Human Eye Location for Quantifying Eye Muscle Palsy

Graham Robertson*, Ian Craw and Blair Donaldson †
Department of Mathematical Sciences
University of Aberdeen, Aberdeen AB9 2UB

Abstract

We describe a system for measuring the angle of eye gaze error on people with a squint. The program to analyse an eye image models each eye simply as a circular iris within a sphere rotating about its centre. A combination of edge detection, image transformation and template matching are then used to best fit the model to the image of the eye. The algorithm behaves correctly on 95% of eye images, over a range of gaze directions of $\pm 30°$ horizontally and $\pm 25°$ vertically, and enables the overall system to achieve a useful level of accuracy.

1 Introduction

We describe the computer vision component of a system to quantify the size and position of a squint in a person with eye muscle palsy. The measurement depends on being able to locate the centre of the iris in an image of the eye, and this is done effectively using relatively simple computer vision techniques. In use, the instrument provides data that assists a clinician in diagnosis, and helps determine, for example, which of the six muscles controlling eye movement, are causing the imbalance. Subsequent measurements assist in assessing the success of any treatment.

There have been many attempts to locate human eyes; as a task in its own right, as part of a more general system to locate face features, as an aid to recognition, or in order to make measurements such as gaze direction or blink rate; and there is little agreement on an appropriate method. Our need is more specialised; we must locate the centre of the iris, but even when this is not *per se* a goal, it is often taken as a characteristic feature, so although we ignore net based methods such as (Linggard, Myers and Nightingale 1992), there is still a wide choice of method. Nixon (1985), used a Hough transform to locate the eye, detecting both the eyelids and the limbus. Yuille, Cohen and Hallinan (1989) developed a detailed eye model, in which the limbus was modeled by a circle of varying radius and position and the eyelid by two elliptical arcs of varying size and position. A deformable template methodology then compares the model to edges and intensity peaks and valleys on the image optimising the parameters to obtain the best match. Such a model was used by Shackleton and Welsh (1991) for image based recognition, and a variant by Craw, Tock and Bennett (1992) for more general feature location.

*This work was funded by Grampian Enterprise, AURIS research and Aberdeen Royal Hospitals NHS Trust. GR is supported by SERC Research Grant Number GR/H75923.

†Aberdeen Royal Hospitals NHS Trust, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, AB9 2ZB

BMVC 1994 doi:10.5244/C.8.35

What emerges from a study of this work is the need to tailor the method of eye location rather closely to the expected set of images, and the particular task. In our application we have many advantages over most other systems; we can assume that the images are of consistent quality, and at relatively high resolution. We can perform calibrations on each eye to be subsequently processed, and can position the eye within the image so that problems of scale and position invariance are much reduced. However, since the system is designed to measure gaze direction, an assumption that the gaze is straight ahead, or that the iris always appears circular, is inappropriate. An additional constraint is the need to provide feedback to the clinician during tests; as such, the processing time, on relatively modest hardware, can be no more than a few seconds.

These hardware restrictions mean we cannot take advantage of methods in which the eye is tracked continuously (Harris and Stennett 1990), (Tock and Craw 1994), even though it is always in view. In fact we adopt a model based approach, working *ab initio* each time on a single frame. Despite following Nixon (1985) in using a Hough transform, the work we describe here is perhaps closest in spirit to that of Yuille et al. (1989) or Hallinan (1991). Fortunately special considerations mean that our model is effectively two dimensional (as opposed to their eleven parameters), despite our need to model the eye in three dimensions. Thus we avoid the need for search optimisation algorithms such as simulated annealing.

There is another human restriction which is perhaps of interest. Cost considerations meant that the complete vision system described here, including the user interface, was assembled by a single person, in a period of two months. It has subsequently functioned usefully in the hands of a clinician who had no expertise in computer vision. It thus provides another illustration of the gradual move of vision systems from the laboratory to effective, economic application.

2 Equipment setup and procedure

Our main interest is in the vision component of the eye squint measuring system. However, we give some details of the full system, so the constraints imposed on the vision component can be better understood. The apparatus consists of a head clamp at a suitable height, positioned in such a way that when the patient is constrained by this clamp, he can see a stimulus screen in front of him, on which a cursor can be displayed. The screen is located so that, when looking straight ahead, the patient's gaze is at the centre of the screen. The distance from the patient to the screen (about 30 cm) is determined by the need to require gaze angle deviations of up to 30° from the centre when looking at each of the four edges of the screen. Interrupting the patient's gaze are two half-silvered mirrors which each provided an eye image to two synchronised cameras. Each image is digitised and made available for subsequent processing. Control is achieved using a console, in which images of each eye are displayed, and in which, in addition there are a pair of replicas, one for each eye, of the stimulus screen. The hardware setup for the system[1] is given in Fig. 1.

