Study on Eye Gaze Estimation

Jian-Gang Wang and Eric Sung

***Abstract*—There are two components to the human visual line-of-sight: pose of human head and the orientation of the eye within their sockets. We have investigated these two aspects but will concentrate on the eye gaze estimation in this paper. We present a novel approach called the "one-circle" algorithm for measuring the eye gaze using a monocular image that zooms in on only one eye of a person. Observing that the iris contour is a circle, we estimate the normal direction of this iris circle, considered as the eye gaze, from its elliptical image. From basic projective geometry, an ellipse can be back-projected into space onto two circles of different orientations. However, by using a geometric constraint, namely, that the distance between the eyeball's center and the two eye corners should be equal to each other, the correct solution can be disambiguated. This allows us to obtain a higher resolution image of the iris with a zoom-in camera, thereby achieving higher accuracies in the estimation. A general approach that combines head pose determination with eye gaze estimation is also proposed. The searching of the eye gaze is guided by the head pose information. The robustness of our gaze determination approach was verified statistically by the extensive experiments on synthetic and real image data. The two key contributions in this paper are that we show the possibility of finding the unique eye gaze direction from a single image of one eye and that one can obtain better accuracy as a consequence of this.**

***Index Terms*—Circle/ellipse, distance constraint, eye gaze, head pose, human-machine interaction, iris contour, monocular vision, one-circle, point-of-regard.**

I. INTRODUCTION

TRADITIONAL human-computer interaction revolves around typing at a keyboard, moving and pointing with a mouse, or selecting from menus and searching through manuals. More natural ways of dealing with computer could be via talking, gesturing, hearing, and seeing. The estimation of head pose and eye gaze is important in applications such as virtual reality, video conferencing, and special human-machine interface/controls. Interacting with the computer through the head pose and eye gaze is natural for human-machine interaction since this involves passive sensing the human gestures of the face that can be recognized without any discomfort for the user. We can imagine two general uses of eye gaze tracking in man-machine interfaces [1]. First, one could collect data of the eye movements in order to get online measurement of the user's focus of attention. Second, one could use specific eye movements as an input device for the interface.

There are two components to the line-of-sight: pose of human head and the orientation of the eyes within their sockets (eye-gaze). We have investigated these two aspects. We found that the domain knowledge of the human face is important and essential for determining the head pose and eye gaze utilizing only minimal robust features and under real-time requirement. In our work, the domain knowledge used is not merely from facial features but from more anatomical properties. For instance, we found that the eye gaze can be estimated using the normal to the iris contour, which has an approximate fixed angle with the true gaze. Hence, we have developed a novel approach called the "one-circle" algorithm for measuring eye gaze using monocular image that zooms in on only one eye of a person. In addition, we make an observation that the eye lines (connecting the two far eye corners and the two neighboring eye corners, respectively) are parallel to the mouth line (connecting the two mouth corners). This domain knowledge led us to develop a new method for determining head pose fast and robustly using the vanishing point in the image formed by the eye lines and mouth line. The details of this pose determination paradigm can be found in [37]. Two alternative proofs of this approach can be found in [31] and [34]. In this paper, we will "focus" on the eye gaze determination.

A good survey of eye-gaze (eye-movement) tracking techniques (most of them are active methods) is provided in [40]. The techniques include electro-oculography [17], limbus, pupil and eyelid tracking [3], [4], [8], [15], [16], [21], [30], contact lens method, corneal and pupil reflection relationship [8], [15], [16], Purkinje image tracking [5], artificial neural networks [30], morphable models [25], and other methods [13], [19], [22].

In most existing approaches [6], [7], [18], [19], [39], the iris contours on the image plane are simplified to be circles; therefore, the felicitous circular geometry is utilized, and iris outer boundaries (limbus) are detected using a circle edge operator. For instance, the center of the iris is detected using the circular Hough transform in [19]. In [18], the iris is located by matching the left and right curvatures of the iris (circle) candidate with those of the iris to be detected in the edge image.

Zelinsky *et al.* [19], [22] presented an eye gaze estimation in which the eye corners are located using a stereo vision system. Then, the eyeball position can be calculated from the pose of the head and a three-dimensional (3-D) "offset vector" from the midpoint of the corners of an eye to the center of the eye. Consequently, the radius of the eyeball can be obtained. However the "offset vector," in addition to the radius of the iris, needs to be manually adjusted through a training sequence where the gaze point of the person is known.

Manuscript received August 22, 2000; revised August 20, 2001. This work was supported by a Ph.D. scholarship from Nanyang Technological University. This paper was recommended by Associate Editor L. O. Hall.

J.-G. Wang is with the Centre for Signal Processing, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: ejgwang@ntu.edu.sg).

E. Sung is with the Division of Control and Instrumentation, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: eericsung@ntu.edu.sg).

Publisher Item Identifier S 1083-4419(02)03006-6.

It is difficult to determine the eye gaze by analyzing the eyeball rotations from a typical image with low resolution for eye region [13], [32]. The iris is partially occluded by the upper and lower eyelids; therefore, it will be difficult to fit its contour consistently and reliably. For example, in [18], the field of the view of the camera is set to capture the whole face in the image, the width of an eye is only about 30 (pixels), and the radius of the iris in the image plane is only 5 (pixels) in a typical situation. Therefore, it is hard to determine the gaze in a 3-D scene by using the iris information in such a typical image.

In this paper, we observe that the iris contour (not the iris) is a circle. The gaze, which is defined as the normal to this iris circle, can be estimated from the ellipse/circle correspondence. However, it will result in two possible solutions of the normal. We propose a "one-circle" algorithm to disambiguate the solutions. The unique supporting plane was obtained based on a geometric constraint, namely, that the distance between the eyeball's center and the two eye corners should be equal to each other. We will refer to this as the "distance constraint."

We have previously proposed a method called the "two-circle" algorithm [34], [36] to determine the gaze based on the two iris contours. There, we apply the principle that from the image of two space circles lying on parallel planes, the resulting two ellipses can be used to deduce the unique normal to these planes. Each ellipse gives rise to two possible space circles. It has been proven that one of the normals will be common to each of the two sets of solutions, whereas the other two remaining are spurious. We thus can disambiguate the gaze results based on the fact that the left and right iris boundaries are reasonably parallel in three dimensions. In practice, however, the two "correct" normals may not be exactly equal each other due to errors and noise. Hence, we disambiguate by treating the two normals from each set that are closest to each other as the correct match. The difference of the normal to the supporting plane of the two irises should be minimal, irrespective of eyeball rotations and head movement, and we refer to it as the "normal constraint."

We extend our investigation by proposing the "one-circle" algorithm to be discussed here, which aims to relax the limitation that two irises are required to disambiguate the normal solutions. Consequently, the field of view of the camera can be narrowed further by focusing on only one eye. With the improvement of the iris resolution, higher precision and robustness can be expected and has been achieved. The robustness of this approach was statistically verified by extensive experiments on synthetic and real image data.

Domain knowledge played a key role in the success of the one-iris approach. Although the eyeball center cannot be seen, its location can be inferred from the pose of the head. This is because its average 3-D location relative to the observed features is very close to a generic constant and can be fixed during model acquisition [22]. In this paper, the ratio of the radius of a person's iris and the radius of his/her eyeball in 3-D space is found to possess very low ensemble variance, and consequently, we can fix the ratio as the generic average. The small variation from person to person of this ratio thus has no significant effect on the results. Hence, the eyeball center can be located once the radius of the iris has been calibrated.

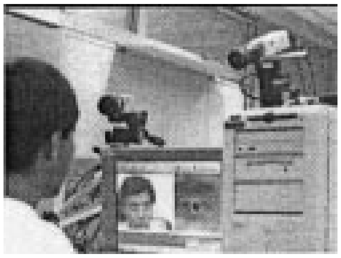


Fig. 1. System configuration.

In summary, our method differs from others in the following respects. We treat the image of the iris contour not as a circle but, correctly, as an ellipse. Hence, our approach is more realistic than the existing approaches. The other difference is that our method is more accurate because it can zoom in onto one eye, thereby allowing a larger iris image for detection. However, in doing so, we need to address two issues that emerge. Zooming and tracking a single eye poses a problem. The "gaze" camera needs to locate the eye. The necessary information can be obtained from the head pose estimation subsystem (which has been accomplished and reported in [37]). Second, the estimated eye gaze direction is only with respect to the head. Thus, the relative eye gaze direction must be combined with the head pose direction to yield the absolute eye gaze direction.

In all, this means that for a viable one-iris paradigm, the two subsystems—the head pose and the eye gaze estimators—must be integrated. In order to obtain the higher resolution image of the iris, a zoom-in "gaze" camera is used. It provides sufficient resolution to measure accurately the rotation of eyeball. A general approach that combines head pose determination with eye gaze estimation is proposed. The system configuration can be seen in Fig. 1. The pose camera (left) is mounted on a fixed tripod, while the gaze camera (right) is mounted on a computer-controlled pan-tilt unit. The problem of having possible out-of-field views can be settled by guiding the gaze camera by the head pose estimation results. The pose of the human head, including the 3-D locations of the eye corners, mouth corners, and the orientation of the face, can be obtained from a second "pose" camera. The 3-D coordinates of the eye corners (respect to the pose camera) informs the gaze camera system on how to focus on the eye region; it could also be used to calculate the distances between the eyeball's center and the eye corners (serve the "distant constraint").

The integration of the eye gaze subsystem with the head pose estimation module together offers great potential especially in the applications mentioned earlier. It is important to note that our method is nonintrusive, fast, and robust. It is robust because the iris contour is one of the simplest and most robust facial features to be extracted.

The approach to estimate the eye gaze is discussed in Section II, and the integration of head pose and eye gaze is discussed in Section III. The mathematical details behind the integration are given in Appendixes A and B. Experimental results on simulated data as well as on real images are given in Section IV. The limitations of the method and conclusion are discussed in Sections V and VI, respectively.

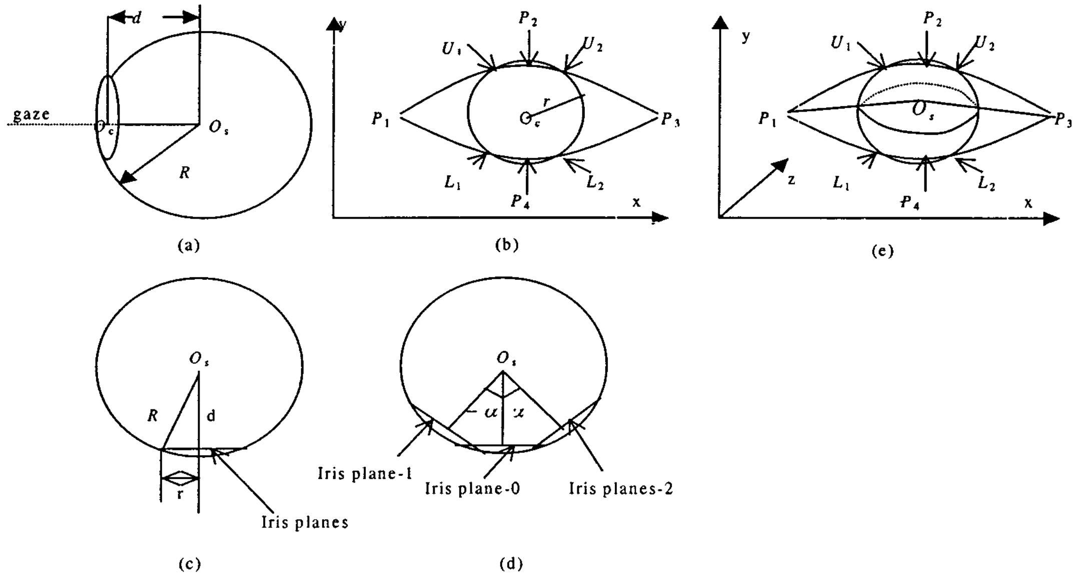


Fig. 2. Eye model. (a) Definition of the eye gaze. (b) Fitting edges (U_1L_1 and U_2L_2) are defined by the eyelids ($P_1P_2P_3$ and $P_1P_4P_3$) and the iris contour (centered O_c with radius r). (c) Relationship among R, r, and d, where O_s is the center of the eyeball. (d) Changing of the eye gaze. (e) "Distance constraint."

II. Gaze Positioning

Gaze can be derived from image iris contour (ellipse) and known model of the iris (being a circle having known radius). In order to determine the gaze, we introduce a simplified eye model and calibrate the eye dimension such as the radius of the iris. With the contour of the iris on the image, it will be shown that the prior knowledge of the eye model and the equation of the ellipse lead to the unique solution of the eye gaze.

A. Eye Model

A simple eye model is defined (see Fig. 2). The eyeball is assumed to be a sphere with radius R. Actually, it is not quite a sphere, but this discrepancy does not affect our methodology. The iris is located at the front of eyeball, and its contour is modeled as a circular ring of radius r [see Fig. 2(a) and (b)]. The distance from the center of the eyeball to the iris plane is d. The relation among R, r, and d [see Fig. 2(c)] is

$$R^2 = r^2 + d^2. \tag{1}$$

The optical axis of the eye is the line passing through the center of the eyeball and the center of the iris; it is defined as the *eye gaze*. By changing the eye gaze, the eyeball rotates around its center [see Fig. 2(d)]. The eye gaze we defined keeps a nearly fixed angle k (k is the angle between the visual and the anatomical axis of the eye) with the central gaze vector that is determined by the eye lens. It is only a matter of calibration relationship between the gaze and the central gaze vector.

The radius of the iris is very close to an anatomical constant (around 7 mm) [22], and the radius of the eyeball ranges from 12 to 13 mm according to the anthropometric data [18]. Hence, the ratio of the radius of a person's iris and the radius of his/her eyeball in 3-D space is stable with a small standard derivation. In our eye model, the average ratio of the radius of a person's iris and the radius of his/her eyeball in 3-D space, i.e., R/r, is assumed to be a generic constant. This assumption is justified because of the very small deviations over the ensemble. Hence, calibration of r (anatomical constant assumed in [22]) is required in our algorithm. Once r has been calibrated, both R and d can consequently be obtained. The computation cost involved in our calibration is lighter than the method presented in [19]. In [19], both the radius of the iris and radius of the eyeball are determined by the manual adjustment through a training sequence where the gaze point of a person is known. We will discuss the sensitivity of our gaze determination algorithm to the ratio in Section II-D.

