Specialised Hough Transform and Active Contour Methods for Real-Time Eye Tracking

David Young, Hilary Tunley and Richard Samuels*

School of Cognitive and Computing Sciences
University of Sussex
Falmer, Brighton BN1 9QH
UK

tel. 01273 678483
fax. 01273 671320
email: d.s.young@cogs.susx.ac.uk

CSRP no. 386
July 1995

Abstract

It is useful to find the elliptical outlines of the pupil and iris in images of a human eye, obtained from a head-mounted camera, in order to automate one type of eye tracker. This task becomes highly constrained when a simple model of the eye is used. After calibration, the model has only two degrees of freedom corresponding to pan and tilt movements of the eye. We show how the model's constraints can be built into Hough transform and active contour methods at the lowest level, allowing high performance in speed, reliability and accuracy.

Keywords: Hough transform, active contour, eye tracking.

---

* Now at: Philosophy Department, Rutgers University, New Brunswick, NJ 08901, USA

INTRODUCTION

Finding the outlines of the pupil and iris in an image of a human eye (Figure 1) is a task of considerable practical usefulness, and also provides an interesting example of how general purpose image analysis techniques can be specialised to deal with a particular problem. In this case the problem is characterised by the availability of a good model of the object (that is, the eyeball), with only two degrees of freedom, so that in principle only a two-dimensional search is required. To carry this out as efficiently as possible requires that the model be exploited to the full at the lowest level of processing. We have therefore designed a Hough transform and an active contour method which are tailored to incorporate the model's constraints, allowing fast and robust identification of the features of interest. This paper is mainly concerned with these techniques, but we first briefly discuss the motivation for analysing this type of image in this way.

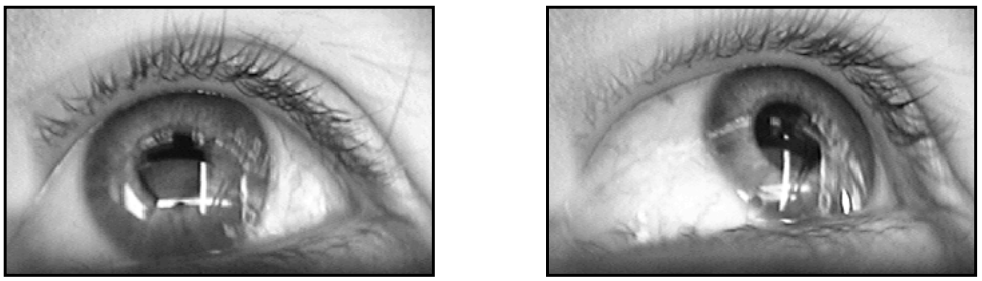

Figure 1. Views of an eye, produced by the equipment shown in Figure 2. The line of sight or direction of gaze relative to the head can be inferred from the position of the iris and pupil in these images, once the system has been calibrated. The linear highlights are accidental reflections and are not used by the system.

Measurements of eye-movements are valuable in understanding human vision and cognitive processes. (There is a large literature, but see as examples Yarbus' classic book[1] and the recent study by Land[2]. There is much current interest in the rôle of eye movements in schizophrenia and dyslexia.) To study eye movements, it is necessary to measure both the orientation of the head in a world coordinate frame, and the orientation of one or both eyes relative to the head. Here we are concerned with the second problem. Our approach is to use a head-mounted camera to view the eye and to identify and locate features such as outlines of the iris and pupil in the resulting image. The equipment (Figure 2) was designed by M. F. Land of the School of Biological Sciences at the University of Sussex.

Various other methods for monitoring eye movements exist, and some of them are available as commercial systems. They include electro-oculograms (EOGs), contact lenses fitted with mirrors, and infra-red reflections. EOG methods only yield the horizontal component of gaze direction, are subject to problems of drift, and are intrusive and cumbersome to set up since they involve placing electrodes on the face. Contact lens methods were important in the past, but are also very intrusive, and can hardly be used outside the laboratory. Infra-red reflection methods are convenient, although there are safety issues since the eye is illuminated with infra-red radiation, and they are limited in both accuracy and angular range. Land's approach using imaging of the eye has several advantages: it can be made relatively unintrusive, particularly as very small TV cameras are now available; it uses available illumination; it is portable; and it incorporates monitoring of head movements. Although Land analysed the images of the eye manually for his studies[2,3], this process is extremely time consuming, and cannot be used

where results are needed for immediate feedback.

Robertson, Craw & Donaldson[4] have recently described a similar system to ours for analysing images of the eye. A major difference between their work and ours is that they used a simple Hough transform for circles to detect the iris outline, whilst we have developed a specialised Hough transform operating directly in the parameter space of the eye model.

The eye has two degrees of freedom, which in cinematographic terms correspond to pan and tilt. (In principle, rotation about the line of sight would give a third degree of freedom, but in fact this torsional rotation is tightly coupled to the other two motions[5], and in any case could not be accurately measured from images of the eye.) The problem therefore lends itself to approaches in which a model with two free parameters is matched to the image. A method based on the Hough transform is well suited to finding the iris position in a single image, whilst one based loosely on active contours is useful in tracking the iris through a series of images. In both of these methods, the ellipses corresponding to the iris or pupil boundaries in the image are detected; once these are found, the direction of gaze relative to the head coordinate frame follows immediately. These methods exploit the fact that once the eye model has been established, the ellipse for, say, the iris is constrained to be a member of a family with two parameters: if the position of the iris centre is known, then the eccentricity and orientation of the major axis of the ellipse are fixed. Although both iris and pupil outlines can be utilised, we here restrict attention to the iris.