[1]The equipment was constructed in Department of Medical Physics at the University of Aberdeen and is subject to patent application PCT/GB93/00715.

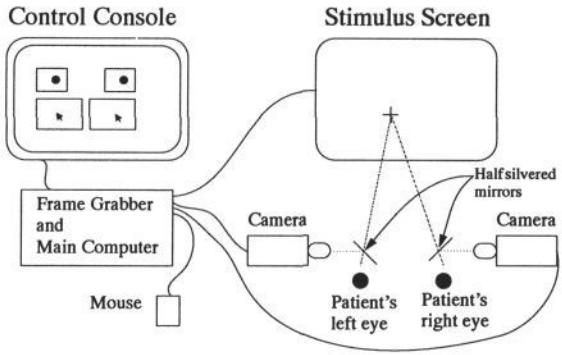

Figure 1: *Equipment setup of the eye squint measuring system.*

In use, the patient's head is first constrained by the clamp. The need to keep the head still, and so inevitably constrained, for the duration of the testing means that a single test session should not exceed about 10 minutes; this in turn imposes time constraints on each step of the vision processing. The system is first calibrated; we describe this in detail in Section 4, and then the testing proper begins. The clinician moves the cursor on the stimulus screen to a point at which gaze function is to be tested, and the patient is requested to look directly at this stimulus. When the gaze is steady, images are acquired, and the vision component measures the achieved gaze angle. The result of this processing is displayed as an overlay (see Fig. 7) on the corresponding eye image in the console; the clinician judges the success of the location, and if necessary retakes the measurement, perhaps with manual intervention to avoid occlusion by an eyelid. As each gaze angle is measured, it is compared with the target gaze angle, computed from the location of the stimulus cursor, and the geometry of the instrument. The difference between these two angles is then displayed on the appropriate replica of the stimulus screen. This feedback is used by the clinician to determine areas of the visual field which need detailed exploration. In particular significant errors can indicate a *squint*, a malfunctioning of one or more of the muscles controlling the position of the eye. Subsequent measurements, perhaps taken months later, can then indicate the progress of any disease or its modification by treatment.

3 Modeling the Eye

The geometric task of the gaze measuring system is to deduce, from geometric information obtained from an image such as those shown in Fig. 2, the gaze direction of the eye at the time the image was acquired. We use a model based approach; our model of the eye itself is relatively simple, although three dimensional. We assume the eyeball is a sphere, rotating about its centre, which we take to be fixed throughout. We consider the iris and pupil to be concentric discs positioned in such a way that the boundary of the iris, known as the *limbus* lies on the surface of the eyeball. In practice, the pupil itself proved to be of little help in our task because of the low contrast caused by the mirror and lighting arrangements, and

is ignored from now on. The ray from the centre of the sphere representing the eyeball, which passes through the centre of the iris, is the *optical axis*; it is this direction which we describe as the *gaze direction* of the eye.

When the eye is viewed from a point on the optical axis, the iris appears as an almost circular disc, but from any other angle the disc appears elliptical. This change can be significant since we consider wide variations in gaze direction; as such we are forced to adopt a three dimensional model. As the iris is not exactly circular we model it as an ellipse, even when viewed along the optical axis. The eyeball model assumes that the iris is planar and moves so that the centre of the limbus ellipse is at a fixed distance R from the fixed centre of the eyeball. The distance R, which we shall describe loosely as the radius of the eyeball is to be determined during calibration, as is the eccentricity and the maximum diameter of the limbus ellipse. Of course this model is an approximation; we discuss this further in Section 5.