The upper and lower eyelids are modeled as two parabolas [in Fig. 2(b)]. The upper eyelid passes through points $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, and $P_3(x_3, y_3)$, and the lower eyelid passes through points $P_1(x_1, y_1)$, $P_4(x_4, y_4)$, and $P_3(x_3, y_3)$. The equation of an eyelid is of the form

$$y = a(x - b)^2 + c \tag{2}$$

which for the upper eyelid yields

$$a = \frac{-y_2}{(x_1 - x_2)^2},\ b = x_2,\ c = y_2 \tag{3}$$

whereas for the lower eyelid yields

$$a = \frac{-y_4}{(x_1 - x_4)^2},\ b = x_4,\ c = y_4. \tag{4}$$

The upper and lower eyelids in real face images occlude parts of the iris contours. Only the unoccluded iris edges can be used to fit the iris contour in the image plane. Consistently, the synthetic iris contour edges that lie between the upper and lower eyelids are located and used to fit the elliptical contour in our simulations. For instance, curves U_1L_1 and U_2L_2, which are shown in Fig. 2(b), are the fitting edges we want. These fitting edges can be located by using the equations of iris and eyelids. We will investigate the performances of our method using the eye model of dimensions close to that for human in the following section.

B. "One-Circle" Algorithm

From the observed perspective projection of a circle having known radius, it is possible to infer analytically the supporting plane, on which the circle lies, as well as where the center of the circle lies. The problem has been extensively investigated, and there are many papers concentrating on 3-D location of circular objects [4], [12], [29], [34], [35]. We adopted the monocular camera-positioning algorithm proposed in [29], [34], and [35]. It was noted that two solutions of the 3-D position of the iris plane will be obtained from circle/ellipse correspondent, and we disambiguate the solution using "one-circle" algorithm, which will be discussed later. The origin of the camera coordinate system is at the lens center and the Z-axis coincides the optical axis of the camera. The Y-axis is vertical, and the X-axis is horizontal while keeping a right-handed system.

The distance between the two corners of an eye and the center of the eyeball should be equal to each other [see Fig. 2(e)]

$$O_sP_1 = O_sP_3. \tag{5}$$

Consider an iris contour $\mathbf{Q}$. The two solutions of the normal of $\mathbf{Q}$ are $\mathbf{n}_1 = (\cos\alpha_1, \cos\beta_1, \cos\gamma_1)^T$, $\mathbf{n}_2 = (\cos\alpha_2, \cos\beta_2, \cos\gamma_2)^T$, and the corresponding solutions of the center of the iris contour are $O_{c1}(x_{01}, y_{01}, z_{01})$ and $O_{c2}(x_{02}, y_{02}, z_{02})$, respectively. Using the eye model defined in Section II-A, the center of the eyeball O_{si} can be calculated as

$$\begin{aligned} x_{si} =& x_{0i} + d\cos\alpha_i, \quad y_{si} = y_{0i} + d\cos\beta_i \\ z_{si} =& z_{0i} + d\cos\gamma_i \end{aligned} \tag{6}$$

where $i = 1, 2$, d is the distance from center of the eyeball to the iris plane [see Fig. 2(a) or 2(c)]

$$d = \sqrt{R^2 - r^2} = r\sqrt{c^2 - 1} \tag{7}$$

where c is a constant.

After that, the solutions of the two eye corners are projected from the pose camera coordinate system to the gaze camera coordinate system. The distances between the center of the eyeball and the two eye corners are compared. Due to the image noise, the unique solution of the iris plane should be the one that satisfies

$$O_sP_1 \approx O_sP_3. \tag{8}$$

In our algorithm, we calculate $O_{s1}P_1$, $O_{s1}P_3$, $O_{s2}P_1$, and $O_{s2}P_3$. If

$$|O_{s1}P_1 - O_{s1}P_3| \leq |O_{s2}P_1 - O_{s2}P_3| \tag{9}$$

then $(\mathbf{n}_1, O_{c1})$ is the solution we want; else, $(\mathbf{n}_2, O_{c2})$ is the solution.

C. Degenerate Cases of "One-Circle" Algorithm

The "one-circle" algorithm degenerates when the following condition is satisfied: The iris contour is symmetrical about the Y-Z plane of the camera, where the Y- and Z- axis is the vertical and the optical axis of the camera, respectively.

Actually, this condition corresponds to the case that user is facing front, whereby the iris is symmetrical about the optical axis. Thus, it is impossible to distinguish, from the ellipse, whether the person is looking upwards or downwards. Fortunately, this degenerate case can be prevented by comparing the Y-position (vertical) of the iris centers with the one of the eye corners. If the Y coordinate of the iris center is greater, then the person is looking upwards; else, the person is looking downwards.

D. Sensitivity of the "Distance Constraint" to the Ratio

In our method, the average ratio of the radius of a person's iris and the radius of his/her eyeball in 3-D space is assumed to be a generic constant (refer to the eye model presented in Section II-A). The observed iris radius is calibrated, and the radius of the eyeball is calculated using the ratio. The center of the eyeball is located along the line-of-sight. The angle between the two normal solutions is large if the head pose is not close to the frontal view (degenerate case). Consequently, the separation of the two resulting eyeball centers is large enough to disambiguate the solutions based on the "distance constraint." In our experiments, we found that the algorithm based on the "distance constraint" is robust to the ratio. The same unique solution can be obtained for different gazes, even when the ratio is varied $\pm 50\%$.

E. Iris Detection

The most prominent and reliable features within the eye region are edges of the iris. Because we wish to model iris contours on the image plane as an ellipse, obviously, the existing iris detection methods using circular edge operators cannot be used. Instead, we detect the iris edge (bright-to-dark and dark-to-bright step edge) using a (3 × 3) vertical edge operator, which detects and emphasizes vertical edges and a (3 × 3) morphological "open" operation. Since the 3-D positions of the corners of the eyes are already known in our pose determination [37], the locations of the eye corners in the gaze image are known. Hence, the iris detection can be executed on a small region between them. Because of the high contrast between the eyeball and the eye white, eye image is easily segmented based on a threshold that was automatically selected in the histogram [23] of the eye region. It is known that opening an image breaks narrow isthmuses [14]. Hence, the morphological "open" operation is applied to separate the iris from the eyelid. In our experiments, we found that just one "open" operation is needed. After that, the (gradient) edge operator "Canny" is used to detect the edges of the iris, and the edges with direction $90° \pm 5°$ were retained. All of the vertical edge segments are then tracked, respectively, using "edge following" technique. The lengths of the edges are obtained. The two sides of the iris are the two longest vertical edges.

Once the iris edges are obtained reliably, iris contour is fitted to an ellipse [2]. An example of the iris detection is shown in Fig. 3. An original eye image is shown in Fig. 3(a). In the segmented eye image, some of the eyelids are connected with the iris [see Fig. 3(b)]. Applying an "open" operation cuts the connection (narrow isthmuses); see Fig. 3(c). A vertical edge operator is applied to the Fig. 3(c), resulting in Fig. 3(d). The two longest edges are deemed the iris edges for fitting the iris contour [see Fig. 3(e) or 3(f)]. Actually, the spurious edges, e.g., the edges in the horizontal direction in Fig. 3(d), are not connected

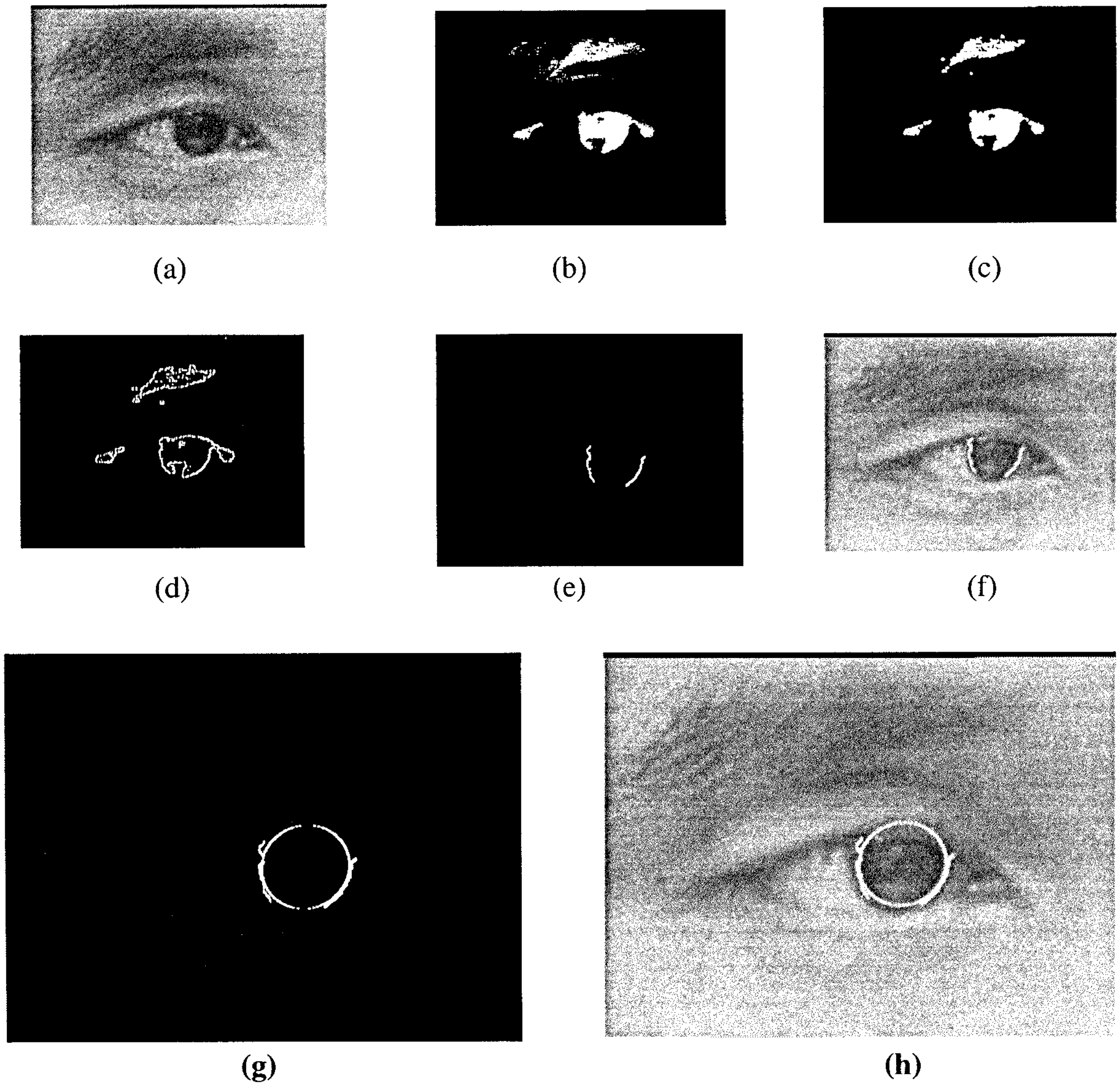


Fig. 3. Iris detection. (a) Original image (lens 55 mm). (b) Thresholding results. (c) Morphological "open" operation. (d) Vertical edges. (e) Two longest edges obtained by applying the edge-following technique. (f) Overlay of the edges onto the original image. (g) Edges and the fitted ellipse. (h) Overlay of the edges and the fitted ellipse onto the original image.

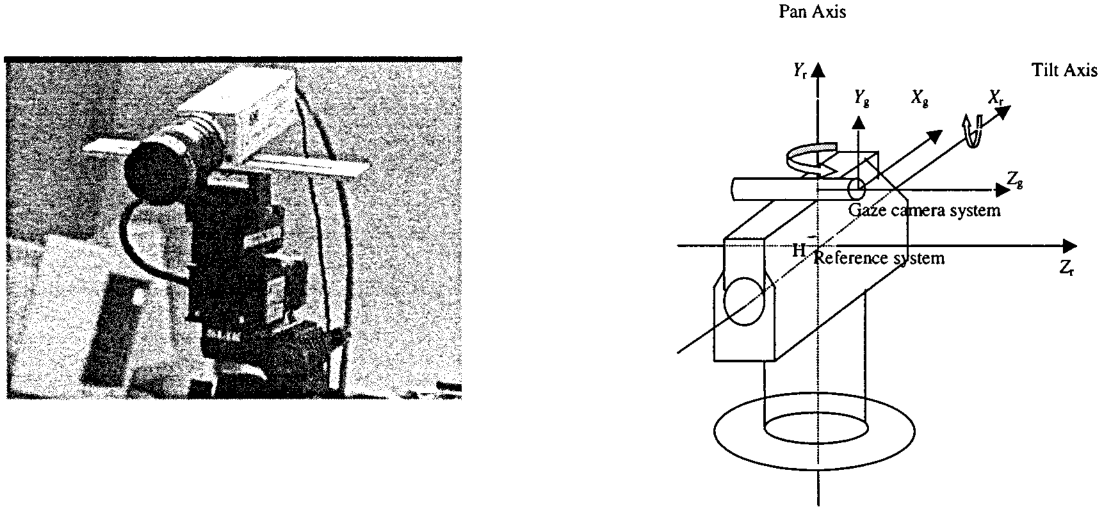
to each other, although it looks that way. Hence, the two sides of the iris [see Fig. 3(e)] can be found from Fig. 3(d). The fitted ellipse can be seen in Fig. 3(g) or 3(h).

III. Integration of Pose and Gaze

A zoom-in camera is used in our gaze determination approach. The drawback of this approach is the problem caused by the camera's narrow field of view and that the iris is easily out of the field of view. Hence, a general approach that combines head pose determination with gaze estimation is proposed. The gaze camera can focus on the iris region provided by the pose camera. On the other hand, the eye corners need to be transformed to the gaze camera coordinate system in order to disambiguate the gaze solutions using the "distance constraint." This creates an additional problem of having to relate the camera for estimating the head pose to the gaze camera.

