Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Vision and Image Understanding 98 (2005) 83–103
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Estimating the eye gaze from one eye

Jian-Gang Wang[a,*], Eric Sung[b], Ronda Venkateswarlu[a]

[a] *Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613*
[b] *School of EEE, Nanyang Technological University, Singapore 639798*

Received 27 July 2004; accepted 27 July 2004
Available online 5 October 2004

Abstract

Eye gaze estimation via images has been well investigated and all methods worked on images with the full face. This makes the eye details small and accuracy is affected. We address this problem by zooming in on a single eye. In this paper, we will show the validity of the method and investigate the performance with controlled synthetic data and also with real images. The principle is to rely on the fact that the outer boundary of the iris is a circle. With a fully calibrated camera, its elliptical image can be back-projected into the 3D space yielding two possible circles. To disambiguate, the solution is found by making use of anthropomorphic knowledge of the structure of the eyeball. Hence, getting a larger eye image with a zoom camera enabled us to achieve higher resolutions and thereby higher accuracies. The robustness of the algorithm was verified by extensive statistical trials conducted on synthetic data and real images. The two key contributions in this paper are to show that it is possible to estimate eye gaze with only one eye image and that consequently this achieves higher accuracy of eye gaze estimation.

Keywords: Eye gaze; Head pose; Iris; One-circle; Circle/ellipse; Zoom camera

* Corresponding author.
E-mail addresses: jgwang@i2r.a-star.edu.sg (J.-G. Wang), eericsung@ntu.edu.sg (E. Sung), vronda@i2r.a-star.edu.sg (R. Venkateswarlu).

1077-3142/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2004.07.008

1. Introduction

Estimation of eye gaze is important in applications such as in human–computer interaction, and surveillance and has received attention in the research community in more recent years. Human visual line-of-sight consists of the pose of human head and the orientation of the eye within its sockets. We have worked on both aspects. However, this paper will only focus on the latter aspect. For head pose estimation, more details can be found in [1,2].

Most early eye-gaze and eye-tracking techniques are active methods and a good account of them is provided by Young and Sheena [3]. Techniques include electro-oculography [4], limbus, pupil, and eyelid tracking [5–7], contact lens method, corneal, and pupil reflection relationship [6], Purkinje image tracking [8], artificial neural networks [7], morphable models [9], and geometry [10–12].

More recent approaches tend to apply passive computer vision techniques [11,13–15]. Frequently, the iris contours are detected and used in various ways to determine eye gaze. However, the extraction of the outer boundary of the iris (limbus) is based on modeling it as a circle using a circle edge operator. For instance, Matsumoto and Zelinsky [11] detected the center of the iris using the circular Hough Transform. Kim and Ramakrishna [15] located the iris 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. [11] 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 3D "offset vector" from the mid-point 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" and the radius of the iris have to be manually adjusted through a training sequence where the gaze point of the person is known.

Our approach differs in that we model the image of the limbus as an ellipse and apply a simple projective geometry to derive two possible circles in 3D space. Originally, we made use of two irises and the common solution is taken to be the true normal direction of the eye gaze [16]. Having two irises in the same image resulted in small iris images. It is difficult to determine the eye gaze by analyzing the eyeball rotations from a typical image with low resolution for the eye region [10,17]. The iris is partially occluded by the upper and lower eyelids so it is difficult to fit its contour consistently and reliably. For example, in [15] 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 3D scene accurately with such low resolution. This led us to the idea of using only one iris and implies that we can zoom in on it to obtain a larger iris image. As 3D data are needed, the zoom camera use must be fully calibrated. The problem is then to find a way to disambiguate the unwanted solution from the true.

The principle of disambiguation is based on two facts. The first is that an ellipse on an image plane when back-projected into 3D space can originate from two possible circles. The second fact is that an anthropomorphic measure of the eyeball structure could be used to find the true circle of the limbus. In this paper, the unique supporting plane

was obtained based on a geometric constraint that is the distances between the eyeball center and the two eye corners should be equal to each other. We will refer to this as the "distance constraint." We will refer to this approach as the "one circle" method.

Domain knowledge played a key role in the success of the one-circle approach. Although the eyeball center cannot be seen, its location can be inferred from the pose of the head. This is because its average 3D location relative to the observed features is very close to a generic constant and can be fixed during model acquisition [12]. The ratio of the radius of a generic person's iris to his/her eyeball radius is found to possess very low ensemble variance and consequently we can fix this ratio at the generic average. The small variations from person to person of this ratio thus had no significant effect on the results as we will show later. The eyeball center can be located once the radius of an individual's iris is calibrated.

Our "one-circle" algorithm finds the eye gaze in the following way. The (vertical parts of the) limbus is extracted by standard edge detection (Canny) and morphological operations. Least-squares elliptical fit is applied to obtain the ellipse of the limbus. By projective geometry, we can find two circular solutions, the disambiguation of which is made possible by the "distance constraint". The correct circle is the normal direction to the limbus plane in 3D. We have to calibrate an individual's iris radius and with that the center of the eyeball is located. This together with the normal direction defines the eye gaze.

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 in the image. The other difference is that our method is more accurate since our method, needing an image of only one eye, allows us to zoom in on it achieving higher resolution. However, in doing so, we need to address an issue that emerges. Zooming and tracking a single eye poses a problem. The problem of having possible out-of-field views can be settled by guiding the "gaze" camera to locate the eye by the head pose estimation results. Considering that, we have developed a general approach that combines head pose determination with eye gaze estimation. The pose of the human head [2], including the 3D location of the eye corners, mouth corners and the orientation of the face, can be obtained from the head pose estimation subsystem. Therefore, our eye gaze estimation must work in tandem with the head pose system. The details of this integrated approach can be found in [1].

The approach to estimate the eye gaze is discussed in Section 2 and the iris detection in Section 3. Experimental results are given on simulated data as well as on real images in Section 4. The conclusion is discussed in Section 5.

2. Gaze positioning

As mentioned above, our gaze determination methodology relies on estimating the ellipse of the iris contour and by projective geometry, this leads to two normal solutions. To discard the false normal, we made use of the anthropomorphic ratio of the eyeball radius to the iris circle (limbus) radius. An algorithm box of the approach is given in Fig. 1.

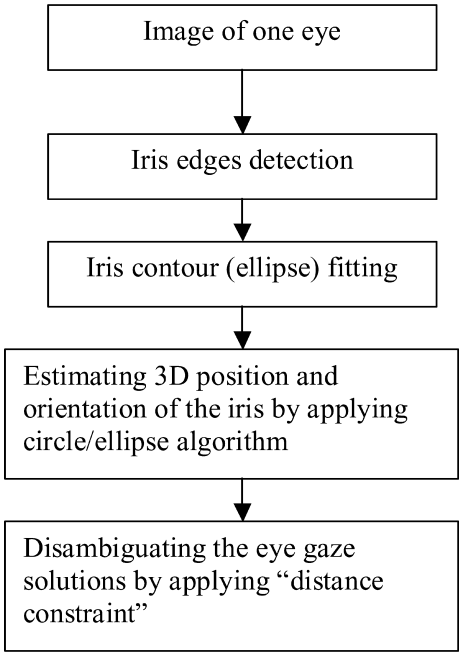

Fig. 1. Diagram of the gaze estimation approach.

