Springer Verlag, 2000.

[28] C.P. Stein. Accurate Internal Camera Calibration Using Rotation, with Analysis of Sources of Errors. In *ICCV*, pages 230–236, Jun. 1995.

[29] R. Stiefelhagen, J. Yang, and A. Waibel. A Model-Based Gaze-Tracking System. *International Journal of Artificial Intelligence Tools*, 6(2):193–209, 1997.

[30] R. Swaminathan, M.D. Grossberg, and S.K. Nayar. Caustics of Catadioptric Cameras. In *IEEE ICCV 01*, volume II, pages 2–9, 2001.

[31] K-H. Tan, D.J. Kriegman, and N. Ahuja. Appearance-based Eye Gaze Estimation. In *WACV*, pages 191–195, 2002.

[32] S. S. Tomkins. *Affect, imagery, consciousness*. Springer, New York, 1962.

[33] N. Tsumura, M.N. Dang, T. Makino, and Y. Miyake. Estimating the Directions to Light Sources Using Images of Eye for Reconstructing 3D Human Face. In *Proc. of IS&T/SID's Eleventh Color Imaging Conference*, pages 77–81, 2003.

[34] H. von Helmholtz. *Physiologic Optics*, volume 1 and 2. Voss, Hamburg, Germany, third edition, 1909.

[35] J-G. Wang, E. Sung, and R. Venkateswarlu. Eye Gaze Estimation from a Single Image of One Eye. In *IEEE ICCV 03*, pages 136–143, 2003.

[36] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[37] G. Westheimer. *Medical Physiology*, volume 1, chapter 16 The eye, pages 481–503. The C.V. Mosby Company, 1980.

[38] L.B. Wolff. Polarization-based Material Classification from Specular Reflection. *IEEE TPAMI*, 12(11):1059–1071, 1990.

[39] L.B. Wolff and T.E. Boult. Constraining Object Features Using a Polarization Reflectance Model. *IEEE TPAMI*, 13(7):635–657, 1991.

[40] L-Q. Xu, D. Machin, and P. Sheppard. A Novel Approach to Real-time Non-intrusive Gaze Finding. In *BMVC*, pages 58–67, 1998.

[41] L.R. Young and D. Sheena. Survey of Eye Movement Recording Methods. *Behavior Research Methods and Instrumentation*, 7(5):397–429, 1975.