In Appendixes A and B, we will explain the mathematical detail behind the integration of head pose and eye gaze. The pan-tilt unit[1] (PTU-46-17.5) used in our system [see Fig. 4(a)] has a special feature that the axis of the pan and the axis of the tilt always meet at one point in space. This will simplify our integration computation. A reference coordinate system (H, X_r, Y_r, Z_r) is defined [see Fig. 4(b)] in the pan-tilt unit:

[1] The unit is quite rigid and has very low backlash. It uses precision brass and stainless worms and precision stepper motors and drivers. The resolution of the unit is 3.086 arc minute (0.0514°). One of the features of the unit is the precise control of position, speed, and acceleration. The accuracy tests on the PTU-46-17.5 was run unloaded, and it can be found that the worst case error was no more than 40% of resolution; that is, accuracy was $0.012\,857^\circ \pm 0.005\,142\,9^\circ$. When moving from the same direction and speed, it can be found that accuracy equaled resolution. Hence, its affects on the gaze estimation were very small. Owing to our use of stepper motors, worm gears, and noncontact precision optical sensors, the unit has extremely high repeatability, except for the case of overload (maximum rated payload is over 4 lb). Any very small positional error would be noncumulative, owing to very small mechanical compliance in the linkage.



(a) (b)

Fig. 4. (a) PTU-46-17.5 pan-tilt unit. (b) Reference coordinate system is defined in the pan-tilt unit.

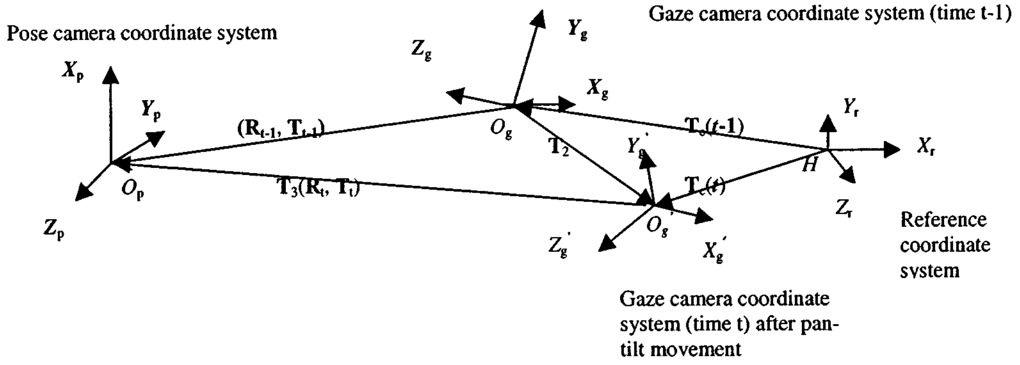

Pose camera coordinate system
X_p
Y_p
O_p
Z_p
$(\mathbf{R}_{t-1}, \mathbf{T}_{t-1})$
$\mathbf{T}_3(\mathbf{R}_t, \mathbf{T}_t)$
Z_g
Y_g
X_g
O_g
$\mathbf{T}_2$
Gaze camera coordinate system (time t-1)
$\mathbf{T}_c(t\text{-}1)$
Y_r
X_r
H
Z_r
$\mathbf{T}_c(t)$
Y_g'
O_g'
Z_g'
X_g'
Reference coordinate system
Gaze camera coordinate system (time t) after pan-tilt movement

Fig. 5. Update of the relationship between the pose and gaze cameras after pan-tilt movement.

The origin is at the intersection of the axis of pan and tilt rotation; the Y_r-axis aligns with the axis of the pan rotation, and the Z_r-axis aligns with the axis of the tilt rotation. The gaze camera coordinate system (O_g, X_g, Y_g, Z_g) is related with (H, X_r, Y_r, Z_r) by a homogenous transform $\mathbf{T}_c$ [20]. Initially, both pan and tilt of the pan-tilt unit are set to be zero; the X_g-Y_g-, and the Z_g-axes are set to be parallel to the X_r-, Y_r-, and Z_r-axes, respectively.

Considering two consecutive frames, which corresponds to the case where the gaze camera is moved from (O_g, X_g, Y_g, Z_g) to (O_g', X_g', Y_g', Z_g'), see Fig. 5. Coordinate systems (O_g, X_g, Y_g, Z_g) and (O_g', X_g', Y_g', Z_g') can be related through the reference coordinate system

$$\mathbf{T}_2 = T_c^{-1}(t-1)\mathbf{R}_Y(\alpha)\mathbf{R}_X(\gamma)\mathbf{T}_c(t-1) \tag{10}$$

where α and γ are the rotation angles around the pan and tilt axes, respectively, by which the gaze camera is moved from (O_g, X_g, Y_g, Z_g) to (O_g', X_g', Y_g', Z_g') to focus on the iris.

At time $t-1$, assume that the relationship between the pose camera coordinate system (O_p, X_p, Y_p, Z_p) and the gaze camera coordinate system (O_g, X_g, Y_g, Z_g) is $(\mathbf{R}_{t-1}, \mathbf{T}_{t-1})$, where $\mathbf{R}_{t-1}$ represents the rotation matrix, and $\mathbf{T}_{t-1}$ represents the translation vector. At time t, the relationship between the two cameras, which is represented as $\mathbf{T}_3(\mathbf{R}_t, \mathbf{T}_t)$ in Fig. 5, will be

$$\mathbf{T}_3 = \mathbf{T}_2^{-1}\begin{bmatrix} \mathbf{R}_{t-1} & \mathbf{T}_{t-1} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{11}$$

To speed up the process, we generate a look-up table that is addressed with the pan-tilt values. The values stored in the table cells are the relative transformation before and after pan-tilt motion. Here, the pan-tilt angles are angles that are relative to the initial gaze camera coordinate system.

The initial relationship between the pose and gaze cameras is

$$\mathbf{R_0} = \mathbf{R}_2\mathbf{R}_1^{-1} \tag{12}$$

$$\mathbf{T_0} = \mathbf{t}_2 - \mathbf{R}_2\mathbf{R}_1^{-1}\mathbf{t}_1 \tag{13}$$

where $\mathbf{R}_1$, $\mathbf{t}_1$ are, respectively, the rotation and translation of the pose camera with respect the world coordinate system (O_w, X_w, Y_w, Z_w) and $\mathbf{R}_2$, and $\mathbf{t}_2$ are the rotation and translation of the gaze camera with respect the same world coordinate system; see Fig. 6. By applying $\mathbf{R_0}$ and $\mathbf{T_0}$, we can project the 3-D points in the pose camera coordinate system to the gaze camera coordinate system.

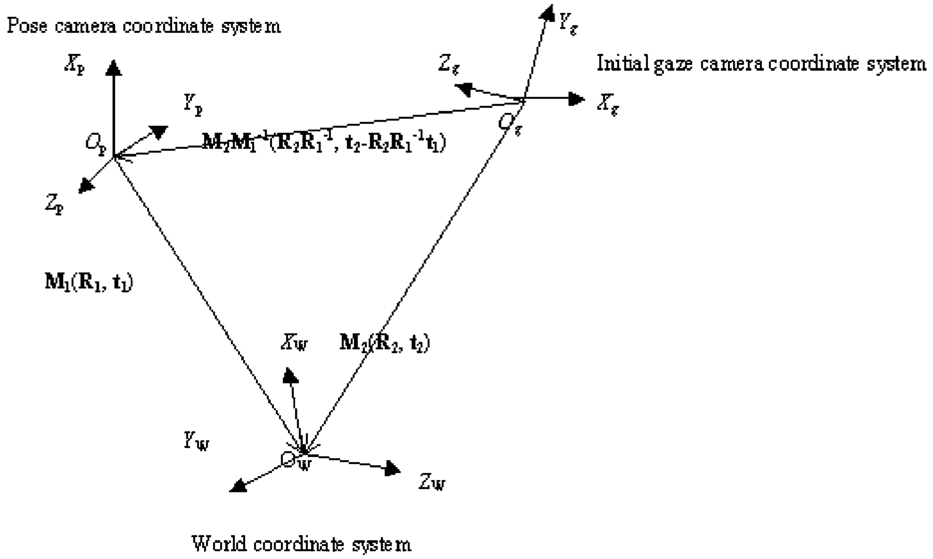

Pose camera coordinate system
X_p
Y_p
O_p
Z_p
M_2M_1^-1(R_2R_1^-1, t_2-R_2R_1^-1t_1)
Y_g
Z_g
Initial gaze camera coordinate system
X_g
O_g
M_1(R_1, t_1)
X_W
M_2(R_2, t_2)
Y_W
O_W
Z_W
World coordinate system

Fig. 6. Initial calibration of the pose and gaze cameras.

IV. Experimental Investigation and Results

We have tested our gaze estimator using a Pentium 450 PC with 128 MB of RAM. The algorithm was implemented in Image-Pro Plus/SDK software with Matrox Meteor-II imaging board. The experiments on synthetic and real data have been done. Good results obtained verified the accuracy and robustness of the method. The experiments using synthetic data are discussed as follows. The experimental results on real images will be discussed in Section IV-B.

A. Experimental Results on Synthetic Data

We test the algorithm using two sets of data: exact and noisy. By observing the predefined targets, the accuracy of the algorithm can be measured. We will present these issues in Sections IV-A1, 2, and 3, respectively.

1) Simulations on Exact Data: The coordinate system of the gaze camera is defined; the image plane is defined as the X-Y plane, and the Z-axis is along the optical axis of the camera and pointing toward the frontal object. The initial plane is where the iris is positioned parallel to the X-Y plane of the camera. The coordinate system of the eyeball is defined as follows: The center of the eyeball is set to be the origin, and the X-, Y-, and Z-axes are set to be parallel to the X-, Y-, and Z-axes, respectively, of the gaze camera coordinate system. The projective ellipse can be fitted by the method presented in [2]. The 3-D location of the iris (six degrees of freedom: three for translation and three for rotation) is obtained from the ellipse/circle correspondence. The errors between the calculated results and the corresponding original synthetic data are recorded under different poses. An example is given as follows.

The size of the image is set to be 640 × 480. The intrinsic parameters of the camera are set as

$$u_0 = 320,\ v_0 = 240,\ f_x = f_y = 5500 \tag{14}$$

where (u_0, v_0) are the coordinates of the principle point, and f_x and f_y are the scale factors of the camera along the X- and Y-axes, respectively. The settings of f_x and f_y here imply that the camera (zoom-in) requires a larger focal length in order for the eyes to appear big enough on the image. The higher resolution images can be obtained compared with the images in the "two-circle" algorithm. A focal length of 50–65 mm is used in our real image experiments, which keep a distance of about 60–100 cm between the human face and the camera. The 3-D coordinate (with respect to the gaze camera coordinate system) of the eyeball center, given in centimeters, is

$$(x_s, y_s, z_s) = (0, 0, 60)\ \text{cm} \tag{15}$$

i.e., the initial contour of the iris (circle) lies parallel to the image plane. The radius of the iris is set as

$$r = 0.65\ \text{cm}. \tag{16}$$

The 3-D position of any point that lies in the iris contour can be obtained using following parametric equations:

$$x = x_c + r\cos(\theta),\ y = y_c + r\sin(\theta),\ z = z_c \tag{17}$$

where $0^\circ \leq \theta \leq 360^\circ$, (x_c, y_c, z_c) is the 3-D coordinate of the iris center.

The distance between the two extreme corners of the parabola (see Fig. 2) is assumed to be

$$P_1P_3 = 3.5\ \text{cm}. \tag{18}$$

The ratio of the radius of the eyeball and the radius of the iris is assumed to be 2. Hence, the radius of the eyeball is

$$R = 2r = 1.3\ \text{cm}. \tag{19}$$

The distance from eyeball center to the iris plane will be

$$d = \sqrt{R^2 - r^2} = 1.13\ \text{cm} \tag{20}$$

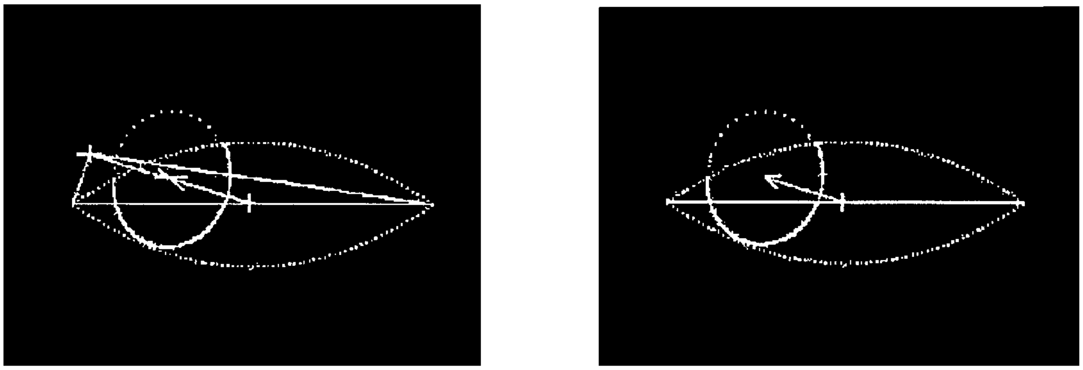
Then, the center of the iris is obtained based on (15) and (20)

$$(x_c, y_c, z_c) = (0, 0, 58.87)\ \text{cm}. \tag{21}$$

The top and bottom points of the parabola are symmetric about the initial iris center (see Fig. 2), and the distances are

$$P_2O_c = P_4O_c = 0.6\ \text{cm}. \tag{22}$$

We consider the possible gaze using the synthetic images of a model eye in which the iris is rotated to every possible direction. Let the head look toward the camera. The eyeball is rotated around its center (x_s, y_s, z_s) about their own Y-(azimuth) following the X-axis (elevation). Consequently, the iris contour



(a) (b)

Fig. 7. Example of the gaze disambiguation using "distance constraint" (the rotation about the Y-axis is $-30°$ and about the X-axis is $-10°$) (a) Fitted iris contour; two estimated eyeball centers (cross points); two estimated eye gaze are represented as two arrows starting from the resulted eyeball's center and ending at the iris' center, respectively. (b) Unique solution of the eye gaze, iris center, and eyeball center (cross point).