The following explains how this is achieved.

2.1. Eye model

A simple eye model is defined (Fig. 2). We take the eyeball to be a sphere with radius R even though it is not quite a sphere. But this discrepancy does not significantly affect the final results. The iris is at the front of eyeball and its contour (limbus) is a circular ring of radius r (Figs. 2A and B). The distance from the center of the eyeball to the iris plane is d. The relation between R, r, and d (Fig. 2D) is:

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

We take the eye gaze to be the line passing through the center of the eyeball and the center of the iris. Note that this direction is the normal to the limbus plane. When changing the eye gaze, the eyeball rotates around its center (Fig. 2E). Our defined eye gaze keeps a nearly fixed angle (k, the angle between the visual and the anatomical axis of the eye) with the central gaze vector that is determined by the eye lens. By a simple calibration, the central gaze vector can be obtained from our estimated eye gaze.

The radius r of the generic iris is very close to an anatomical constant (around 7 mm) [12], and the radius R of the eyeball ranges from 12 to 13 mm according to anthropomorphic data [15]. Hence, this ratio R/r is stable with a small standard der-

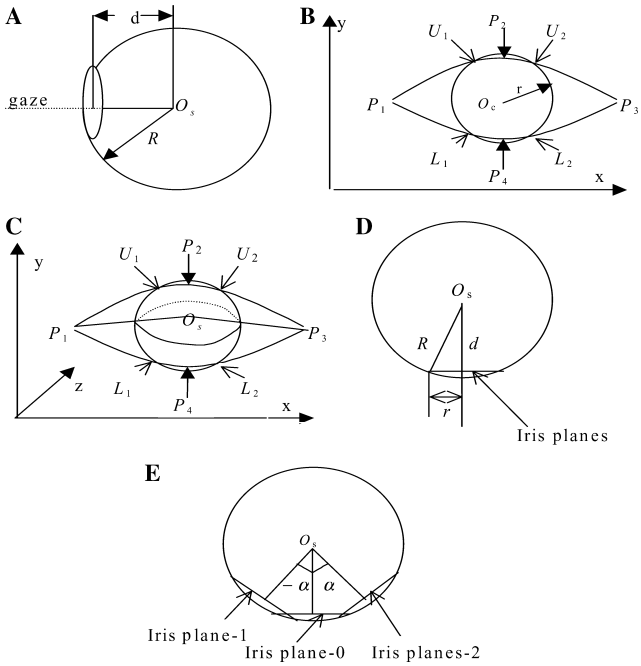

Fig. 2. The eye model.

ivation. In our method, we take this ratio to be the generic average based on the finding in [12]. Once r is found (manually), both R and d can be obtained consequently. In [11], both r and R are determined manually from a training sequence whereas our method only needs a one-time prior calibration to compute the individual's iris radius r. We will discuss the sensitivity of our gaze determination algorithm to the ratio R/r in Section 2.4.

To extract the iris ellipse in the unoccluded parts of the iris, we first have to detect the upper and lower eyelids which are modeled as parabolas (Fig. 2B). Let the upper eyelid pass 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 pass 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:

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

while for the lower eyelid yields:

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

In our simulations, the iris contour edges that lie between the upper and lower eyelids are consistently located and used to fit the elliptical contour. For instance, curves U_1L_1 and U_2L_2 shown in Fig. 2B are the fitting edges that we want. These fitting edges can be located by using the above equations of the iris and eyelids.

2.2. The "one-circle" algorithm

The problem of finding the two 3D circles from the image of an ellipse has been extensively investigated [1,18–20]. Our earlier work solved the problem of ambiguity by using the "two-circle" algorithm [16] where an image of two irises is used. Here, we show how we disambiguate by the "one-circle" algorithm which assumes that the camera is fully calibrated and that we are able to calibrate an individual's r value.

Assume the normalised camera coordinates where the real world coordinates coincide with the camera coordinates with origin at the optical center. Having found the optimal ellipse $\boldsymbol{Q}$ of the limbus, we can obtain two possible circular solutions, namely $\boldsymbol{n}_{\text{true}}$ and $\boldsymbol{n}_{\text{false}}$ being the unit normals to the circles C_{true} and C_{false} in space. By projective geometry, once r is known, this allows us to define the circle of the limbus. Since the center of the ellipse, whose coordinates can be read, must pass through the center of the iris circle, the 3D coordinates of the iris center $\boldsymbol{O}_{\text{c1}}$ is found. The reasoning is that for a cone, there can only be one circle of fixed size that can fit into the cone. Also, with r known, R is known from the generic average of R/r. And from Eq. (1), d is found. This leads to $\boldsymbol{O}_{\text{s}}$, the 3D coordinates of the eyeball center since

$$\boldsymbol{O}_{\text{s}} = \boldsymbol{O}_{\text{c1}} - d\boldsymbol{n}, \tag{5}$$

where $\boldsymbol{n} \in \{\boldsymbol{n}_{\text{true}}, \boldsymbol{n}_{\text{false}}\}$.

In our eye gaze estimation system, the 3D coordinates of the eye corners, $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, are from the head pose system [2]. However, the pose can be determined by any other approaches available. This is where we can disambiguate the false solution by invoking the "distance constraint" that was mentioned earlier. This requires

$$\|\boldsymbol{p}_1 - \boldsymbol{O}_{\text{s}}\| = \|\boldsymbol{p}_2 - \boldsymbol{O}_{\text{s}}\| \tag{6}$$

or

$$\|\boldsymbol{p}_1 - \boldsymbol{O}_{\text{s}}\| - \|\boldsymbol{p}_2 - \boldsymbol{O}_{\text{s}}\| = \varepsilon \tag{7}$$

taking into account errors due to noise. Two errors are computed, $\varepsilon_{\text{true}}$ and $\varepsilon_{\text{false}}$, respectively, for $\boldsymbol{n}_{\text{true}}$ and $\boldsymbol{n}_{\text{false}}$. It is expected that $\varepsilon_{\text{true}}$ $\varepsilon_{\text{false}}$ and our experiments verified this. In fact, we have found that the correct disambiguation is very tolerant to variations of R/r even up to 50% deviations from the generic average. The center of the eyeball is located along the line-of-sight. The angle between the two normal solutions, $\boldsymbol{n}_{\text{true}}$ and $\boldsymbol{n}_{\text{false}}$, is large if the head pose is not close to the (degenerate) frontal view (Fig. 3). Consequently, the separation of the two resulting eyeball centers is large enough to disambiguate the solutions based on the "distance constraint."