Coordinate Systems We fix (in space) a cartesian eye-centered co-ordinate system with the origin at the centre of the eyeball. Strictly we have one system for each eye, but we concentrate on just one at present. For ease of description, we assume that the eye, in its rest position, is looking horizontally. Our axes then consist of a horizontal x-axis, the y-axis vertically upwards, and the z-axis pointing out from the centre of the eye 'straight ahead', along the optical axis of the eye. With this co-ordinate system, the camera is assumed positioned at infinity on the positive z-axis, looking towards the eye. There is then a parallel projection from the eye onto the image plane of the camera, which we can thus confuse with the $x-y$ plane of the eye co-ordinate system, referred to as *image co-ordinates.*

When the eyeball rotates, the optical axis no longer co-incides with the z-axis. Since we are interested in gaze calculations, it is convenient to specify the position of the optical axis using a system of modified spherical polars (r, θ, φ) whose origin co-incides with the origin of the cartesian co-ordinate system. The point (r, θ, φ) defines a vector $\mathbf{r}$ which makes an angle φ with the x-z plane. The projection of $\mathbf{r}$ onto the $x-z$ plane then makes an angle θ with the z-axis. The choice of angles is convenient in this application, since for a real eye, $-\pi/2 < \theta, \varphi < \pi/2$. Knowing the position (x_0, y_0) of the centre of the projection onto the image plane of the limbus ellipse, then gives the gaze direction (θ, φ) by

$$R \sin \varphi = y_0, \quad R \cos \varphi \sin \theta = x_0. \tag{1}$$

We refer to this as the *gaze equation.* With the eye at rest, so the optical axis co-incides with the z-axis, we assume the limbus is elliptical. If the eye is rotated by (θ, φ), the parametric equation of the limbus ellipse becomes

$$\begin{aligned} x(t) &= a \cos t \cos \theta + R \sin \theta \cos \varphi - b \sin t \sin \theta \sin \varphi \\ y(t) &= b \sin t \cos \varphi + R \sin \varphi \end{aligned} \tag{2}$$

where we have suppressed the corresponding z-value; this is then the equation of the ellipse which is the projection of the limbus ellipse onto the image plane; we describe it as the *image ellipse.*

4 Eye location

The geometry of the previous section enables us to compute the gaze direction provided we can determine the center of the image ellipse. In this section we give details of the algorithm used to do this. Rather than attempting to determine the centre point directly, we fit a suitable ellipse to the image of the limbus, and obtain the centre by calculation.

Typical images are shown in Fig. 2; they are 192×144 pixels at 8 bits per pixel, acquired with an S2200 framegrabber. The lighting is controlled, but its interaction with the vision part of the system is secondary to ensuring that the patients can comfortably perform the tracking task. Thus the images have lower contrast than would be ideal from the vision standpoint, and there is a noticeable specularity in the image. During testing, the iris moves freely around the image, although initialisation ensures that at rest the iris is central within the image. Manual manipulation of the lower eyelid is possible during testing to avoid severe occlusion problems. Because each individual head is fixed during testing, scale variations are relatively limited.

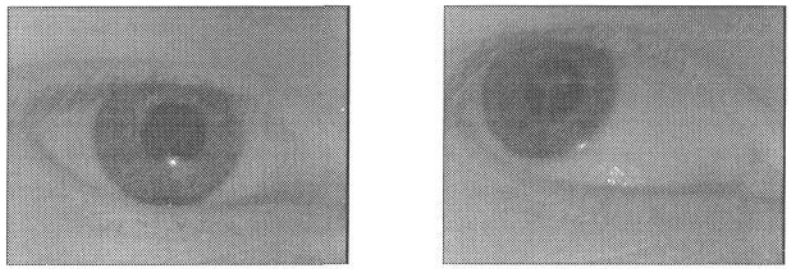

Figure 2: *Representative eye images with differing iris positions.*

This control over the imaging conditions enables use to be made of constructions which appear very image dependent. In outline, the algorithm we use is:

- an initial prediction for the centre of the image ellipse is made using relatively naive methods;
- a fairly sparse collection of edgels, chosen because they are expected to lie on the image ellipse, is marked;
- a model ellipse is calculated based on Equation 2 and matched with this set of edgels; the position of the centre is then adjusted within a small region to find the best such match;
- the process is now repeated using the new centre; and
- the centre of the ellipse resulting from this second adjustment is taken as the required centre.

Our initial estimate of the centre of the image ellipse can be relatively inaccurate; the subsequent refinement will succeed, providing the estimated centre lies within, or very close to, the actual pupil. A simple Hough transform, with a small accumulator array is effective in most situations, and only fails when the gaze is downwards, the eye begins to shut, and the eyelid dominates. Accordingly we

accept the predicted centre if it lies in the upper half of the image, and use a different method otherwise. This second method is even simpler; the predicted centre is the pixel with the lowest intensity in the bottom half of the image.