TABLE I
DETERMINING THE GAZE BY APPLYING "DISTANT CONSTRAINT"

n	O_c (cm)	O_s (cm)	D_R (cm)	D_L (cm)	D (cm)	
(-0.501,-0.151,0.852)	(0.560, 0.164, 59.033)	(-0.006, -0.006, 59.996)	2.076	2.087	0.012	T
(0.517, 0.156, 0.842)	(0.554, 0.163, 59.033)	(1.138, 0.338, 59.985)	3.114	1.316	1.798	F

TABLE II
ERRORS OF THE GAZE AND IRIS CENTER

	True	Estimated	Error
n	(0.922, 0.277, 0.271)	(-0.501, -0.151, 0.852)	0.077^0
O_c (cm)	(0.565, 0.170, 59.036)	(0.560, 0.164, 59.033)	0.01
O_s (cm)	(0, 0, 60)	(-0.006, -0.006, 59.996)	0.19

will rotate to the position that corresponds to the expected gaze, following the eyeball. We project the two eyelids and the rotated iris onto the image plane according the assumed camera parameters and then estimate the gaze from the projections. Only the edges between the two eyelids are considered to compute the eye gaze.

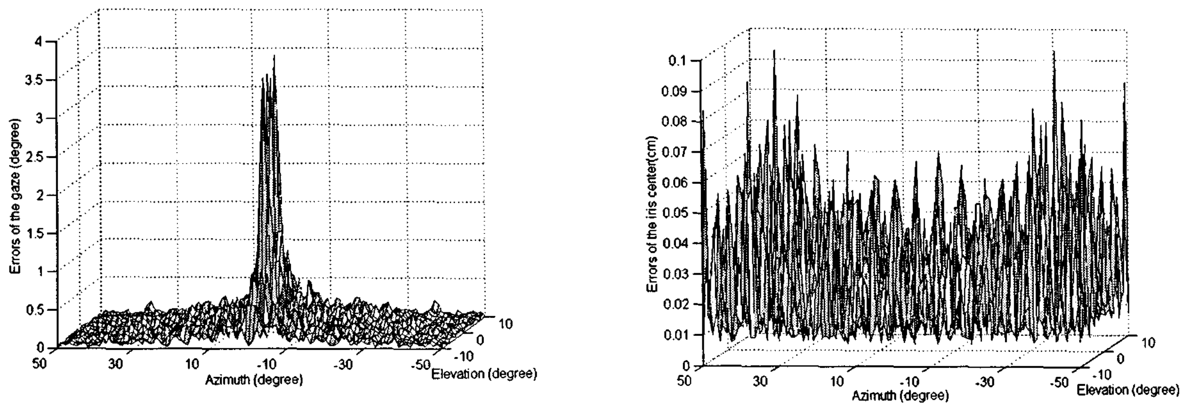
In our simulations, the eyeball is rotated about the vertical axis from $-50°$ to $50°$ in steps of $1°$ (azimuth) and rotated about the horizontal axis from $-10°$ to $10°$ in steps of step $1°$ (elevation) to form a set of synthetic images. The performances of the "one-circle" algorithm are tested on these synthetic images.

The image, for instance, when the rotation angle about the Y-axis is $-30°$ and the X-axis is $-10°$, is shown in Fig. 7. The iris contour (dotted ellipse) is fitted using the iris edges that lie between the two eyelids; see Fig. 7(a). Two solutions for the normal to the supporting plane of the iris contour and the center of the iris contour are given in Table I. Two eyeballs' centers are estimated [see the cross points in Fig. 7(a)] based on the two solutions and the known d [see (20)]. Two arrows from the resulting eyeball centers to the iris centers, i.e., gaze according to our definition, are shown, respectively, in Fig. 7(a). The unique solution of the normal and the center of the iris, which satisfies the "distance constraint," are shown in Fig. 7(b). In Table I, D_R and D_L are the distance from the eyeballs center to the two eye corners, respectively, and D represents the difference between D_R and D_L. The unique solution is marked as "T" in the last column of Table I. As for this example, the error of the eye gaze and the iris center is listed in Table II.

The errors of the eye gaze over the testing synthetic images are shown in Fig. 8(a). We can see that the accuracy tends to fall greatly when the face is around the frontal view and falls less as the gaze turns away from the fronto-parallel position. This is because the iris contour becomes nearly a circle instead of an ellipse when the camera direction is approximately fronto-parallel. Fortunately, this degenerate case can be prevented in our application by simply putting the camera slightly skewed to the face; see Fig. 1. Hence, we will consider the errors of the eye gaze in the following excluding this degenerate case. The errors of the iris center are shown in Fig. 8(b).

The experimental results show that the precision of the eye gaze is improved significantly compared with the results obtained in the "two-circle" algorithms [36]. Using the "one-circle" method, the maximum error of the gaze due to the eyelids' occlusion is $0.3°$, whereas the maximum error of the center of the iris is 0.1 cm. However, the experiments of the "two-circle" algorithm [36] showed that the maximum error of the gaze is $2.5°$, and the maximum error of the center of the iris is 0.5 cm.

2) Robustness to Geometrical Disturbances: Using the same settings of the parameters of the camera and the dimensions of the iris, as illustrated in Section IV-A1, we tested the performance of the algorithm when subjected to geometrical disturbances. Corrupting the locations of the imaged features



(a) (b)

Fig. 8. Errors of the eye gaze and iris center over the possible view directions. (a) Eye gaze. (b) Iris center.

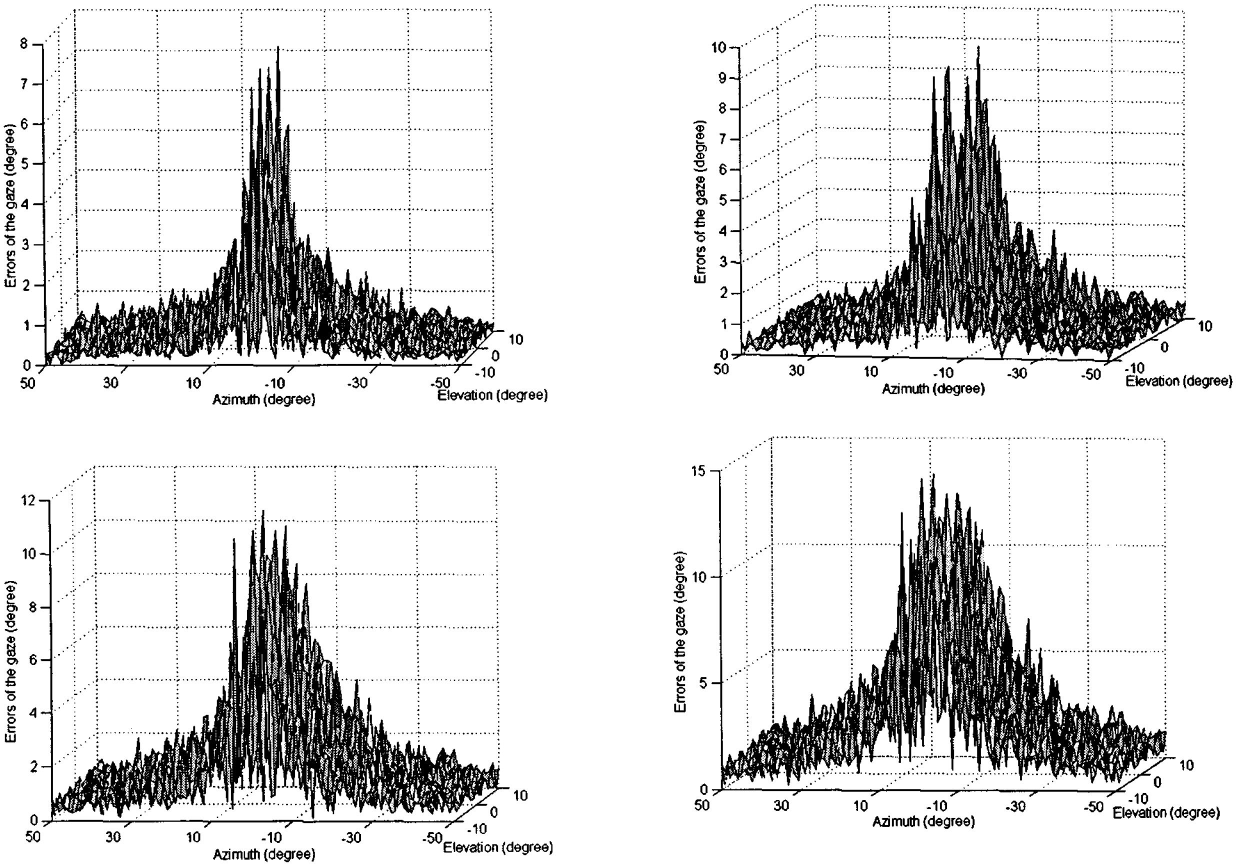


Fig. 9. Mean errors of the gaze for the synthetic images. Images are perturbed by a standard Gauss noise; standard deviations are, respectively, (Top left) 1, (Top right) 1.4, (Bottom left) 2.0, and (Bottom right) 2.8.

by Gaussian noise can do this. This simulates the condition of imprecise location of the features caused by the feature extraction algorithm, noise, and other processes.

Considering the generation of the synthetic images of the eye, the image coordinate (x_j, y_j) of the jth point of the iris contour will be disturbed as

$$\begin{aligned} x_j &= x_j + G_x(M, \sigma) \\ y_j &= y_j + G_y(M, \sigma) \end{aligned} \tag{23}$$

where $G_x(\sigma)$ and $G_y(\sigma)$ are two independent random Gaussian noise generators (mean M and standard deviation σ).

After that, we estimate the gaze from the corrupted image. The mean error from 100 trials for a gaze is computed to indicate the robustness of the gaze under geometric disturbances. The errors of the estimated gaze and the iris center are shown in Figs. 9 and 10, respectively. The different levels of noise, namely, with standard deviations of 1, 1.4, 2.0, and 2.8, are tried. The results showed that the algorithm is robust. The algorithm can work even when the noise is increased significantly, with the resulting

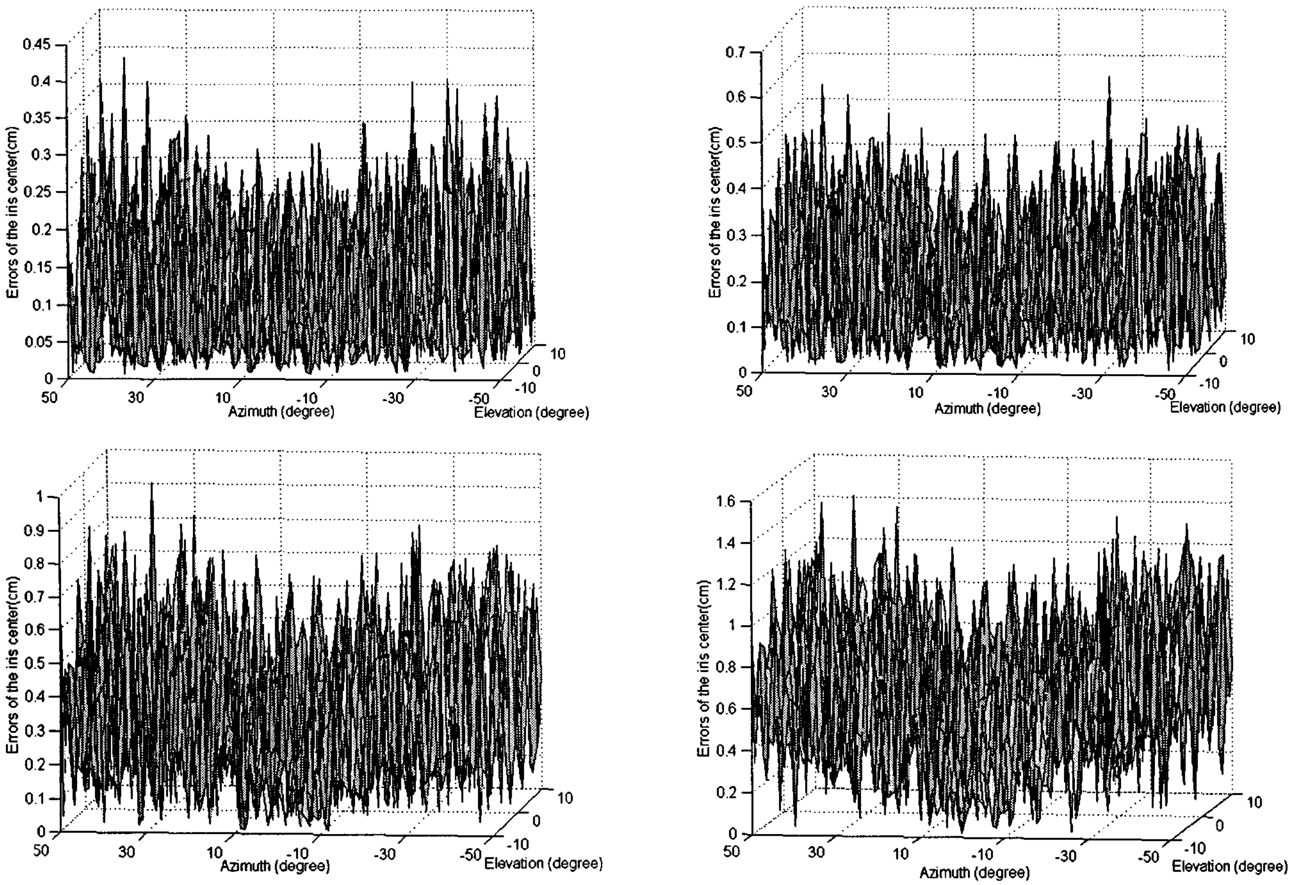


Fig. 10. Mean errors of the center of the iris for the synthetic images; the images are perturbed by a standard Gauss noise, and the standard deviations are, respectively, (Top left) 1, (Top right) 1.4, (Bottom left) 2.0, and (Bottom right) 2.8.