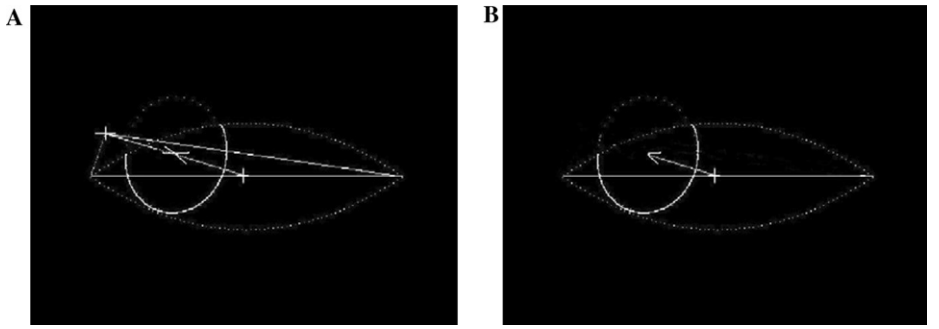

Fig. 3. Disambiguating the gaze solutions using the distance constraint when the rotation angle around Y-axis $-30°$ and X-axis $-15°$. (A) Two eyeball centers (cross points) are found. (B) The unique solution of the normal and iris center is found by the distance constraint.

2.3. Degenerate cases of "one-circle" algorithm

When will the "one-circle" algorithm not able to distinguish between the two candidate solutions? Such a situation arises when the iris contour is symmetric about the Y–Z plane of the camera, where 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 solved by simplifying checking on 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.

2.4. Sensitivity of the "distance constraint" to the ratio

As our method depends on the ratio R/r to be equal to the generic average, we should check the sensitivity of the algorithm to its deviations. 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 by $\pm 50\%$.

3. Iris detection

Our algorithm picks the most reliable features within the eye region which are outer edges of the iris. The popular iris detection methods using circular edge operators obviously cannot be adopted for our method. Instead, the bright-to-dark and dark-to-bright iris edges are detected using a 3×3 vertical edge operator and a 3×3 morphological "open" operation. The locations of the eye corners have to be located in the prior pose determination algorithm [2]. Hence, the iris detection search can be

executed on a small region between them. Because of the high contrast between the iris and sclera, the eye image is easily segmented by a thresholding value that is automatically selected from the histogram [21] of the eye region from frame to frame. "Opening" an image breaks narrow isthmuses. 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 sufficient. After that, the Canny edge operator is used to detect the edges of the iris and the edges with directions $90° \pm 5°$ are retained. All of the vertical edge segments are then tracked, respectively, using an "edge following" technique. The lengths of the edges are obtained. The two sides of the iris are the ones with the two longest vertical edges.

Once the iris edges are obtained reliably, the iris contour is fitted to an ellipse. It is important to get an accurate estimate of the ellipse since it is the pivotal point of our algorithm. Hence a robust ellipse fitting method is needed and we have adopted the method of Fitzgibbon et al. [22].

The steps in the iris detection process that we have implemented are shown in Fig. 4. An original eye image is shown in Fig. 4A. First, the original image with a single eye (Fig. 4A) is adaptively segmented by the histogram to yield the binarised image (Fig. 4B). Note that some of the eyelids are connected with the iris. This necessitates the application of a morphological "open" operation to cut the connection (of

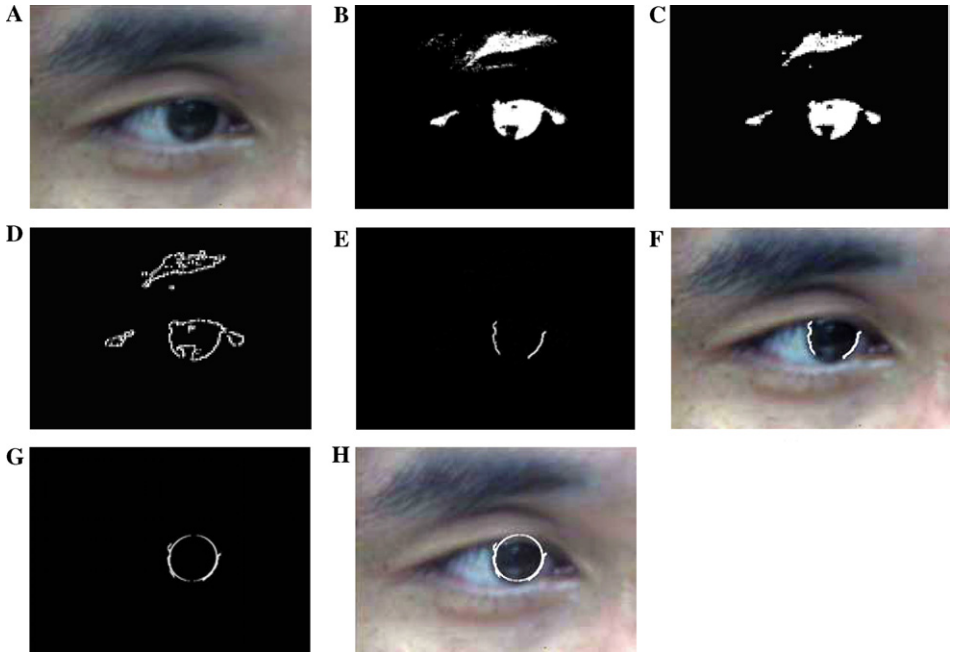

Fig. 4. Iris detection: (A) Original image, (B) thresholding results, (C) morphological "open" operation, (D) vertical edges, (E) two longest edges by region-following, (F) overlay edges onto the original image, (G) edge result and least-squares fitted ellipse, and (H) overlay edge result and least-squares fitted ellipse onto the original image.

narrow isthmuses) resulting in Fig. 4C. A vertical (Canny) edge operator is resulting in Fig. 4D. The two longest edges are deemed the iris edges to be used for fitting the iris contour (Fig. 4E) and their superimposition is shown in Fig. 4F. Hence the two outer boundaries of the iris not occluded by the eyelids are used for ellipse fitting by the algorithm of Fitzgibbon et al. [22]. The resulting ellipse can be seen in Fig. 4G and its superimposition onto the eye in Fig. 4H.

4. Experimental investigation and results

We will test our "one-circle" approach first with synthetic data and then with real images. The purpose of the former is to ensure that under noiseless conditions, we should expect perfect results in order to verify the principle of the method. Due to pixel discretisation, some small errors were noted. Next, the synthetic data was perturbed by Gaussian noise at different levels to simulate the geometric shifts arising from imperfect image capture and image processing algorithms. Also, the expected accuracies of the eye gaze wrt the point-of-regard were also investigated. Finally, we put our algorithm to the test with real images.

All the experiments on synthetic data have been done extensively and are statistically valid. Our results verified the accuracy and robustness of the method. The various experiments are presented below.