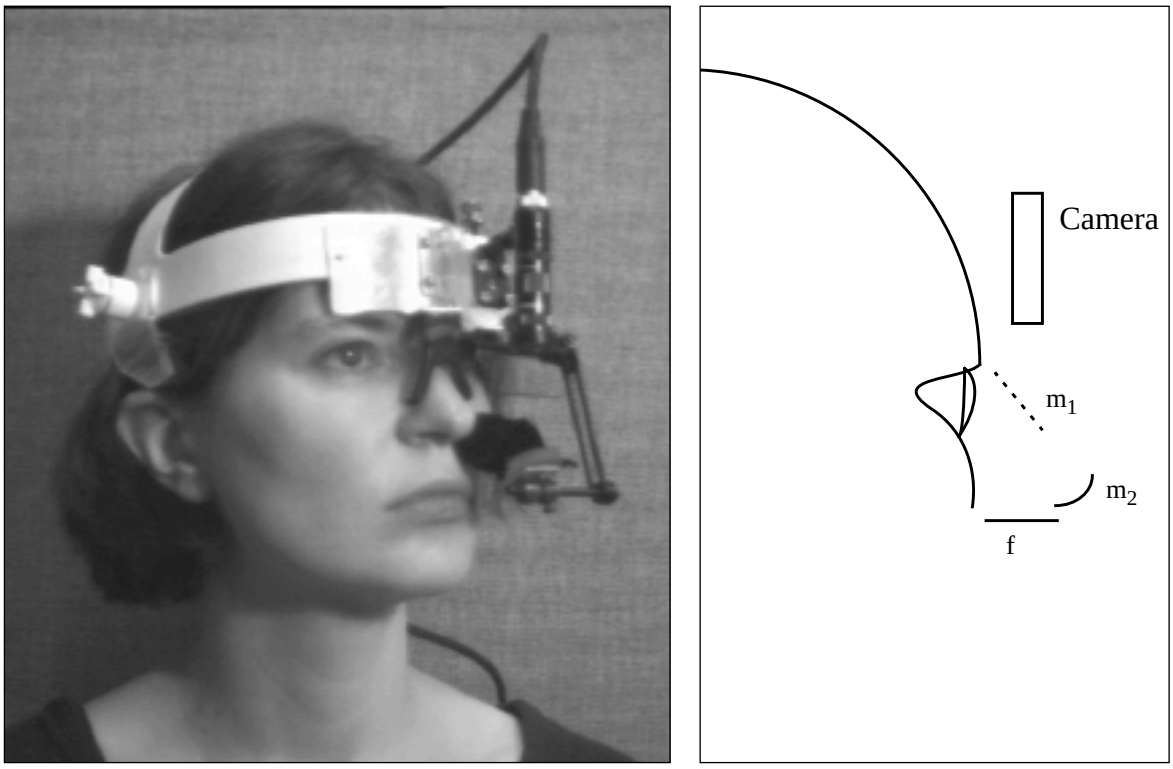

Figure 2. The equipment used to produce the images. In the schematic diagram at the right, the camera is looking downwards. The half-silvered mirror m_1 gives the camera a view of the scene in front of the person (as from a virtual camera in the orbit); the black felt *f* prevents the camera seeing the ground. The concave mirror m_2 gives the camera a view of the eye surface in part of its field. The camera and mirrors are all mounted rigidly on the head.

The speed required for the iris-finding operation depends on several factors. Analysing a video recording frame by frame may be possible, and in this case speed of processing is largely a matter of convenience. However, it may be essential for some applications to have real-time access to eye movement information, for example when an experiment requires that a display be updated in a way depending on gaze direction. Virtual reality and head-up information systems may also be able to exploit eye movements given sufficiently rapid measurements. For observations of eye movements in everyday activities, an important application of Land's equipment, the speed of operation required will depend on the particular task. When driving, a task requiring rapid dynamic pickup of visual information, saccades occur about 3 times a second[2], so a system will need to operate in appreciably less time than 1/3 s if useful information about individual fixation points is to be extracted. The algorithms described here were intended to be fast enough to be useful in applications that require some form of real-time processing. The speed achievable depends on the details of the hardware and on the preprocessing method, and there is also a trade-off between speed and accuracy. However, it is clear that the Hough transform computation would not be a limiting factor in a system required to make 3-10 measurements per second, whilst the active contour computation would not be a limiting factor in a system operating at normal video frame rates or faster.

THE EYE MODEL

The eye model used is very simple, but is sufficiently accurate to allow good tracking. The discussion is given in terms of the iris boundary, which is the most easily identified feature, but can equally be applied to the pupil boundary.

The eye is modelled as a sphere which can rotate about its centre, and the iris boundary as a circle on its surface. The imaging system is modelled by orthogonal projection, which is reasonable given that in the equipment used the optical distance from the eye to the camera lens is about 120 mm and the greatest depth change of the iris boundary is less than 10 mm. (Assuming orthogonal rather than weak perspective projection simplifies the discussion without loss of generality.) The model geometry is shown in Figure 3a. The iris boundary projects to an ellipse (Figure 3b), and the problem of estimating the gaze direction reduces to the problem of locating this ellipse in the image.

The model has four fixed parameters and two time-dependent parameters. The four fixed parameters are:

(X_c, Y_c): the position of the image of the centre of the eye;

R_i: the radius of the outer boundary of the iris;

R_e: the distance from the centre of the eye to the plane of the iris boundary.

Under orthogonal projection, we can take all dimensions to be in units of pixels. These four parameters are fixed in the sense that they will not change during a given recording or experiment if the apparatus does not move on the head, but they need to be measured on each occasion the system is used.

The two varying parameters specify the position in the image of the iris centre's projection. Once the equipment is calibrated, this specifies the direction of gaze in a coordinate system

