Model-Based Feature Extraction and Classification For Automatic Face Recognition

by

David E. Benn

A Thesis for submission for the degree of
Doctor of Philosophy

in the

University of Southampton
Faculty of Engineering and Applied Science
Department of Electronics and Computer Science

April 2000

Abstract

Recognising faces through model-based feature extraction and description currently appears to be less popular than statistical or face-based recognition approaches. Certainly there is concern that model-based approaches might not prove reliable in practice. Accordingly, this thesis describes a programme of research for improving model-based recognition through robust feature extraction, selection and combination. First we present a new two stage process for finding eyes. A reformulated evidence gathering process is used to determine the rough location of the eyes by exploiting their natural concentricity. Their location was refined by an improved deformable eye template which does not require internal energy terms and uses few parameters. These parameters were best optimised using a genetic algorithm. The technique produced 91% and 93% successful location rates on face databases of 1000, and 88 faces, respectively. A feature vector composed of 29 geometric, 6 colour and 55 forehead contour measures, was extracted from 44 faces from the XM2VTS database. To achieve this, the skin boundary was extracted by region growing using a sample of skin below the eyes. Other features such as the nose, mouth and eyebrows were then located by noting that these features are enclosed by skin but exhibit different statistical properties. A new method, based on intrinsic feature variance, is presented for combining and selecting features which are of potentially disparate magnitude and/or independent sources. Our method provided an increased variance in the classification matrix and facilitated identification of the most discriminating features. Surprisingly, although the eyes were a good initialiser in the search for other face features, their template parameters offered low discriminatory power. Much higher discriminatory power was available through the normalised Fourier descriptors of the forehead contour. We simulated the effect of measurement noise on classification performance and found that errors of 6 pixels on the geometric features resulted in up to 43% classification error. Recognition rates of 77% and 72% were experienced using manual and automatic geometric measures. However, when we combined the geometric measures with perfectly extracted contour measures from its first eight Fourier descriptors we achieved 100% classification. Our work indicates that model-based face recognition is achievable and suggests relative importance of the components of a feature vector. This information is clearly of interest given the high computational requirements of such an approach.

Acknowledgements

I wish to thank my supervisors, Dr. Mark Nixon and Dr. John Carter for their guidance in the work contained in, and the preparation of this thesis.

Many thanks to Dr. J. Matas from Surrey University for his brief, but informative introduction to colour image processing.

Ta mucho to the VSSP group for their useful discussions, in particular Dr. Alberto Aguado, Dr. Steve Gunn.

I gratefully acknowledge financial support afforded by the Home Office (UK).

To my mum, my family and my friends:-

Thanks for your support and encouragement.

1. Introduction

There is an ever increasing commercial demand for reliable personal identification systems which has driven research into areas such as voice recognition [11], fingerprint analysis [9], gait [18], and signature analysis [66]. Currently, face recognition systems are emerging as a powerful tool within the biometric recognition community. Closed circuit television (CCTV) systems are now common place in retail shops and society in general and can be used to capture images of suspects. An attractive benefit of automatic face recognition is that, unlike finger print analysis, it can be performed without the observed target being made aware. For obvious reasons, these benefits appeal to organisations involved in security or covert operations. Face recognition systems are now able to provide a significant contribution in forensic security and law enforcement applications. These applications currently use human descriptions of faces as a system input; parameters may include, colour of the eyes, shape of the nose, jaw and eyebrows, size of the mouth, colour and texture of skin and hair for example. The recognition system is then tasked with retrieving faces from a database which best match the data input from the human description. In this type of application the natural features of the face (eyes, nose, mouth etc) are extracted and compared in order to effect recognition; such systems are known as *feature-based.* A feature-based face recognition system may have just one instance of a particular subject's face which needs to be located on a potentially large database. In feature-based systems we need to develop techniques which are robust enough to locate the major face organs and landmark points which are common to most faces, over the full range of the database. After locating the common features, we seek metrics which emphasise the differences in the extracted features, yielding the maximum discriminatory measure between all the faces in the database. Unlike gait recognition, we cannot rely on powerful temporal information to initialise our search for features [18]. The only assumption that we can reasonably make is that there is a face somewhere within the image. Given this limited information, clearly, feature-based face recognition is fundamentally a difficult task for computer vision techniques. On the opposing end of the face recognition spectrum, the *holistic-based* methods are typically used for security access systems. In these methods, the whole face (indivisibly from its components) is used to train a recognition system. The face must be segmented from the background in order to minimise the effect of variations in backgrounds or other extraneous artefacts on recognition. Then after training, the system can either grant or refuse access to the requested resource. Typically, holistic systems require a large number instances of each training image with a small set of distinct images. For example, the eigenface based system of Turk and Pentland [78] used a total of 2500 training image of 16 different subjects. After training, which may be performed as a background task, the system was able to perform in near real-time, i.e. considerably faster than a feature-based system. Our interest in face recognition systems is biased towards realising the aspirations of our sponsor's requirements which are detailed in the following section.