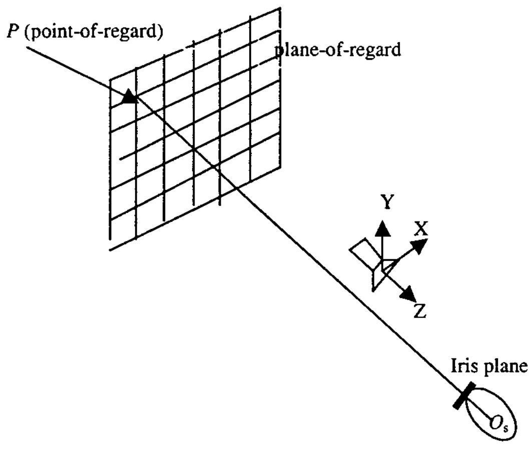


Fig. 11. Experiments on the accuracy of the eye gaze versus the points-of-regard.

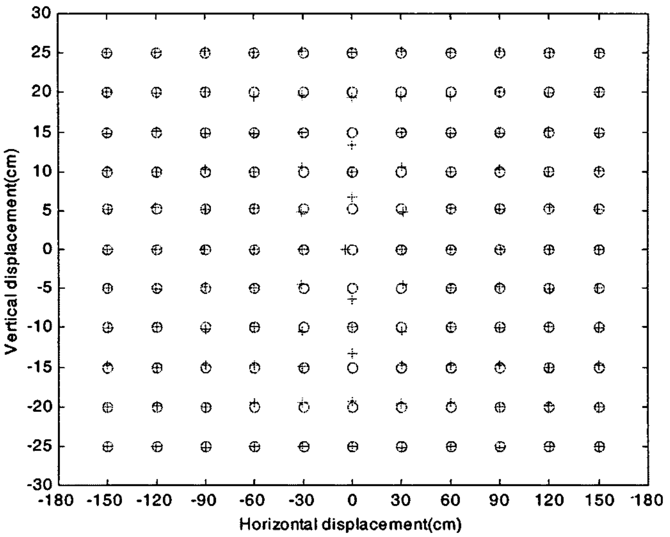


Fig. 12. Errors of the points-of-regard (true points-of-regard are marked "*o*," and estimated points-of-regard are marked "+").

accuracy degrading gracefully. When we disturb the iris edges with a standard Gaussian noise (zero mean and standard deviation one pixels), the errors of the gaze over the testing range are less than 1°, and the errors of the iris center are less than 0.3 cm.

Eye gaze could be used as the cue for human-machine interaction because point-of-regard reflects a subject's interest. Hence, the accuracy of the point-of-regard is a crucial factor for considering the applicability of an eye-gaze estimation method in HCI applications. In the following section, we will discuss the accuracy of our eye gaze estimation method with respect to the point-of-regard. In Section IV-B, measurement of the accuracy of the eye gaze with respect to the point-of-regard will be discussed on the real images.

3) Accuracy of the Eye Gaze Versus Point-of-Regard: The point-of-regard (fixation in 3-D space) can be obtained in the

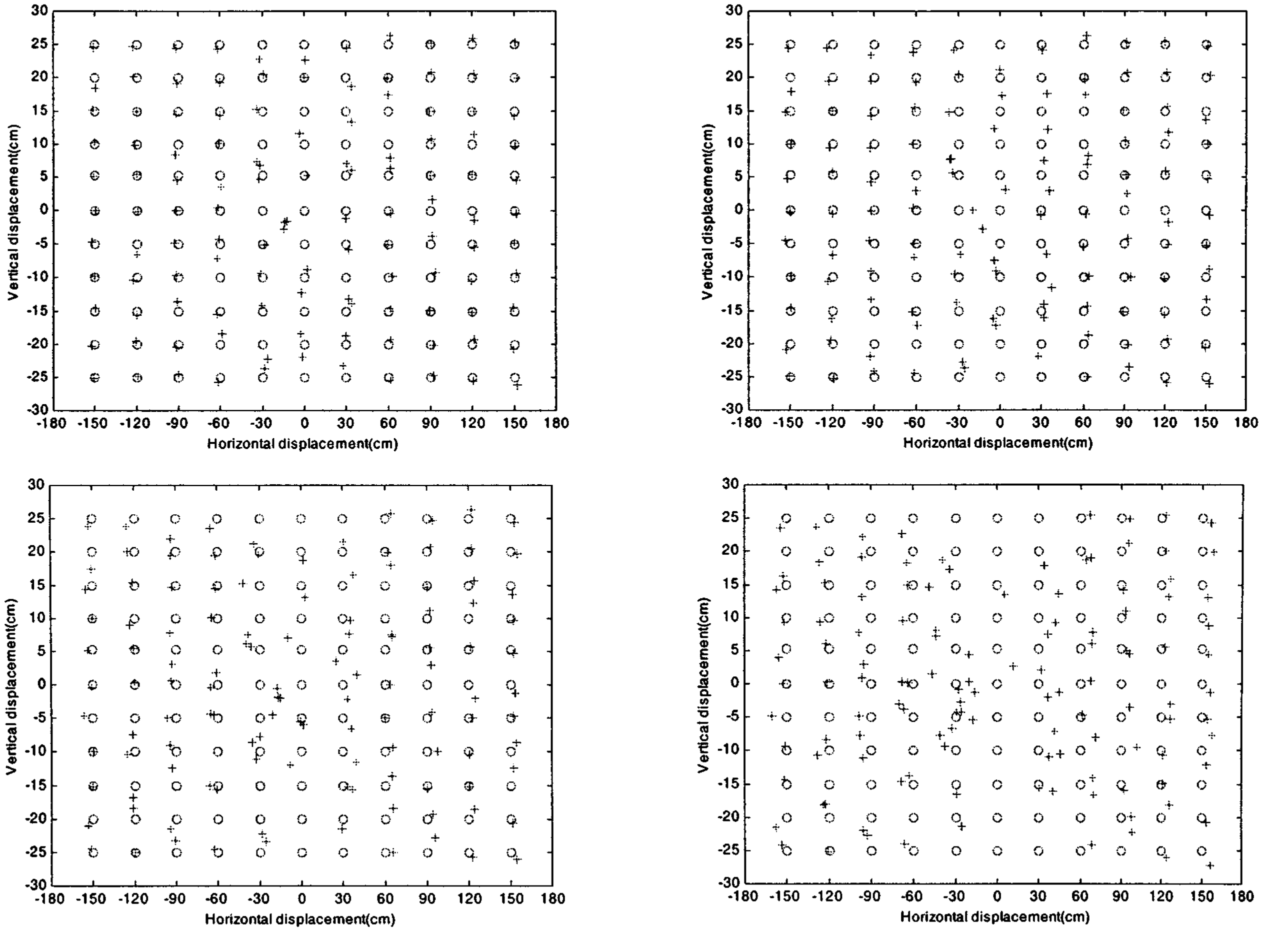


Fig. 13. Errors of the focus-of-point for the synthetic images when the image is perturbed by Gaussian noise, the standard deviation of the noise are respectively: (Top left) 1, (Top right) 1.4, (Bottom left) 2.0, and (Bottom right) 2.8.

"two-circle" method from the line-of-sight of the two eyes [34], [36]. The "one-circle" method proposed here would be able to determine the eye gaze but cannot detect point-of-regard from one image. However, the "one-circle " method can lead to the computation of a point-of-regard from two images of the subject observing the point-of-regard from two viewpoints. This is because the line-of-sight for the two viewpoints is each uniquely located in "one-circle " method from the eye gaze and the 3-D position of the iris; the point-of-regard is the intersection of the two lines-of-sight.

In the simulation, we assume a plane-of-regard is positioned in front of the subject. The position of the plane-of-regard with respect to the gaze camera is known. This assumption simplifies the computing because the point-of-regard can be located by intersecting the line-of-sight with the known plane-of-regard, which, in some HCI applications, could be the monitor screen. The simulation experiment is shown in Fig. 11. The settings of the camera and the subject are the same as Section IV-A1. The focus plane is defined in the gaze camera coordinate system and set to be

$$Z = -120\ \text{cm} \tag{24}$$

i.e., it is set to be parallel to the image plane and with a distance 1.2 m to the camera. The distance from the subject to the plane-of-regard is, thus, 1.8 m.

For a point-of-regard P on the plane-of-regard, the position of the iris plane is changed in order to focus on the point-of-regard (see Fig. 11). The iris plane is rotated around the center of the eyeball O_s until the line $O_S P$ becomes the normal of the iris plane. After that, the eye gaze and, consequently, the point-of-regard are estimated from the synthetic image.

The results are shown in Fig. 12, where the true and estimated point-of-regard are marked as "O" and "$+$," respectively. The testing range is set to be -150 cm to 150 cm in 30-cm steps for horizontal displacement and -25 cm to 25 cm in 5-cm steps for vertical displacement. The results on the noised data are shown in Fig. 13, Similarly to the tests done in Section IV-A and B, standard derivations of 1, 1.4, 2.0, and 2.8, respectively, are tried.

We can see the accuracy of the point-of-regard is satisfactory. Corrupting the image with Gaussian noise (zero mean, standard deviation one pixel), the error is less than 2 cm within the range of 1.8 m; see the top left of Fig. 13.

The simulations have verified the accuracy and robustness of the algorithm with quantitative performance measures. The next step is naturally to test the performance of the algorithm on the real images.

B. Experiments on Real Images

As can be seen in the thorough simulations discussed in the previous section, we have proved our approach is robust. We also did the experiments on real face images. We will see the satisfactory eye gaze estimates can be obtained from zoom-in iris images, where the iris edges and the consequent contour can be detected reliably.

We have implemented the extraction algorithm of the iris edges in Section II-E. The elliptical contour of the iris in the image needs to be fitted from the extracted iris edges. We adopt the fitting method presented in [2].

A 55-mm lens is used that will guarantee the eye in the image is large enough when the distance between the human face and the camera is within 60–100 cm. The distance is suitable for our human-machine interaction application. Of the 640 × 480 pixels in the image, about 800 edge points are obtained for fitting the iris contour.

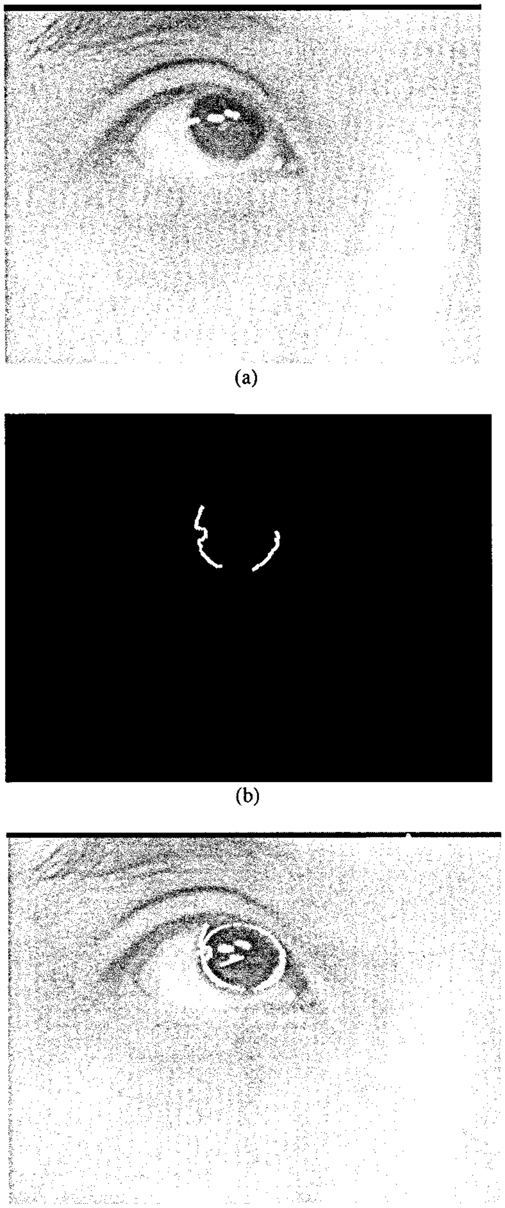
We have demonstrated our gaze estimation approach on real images of three subjects. As we will see in Sections IV-B1 to 3, the experiments on the real images include the iris detection and gaze determination, accuracy of the eye gaze, and the integration of the head pose and eye gaze. The results are satisfactory. The errors of the point-of-regard are less than 1.5 cm within a 1.5-m range, and consequently, the errors of the gaze are less than 1° (see Section IV-B2). The accuracy of the eye gaze is higher because the most reliable facial features such as eye corners and iris contour are used in the approach. The result is found to better than the existing nonintrusive approaches, such as Zelinsky [22]. We will take some examples to discuss the problems encountered, which include iris detection, accuracy of the gaze, and integration, as follows. In the following examples, the radius of the iris contour is 0.63 cm; the ratio between the radius of the eyeball and the radius of the iris is set to be 2.

1) Iris Detection and Eye Gaze Determination: Some of the gaze determination results are shown in Figs. 14–16, including iris edges, fitted iris contour, and eye gaze. We can see that the iris detection technique, which is presented in Section II-E, is efficient.

2) Accuracy of the Eye Gaze: Using the same method presented in Section IV-A3, the accuracy of the eye gaze respect to the point-of-regard is evaluated on real images.

The experiment is described in Fig. 17. The person gazes in four prespecified targets, including two top corners of a whiteboard V_1 and V_4 and two points V_2 and V_3 that divides V_1V_4 into three segments having equal length. The subject maintains his/her head stationary relative to the gaze camera, whereas he/she gazes at each of the four targets. Hence, the coordinates of his/her eye corners remain almost fixed for the four line-of-sight. In Fig. 17(b), $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$, and $\mathbf{n}_4$ correspond to the true eye gaze for point V_1, V_2, V_3, and V_4 respectively.