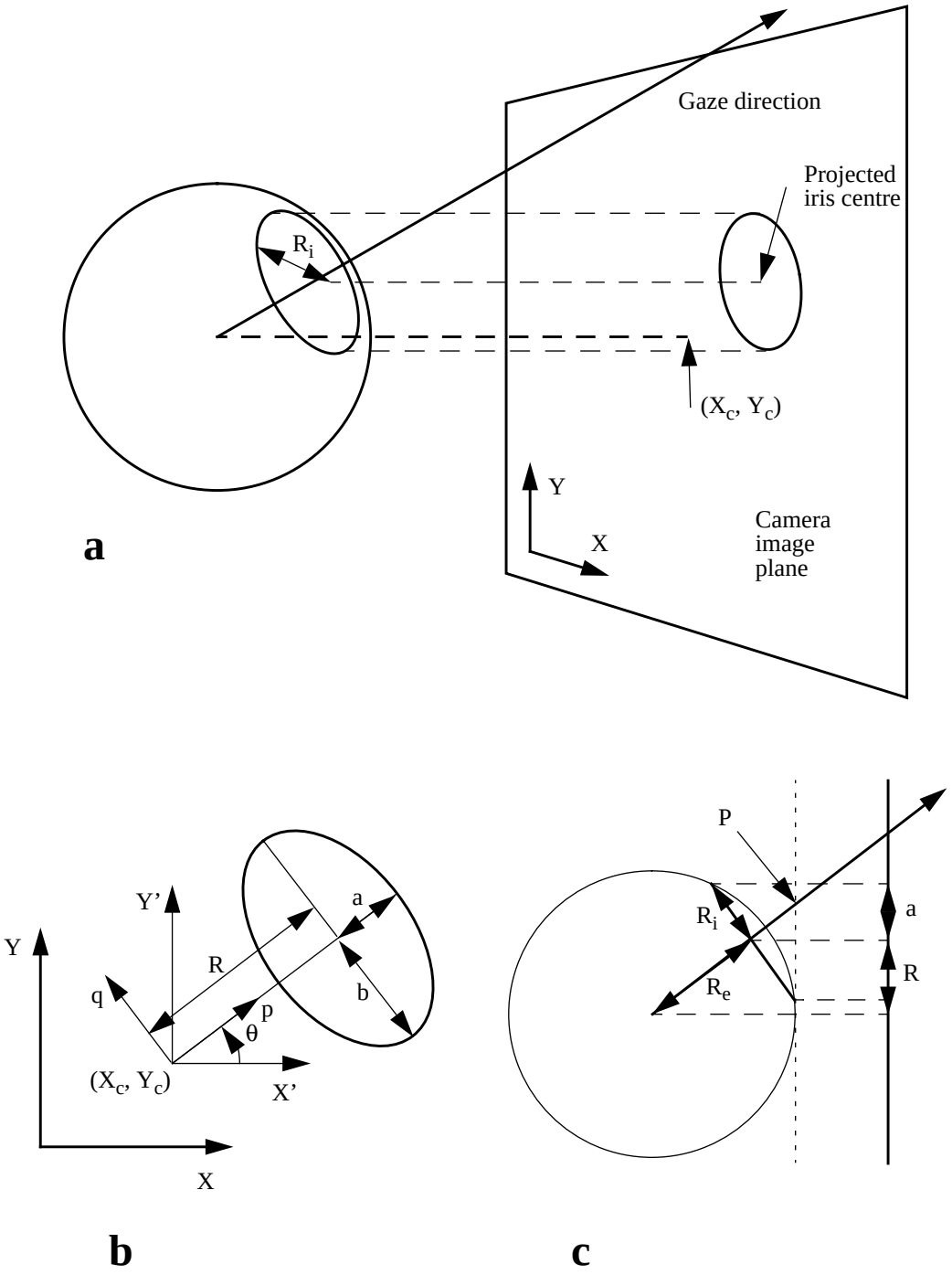

Figure 3. a Geometry for the eye model and its projection. **b** Coordinate systems used to describe the projection of the iris outline in the image plane. **c** Plane normal to the image plane and containing the line of sight, used to calculate the foreshortening of the iris image. The point marked P is the intersection of the line of sight with the tangent plane to the eye (shown dotted).

fixed to the head. We will specify this position relative to (X_c, Y_c), either using orthogonal coordinates (X_i, Y_i), or polar coordinates (R, θ). The projection of the iris centre is at (X_c+X_i, Y_c+Y_i) in image coordinates, and the representations are related by

$$\begin{aligned} R &= \sqrt{X_i^2 + Y_i^2} \\ \tan\theta &= Y_i / X_i \end{aligned} \qquad (1)$$

Determining the two parameters (X_i, Y_i) or (R, θ) for any given image is the task of the programs described below.

The elliptical image of the iris boundary is subject to constraints which follow from the fact that the eye rotates about its centre. The ellipse results from foreshortening of the circular iris boundary along the line joining the image of the iris centre to the image of the eye centre. The projection geometry is shown in Figures 3b and 3c, from which it may be seen that the minor axis of the ellipse lies at an angle θ to the image X-axis, and that the length of the semi-major axis, b, is equal to R_i. The length of the semi-minor axis, a, is related to R by

$$a^2 = R_i^2 - R^2\left(R_i^2 / R_e^2 \right) \qquad (2)$$

Assuming that the line of sight from passes through the centre of the eyeball and the centre of the iris, we can calculate its intersection with a plane tangent to the eyeball using

$$r^2 = R^2 R_e^2 / \left(R_e^2 - R^2 \right) \qquad (3)$$

where (r, θ) are the polar coordinates of the intersection point in the tangent plane. To study eye movements, we usually need to map the line of sight's direction onto the head-mounted camera's view of the scene. We assume that the intersection point in the eye's tangent plane and the position of the line of sight in the scene image are related by an affine transformation whose parameters can be determined by a calibration step. This calibration can be done by asking the subject to look at identifiable targets in the scene, and does not require any measurements of camera or scene parameters.

THE HOUGH TRANSFORM METHOD

Theory

Although a general ellipse has five parameters, there are only two variable parameters in the eye model, assuming that X_c, Y_c, R_i and R_e are fixed. Thus only a 2-D parameter space is necessary, and the Hough transform can be applied in a reasonable length of time.

The (R, θ) representation of the iris position is used; this gives straightforward equations for the Hough transform, and leads to a fairly uniformly filled accumulator array when applied to input features at random positions.

For any feature point in the image, the model gives a constraint equation relating R and θ, and elements of the Hough accumulator array consistent with this equation are incremented. We

find the constraint equation for R as a function of θ as follows.

We define two coordinate systems (Figure 3b). The first, *(X', Y')* is aligned with the image axes and centred on the projection of the eyeball centre. It is related to image coordinates by

$$\begin{aligned} X' &= X - X_c \\ Y' &= Y - Y_c \end{aligned} \tag{4}$$

The second, *(p, q)*, has the same origin but is aligned with the iris ellipse axes, and is related to *(X', Y')* by

$$\begin{aligned} p &= X'\cos\theta + Y'\sin\theta \\ q &= -X'\sin\theta + Y'\cos\theta \end{aligned} \tag{5}$$

The equation for a point on the elliptical contour is

$$\frac{(p-R)^2}{a^2} + \frac{q^2}{b^2} = 1 \tag{6}$$

Substituting Eq. 2 and $b = R_i$ into Eq. 6, and solving for R, yields

$$R = \frac{R_e\left(R_e p \pm \sqrt{\left(R_i^2 - q^2\right)\left(R_e^2 + R_i^2 - p^2 - q^2\right)}\right)}{R_e^2 + R_i^2 - q^2} \tag{7}$$

This is the basis of the Hough method. For an image feature at *(X, Y)* we first find *(X', Y')* from Eq. 4 and then iterate over values of θ. For each θ, we use Eqs. 5 and 7 to find the compatible values of R.

If either of the two terms under the square root is negative, there is no R compatible with the current X, Y, and θ. Otherwise, the two values of R correspond to the feature's being on the outside or inside of the iris (that is, $|R| < |p|$ or $|R| > |p|$ respectively. One of the two values can be discarded on the basis of the grey level gradient, if it assumed that the iris is darker than the surrounding part of the image. It is possible to show that if the local direction of increase in grey level is (g_x, g_y), then the sign to use in Eq. 7 to obtain the useful value of R is the same as the sign of $g_x \cos\theta + g_y \sin\theta$, the projection of the gradient vector onto the p-axis.