1.1 Sponsor's Requirements

This research was part funded by the Home Office (UK) and was therefore a programme of research with perhaps more pre-defined goals and deliverables than other research programmes. The Police Information Technology Organisation (PITO) within the Home Office, have a psychological coding scheme which attempts to encode aspects of facial appearance which are most likely to be remembered by a witness recalling a face. Their scheme is adapted for witnesses searching a database of facial images to retrieve a previously viewed face. They believe this contrasts with facial coding schemes which have been developed for the automatic comparison by computer of target facial images within a database of facial images. Nevertheless, there are common functional units which can be implemented and investigated. The Home Office wish to train artificial neural networks to encode facial components when individual features (eyes, nose, mouth etc.) are presented to the appropriate network. The neural network portion of work is outside the main thrust of our research but ran concurrently with researchers at Warwick University. We (Southampton University) were contracted to identify and isolate facial features for presentation to the networks. The Home Office provided a database consisting mainly of students and staff from Aberdeen University and police officers. The database contains over 1000 faces from different ethnic backgrounds, and genders. The images are typical passport photos: usually the whole head is visible, surrounded by a plain background but in many cases the face fills most of the image so that the head is outside the frame of the image.

Jia and Nixon [42] tested feature extraction techniques which were able to extract geometric features on a small (40 faces) database. In order to interface with the existing Home Office's psychological coding scheme it was initially suggested that we convert the existing FORTAN suite of feature extraction algorithms to be structured as MicroSoft (MS) Windows dynamic link libraries (DLL's), with a well defined and public application programmers interface (API). This would allow third party Windows development tools, for example MS Visual Basic, to be used to handle overall control of the system and provide a customisable Graphical User Interface (GUI) front end. It became apparent [3] that while the algorithms showed interesting feature extraction techniques on a database of 40 faces, they appeared insufficiently robust for feature extraction on a 1000+ database. Part of our research included devising new algorithms capable of addressing feature extraction on a larger scale than previously reported in the literature.

1.1.1 Manually coded face features

Each subject is represented by a front-view full-face image and a profile image. From these images the operator is required to produce a physical coding record of the face, mainly from the front-view image) and a psychological rating on the facial features. The physical measures consists of the location of 38 landmark points on a face, shown in Figure 1.1. Some of the points in the coding scheme (such as eyes nose, mouth) are quite obvious, others require a little more explanation. The landmark points and the measures derived from them is given in section 1.1.1.1 while the parameters for the psychological coding is given in section 1.1.1.2.

1.1.1.1 Physical coding

Points 1, 2. The manual coding process begins by inputting the eyes' points 18 and 19. The system draws lines at 120° to the horizontal axis which enables points 1 and 2 on the outer hair boundary to be input. Points (19, 18, 1) and (18, 19, 2) make an angle of 120°.