For the overall system to function, it is essential that the patient should be able to focus on the stimulus screen. This may necessitate wearing glasses throughout the tests, and the image processing routines must continue to function. For images in which glasses are *not* being worn, the second prediction method described above could be used throughout the image, since the pixel with the lowest intensity lies within the pupil of the eye. If glasses are present, such a minimum can occur on the glasses frame; but since glasses only interfere with the top half of the image (see Fig. 3) the method we have given is effective. However, this problem does mean that we need both parts of the prediction; neither is reliable alone.

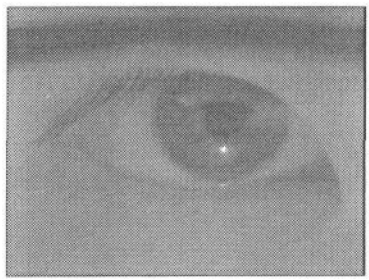

Figure 3: *An eye image in which glasses are being worn. The frame only appears in the top half of the image.*

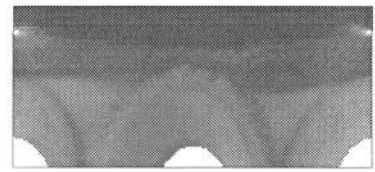

Figure 4: *The polar form of an eye image. The located centre of the iris appears at the* top.

In the next stage of the algorithm a sparse collection of edgels is generated, designed to mark the image of the limbus. This boundary is between the dark iris and lighter sclera, or white of the eye. We aim to do some edge following once suitable edgels have been located, but following a (roughly) circular edge can be computationally intensive. To avoid this overhead we do a polar transformation of the image about the predicted iris centre, to obtain an image such as that in Fig. 4. Each column in this polar image, displayed with low values of r at the top, thus corresponds to a ray in the original image emanating from the predicted centre of the limbus.

We work with these polar images when marking edgels. Each column of the image is examined in turn, and divided into intervals in which the intensity change is monotone. The two largest such intervals, in which the monotone intensity change is of the same sign as that at the limbus, are then selected, and the point at which the peak intensity gradient occurs within each interval is marked as an edgel. This necessarily produces a sparse map, while the selection of large intervals for subsequent processing gives a measure of noise suppression. A difficulty occurs with illumination; specularities give rise to unwanted edgels unrelated to the limbus. These edgels usually have the highest intensity gradient of the whole image, and most are removed simply by deleting edgels co-inciding with the 0.25% of pixels having the strongest intensity gradients.

In a second stage of this marking process, each edgel, found as above, is used

as a seed, from which edge growing for a maximum of four columns on either side is attempted. In each column, the 11 pixels nearest the seed edgel are examined, and the one with maximum gradient in the correct sense is marked as an edgel. If no point has a gradient of the correct sense, the growing ceases. Fig. 5 shows an edge image overlayed on a polar image, while the corresponding cartesian image is Fig. 6.

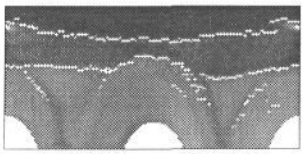

Figure 5: *A polar image with horizontal edgels marked.*

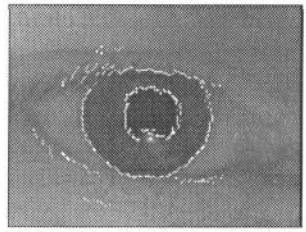

Figure 6: *The cartesian image which corresponds to that in Fig. 5.*

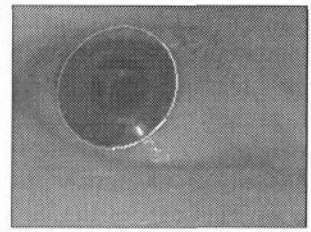

Figure 7: *The final ellipse overlayed on the original image.*

Working back in the cartesian image, the next stage involves finding the best fit between a model ellipse and these edgel data. The model is generated using the prediction of the centre of the image ellipse which was made above. Fixing the centre determines the gaze direction, and hence the ellipse is completely specified by Equation 2. If the predicted centre is correct, many of the edgels will lie on this ellipse; by searching within different centres with a neighborhood (our timing data were derived using a neighbourhood containing ~ 500 pixels) of the predicted one, we seek to maximise the fit between prediction and data. This part of the algorithm is computationally intensive, and so rather than recalculating, we use the same shaped ellipse, and simply vary the centre. This approximation to the ellipse shape means we can still get an improved fit. The 'best' centre, as derived above is used as the new predicted centre of the image ellipse, and the whole fitting process is repeated; the resulting updated centre is accepted as correct. We show the type of fit obtained in Fig. 7, although we are only concerned about the position of the centre of this final ellipse.