Practice

The camera's field of view includes more than the image of the eye, but processing is restricted to the region occupied by the projection of the eyeball. In the examples in Figure 4 and the left side of Figure 5, this is 288 pixels square, whilst in the examples on the right of Figure 5 it is 160 pixels square. The difference is partly due to different adjustment of the apparatus, but mainly a result of digitising full interlaced video frames for laboratory tests but a single video field for driving experiments, when rapid saccadic eye movements are expected.

Features for input to the Hough transform can be located by a variety of methods. Tunley & Young[6] discuss this in more detail, but for the work reported here an edge detector based on Canny's ideas[7] was used. The Gaussian-smoothed first derivatives in the *X* and *Y* directions are obtained and hence the total grey-level gradient is estimated at each image pixel. Non-maximum suppression along the gradient direction is applied, but a simple threshold without hysteresis is adequate as the final stage of pre-processing. The results of this process applied to eye images are shown in the upper part of Figure 4.

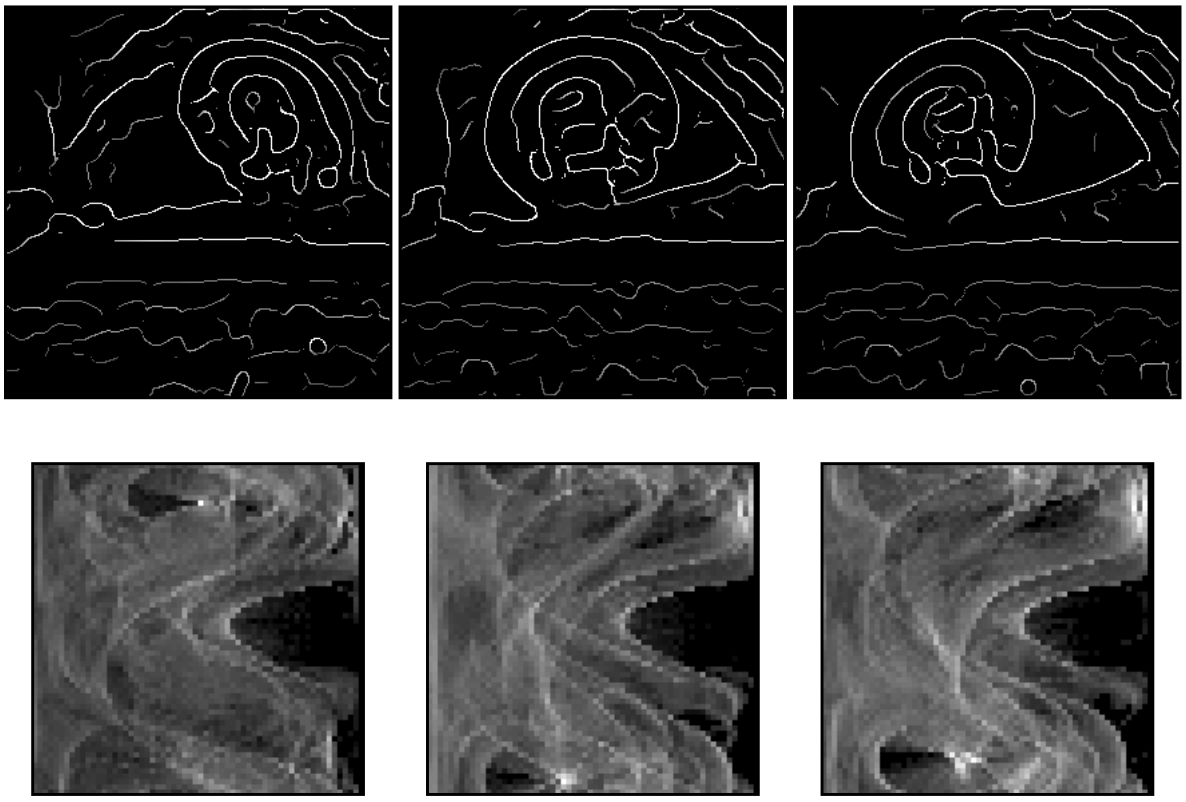

Figure 4. Top, three edge maps obtained by Canny-type preprocessing of eye images. The projections of the eyeball centres are at the centres of the images. Bottom, the corresponding Hough transform arrays after accumulation; brighter areas correspond to greater numbers of votes. In these arrays, *R* runs from left to right and θ from bottom to top, wrapping round with the top and bottom sides at $\theta = 90^{\circ}$.

The Hough accumulator array is set up to map onto $[0\ R_e]$ in *R* and $[0\ 2\pi]$ in θ. The number of cells in the array depends on the resolution in parameter space permitted by the data; for the examples of Figure 4 a 60×60 array was found to give good results, but this is not critical. For the lower resolution examples in Figure 5, a 30×30 array was used. For each non-zero point in the thresholded gradient array, θ is varied over the range 0 to π with the appropriate increment.

For each θ, *R* is calculated (where possible) using Eq. 7 with the sign determined by the gradient direction, and the corresponding cell in the accumulator array is incremented. When a negative value of *R* occurs, the cell incremented is that corresponding to $(-R, \theta + \pi)$. The magnitude of the grey level gradient is taken as the weight of evidence for the existence of a feature at *(X, Y)*, and the accumulator array values are increased by this quantity. It is sometimes useful to place a ceiling on the increment to reduce the effect of bright highlights.

After accumulation, the global maximum in the array indicates the values of *R* and θ for the

iris. The lower part of Figure 4 shows accumulator arrays for typical cases. The actual values of R and θ are refined by finding the centroid of a small region round the accumulator cell containing the maximum; a 3×3 region gives a small improvement in accuracy.

The procedure thus entails setting the following parameters before it can be applied to a given set of eye images:

X_c, Y_c, R_i, R_e: eye position and size;

σ: the Gaussian smoothing parameter for the Canny edge detector;

thresh: the threshold for gradient edge detection;

n_R, n_θ: the dimensions of the accumulator array;

m_R, m_θ: the dimensions of the local averaging region for peak position estimation.

The first four of these need to be measured for each sequence of images. A variety of methods can be applied. For instance X_c, Y_c and R_i can be found by fitting a circle to the iris outline when the subject is looking straight at the camera's image in the mirror. Alternatively, the images can be segmented and elliptical boundaries extracted using any general method; once the iris boundaries have been identified, R_i can be estimated from the lengths of the major axes, X_c and Y_c from the intersection of the minor axes, and R_e from the lengths of the minor axes. It is assumed that these operations are not time-critical. For the purposes of this paper, these parameters were found by adjusting them manually, searching for values that allowed an ellipse to be displayed that coincided with the iris boundaries for a number of images with the eye in very different positions. This can be done quickly, and the values found are accurate enough to allow an ellipse to be drawn round the iris without visible error in every image observed.