4.1. Experimental results on synthetic data

We test the algorithm on two sets of synthetic data: exact and noisy. By observing the pre-defined targets, the accuracy of the algorithm can be measured.

4.1.1. Simulations on exact data

An eye model was simulated in which the iris can be rotated in every possible direction. The head was fixed facing directly at the camera. Then, without moving the head, the "eye" is made to gaze in various directions. For each gaze, the eyeball is made to rotate accordingly about its center. Consequently, the iris contour will be rotated to the position which corresponds to the expected gaze, following the eyeball movement. A camera coordinate system is set up and the iris contour and the two eye-lids are projected onto the pre-defined image plane. Then the gaze is estimated as per our algorithm. Only the approximately vertical edges between the two eyelids are considered for computing the eye gaze.

In our simulations, the initial contour of the iris (circle) lies parallel to the image plane at a distance of 60–100 cm. The radius of the iris is set as 0.65 cm. The ratio of R/r is fixed at 2. The size of the image is set to be 640×480. The intrinsic parameters of the camera are set as:

$$u_0 = 320, \quad v_0 = 240, \quad f_x = f_y = 5500,$$

where (u_0, v_0) are the coordinates of the principle point, f_x and f_y are the scale factors of the camera along the X- and Y-axis, respectively. The settings of f_x and f_y here

imply that the zoom camera requires a shorter focal length in order for the eyes to appear big enough on the image. 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 $1°$ (elevation) to form a set of synthetic images. The performances of the "one-circle" algorithm are tested on these synthetic images

The errors of the eye gaze over the synthetic test images are illustrated in Fig. 5A. We can see that the accuracy tends to fall greater when face is around the frontal view ($\pm 2°$), and falls lesser as the gaze turns away from the fronto-parallel position. This is expected 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, e.g., $5°$. Hence, we will consider the errors of the eye gaze in the

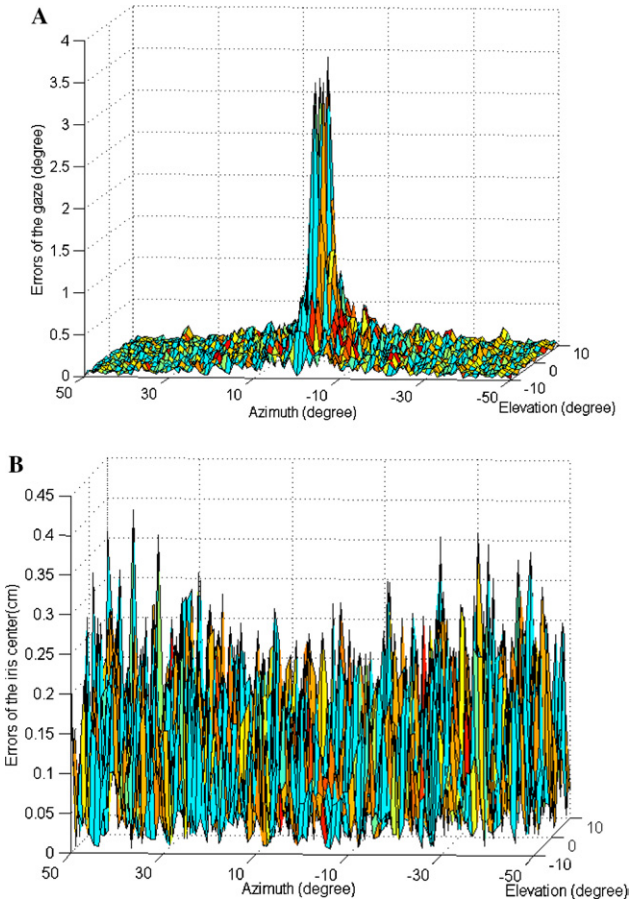

Fig. 5. The errors for different poses for the case of noiseless synthetic data: (A) the normal of the iris and (B) the center of the iris.

following by excluding this degenerate case. The errors of the iris center are shown in Fig. 5B.

Actually, the errors should be zero but because of the pixel discretisation, there would be errors arising from the estimation of the ellipse. The experimental results for noiseless data show that using the ''one-circle'' method the maximum error of the gaze due to the eyelids' occlusion is 0.3°, while the maximum error of the center of the iris is 0.1 cm, this can be seen in Figs. 5A and B, respectively. This verifies the principle of the one-circle method and at the same time gave an indication that the lower bound of errors is dictated by the resolution of the image.

4.1.2. Robustness to geometrical disturbances

As the location of the iris contour points cannot be perfectly derived, it is necessary to find out how sensitive the algorithm will be to incorrect pixel locations. So the algorithm is tested under the case when the data are subjected to geometric disturbances (i.e., pixel shifts) and it is done by corrupting the locations of the iris contour edges by Gaussian noise of different standard deviations. This simulates the condition of imprecise location of the features caused by the feature extraction algorithm, noise and other processes.

The mean error from 100 random trials for gaze per standard deviation is computed to indicate the robustness of gaze under geometric disturbances. Different levels of Gaussian noise with standard deviations of 1, 1.4, 2.0, and 2.8 pixels were tested. The results showed that the algorithm is robust. The algorithm can work even when noise is increased significantly, with the resulting accuracy degrading gracefully. With a standard deviation of one pixel, the errors of the estimated gaze over the testing range are less than 1°, and the errors of the estimated iris center are less than 0.3 cm.

4.1.3. Accuracy of the eye gaze versus point-of-regard

The ''one-circle'' method proposed here would be able to determine the eye gaze, but it cannot detect point-of-regard from one image. To do the latter, two views would be needed. However, the point-of-regard can be determined by one view with the ''one-circle'' method if we assume that the person is looking at a point on a fixed plane.

In the simulation, we assume a plane-of-regard 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 setup is shown in Fig. 6. The settings of the camera and the subject are the same as those in Section 4.1.1. The focus plane is defined in the gaze camera coordinate system, and 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. 6): the iris plane is rotated around the center of the eyeball O_s until the line O_sP becomes the normal of the iris

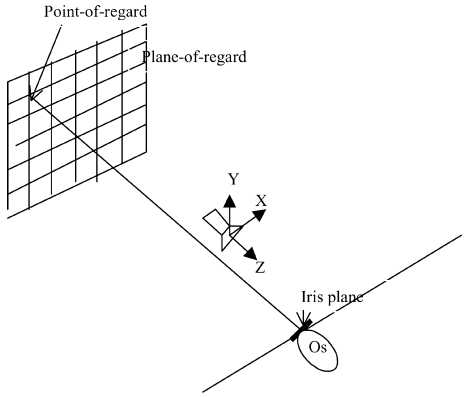

Fig. 6. Experiments on the accuracy of the eye gaze versus the focus points.

plane. From that, the eye gaze and consequently the point-of-regard are estimated from the synthetic image. The sensitivity of the iris pose to gaze accuracy is tested.