Calibration In order to use the gaze equation we need the three constants (R, a and b) associated with the eye model. Since these vary between patients, they are obtained in a calibration stage before any other measurements are made.

The patient is first attached to the equipment, in such a way that his head remains motionless during the remainder of the tests. A stimulus point is displayed, placed in such a position on the stimulus screen that the patient looks directly ahead; so the gaze angle is $\theta = \varphi = 0$. The position of the camera is then adjusted, with motorised slides in two orthogonal directions, until the centre of the camera image co-incides with the centre of the pupil in the image. The camera then remains fixed until a new patient is to be tested.

As a first step, the parameters of the limbus ellipse (a and b) are determined using the special case of Equation 2 with $\theta = \varphi = 0$. In this case, the generated

ellipse is independent of the eyeball radius R, and a and b become the semi-major and semi-minor axes of the ellipse as seen by the camera. The fitting procedure to locate the image of the limbus ellipse is exactly as described above, except that now the centre of the ellipse is known (at $\theta = \varphi = 0$), and we search instead on pairs (a, b) in order to find the best fit. In the setup described here, both major and minor axes range between 60 and 90 pixels.

A variant of this procedure is now used to determine R, based on the values a and b which have just been measured. The clinician first locates a number of points on the stimulus screen at which the patient can accurately direct his gaze. For each such point, the actual gaze angle can then be calculated, rather than measured. As such we can apply the above procedure for locating the limbus, searching over different values of R, with a known gaze direction. Our relative scale invariance means that the initial value of R (85 pixels) is accurate to within 20%. This part of the initialisation requires the existence of points at which there is no gaze error; identification of such points will not pose a problem in practice. A number of determinations of R are made, each using a single point at which the gaze can be directed accurately; the average radius is then used during the subsequent processing.

Taking Measurements In use the clinician sets a target or expected gaze angle, and a cross is displayed on the stimulus screen. This is positioned so that, provided the patient has accurate gaze control, an attempt to gaze at the stimulus cross will result in the optical axis of the eye assuming the target gaze angle. Of course the whole interest is to explore those gaze directions for which this fails, and thus for which the measured gaze angle does not co-incide with the test gaze angle. This *gaze error* can be displayed in a number of ways; we chose to display the horizontal angular error, vertical angular error or the great circle angular error.

5 Results

Running on a SparcStation LX, each measurement of the gaze angle and the associated error display takes approximately three seconds. Measuring both eyes, and allowing time to assess the results, and decide on the next area to be explored, gives a cycle time of about ten seconds. Calibration itself takes roughly a minute, so in a ten minute session approximately 50 readings can be made. This is more than adequate for the mapping task.

The limbus location program was tested on 105 images from each of four different eyes, a total of 420 eye images. The images were obtained for a range of eye orientations from the center of up to 30° in the horizontal direction and up to 25° in the vertical direction. Based on the type of visual criteria likely to be adopted by a clinician the limbus location was judged acceptable in 95% of trials. Errors occurred most often at extreme angles, typically because the eyelashes interfered with the edge finding algorithm; even in these cases it was possible to repeat the reading to get an acceptable measurement.

Given that the location was regarded as acceptable, there remain a number of sources of error. Our model of the eye is simplistic; real eyes are not spherical, while the fixed centre of rotation is a crude description of how the eyeball moves

in its socket. This causes systematic errors varying with the gaze angle. Just as important is the need to keep the head fixed rigidly for the duration of all the measurements; in practice some movement occurs, with significant effect on the geometry. Inaccurate calibration gives rise to systematic error; an analysis shows the final results to be very dependent on the accuracy of R. A further source of error is when the centre of the limbus ellipse is inaccurately identified, perhaps because only a small part of the limbus can be seen, despite the fit being acceptable to the clinician.