Any eye monitoring system must deal with the problem of how to detect blinks. In the present system, blinks cause the value of the peak in Hough space to fall because no clear iris outline is visible. A threshold on the number of votes at the peak can be used to detect when a blink is taking place.

Results

The algorithm outlined above was implemented in Pop-11 and C. Accumulation of the Hough array takes about 270 ms on a single-processor SPARCstation 10, for a 60×60 Hough array and processing a 288×288 pixel region of the image. For smaller Hough arrays and image regions, normally adequate for real experiments, the time is appreciably less, typically below 100 ms. Time for image preprocessing is needed in addition to this, but the amount depends on the choice of edge detector.

Figure 5 shows some results. In tests of the system, the qualitative performance depended critically on the presence or absence of strong highlights caused by corneal reflections. When the lighting is such that highlights are absent or weak, the algorithm finds the iris correctly (in the sense that it corresponds to where one would locate it manually) in every case. When highlights are present, they can lie along possible iris contours and cause gross errors. This

problem can largely be dealt with by suitable preprocessing of the image, at the cost of computation time, and also by carrying out multiple transforms on pre-segmented contours. These techniques were discussed more fully in Tunley & Young[6]; here we concentrate on the underlying algorithms.

The accuracy was tested more formally by asking subjects to view a calibration scene, as shown in the upper part of Figure 6. The subjects were required to look at each of the targets in a random order, and the iris position was estimated from a single frame for each target. For the same frames, the target position in the camera's view of the scene was found as a peak in the grey level. The affine transformation between the line of sight coordinates in the eyeball tangent plane and in the scene image plane was estimated by least squares. The residuals were then used to estimate the variance of the direction estimates in terms of scene image pixels. In order to convert these to angular measures, the target separation and distance from the camera were measured to provide a camera calibration.

Tested on one of the authors, the RMS residual gave an estimate of angular error of 0.95˚ laterally and 1.75˚ vertically, for a gaze angle range of about 40˚. The residual vectors are shown in the lower part of Figure 6. This error includes, of course, any inaccuracies in the actual gaze direction relative to the target point, since we assume that the subject is exactly fixating the target. Fitting ellipses to the iris outline by hand gave residual RMS errors of 0.8˚ horizontally and 1.68˚ vertically.

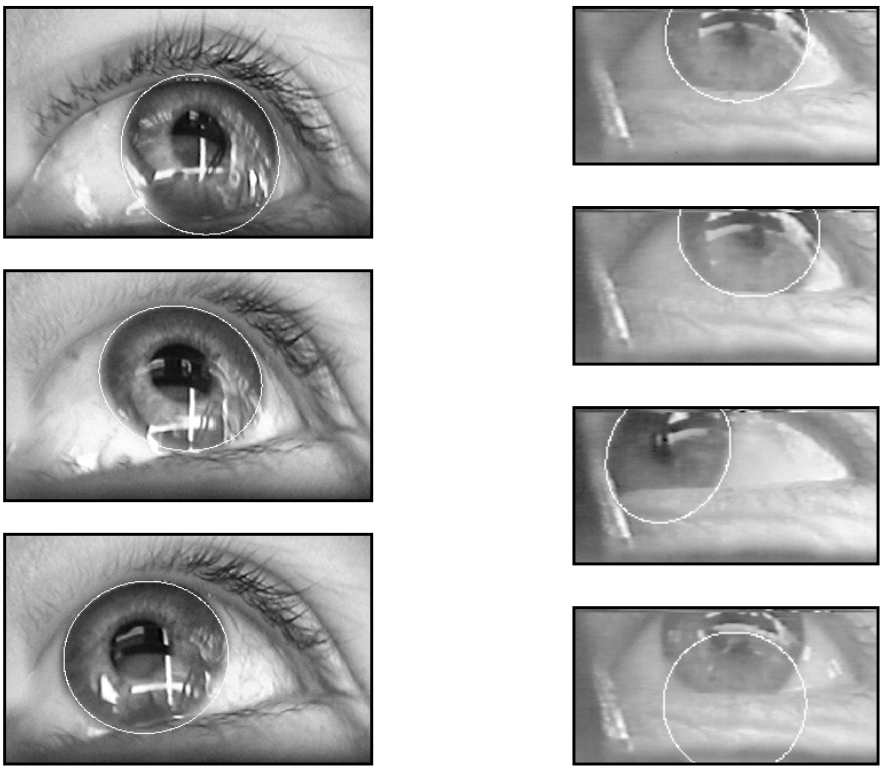

Figure 5. Iris ellipse positions found by the Hough transform technique. Left hand column is of high resolution images obtained during system calibration. Right hand column is of low resolution images obtained during an experiment in which the subject was driving a car. Due to misalignment of a mirror, only part of the eye is visible. In the bottom image on the right, a highlight has mimicked a possible iris boundary and the wrong peak in the Hough accumulator has dominated.

THE ACTIVE CONTOUR METHOD

Theory

"Active contours"[8] are analytic curves whose parameters are varied in order to make them lie close to boundaries in an image. Usually, such curves have many parameters and can represent complex shapes. The general idea can, however, be applied to the 2-parameter ellipse representing the iris boundary. The value of the active contour approach is that updating of the contour position is done only on the basis of image information close to the predicted position of the contour, thereby greatly reducing processing time. Indeed, the more accurately the position can be predicted in each frame of a sequence, the smaller can be the region of the image that is processed, so that high frame rates become possible as the resulting improvements in prediction accuracy reduce processing time.