A camera is put between the subject and the whiteboard. The 3-D coordinates of the whiteboard with respect to the gaze camera are known. Consequently, the coordinates of V_1-V_4 in the gaze camera system are known. We take the coordinates of the four targets as the reference coordinates for the true point-of-regard. The errors of the algorithm are computed by comparing the estimated point-of-regard and the reference coordinates. The results of the gaze estimation are shown in Fig. 18.


(a)

(b)

(c)

Fig. 14. Example of the eye gaze determination. (a) Original image. (b) Iris edges. (c) Iris edges, fitted ellipse, and the gaze.

The coordinates of the eye corners, which are obtained by projecting the head pose results [37] to the gaze camera using (12) and (13), are listed in Table III. The eye gaze ($\mathbf{n}$), the center of the iris (O_c), and, consequently, the center of the eyeball (O_s) are listed in Table IV. D_L and D_R are the distances from the O_s to the two eye corners (R and L in Table I), respectively

$$D = |D_L - D_R|. \tag{25}$$

Applying the "distance constraint" to disambiguate the solutions of the normal and the center, the results are listed in Table IV, where the unique solution is marked as "T."

The errors of the point-of-regard E_p are then estimated and listed in Table V. The errors of the eye gaze are also estimated by

$$E_g = \tan^{-1}\left(\frac{E_p}{D_t}\right) \tag{26}$$

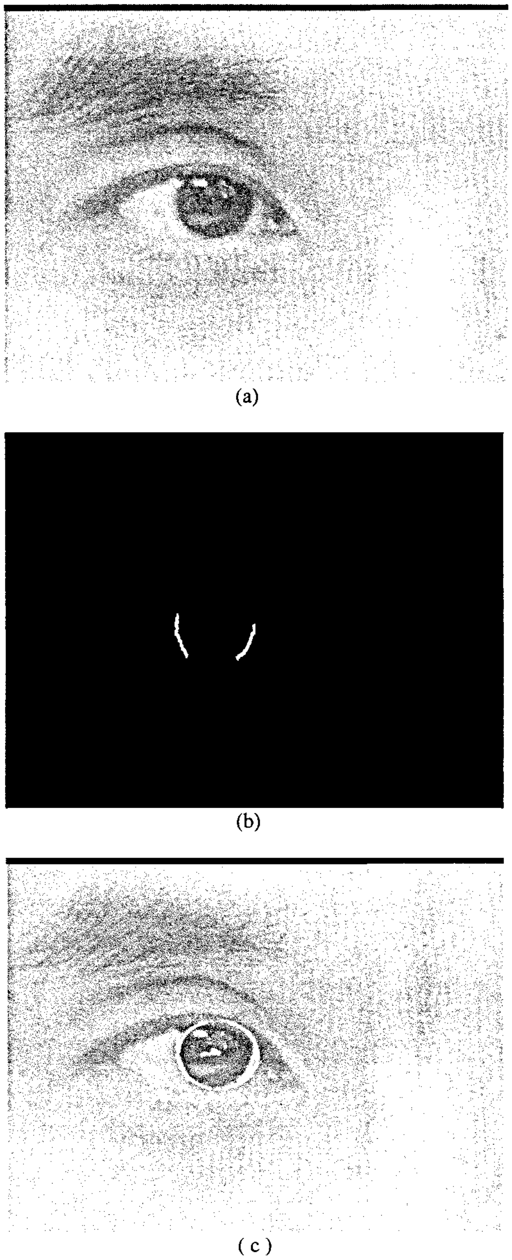


Fig. 15. Example of eye gaze determination. (a) Original image. (b) Iris edges. (c) Overlay the iris edges, fitted ellipse, and the gaze onto the original image.

where D_t is the distance from the subject to the plane-of-regard; see Fig. 17(a).

We can see the errors of point-of-regard are less than 1.5 cm within the 1.5 m (D_t) range, and consequently, the errors of the gaze are all less than 1°.

In Zelinsky's approach [22], the eye gaze is determined using stereo vision. A total of four eyes in the stereo image pair were used to compute the eye gaze. However, the resolution of the images is low since the width of an eye is only 30 pixels. Each measurement is not sufficiently accurate to determine the gaze point. Hence, four gaze vectors were averaged to generate a single gaze. Accuracies of ±3.5° were reported.

Nonetheless, we acknowledge that Zelinsky's work is a fully operational system, which is a rare achievement in computer vision.

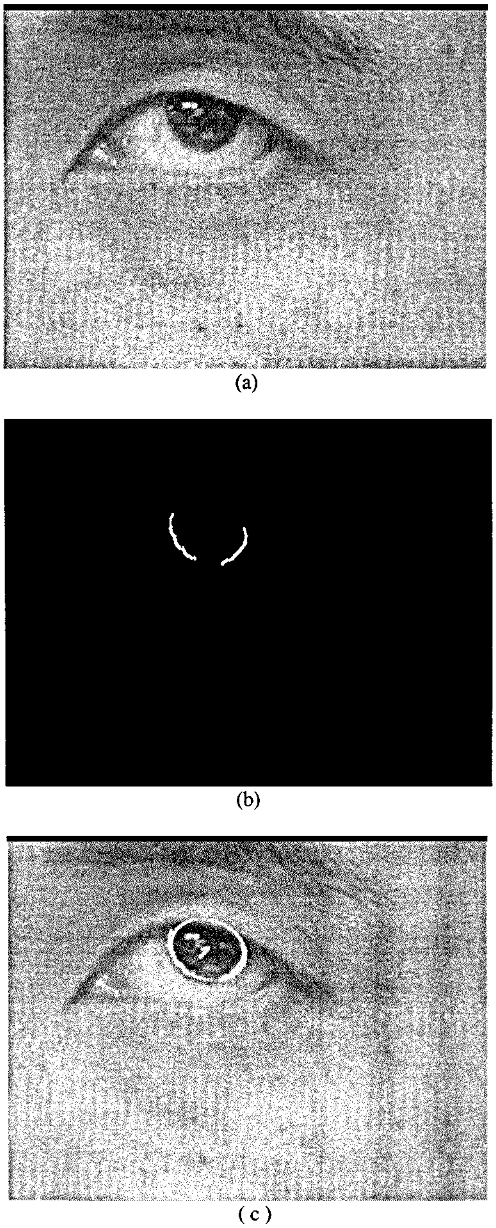


Fig. 16. Example of eye gaze determination. (a) Original image. (b) Iris edges. (c) Overlay the iris edges, fitted ellipse, and the gaze onto the original image.

3) Integration: An example to illustrate the integration of the pose and gaze is shown in Fig. 19. The pose image is shown in Fig. 19(a). The head pose results, including the eye and mouth corners and the facial normal, are determined using our method of vanishing point [34], [37] and shown in Fig. 19(b). The eye gaze image, where the iris is focused based on the two corners of the left eye provided by the pose determination results, is captured and shown in Fig. 19(c). The iris edges and the fitted iris ellipse are shown in Fig. 19(d) and (e). The two corners of the left eye are projected to the gaze camera coordinate system. The eye gaze, disambiguated by applying "distance constraint," is shown in Fig. 19(e).

V. Limitations of the Method

We mention a couple of difficulties encountered in our algorithm. Developing a practical eye gaze and head pose esti-

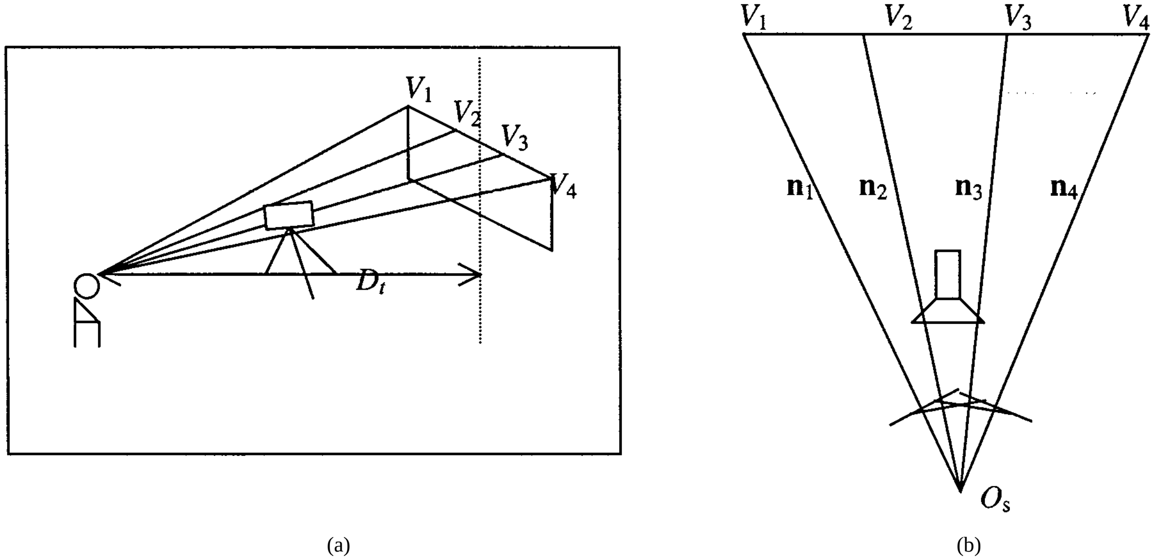


Fig. 17. Test the accuracy of the eye gaze versus the point-of-regard. (a) Side view. (b) Top view.

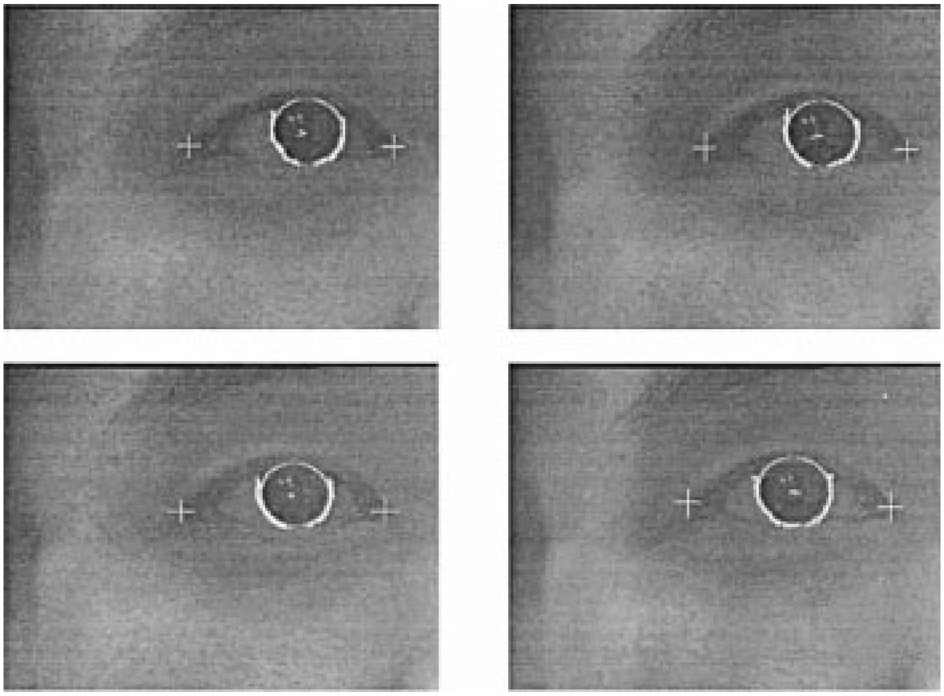


Fig. 18. Computed eye gaze correspoding to the four target points. (Top left) V_1. (Top right) V_2. (Bottom left) V_3, and (Bottom right) V_4 .

TABLE III
EYE CORNERS WITH RESPECT TO THE GAZE CAMERA

	Eye corners (cm)
V_1, V_2, V_3, V_4	R: (-4.010, 0.030, 59.100), L: (-2.501, .028, 59.301)

mation system is a difficult task. The face can be deformed by facial expressions, making it difficult to extract the facial features, such as eye and mouth corners and iris edges. The face detection and facial feature extraction are seriously affected by variations in pose and lighting conditions and even by factors like glasses, beard, age, and change in facial hair. This leads to a need for employing complex algorithms that are used to tackle these problems.

Another difficulty is in the elliptical estimation. We use the method developed by Bookstein [2] to fit the ellipse. The fitted ellipse seems to be small compared with the edge segments extracted. We can see this from the results shown in Figs. 14–16. This is an artifact of the small number of pixels that participate in the estimation of the ellipse. Although zoom-in camera is used, the number of pixels on the iris edges is small due to the eyelids naturally occluding the iris. This could be a serious problem when the person has large facial expressions, such as anger or looking down. In such a case, the occlusion of the eye-

TABLE IV
DETERMINING THE GAZE BY APPLYING THE "DISTANT CONSTRAINT"

	$\mathbf{n}$	O_c (cm)	O_s (cm)	D_R (cm)	D_L (cm)	D (cm)	
V_1	(-0.392, -0.092, 0.915)	(-2.854, 0.097, 58.814)	(-3.297, -0.007, 59.844)	0.964	1.024	0.059	T
	(0.260, 0.096, 0.961)	(-2.858, -0.095, 58.813)	(-2.565, 0.204, 59.895)	0.632	1.653	1.020	F
V_2	(-0.148, -0.107, 0.983)	(-3.123, 0.101, 58.662)	(-3.290, -0.020, 59.769)	0.919	0.976	0.057	T
	(0.001, 0.119, 0.993)	(-3.124, -0.099, 58.062)	(-3.123, 0.233, 59.780)	0.820	1.134	0.314	F
V_3	(-0.260, 0.103, 0.960)	(-3.406, 0.010, 58.477)	(-3.699, 0.215, 59.558)	1.246	0.588	0.657	F
	(0.105, -0.108, 0.989)	(-3.409, -0.101, 58.477)	(-3.290, -0.021, 59.590)	0.842	0.862	0.020	T
V_4	(-0.510, 0.094, 0.855)	(-3.666, 0.094, 58.567)	(-4.240, 0.200, 59.529)	1.767	0.531	1.236	F
	(-0.359, 0.093, 0.929)	(-3.671, 0.095, 58.566)	(-3.268, -0.010, 59.612)	0.829	0.893	0.064	T

TABLE V
ERRORS OF THE POINT-OF-REGARD AND THE EYE GAZE

	True (cm)	Estimated (cm)	E_p (cm)	E_g
V_1	(60, 15, -90)	(61.056, 14.992, -90)	1.056	0.40^0
V_2	(20, 15, -90)	(19.277, 16.290, -90)	1.479	0.56^0
V_3	(-20, 15, -90)	(19.249, 16.294, -90)	1.496	0.57^0
V_4	(-60, 15, -90)	(-61.020, 14.991, -90)	1.025	0.39^0

lids becomes serious. The accuracy of the ellipse fitting affects the gaze computing; hence, an unbiased ellipse fitting method is expected in the future work.