The results, represented in terms of distance (cm) and angle (degree), are shown in Figs. 7A and B, respectively, where in Fig. 7A the true and estimated point-of-regards are marked as "O" and "+," respectively. The testing range on the plane-of-regard is set to be −150 to 150 cm in 30 cm steps for the horizontal displacement and −25 to 25 cm in 5 cm steps for the vertical displacement. The mean error from 100 random trials for gaze per standard deviation is computed and shown in Figs. 8A and B, where the results are presented in terms of distance (cm) in Fig. 8A) and in terms of angle (degree) in Fig. 8B. Similarly to the tests done in Section 4.1.2, noise levels of standard derivations of 1, 1.4, 2.0, and 2.8 pixels were used. The algorithm can work even when the noise level is increased significantly. Hence the error degrades gracefully.

The accuracy of the point-of-regard is found to be satisfactory. Corrupting the image with Gaussian noise (zero mean, standard deviation one pixel), the error is less than 1 cm within the range of 1.8 m (Fig. 8).

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.

4.2. Experiments on real images

We demonstrate our gaze estimation approach on real images of 10 subjects, including male, female, children, and wearing of eye glasses. As we shall see in the following sections, the experimental results on the real images 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°. Higher accuracy of the eye gaze determination has been attained because of the higher image resolution made possible and because we used reliable facial features such as eye corners and iris contour. The result is found to be better than existing non-intrusive

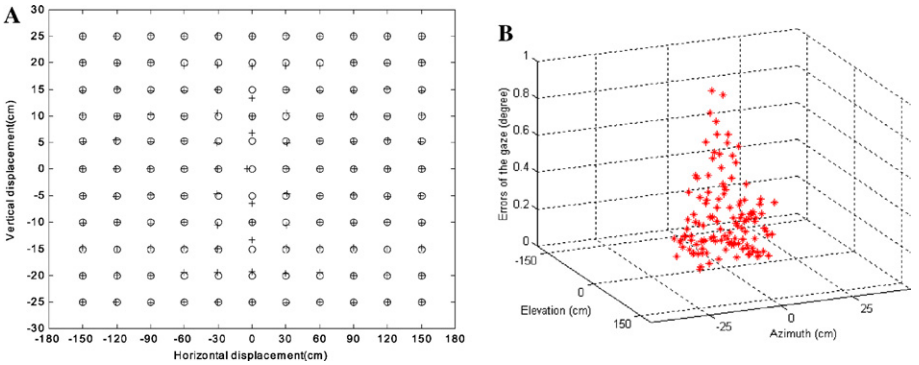

Fig. 7. The errors of the point-of-regard, in terms of (A) cm; (B) degree.

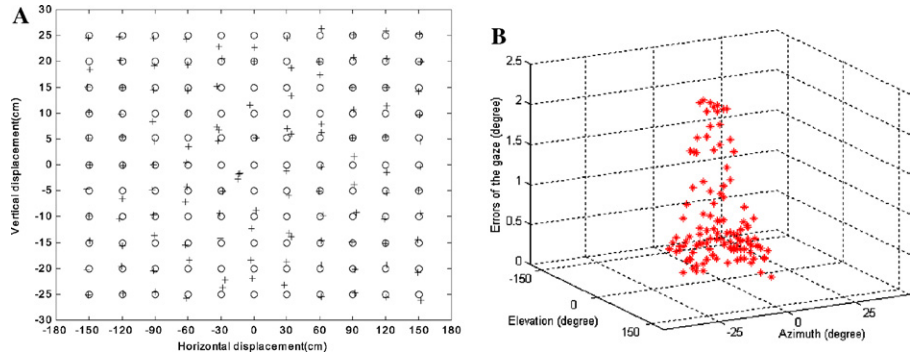

Fig. 8. The mean errors of the point-of-regard for the synthetic images from 100 random trials when the image is perturbed by a Gaussian noise of the 1 pixel standard deviation, in terms of (A) cm; (B) degree.

approaches, such as that of Zelinsky [12]. We will present some examples to discuss the problems encountered in iris detection, gaze determination, and in integration.

In the following examples, the radius of the iris contour is calibrated for each person. For instance, it is 0.63 cm for subject #1 shown in Fig. 10. The ratio between the radius of the eyeball and the radius of the iris is set to be 2.

Using the same method presented in Section 4.1.3, the accuracy of the eye gaze with respect to the point-of-regard is evaluated on real images.

The experimental setup is shown in Fig. 9. The person gazes in four pre-specified 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 equal segments. The size of the board is 150 cm (horizontal) $\times$ 50 cm (vertical). The effective area is 120 cm $\times$ 40 cm. The subject maintains his/her head stationary relative to the gaze camera while he/she gazes at each of the four targets in turn. Hence, the coordinates of his/her eye corners remain almost fixed for the four lines-of-sight. In Fig. 9B, $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, $\boldsymbol{n}_3$, and $\boldsymbol{n}_4$ correspond to the true eye gaze at points V_1, V_2, V_3, and V_4, respectively.

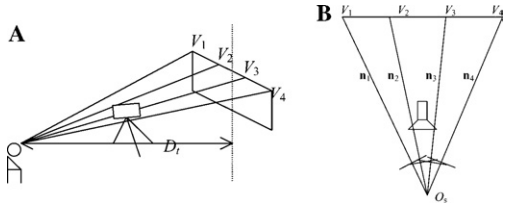

Fig. 9. Testing the accuracy of the eye gaze versus point-of-regard: (A) side view and (B) top view.

A camera, with 55 mm lens, is placed between the subject and the whiteboard. The resolution of the image is 640×480, where about 350 iris edges pixels can be detected and used to fit the iris contour. If we allow that the best the camera can give us is half pixel resolution, then the angular resolution of our method is $\arctan(0.5 * 0.078/55)$, i.e., 0.04°. The 3D 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 of the subject #1, #4, #5, and #7 are shown in Figs. 10–14.

The image coordinates of the eye corners are found from our previous work on head pose estimation [1,2]. This involves using another camera for estimating the head pose. The relationship between the gaze camera and the head pose camera is

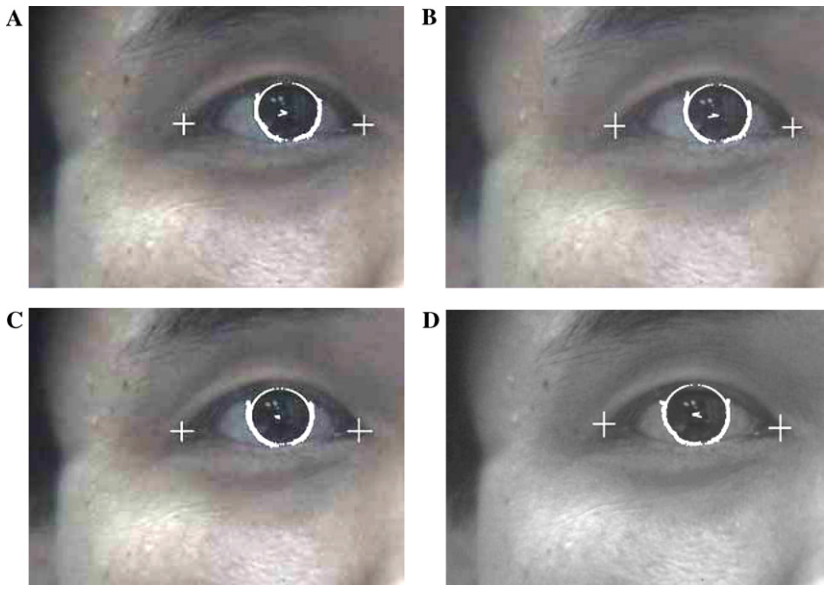

Fig. 10. Computed eye gaze (subject #1) for the four target points: (A) V_1, (B) V_2, (C) V_3, and (D) V_4.