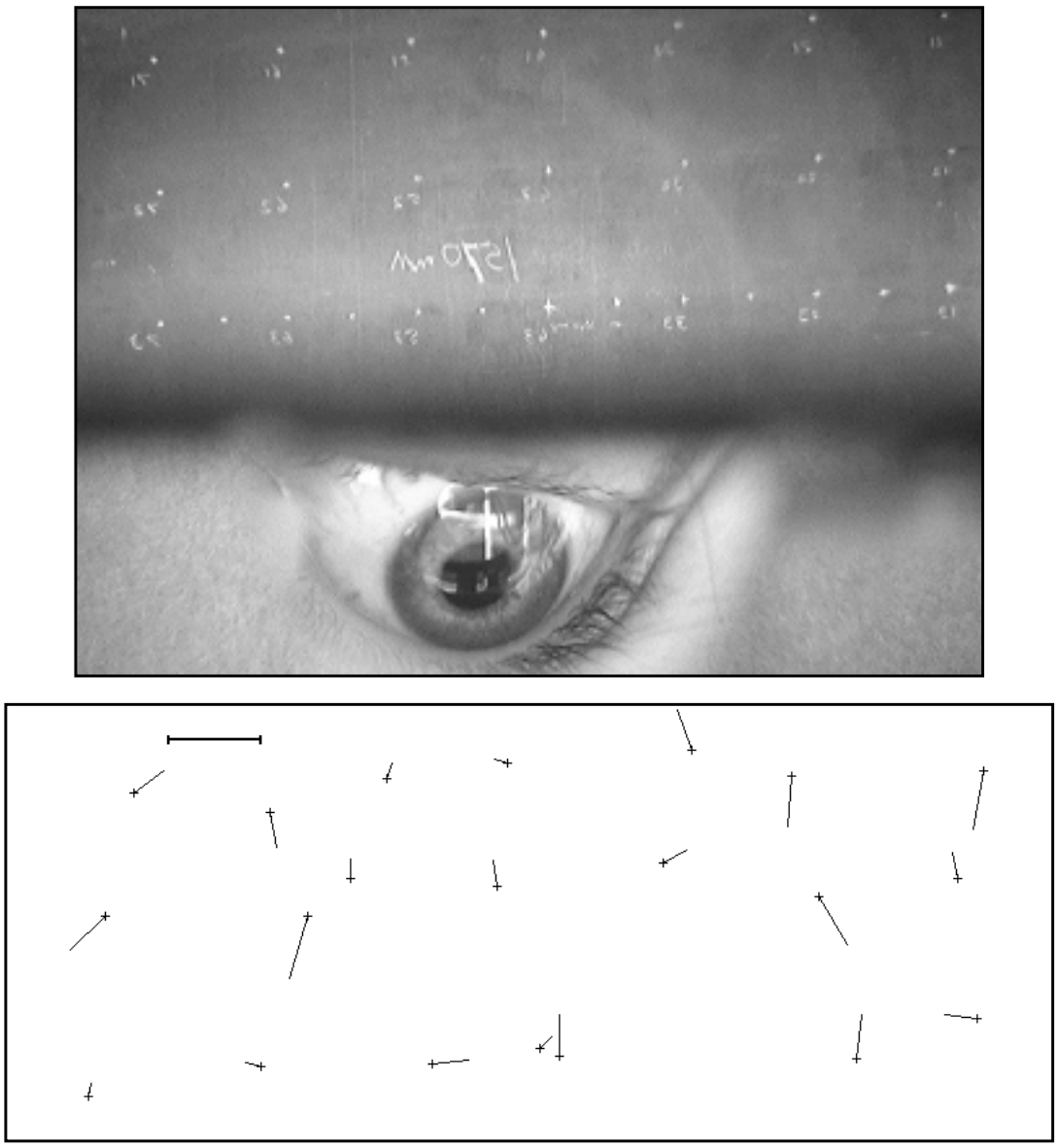

Figure 6. Top, the calibration scene used to assess system accuracy. Bottom, residual vectors after calibration. The crosses mark the target points in the scene images; they are not regularly spaced because their positions in individual images are affected by head movements. The horizontal bar indicates 5˚.

The algorithm begins with an ellipse with parameters X_i, Y_i, which represents an estimated position for the iris in the current image. One or more iterations then take place, at each of which X_i and Y_i are adjusted to move the ellipse closer to the actual iris position in the image. Since the iris boundary is represented by a sharp change in grey-level, it is appropriate to find explicitly the positions of high grey-level gradients close to the estimated iris position, and to make an adjustment to reduce the mean square distance between the ellipse and these edge positions. (An alternative would be to move the ellipse so as to reduce the value of an energy function defined using spatial derivatives of the grey levels on the ellipse.)

We define a number of control points on the ellipse, and for each one find the position of a nearby maximum in the grey level gradient (Figure 7a). In fact, since we can safely assume that the iris is darker than its surround, we only need consider gradients such that the grey level increases going from the interior to the exterior of the ellipse. Suppose the maximum for the j'th point is found at a distance n_j from the current ellipse, measured along the normal, with positive values in the outward direction. An adjustment to the ellipse parameters X_i and Y_i will produce a change in n_j given approximately by

$$\hat{n}_j = \frac{\partial n_j}{\partial X_i}\delta X_i + \frac{\partial n_j}{\partial Y_i}\delta Y_i \qquad (8)$$

For N points on the ellipse, we obtain estimates of the δX_i and δY_i required to move the ellipse close to the gradient maxima by minimising $\Sigma w_j \left(n_j - \hat{n}_j \right)^2$ where w_j is a weighting factor and the sum ranges over j from 1 to N. This gives the condition for the best least-squares fit:

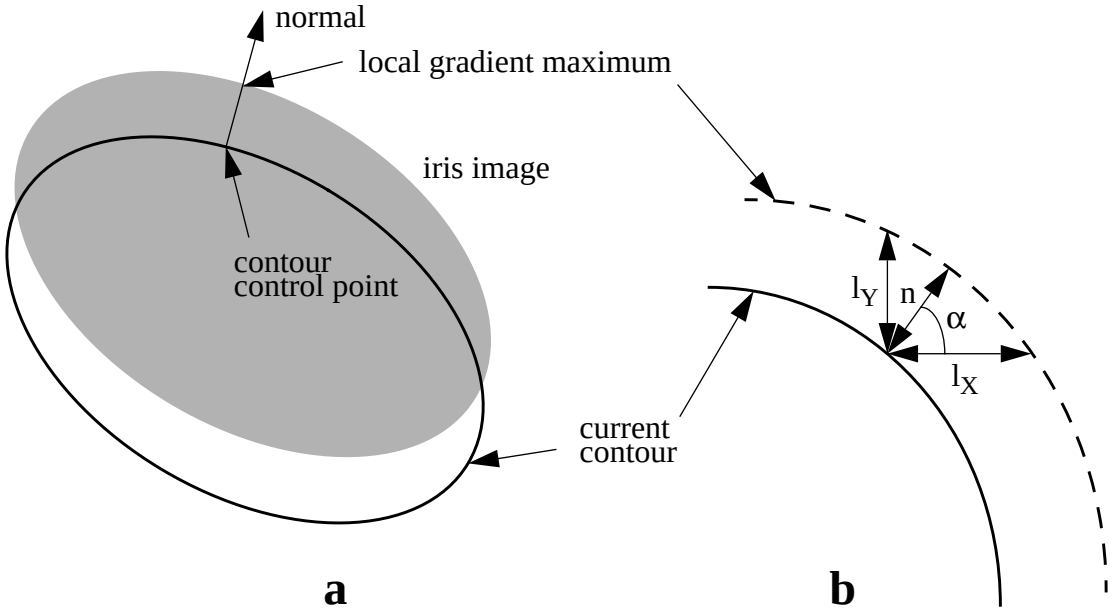

Figure 7. a Adjustment of contour position occurs by minimising the distances between the contour and a local peak in the grey level gradient. The sum of the squared distances along the normals for a number of control points is minimised, whilst the contour is simultaneously constrained to be a projection of the model iris. **b** The distance n and the angle α for a single control point. It is fastest to search for gradient maxima parallel to the image axes, so an approximation to n is estimated from l_X and l_Y.