Points 5, 6. The system extends a vertical line mid way between the eyes, representing their axis. The operator is prompted to input the co-ordinates where this line intersects the inner and outer hair boundary.

Points 33, 34. The system prompts the operator for points 34, 32, 35 on the mouth. The system extends a line through points 34 and 35 and prompts the operator to input points 33 and 36 on the jaw. Our sponsors define points 33, 34, 31, 35, 36 as the mouth line.

Points 30, 32, 37. The system extends a line perpendicular to the mouth line through the centre of the mouth, point 31and prompts the operator for the upper lip, lower lip, and chin.

Points 3, 4. The system extends lines from the centre of the mouth to intersect the left and right jaw. The angle subtended by points 33, 31, 3 and point 36, 31, 4 is 30° in both cases.

Points 7, 8, 9, 10. The system prompts the operator for point 38 which is the highest point on the left eyebrow. The system then extends a line through this point to intersect the inner and outer hair boundaries on the left and right sides of the face.

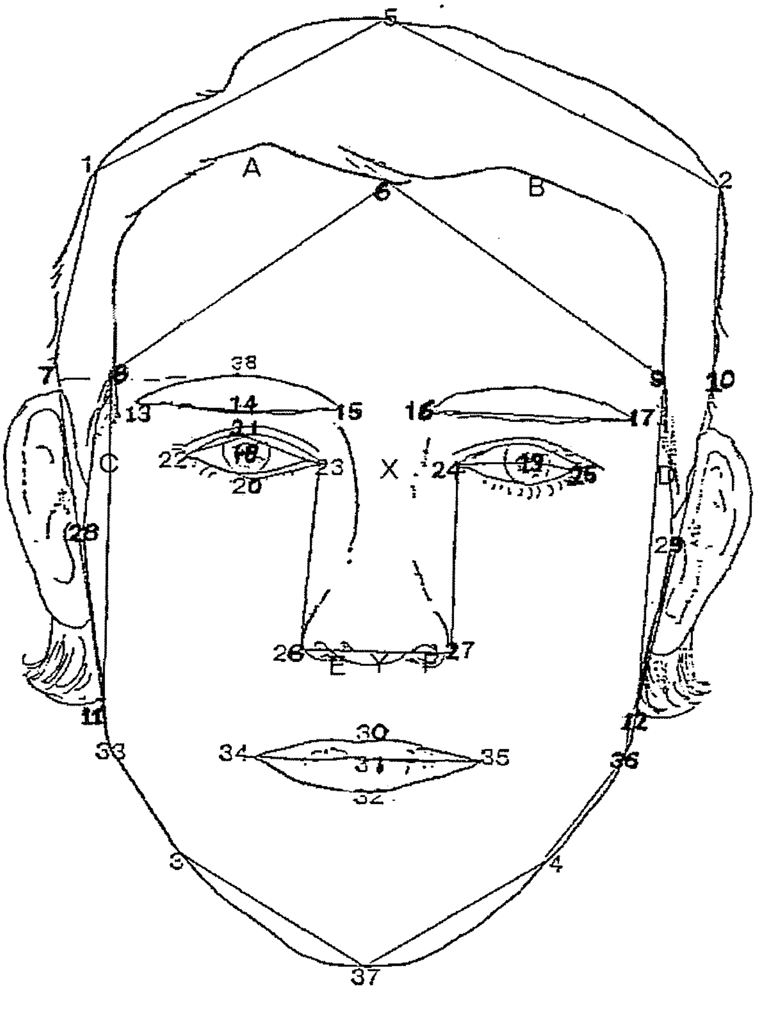


Figure 1.1 *Landmark points used in PITO coding scheme.*

Our sponsors use each set of face points to calculate three types of measures:

- Distance measures D, between two points $p1, p2$ with co-ordinates (x_1, y_1) and (x_2, y_2)

$$D = \left((x_1 - x_2)^2 - (y_1 - y_2)^2\right)^{\frac{1}{2}} \tag{1.1}$$

- Angle measures defined by the angle subtended by three points, $p1, p2, p3$.
- Area measures A, in which the features of interest are represented as an irregular polygon.