There has been continued interest in the fitting of ellipse to image data because the ellipse, being the perspective projection of the circle, is of great importance in pattern recognition and computer vision. Some good surveys of ellipse fittings can be found in [26], [27], and [33]. The methods on ellipse fitting can be divided into two broad techniques: clustering (e.g., Hough-based [41]) and least squares fitting. The least squares method, which does effort to find the parameters for describing the ellipse, depends on minimizing some distance measures between the data points and the ellipse. In the case of moderate occlusion or noise, they may yield unbounded fits to hyperbolae. Fitzgibbon *et al.* [10], [11] developed a least squares fitting method that is specific to ellipse rather than general conics. The advantage of the method is that even bad data always return an ellipse.

Important research focuses on the incorporation of the above-mentioned ellipse fitting methods into bias-correction methods. There are better methods that give unbiased estimates, e.g., Porrill [24]. Porill's algorithm, which is a Kalman filter approach, naturally includes prior information about the ellipse and eliminates the curvature bias associated with the Bookstein algorithm [2].

VI. CONCLUSION

In this paper, we presented a nonintrusive method of robustly estimating the eye gaze by zoom-in iris imaging. The motivation of our approach is the estimate of the eye gaze robustly in real time and with satisfactory accuracy from a single image frame. The use of the domain knowledge of the human face is crucial, and this makes our paradigm original and novel. Actually, an ellipse can be backprojected into the space onto two possible circles. The principle is applied to the eye-gaze by observing that the contour of the iris (3-D) is circular, and hence, it is the circle that we are looking for. By using a geometric constraint, namely, that the distance between the eyeball's center and the two eye corners should be equal to each other, the unique solution can be disambiguated. We improve on current eye-gaze determination methods by achieving higher resolutions. This comes about mainly by the algorithm requiring focus on a single eye; therefore, the estimation of the iris ellipse is more accurate. We proposed a general approach that combines head pose and eye gaze determination. The absolute eye gaze can be obtained accurately with the help of a second pose camera. Another novel feature is that we make use of the circle/ellipse relation to determine the unique gaze direction. Others have not tried to use the iris contour in this way before.

In conclusion, head pose and eye gaze estimation are important in applications such as virtual reality, video conferencing, and human-machine interface/controls. The eye gaze method above is integrated with a head pose estimation module and, together, will offer great potential, especially in these mentioned applications. It is important to note that our method is noninvasive, fast, and robust. It is robust because the iris contour is one of the simplest and most robust facial features to extract.

APPENDIX A
INTEGRATION OF THE HEAD POSE AND EYE GAZE

In Section III, a general approach that combines the head pose and eye gaze is described. The integration can be described as following steps.

Step 1) Calibrate the pose camera (O_p, X_p, Y_p, Z_p) and the gaze camera (O_g, X_g, Y_g, Z_g) under the same world coordinate system (O_w, X_w, Y_w, Z_w); see Fig. 6. This results in the perspective projection

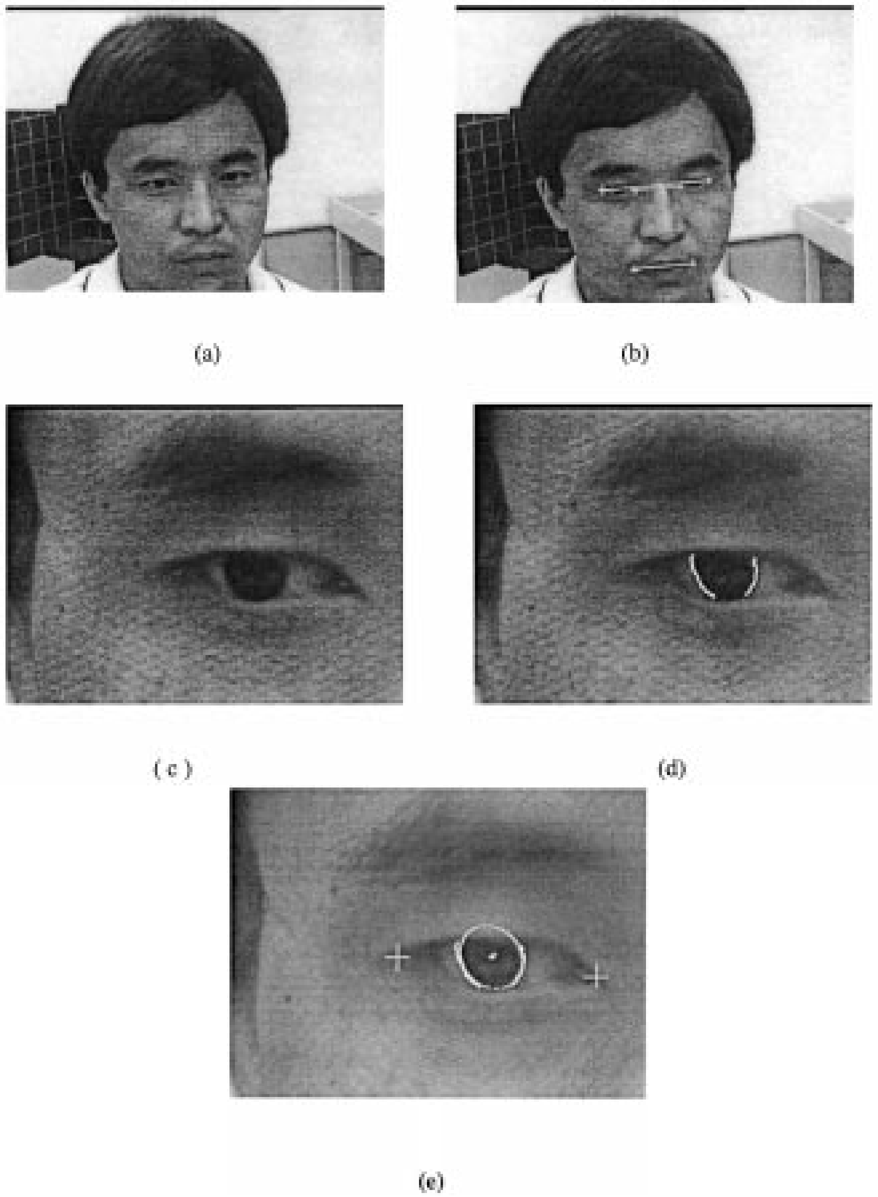


Fig. 19. Example of integration of the pose and gaze. (a) Original pose image. (b) Pose determination results. The orientation of the face is represented as an arrow staring from the center of the two far-eye corners. (c) Original gaze image. (d) Iris edges of the gaze image. (e) Iris edges, the fitted ellipse, and the gaze. The gaze is represented as an arrow starting from the estimated eyeball's center and ending at the estimated iris's center.

matrixes $\mathbf{M}_1$ and $\mathbf{M}_2$, respectively. $\mathbf{M}_1$ and $\mathbf{M}_2$ are the 3×4 matrixes.

Step 2) Decompose $\mathbf{M}_1$ and $\mathbf{M}_2$ using the method presented in [9] and [38]; we obtain the intrinsic parameters of the camera, including the image coordinate of the principal point and the scale factors along the x- and y-axes. The rotation matrixes and translation vectors are also obtained by the decomposition, resulting in $\mathbf{R}_1$ and $\mathbf{t}_1$ for the pose camera and $\mathbf{R}_2$ and $\mathbf{t}_2$ for the gaze camera.

Step 3) The initial relationship between the pose and gaze cameras is

$$\mathbf{R_0} = \mathbf{R}_2\mathbf{R}_1^{-1} \tag{A.1}$$

$$\mathbf{T_0} = \mathbf{t}_2 - \mathbf{R}_2\mathbf{R}_1^{-1}\mathbf{t}_1. \tag{A.2}$$

$\mathbf{R_0}$ and $\mathbf{T_0}$ can project the 3-D points in the pose camera coordinate system to the gaze camera coordinate system.

Step 4) The gaze camera is related to the reference coordinate system by a homogenous transform $\mathbf{T}_c$

$$P_r = \mathbf{T}_c P_c \tag{A.3}$$

where P_c is the point in the gaze camera coordinate system, and P_r is the point in the reference coordinate system. Any point in the reference coordinate system can be represented in terms of the gaze camera coordinate system before and after pan-tilt motion

$$P_r = \mathbf{T}_c(t)P_c(t) = \mathbf{T}_c(t-1)P_c(t-1). \tag{A.4}$$

The initial tilt from the level θ is known from calibrated $\mathbf{R}_2$. Hence, the initial $\mathbf{T}_c(0)$ is [20]

$$\mathbf{T}_c(0) = \mathbf{R}_{\mathbf{X}}(\theta)\mathbf{T}_{c|_{\theta=0}} = \begin{bmatrix} 1 & 0 & 0 & \rho_X \\ 0 & \cos\theta & -\sin\theta & \rho_Y\cos\theta - \rho_Z\sin\theta \\ 0 & \sin\theta & \cos\theta & \rho_Y\sin\theta + \rho_Z\cos\theta \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.5}$$

where

$$\mathbf{T}_{c|_{\theta=0}} = \begin{bmatrix} 1 & 0 & 0 & \rho_X \\ 0 & 1 & 0 & \rho_Y \\ 0 & 0 & 1 & \rho_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A.6}$$

where (ρ_x, ρ_y, ρ_z) is a vector that corresponds to the case of the gaze camera is in the initial status. In the initial status, both pan and tilt of the pan-tilt unit are set to be zero; the X_g-,Y_g-, and Z_g-axes are set to be parallel to the X_r-, Y_r-, and Z_r-axes, respectively. (ρ_x, ρ_y, ρ_z) is the vector starting from the center of the reference frame and ending at the optical center (P_0) of the gaze camera. The calibration of (ρ_x, ρ_y, ρ_z) will be given in the Appendix B.

Let time $t = 1$.

Step 5) Capture an image using the pose camera. Determine the pose using our vanishing point-based method [34], [37]; the 3-D coordinates of the eye and mouth corners are obtained under the pose camera coordinate system. Assuming the two corners of the left eye are (X_{e1}, Y_{e1}, Z_{e1}) and (X_{e2}, Y_{e2}, Z_{e2}), respectively, the midpoint of the two eye corners is computed as

$$\begin{aligned} X_m &= \frac{X_{e1} + X_{e2}}{2} \\ Y_m &= \frac{Y_{e1} + Y_{e2}}{2} \\ Z_m &= \frac{Z_{e1} + Z_{e2}}{2.} \end{aligned} \tag{A.7}$$

Step 6) Project (X_m, Y_m, Z_m) to the gaze camera by applying $\mathbf{R}_{t-1}$ and $\mathbf{T}_{t-1}$ results in (X_1, Y_1, Z_1)

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \mathbf{R}_{t-1} \begin{bmatrix} X_m \\ Y_m \\ Z_m \end{bmatrix} + \mathbf{T}_{t-1}. \tag{A.8}$$

Step 7) The gaze camera is rotated in order so that the middle point, (X_1, Y_1, Z_1) is imaged at the principal point of the gaze image. The angles of the pan (α) and tilt (β) to be rotated will be

$$\alpha = \tan^{-1}\left(\frac{X_1}{Z_1}\right) \tag{A.9}$$

$$\beta = \tan^{-1}\left(\frac{Y_1}{Z_1}\right). \tag{A.10}$$

Step 8) Capture an image using the gaze camera. The relationship between the pose and gaze cameras, which are represented as $\mathbf{R_t}$ and $\mathbf{T}_t$ in Fig. 5, is updated as follows.

Since $\mathbf{T}_c(t)$ is the result after applying pan and tilt rotations to $\mathbf{T}_c(t-1)$, we have

$$\mathbf{T}_c(t) = \mathbf{R}_Y(\alpha)\mathbf{R}_X(\beta)\mathbf{T}_c(t-1). \tag{A.11}$$

The relationship between the two gaze coordinate systems (before and after motion) will be

$$\mathbf{T}_2 = \mathbf{T}_c^{-1}(t-1)\mathbf{T}_c(t) = \mathbf{T}_c^{-1}(t-1)\mathbf{R}_Y(\alpha)\mathbf{R}_X(\gamma)\mathbf{T}_c(t-1). \tag{A.12}$$

The relationship between the pose and gaze cameras will become

$$\mathbf{T}_3 = \mathbf{T}_2^{-1} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{A.13}$$

Now, $\mathbf{R_t}$ and $\mathbf{T_t}$ can be obtained by decomposing $\mathbf{T}_3$ [9], [38].

Step 9) The two eye corners, i.e., (X_{e1}, Y_{e1}, Z_{e1}) and (X_{e2}, Y_{e2}, Z_{e2}) are projected to the gaze camera using $\mathbf{R}_t$ and $\mathbf{T}_t$. Apply the "distance constraint" to disambiguate the eye gaze.

Step 10) Let $t = t + 1$; go to step 5.