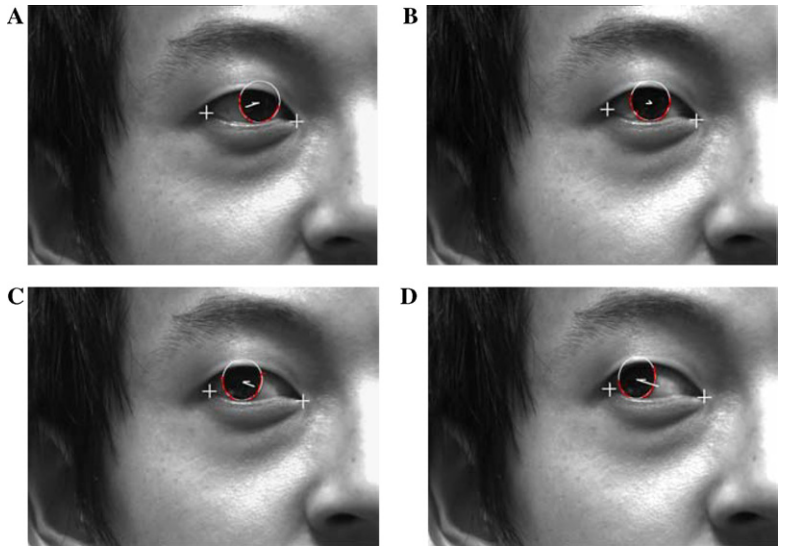

Fig. 11. Computed eye gaze (subject #4) for the four target points: (A) V_1, (B) V_2, (C) V_3, and (D) V_4.

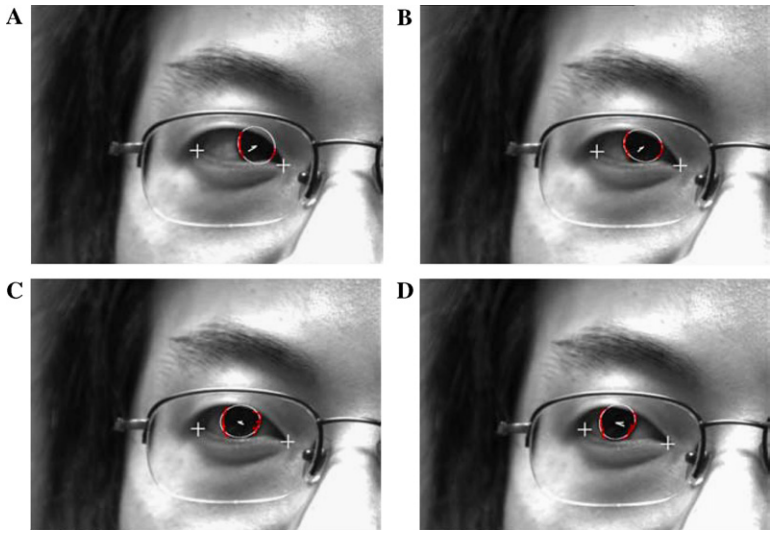

Fig. 12. Computed eye gaze (subject #4, wearing glasses) for the four target points: (A) V_1, (B) V_2, (C) V_3, and (D) V_4.

calibrated. Hence, the head pose can be projected onto the gaze camera system. In this paper, we will report only on the gaze estimation.

Applying the ''distance constraint'' to disambiguate the solutions of the normal and the center, the results on subject #1 are listed in Table 2, where the unique solu-

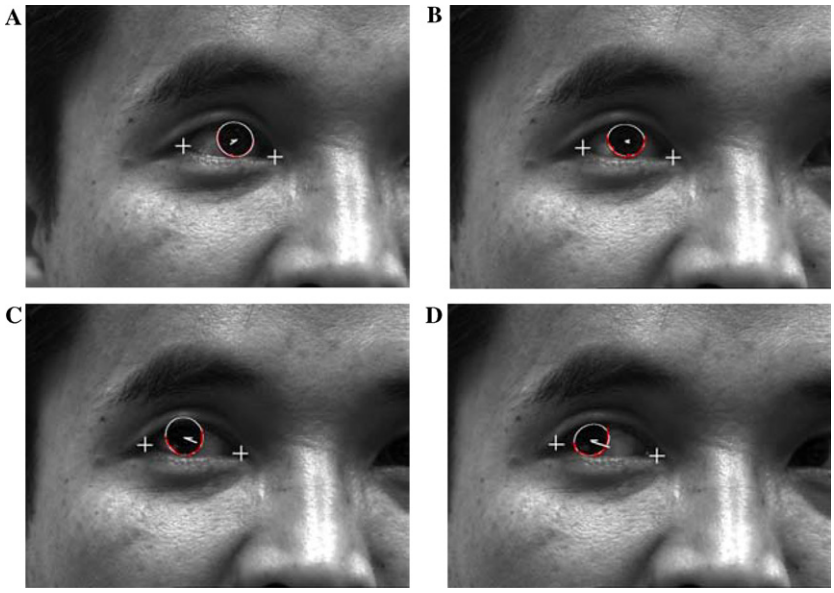

Fig. 13. Computed eye gaze (subject #5) for the four target points: (A) V_1, (B) V_2, (C) V_3, and (D) V_4.