Quantifying these errors individually is hard, but we can get a number of indications of the effectiveness of the complete system. One such is the repeatability of a given measurement; this was tested by presenting a short sequence of target gaze angles a number of times, and comparing different measurements of the same gaze angle. Of more interest is the repeatability across sessions, in which the head clamp is removed from the subject and subsequently re-attached. In this test, the cameras were re-centred at the start of the second session, but the geometry of the eye was not re-determined. This second test gives a measure of the ability of the system to detect changes in a given patient over time, and corresponds closely to the system's intended function. Results of these tests are shown in Table 1; the interval given contains readings within two standard deviations of the mean, indicating the expected repeatability 95% of the time. A total of 20 readings on two different subjects contribute to each data point, and angles on both sides of 0° were tested. For simplicity we report here only the horizontal angular error; the pattern of results is the same for the vertical or great circle angular errors.

Target angle	0°	10°	20°
Repeatability within a single session	±1.2°	±1.2°	±1.9°
Repeatability across sessions	±1.6°	±2.0°	±2.4°

Table 1: Repeatability at fixed target gaze angles (2 standard deviations).

A more stringent test is to determine the absolute accuracy of the whole system; in other words to consider the means achieved in Table 1 as well as the standard deviations. At present there is no other formal way of taking these measures with which our system can be compared; we thus tested subjects with healthy eyes and assume that the gaze angle achieved by the subject was the target angle. Subjects were chosen by testing for visual acuity with a cover test, and stereopsis with a Titmus test. No useful pattern emerges from the results; one subject had mean errors of −0.5°, −3.2° and −4.5°, while another had means of 2.0°, 2.2° and 1.0° at (horizontal) target angles of 0°, 10° and 20° respectively; here a +ve sign means the measured gaze is to the right of the target. Each of these means is significantly different from zero, at a 95% confidence level, when tested with the standard deviations in Table 1, and indeed the total error seems to be large compared to the range of gaze directions. The repeatability results, not affected by an inaccurate value for R, thus seem to confirm the importance of R if absolute measurements are required. However, while of interest — and without some indication of accuracy, we might have been describing a (very repeatable) null system which ignored its input — the main clinical need is to observe changes over time; in this situation, the same value of R can be used on both occasions and

the results are much less affected by the inaccuracy. An additional simplification is to just consider the difference in gaze angles between two eyes looking at the same point; this value is much less affected by head movement.

6 Conclusions

We have presented a system for calculating the gaze error of an eye in order to measure squints. The system was specifically designed to work on high resolution pictures of the eye under controlled conditions, with even lighting. The method, based on a simplified geometrical model operates over a wide range of gaze angles, and proved sufficiently useful to suggest the simple model was worthwhile. Hardware changes, including better lighting to enable the pupil to be matched in the same way as the limbus, and the ability to grab, and hence average results over several consecutive frames, have been identified which would lead to improved accuracy. The development of this system, which took two months, demonstrates that a successful transition from basic computer vision components to a medical tool can be done usefully in a short time. Clinical trials are now taking place at Aberdeen Royal Infirmary on the equipment to assess the capability of the machine in measuring squints on a number of patients.

References

Craw, I., Tock, D. and Bennett, A.: 1992, Finding face features, *in* G. Sandini (ed.), *Proceedings of ECCV-92*, number 588 in *Lecture Notes on Computing Science*, Springer-Verlag, pp. 92–96.

Hallinan, P. W.: 1991, Recognizing human eyes, *Society of Photo-Optical Instrument Engineers - Geometric Methods in Computer Vision.*

Harris, C. and Stennett, C.: 1990, RAPID - a video rate object tracker, *Proceedings of the British Machine Vision Conference*, pp. 73–77.

Linggard, R., Myers, D. and Nightingale, C.: 1992, *Neural networks for vision, speech and natural language*, Chapman and Hall, London.

Nixon, M.: 1985, Eye spacing measurement for facial recognition, *Proceedings of the Society of Photo-Optical Instrument Engineers.*

Shackleton, M. A. and Welsh, W. J.: 1991, Classification of facial features for recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVIP-91)*, pp. 573–579.

Tock, D. and Craw, I.: 1994, Tracking and measuring drivers eyes, Chapter to appear in Real-Time Computer Vision, editors Chris Brown and Demetri Terzopoulos, Cambridge University Press.

Yuille, A., Cohen, D. and Hallinan, P.: 1989, Facial Feature Extraction from Faces using Deformable Templates, *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 104–109.