$$\begin{bmatrix} \sum_j w_j \left(\frac{\partial n_j}{\partial X_i}\right)^2 & \sum_j w_j \frac{\partial n_j}{\partial X_i}\frac{dn_j}{dY_i} \\ \sum_j w_j \frac{\partial n_j}{\partial X_i}\frac{\partial n_j}{\partial Y_i} & \sum_j w_j \left(\frac{\partial n_j}{\partial Y_i}\right)^2 \end{bmatrix} \begin{bmatrix} \delta X_i \\ \delta Y_i \end{bmatrix} = \begin{bmatrix} \sum_j w_j \frac{\partial n_j}{\partial X_i} n_j \\ \sum_j w_j \frac{\partial n_j}{\partial Y_i} n_j \end{bmatrix} \tag{9}$$

Eq. 9 is easily solved for δX_i and δY_i once the sums have been accumulated. This is equivalent to minimising the elastic energy that would arise from connecting each control point to a nearby gradient peak with a spring whose spring constant is given by the weight.

It remains to find expressions for the partial derivatives, and to specify how n_j is to be estimated from the image. To reduce clutter, we now drop the *j* suffix, since all the calculations refer to a single point on the contour.

It is convenient to express the ellipse equation (Eq. 6) in parametric form as

$$\begin{aligned} p &= a\cos\phi + R \\ q &= b\sin\phi \end{aligned} \tag{10}$$

where ϕ specifies position on the contour. It is also helpful to define the quantities

$$\begin{aligned} \rho^2 &= a^2/b^2 = 1 - R^2/R_e^2 \\ A &= (p - R)/\rho \\ B &= q\rho \\ L^2 &= A^2 + B^2 \end{aligned}$$

The normal to the ellipse lies along *(A, B)* in *(p, q)* space, and the angle between the normal and the *X*-axis is therefore given by

$$\begin{aligned} \cos\alpha &= (A\cos\theta - B\sin\theta)/L \\ \sin\alpha &= (A\sin\theta + B\cos\theta)/L \end{aligned} \tag{11}$$

The speed of the intersection of the contour with its normal is given by the projection of the velocity of a point on the contour onto the normal: for a point at *(X', Y')*,

$$\frac{\partial n}{\partial X_i} = \frac{\partial X'}{\partial X_i}\Big)_\phi \cos\alpha + \frac{\partial Y'}{\partial X_i}\Big)_\phi \sin\alpha \tag{12}$$

with a similar expression for $\partial n/\partial Y_i$.

Applying the chain rule for partial differentiation to Eqs. 1, 10 and 5 gives expressions for the partial derivatives on the right of Eq. 12, which when substituted give

$$\frac{\partial n}{\partial X_i} = S\cos\alpha$$
$$\frac{\partial n}{\partial Y_i} = S\sin\alpha \qquad (13)$$

where

$$S = 1 - \frac{RA}{\rho R_e^2}$$

This completes the partial derivative calculation. We finally have to estimate n, which could be done by searching for the maximum grey-level gradient along the normal to the ellipse, starting from each control point. However, it is faster to search along lines parallel to the X and Y axes, giving estimates of l_X and l_Y (Figure 7b). Assuming low local curvature, we obtain two estimates for n from each control point, using

$$n = l_x\cos\alpha = l_y\sin\alpha \qquad (14)$$

Since the two estimates can have different weights, they are treated separately so that each control point makes two contributions to the sums in Eq. 9.

It would have been possible to consider l_X and l_Y as the measured quantities from the start, instead of n. This turns out to be unsatisfactory because l_X and l_Y become very large where the ellipse tangent is parallel to the X-axis and Y-axis respectively, and hence we would have large coefficients in the equations for these regions of the curve. The formulation above is better behaved, because the magnitude of n is everywhere small for a small movement of the ellipse. In the spring model, the springs pull along normals to the contour, as if attached to it by sliders, even though the grey level features are found by moving parallel to the image axes.

Practice

Ideally, N equally spaced points on the ellipse would be used in the calculation, but in fact it is faster to use points at equal intervals of ϕ. We can step round the ellipse incrementing ϕ and using Eqs. 10, 11, 13 and 14 to calculate the contributions of each point to the sums in Eq. 9.

The values of l_X and l_Y are found by moving along rows and columns of the image array, searching for the largest grey-level jump between adjacent pixels, where the jump must be in the correct sense given that the iris is darker than its surround. (The correct sense is indicated by the sign of $\cos\alpha$ for the X direction and the sign of $\sin\alpha$ for the Y direction.) The weight used for each equation is just the magnitude of the grey-level jump between the two adjacent pixels. No smoothing or other preprocessing is used; we rely on the effective averaging round the ellipse of the least-squares procedure to do all the smoothing that is needed.