Appendix B
Calibration of (ρ_x, ρ_y, ρ_z)

We defined a reference coordinate system (H, X_r, Y_r, Z_r) in the pan-tilt unit in Appendix A; see the top of Fig. 20. The vector (ρ_x, ρ_y, ρ_z) with respect to the world coordinate system (O_w, X_w, Y_w, Z_w) is needed for integrating the pose and gaze; see step 4 of Appendix A. We give a method to calibrate (ρ_x, ρ_y, ρ_z) as follows, where the vector is determined by calibrating H and optical center (P_0), respectively, in (O_w, X_w, Y_w, Z_w).

H is the intersection of the pan (Y-axis) and tilt (X-axis) rotation axes. The optical center of the gaze camera can be determined in the world coordinate system by decomposing the transform matrix of the gaze camera. Keeping the pan of the pan-tilt unit zero, the gaze camera is calibrated with respect to the same world coordinate system in the following three positions of the optical center; see the middle of Fig. 20.

1) P_0, where the rotation angle about the tilt axis is zero, and the definition can be seen step 4 of Appendix A;
2) P_1, which rotates the pan-tilt an angle β about the tilt axis from P_0;
3) P_2, which rotates an angle β about the tilt axis from P_1.

Assume that the transform matrixes corresponding to P_0, P_1, and P_2 are $\mathbf{M}_0$, $\mathbf{M}_1$, and $\mathbf{M}_2$, respectively. Consequently, the optical centers P_0, P_1, and P_2 that can be obtained by decomposing $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ are (x_0, y_0, z_0), (x_1, y_1, z_1), and (x_2, y_2, z_2), respectively. Obviously, P_0, P_1, and P_2 are positioned at the same plane (represented as π), namely, the Y-Z plane of the reference coordinate system. β is known, and hence, the length

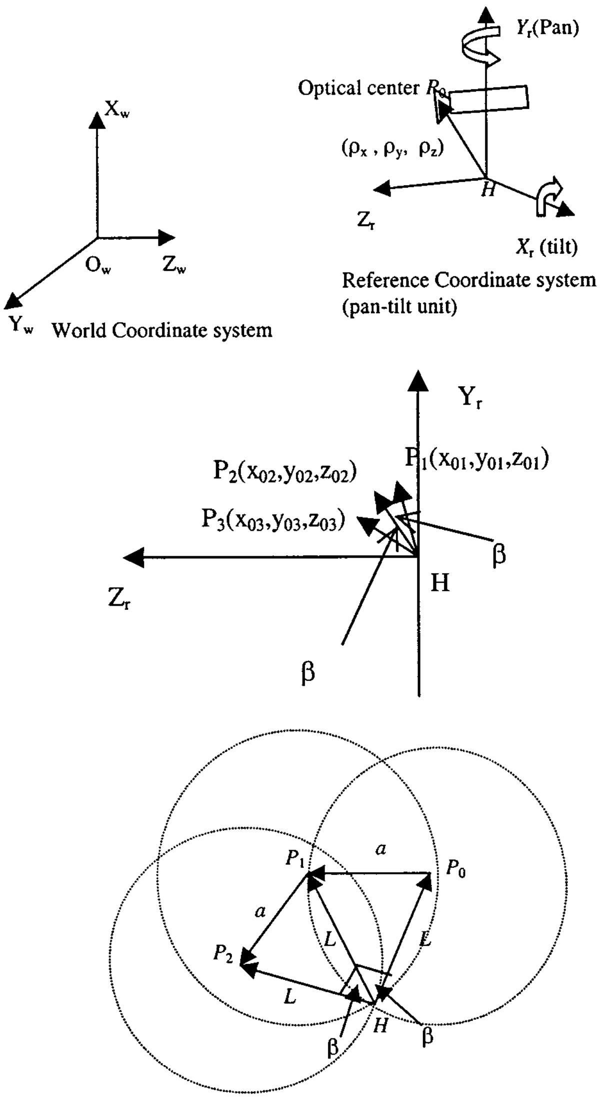


Fig. 20. Calibration of the origin of the reference coordinate system. (Top) Definition of (ρ_x, ρ_y, ρ_z). (Middle) Gaze camera calibrated with respect to the world coordinate system in three positions, respectively. (c) H is intersected by the three circles centered as P_0, P_1, and P_2, respectively, and having the same radius L.

between $H(x_{or}, y_{or}, z_{or})$ and the center of the gaze camera coordinate system can be calculated using the cosine theorem; see the bottom of Fig. 20:

$$L = \frac{a}{\text{sqrt}(2^*(1-\cos(\beta))} \tag{B.1}$$

where

$$a = \text{sqrt}\left((x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2\right). \tag{B.2}$$

In plane π, we derive three circles that, respectively, are centered as P_0, P_1, and P_2 and having the same radius L. H is the intersection of the three circles. This can be seen in the bottom of Fig. 20. Using H and P_0, we have

$$(\rho_x, \rho_y, \rho_z) = (x_0 - x_{0r}, y_0 - y_{0r}, z_0 - z_{0r}). \tag{B.3}$$

Acknowledgment

The authors are grateful to the anonymous reviewers for some useful suggestions that helped strengthen the paper.

References

[1] S. Barattelli, L. Sichelschmidt, and G. Rickheit, "Eye-movements as an input in human computer interaction: exploiting natural behavior," in *Proc. Annu. Conf. IEEE Ind. Electron. Soc.*, vol. 4, Aug.–Sept. 1998, pp. 2000–2005.

[2] F. L. Bookstein, "Fitting conic sections to scattered data," *Comput. Graph. Image Process.*, vol. 9, pp. 56–71, 1979.

[3] C. Collet, A. Finkel, and R. Gherbi, "CapRe: a gaze tracking system in man-machine interaction," in *Proc. IEEE Int. Conf. Intell. Eng. Syst.*, 1997, pp. 577–581.

[4] C. Colombo, S. Andronico, and P. Darrio, "Prototype of a vision-based gaze-driven man-machine interface," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 1995.

[5] T. N. Cornsweet and H. D. Crane, "Accurate two-dimensional eye tracker using first and fourth purkinje images," *J. Opt. Soc. Amer.*, vol. 63, no. 8, pp. 921–928, August 1973.

[6] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1148–1161, Nov. 1993.

[7] J.-Y. Deng and F. Lai, "Region-based template deformation and masking for eye-feature extraction and description," *Pattern Recogn.*, vol. 30, no. 3, pp. 403–419, 1997.

[8] Y. Ebisawa, "Improved video-based eye-gaze detection method," *IEEE Trans. Instrum. Meas.*, vol. 47, pp. 948–955, Aug. 1998.

[9] O. D. Faugers, *Three Dimensional Computer Vision, A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.

[10] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting ellipse," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 476–480, May 1999.

[11] A. Fitzgibbon and R. B. Fisher, "A buyer's guide to conic fitting," in *Proc. British Machine Vis. Conf.*, 1995, pp. 513–522.

[12] D. Forsyth, J. L. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell, "Invariant descriptors for 3-D object recognition and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 971–991, Oct. 1991.

[13] A. Gee and R. Cipolla, "Determining the gaze of faces in images," *Imag. Vis. Comput.*, vol. 12, no. 10, pp. 639–647, Dec. 1994.

[14] R. M. Haralick and L. G. Shapiro, "Computer and robot vision," in *Mathematical Morphology*. Reading, MA: Addison-Wesley, 1993, ch. 5.

[15] B. Hu and M. H. Qiu, "A new method for human-computer interaction by using eye gaze," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 1994, pp. 2723–2728.

[16] T. E. Hutchinson, K. P. J. White, W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, pp. 1527–1533, June 1989.

[17] A. E. Kaufman, A. Bandopadhay, and B. D. Shaviv, "An eye tracking computer user interface," in *Proc. Res. Frontier Virtual Reality Workshop*, Oct. 1993, pp. 78–84.

[18] K.-N. Kim and R. S. Ramakrishna, "Vision-based eye-gaze tracking for human computer interface," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 2, Oct. 12–15, 1999, pp. 324–329.

[19] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Proc. Fourth Int. Conf. Automat. Face Gesture Recogn.*, 2000, pp. 499–504.

[20] D. Murray and A. Basu, "Motion tracking with an active camera," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 449–459, May 1994.

[21] G. A. Myers, K. R. Sherman, and L. Stark, "Eye monitor," *IEEE Comput. Mag.*, vol. 3, pp. 14–21, 1991.

[22] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, "Real-time stereo tracking for head pose and gaze estimation," in *Proc. Fourth Int. Conf. Automat. Face Gesture Recogn.*, 2000, pp. 122–128.

[23] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, pp. 62–66, Jan. 1979.

[24] J. Porill, "Fitting ellipses and predicating confidence envelopes using a bias corrected Kalman filter," *Image Vis. Comput.*, vol. 8, no. 1, pp. 37–41, 1990.

[25] T. Rikert and M. Jones, "Gaze estimation using morphable models," in *Proc. Fourth Int. Conf. Automat. Face Gesture Recogn.*, 1998, pp. 436–441.

[26] P. L. Rosin, "Further five-point fit ellipse fitting," *Graph. Models Image Process.*, vol. 61, pp. 245–259, 1999.

[27] ——, "A survey and comparison of traditional piecewise circular approximations to the ellipse," *Comput.-Aided Geom. Des.*, vol. 16, pp. 269–286, 1999.

[28] R. Safaee-Rad, I. Tchoukanov, K. C. Smith, and B. Benhabib, "Three-dimensional location estimation of circular features for machine vision," *IEEE Trans. Robot. Automat.*, vol. 8, pp. 624–639, Oct. 1992.

[29] H. S. Sawhney, J. Oliensis, and A. R. Hanson, "Description and reconstruction from image trajectories of rotational motion," in *Proc. IEEE Int. Conf. Comput. Vision*, 1990, pp. 494–498.

[30] R. Stiefelhagen and J. Yang, "Gaze tracking for multimodal human-computer interaction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1997, pp. 2617–2620.

[31] E. Sung and J. G. Wang, "Head pose and eye gaze for human-computer interface, invited paper on "Computer vision in HCI"," in *Proc. Sixth Int. Conf. Contr., Automat., Robotics, Vis.*, Singapore, Dec. 2000.

[32] K. Talmi and J. Liu, "Eye and gaze tracking for visually controlled interactive stereoscopic displays," *Signal Process.: Image Commun.*, vol. 14, pp. 799–810, 1999.

[33] R. M. Taylor and P. J. Probert, "Range finding and feature extraction by segmentation of images for mobile robot navigation," in *Proc. IEEE Int. Conf. Robotics Automat.*, Minneapolis, MN, Apr. 1996, pp. 59–100.

[34] J. G. Wang, "Head-Pose and Eye-Gaze Determination for Human-Machine Interaction," Ph.D. disseration, School Elect. Electron. Eng., Nanyang Technolog. Univ., Singapore, 2001.

[35] J. G. Wang, Y. M. Hao, H. K. Cao, and X. P. Xu, "A visual navigation system for autonomous underwater vehicle," in *Proc. MVA IAPR Workshop Machine Vis. Appl.*, Kawasaki, Japan, Dec. 13–15, 1994, pp. 544–547.

[36] J. G. Wang and E. Sung, "Gaze determination via images of irises," in *Proc. 11th Brit. Machine Vis. Conf.*, vol. 1, Bristol, U.K., Sept. 11–14, 2000, pp. 132–141.

[37] ——, "Pose determination of human faces by using vanishing points," *Pattern Recognit.*, vol. 34, no. 12, pp. 2427–2445, 2001.

[38] J. G. Wang and Y. F. Li, "Human assisted environment modeling for robots," *Auton. Robots*, vol. 6, no. 1, pp. 89–103, 1999.

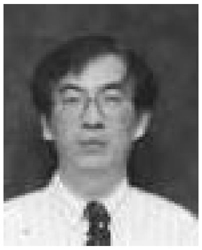
[39] X. Xie, R. Sudhakar, and H. Azhang, "On improving eye feature extraction using deformable templates," *Pattern Recognit.*, vol. 17, pp. 791–799, 1994.

[40] L. R. Young and D. Sheena, "Survey of eye movement recording methods," *Beh. Res. Methods Instrum.*, vol. 7, no. 5, pp. 397–429, 1975.

[41] H. K. Yuen, J. Illingworth, and J. Kittler, "Detecting partially occluded ellipses using the hough transform," *Image Vision Comput.*, vol. 7, no. 1, 1989.

Jian-Gang Wang received the B.E. degree in computer science from Inner Mongolia University in 1985, the M.E. degree in pattern recognition and machine intelligence from Shenyang Institute of Automation, Chinese Academy of Sciences, in 1988, and the Ph.D. from Nanyang Technological University (NTU), Singapore, in 2001. His Ph.D. thesis was on head-pose and eye-gaze determination for human-machine interaction.

From 1988 to 1997, he was with the Robotics Laboratory of the Chinese Academy of Sciences, where he was appointed associate professor in 1995. From 1997 to 1998, he worked as a Research Assistant with the Department of Manufacturing Engineering and Engineering Management of the City University of Hong Kong. He joined NTU in June 1998. He is presently a Research Fellow with the Centre for Signal Processing, School of Electrical and Electronic Engineering, NTU. His research interests include 3-D computer vision, face recognition, autonomous robots, human-machine interaction, and virtual reality.



Eric Sung received the B.E. degree (Honors Class 1) from the University of Singapore in 1971, the M.S.E.E. degree from the University of Wisconsin, Madison, in 1973, and the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 1999, with a thesis on structure from motion from image sequences.

He lectured in the Electrical Engineering Department of Singapore Polytechnic from 1973 to 1978. Subjects taught included control engineering and industrial electronics. In 1975, he was sent on a one-year industrial attachment at the Singapore Senoko Power Station. From June 1978 to April 1985, he worked in design laboratories in multinational organizations such as Philips (Video), Luxor, and King Radio Corporation, designing televison and microprocessor-based communication products. He joined NTU in April 1985, where he is presently an Associate Professor with the Division of Control and Instrumentation of the School of Electrical and Electronic Engineering. His research interests include computer vision and microprocessor applications in automation. His current research interests are in structure from motion, stereovision, face and facial expression recognition, and machine learning.