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

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

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**

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

(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

(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

(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

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]6

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_i the Gaussian smoothing parameter for the Canny sTbge dtercor ;3.3301 -3 Td (etqwgv88t Tway shr) q of

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

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

$$\hat{n}_j \quad \frac{\partial n_j}{\partial X_i}\delta X_i \quad n_j \nu \tag{8}$$

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 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:

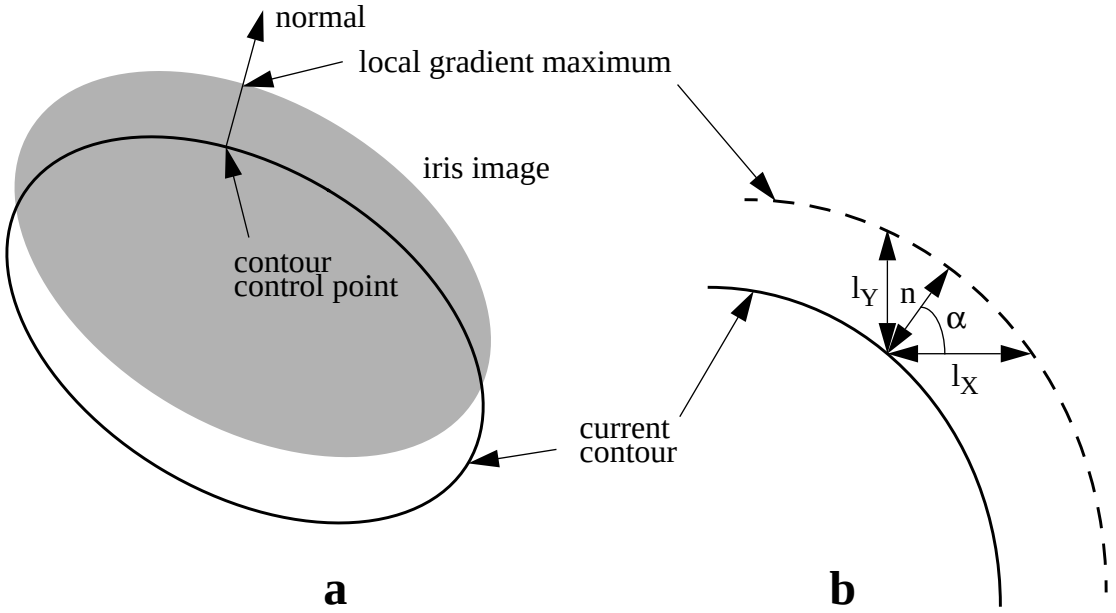

Figure 7. a Adjustment of contour position occurs by minimising the distances between the contour and a local peak in the grey level gradient. The sum of the squared distances along the normals for a number of control points is minimised, whilst the contour is simultaneously constrained to be a projection 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.

(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

(10)

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

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

(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'),

(12)

with a similar expression for

(13)

where

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

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

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

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