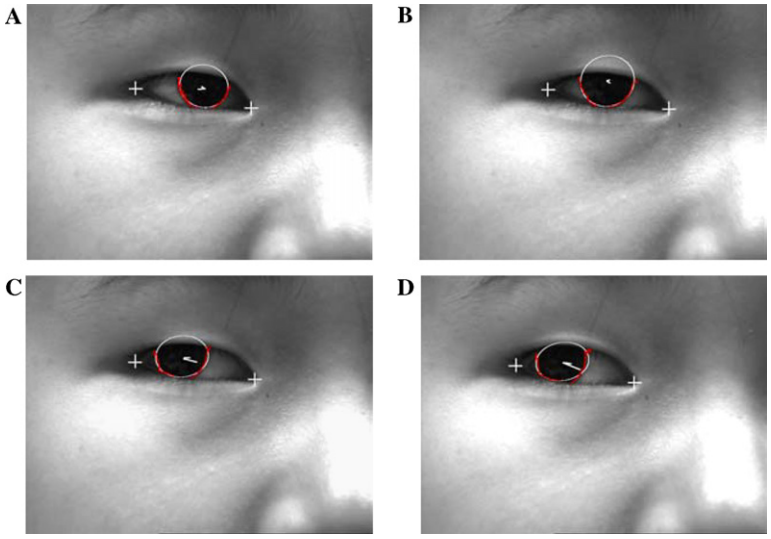

Fig. 14. Computed eye gaze (subject #7) for the four target points: (A) V_1, (B) V_2, (C) V_3, and (D) V_4.

tion is marked as "T." The eye gaze ($\boldsymbol{n}$), the center of the iris (O_c), and consequently the center of the eyeball (O_s) are listed in Table 2. D_L and D_R are the distance from O_s to the two eye corners (R and L in Table 1), respectively.

Table 1
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, 0028, 59.301)

Table 2
Gaze results and the eye gaze disambiguation by applying "distance constraint"

	$\boldsymbol{n}$	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

$$D = |D_L - D_R|. \tag{8}$$

We observe that the errors of point-of-regard are less than 1.5 cm within 1.5 m (D_t) range, and consequently the errors of the gaze are all less than 1°. The errors, E_p, of the point-of-regard are then estimated and are shown in Fig. 15. The errors of the eye gaze are also estimated by

$$E_g = \tan^{-1}(E_p/D_t), \tag{9}$$

where D_t is the distance from the subject to the plane-of-regard, see Fig. 9A.

Similar experiments have been done on testing the accuracy when a subject observes the four corners of a monitor. In general, the farther the observer look away from center point of the effective area, the less will be the number of the fitting points and the lower will be the accuracy of the eye gaze consequently. This can be seen by comparing Figs. 11 and 16. The fitting points is 350 when the subject gaze at a board (Fig. 11), size of 150 cm (horizontal) × 50 cm (vertical), and it became 450 when the subject gaze at a monitor (Fig. 16), size of 40 cm (horizontal) × 33 cm (vertical). The size of the board we used is large enough to test the accuracy of the eye gaze with respect to the iris pose, that means the accuracy on a smaller size board, e.g., an monitor, will be the same or higher than the accuracy reported here.

In the experiment, it was found that the eye remains in view of the camera even with head movement of up to ±35° away from the frontal view.

In Zelinsky's approach [12], 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. A comparison between the Zelinsky's method and our method is given in Table 3.

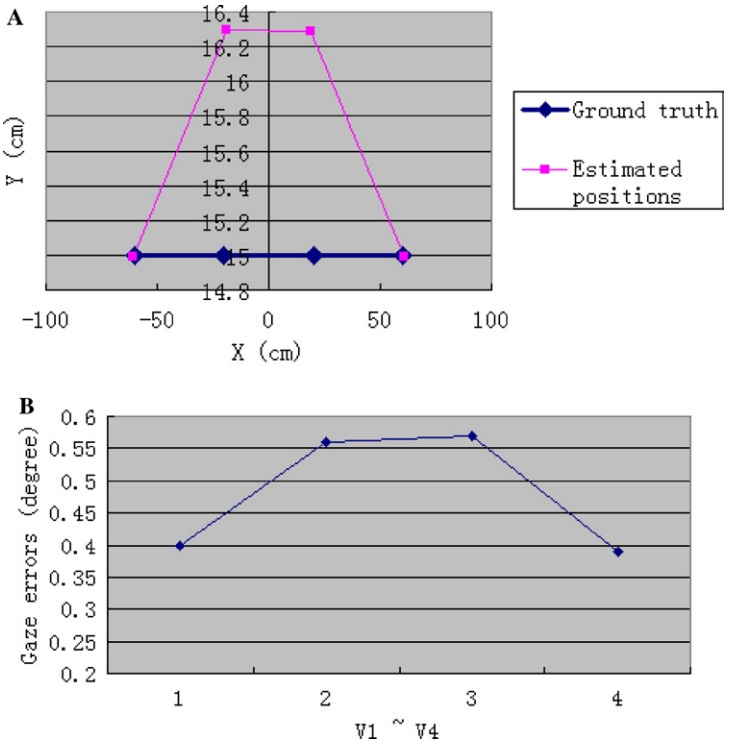

Fig. 15. Errors of the gaze estimation (A) errors of the point-of-regards (B) errors of the eye gaze.

Table 3
Comparison between Zelinsky's method and our method

	Width of eye in image (pixels)	Radius of iris in image (pixels)	Iris contour in image	Gaze accuracy
Zelinsky's method	30	5	circle	3.5°
Proposed method	180	35	ellipse	1°

We found the eye gaze can still be estimated even if the eye is not fully opened (see the examples in Fig. 17). However, the accuracy will be affected by the iris edges detection and the accuracy of the eye gaze depends on the number of the iris fitting points. In the above experimental results, two iris edges were detected and used for fitting the iris contour. The algorithm will fail to find two iris edges when the iris moves to the extreme when only one correct iris edge can be found, as shown in Fig. 18. An inaccurate gaze resulted although the iris contour can still be fitted. In our algorithm, the case can be found by comparing the positions of the iris edges and the position of the eye corner. Error information is reported when the distance from the position of one eye corner to the iris edge is less than a threshold.

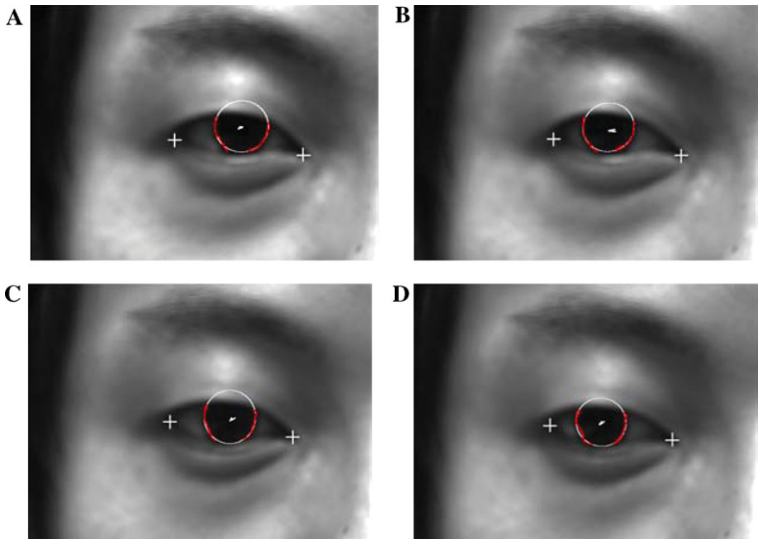

Fig. 16. Computed eye gaze (subject #4) for the four corner points of a monitor: (A) left top, (B) right top, (C) right bottom, and (D) left bottom.