A single iteration of the procedure requires the following parameters to be set:

X_c, Y_c, R_i, R_e: eye position and size;

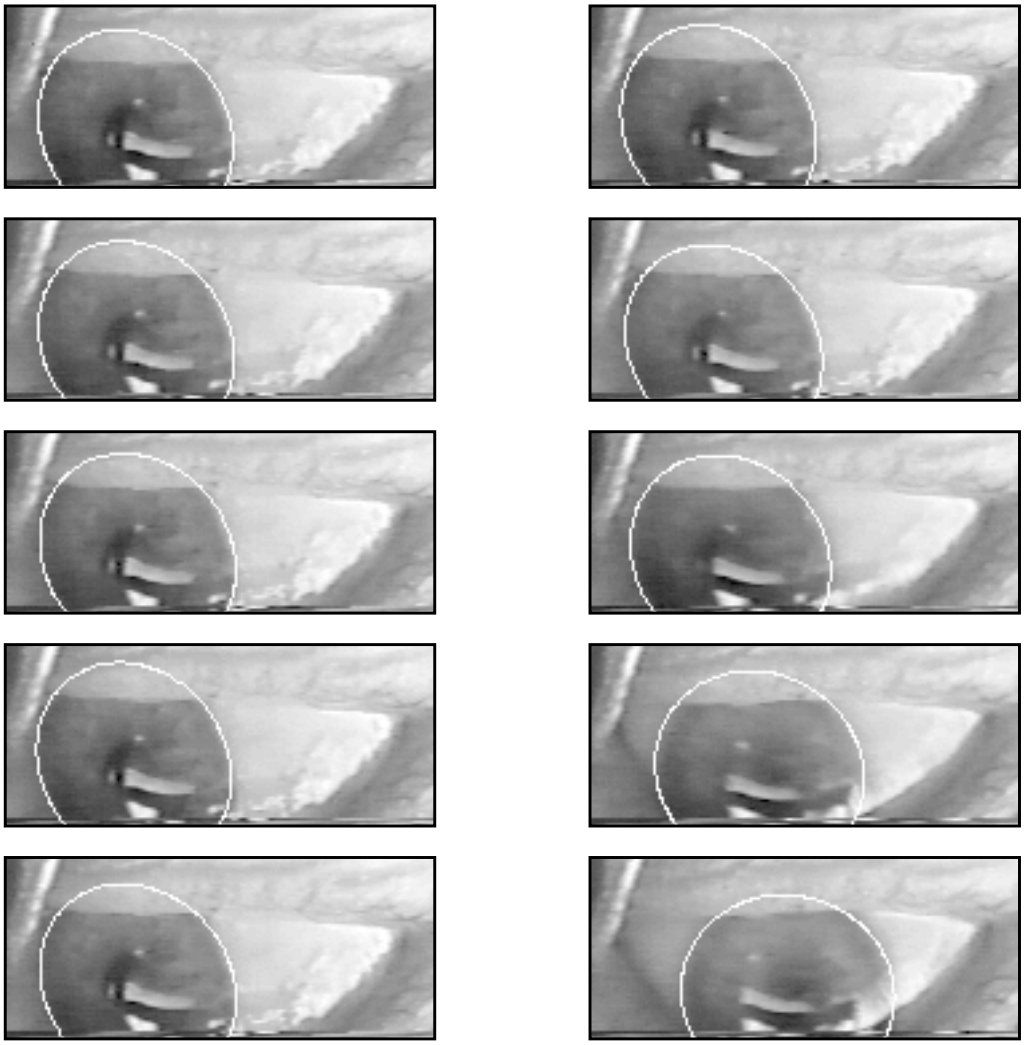

Figure 8. A sequence of iris images tracked using the active contour. The sequence is read down the columns. A rapid eye movement occurs in the final three frames. The interframe interval is 20 ms.

X_i, Y_i: the current estimate of the iris centre position;

N: the number of points to consider on the ellipse boundary;

P: the number of pixels to search in each direction along X or Y when looking for the maximum grey-level jump.

The first four are set up as for the Hough method. X_i and Y_i are found using the Hough method for the first frame of a sequence, and are subsequently taken as equal to the best estimates for the previous frame. A more sophisticated prediction could be used.

Several iterations of the procedure can be applied to one frame to refine the estimates of X_i and Y_i, reducing P each time as the ellipse converges to the real iris boundary. There are two reasons for doing this. First, the assumptions used to derive the linear equations above produce inaccuracies when the adjustment is large. Second, as the search size is reduced, the set of

features used converges to include only gradients very close to the expected iris position, reducing the effects of highlights and other boundaries in the image.

A further refinement which can improve performance is to estimate the grey level gradient over 4 adjacent pixels rather than 2. Given that only increases in gradient are of interest (searching towards the outside of the contour), a robust and simple technique is to subtract the minimum of two adjacent pixel values from the maximum of the next two. This allows the boundary to be blurred on a scale of 1 or 2 pixels and still be detected.

Results

Results for tracking the eye through a sequence of images are shown in Figure 8. Assessed visually, the results are accurate whenever strong highlights are absent. The number of data points on each ellipse was 32, and 4 iterations were carried out on each image, with the search length P equal to 20 pixels on the first iteration, then 10, 5 and 2. With these parameters, processing took 3 ms per image, using a single processor SPARCstation 10. No preprocessing is needed, so the limiting speed factor is the time needed to digitise the images.

CONCLUSION

Two methods have been presented for finding the position of the iris or pupil outline in the image of an eye. They use standard methods in a highly specialised way, and can achieve real-time performance because a very simple model, with only two free parameters, is applicable. The use of the Hough transform and a kind of simple active contour, both constrained by the model, makes it possible to estimate these parameters directly from the image. The Hough transform method provides a reliable way of locating the contour in a single frame. The active contour method is useful when a previous estimate of the contour position is available, and is very fast. Both methods are adversely affected by highlights in the image, which are therefore best avoided, although the problem can be effectively addressed at the cost of computation time.

ACKNOWLEDGEMENTS

We thank Professor Mike Land for the loan of a prototype of the eye movement monitoring equipment, data collection, assistance in building further equipment, and discussions. The work was supported by grants from the ESRC/MRC/SERC Cognitive Science & HCI Initiative and the SERC Image Interpretation Initiative.

REFERENCES

1 **Yarbus, A L** *Eye Movements and Vision*, Plenum Press, New York (1967)

2 **Land, M F '**Predictable eye-head coordination during driving', *Nature*, Vol 359 (1992) pp 318-20.

3 **Land, M F** 'Eye-head coordination during driving', *Proc. Int. Conf. on Systems, Man and Cybernetics*, Vol 3 (1993) pp 490-4

4 **Robertson, G, Craw, I and Donaldson, B** 'Human eye location for quantifying eye muscle palsy', *Proc. 5th British Machine Vision Conf.*, Vol 1 (1994) pp 357-66

5 **Clement, R A** 'An extension of Helmholtz's explanation of Listing's law', *Ophthal. Physiol. Opt.*, Vol 10 (1990) pp 373-80

6 **Tunley, H and Young, D** 'Iris localisation for a head-mounted eye tracker', *Proc. 6th British Machine Vision Conf.* (1995)

7 **Canny, J F** 'A computational approach to edge detection' *IEEE Trans. PAMI*, Vol 8 (1986) pp 679-98

8 **Kass, M, Witkin, A and Terzopoulos, D** 'Snakes: active contour models', *Proc. 1st Int. Conf. on Computer Vision* (1987) pp 259-68