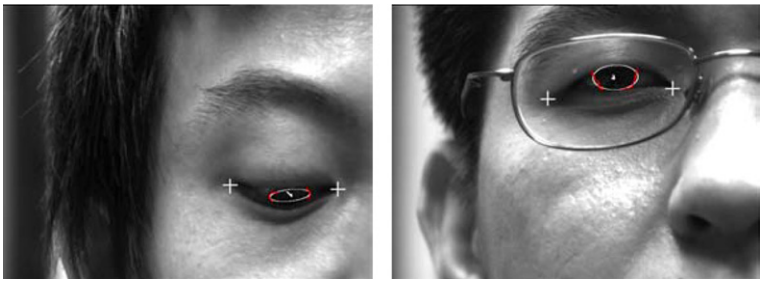

Fig. 17. The eye gaze can be estimated when the eye is not fully opened.

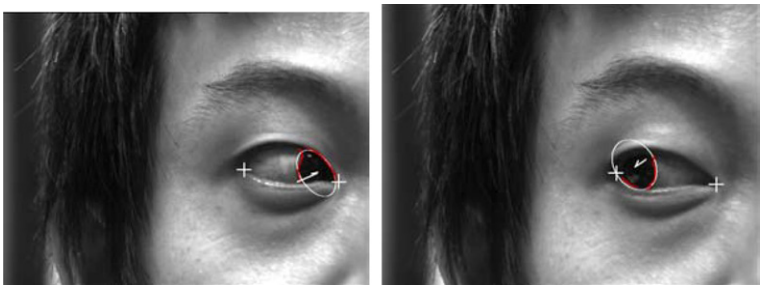

Fig. 18. Inaccurate iris gaze resulted when the eye move horizontally to the extreme.

5. Conclusion

In this paper, we presented a non-intrusive method of robustly estimating the eye gaze from higher resolution image of the iris. The motivation of our approach is to estimate the eye gaze robustly in real-time and with satisfactory accuracy from an image of one eye. The proposed method of eye gaze estimation works in tandem with the head pose system, with the head pose information driving the iris camera to track the eye robustly. The use of domain knowledge of the human face is crucial and this makes our paradigm original and novel. An ellipse can be back-projected into the space onto two possible circles. The principle is applied to the eye-gaze by observing that the (3D) contour of the iris is circular and hence it is the circle that we are looking for. By using the "distance constraint" described, we were able to discard the ambiguous solution. This constraint is based on anthropomorphic properties of the eye. We improve on current eye gaze determination methods by achieving higher resolutions. This comes about mainly because our algorithm is designed to use only a single image of the eye which subsequently means that we can zoom in on it.

In conclusion, eye gaze estimation is 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 potential benefits to these related applications. Of importance to note is that our method is non-invasive, fast and robust. It is robust because the segmentation of the iris contour is one of the simplest and most robust facial feature to extract.

References

[1] Jian-Gang Wang, Head-pose and eye-gaze determination for human–machine interaction, Ph.D. Thesis, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 2001.

[2] J.-G. Wang, E. Sung, Pose determination of human faces by using vanishing points, Pattern Recognition 34 (2001) 2427–2445.

[3] L.R. Young, D. Sheena, Survey of eye movement recording methods, Behavior Research Methods and Instrumentation 7 (1975) 397–429.

[4] A.E. Kaufman, A. Bandopadhay, B.D. Shaviv, An eye tracking computer user interface, in: Proceedings of Research Frontier in virtual reality workshop, IEEE Computer Society Press, 1993, pp. 78–84.

[5] G.A. Myers, K.R. Sherman, L. Stark, Eye monitor, IEEE Computer Magazine 3 (1991) 14–21.

[6] T.E. Hutchinson, K.P. White Jr., W.N. Martin, K.C. Reichert, L.A. Frey, Human–computer interaction using eye-gaze input, IEEE Transactions on System, Man and Cybernetics 19 (1989) 1527–1533.

[7] R. Stiefelhagen, J. Yang, Gaze tracking for multimodal human–computer interaction, in: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP-97, vol. 4, 1997, pp. 2617–2620.

[8] T.N. Cornsweet, H.D. Crane, Accurate two-dimensional eye tracker using first and fourth Purkinje images, Journal of the Optical Society of America 63 (1973) 921–928.

[9] T. Rikert, M. Jones, Gaze estimation using morphable models, in: Proc. Third Int Conf. on Automatic Face and Gesture Recognition, Nara, Japan April 14–16, 1998, pp. 436–441.

[10] A. Gee, R. Cipolla, Determining the gaze of faces in images, Image and Vision Computing 12 (1994) 639–647.

[11] Y. Matsumoto, A. Zelinsky, An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement, in: Proc. Fourth Int. Conf. on Automatic Face and Gesture Recognition, Grenoble, France, March 28–30, 2000, pp. 499–504.
[12] R. Newman, Y. Matsumoto, S. Rougeaux, A. Zelinsky, Real-time stereo tracking for head pose and gaze estimation, in: Proc. Fourth Int. Conf. on Automatic Face and Gesture Recognition, Grenoble, France, March 28–30, 2000, pp. 122–128.
[13] X. Xie, R. Sudhakar, H. Azhang, On improving eye feature extraction using deformable templates, Pattern Recognition 17 (1994) 791–799.
[14] J.G. Daugman, High confidence visual recognition of persons by a test of statistical independence, IEEE Transactions on PAMI 15 (1993) 1148–1161.
[15] K.-N. Kim, R.S. Ramakrishna, Vision-based eye-gaze tracking for human–computer interface, in: Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, Tokyo, JAPAN, 12–15, Oct. 1999, vol. 2, pp. 324–329.
[16] J.-G. Wang, E. Sung, Gaze determination via images of irises, Image and Vision Computing 19 (2001) 891–911.
[17] K. Talmi, J. Liu, Eye and gaze tracking for visually controlled interactive stereoscopic displays, Signal Processing: Image communication 14 (1999) 799–810.
[18] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, C. Rothwell, Invariant descriptors for 3-D object recognition and pose, IEEE Transactions on PAMI 13 (1991) 971–991.
[19] R. Safaee-Rad, I. Tchoukanov, K.C. Smith, B. Benhabib, Three-dimensional location estimation of circular features for machine vision, IEEE Transactions on Robotics and Automation 8 (1992) 624–639.
[20] H.S. Sawhney, J. Oliensis, A.R. Hanson, Description and reconstruction from image trajectories of rotational motion, in: Proc. IEEE Int. Conf. on Computer Vision, 1990, pp. 494–498.
[21] N. Otsu, A threshold selection method from gray-level histogram, IEEE Transactions on Systems, man and Cybernetics, vol. SMC-9, 1979, pp. 62–66.
[22] A. Fitzgibbon, M. Pilu, R.B. Fisher, Direct least square fitting ellipse, IEEE Transactions on PAMI 21 (1999) 476–480.