$$A = 0.5\left(\left(x_1 y_2 + x_2 y_3 + \ldots x_n y_1\right) - \left(y_1 x_2 + x_1 y_2 + \ldots x_1 y_2\right)\right) \qquad (1.2)$$

The area and distance measures of interest are shown in Figure 1.2 and Figure 1.3 respectively. Points 3, 37,4 constitute the chin angle which will vary with the lowest point on the chin, point 37.

Area	Points
Face area	3, 37, 4, 36, 10, 2, 5, 1, 7, 33
Hair area	11, 8, 6, 12, 10, 2, 5, 1, 7
Eye Area	22, 20, 23, 21
Chin area	33, 3, 37, 4, 36
Mouth area	34, 32, 35, 30
Nose area	26, 27, 23, 24

Figure 1.2 *Area measures from landmark points using the PITO coding scheme.*

Measure number	Name	x_1	y_1	x_2	y_2
dm1	Face Height	5	5	37	37
dm2	Face width at brow	7	7	10	10
dm3	Face width at cheek	28	28	29	29
dm4	Face width at mouth	33	33	36	36
dm5	Face width at chin	3	3	4	4
dm6	Hair length	5	5	11	11
dm7	Forehead height	6	8	5	5
dm8	Forehead width	8	7	9	10
dm9	Eyebrow height	14	14	21	21
dm10	Eyebrow width	16	16	17	17
dm11	Eyebrow setting	15	15	16	16
dm12	Eyebrow thickness	38	38	14	14
dm13	Interocular distance	18	18	19	19
dm14	Eye Narrowness	20	20	21	21
dm15	Nose width at bridge	23	23	24	24
dm16	Nose width at base	26	26	27	27
dm17	Nose length	5	23	5	26

dm18	Mouth width	34	34	35	35
dm19	Upper lip thickness	30	30	31	31
dm20	Lower lip thickness	31	31	32	32
dm21	Chin height	32	32	37	37

Figure 1.3 *Distance measures from landmark points using the PITO coding scheme.*

1.1.1.2 Psychological coding

The psychological features of interest and their ranges are listed in below are self explanatory.

Feature	Range
1. face height	short, long
2. face width	narrow, broad
3. face shape	bony, fleshy
4. complexion	fair, dark
5. complexion	pale, florid
6. complexion	unlined, lined
7. complexion	clear, blemished
8. hair length	short, long
9. hair tidiness	tidy, untidy
10. hair type	straight, curly
11. hair volume	bald, full-head
12. hair greyness	no grey, white
13. hair colour	black, brown, red, fair, blond
14. forehead height	low, high
15. forehead width	narrow, broad
16. forehead	straight, sloping
17. eyebrow thickness	thin, thick
18. eyebrow shape	straight, bent
19. eyebrow setting	meet in middle, set far apart
20. eyebrow height	low, high
21. eye size	small, large
22. eye narrow	narrowed, open
23. eye spacing	close set, wide spaced

24. eye setting	deep-set, protruding
25. eye colour	blue, grey, green, hazel, brown
26. ear size	small, large
27. nose (small/large)	small, large
28. nose length	short, long
29. nose width	narrow, broad
30. nose shape	concave, hooked
31. nostril size	small, large
32. nose tip width	narrow, broad
33. mouth size	small, large
34.upper lip thickness	thin, thick
35. lower lip thickness	thin, thick
36. chin size	small, large
37. chin shape	pointed, square
38. chin recession	receding, jutting
39. no facial hair	no, yes
40. moustache	no, yes
41. sideburns	no, yes
42. beard	no, yes
43. squint	no, yes
44. bags under eyes	no, yes
45. scars	no, yes
46. spectacles	no, yes
47. earrings	no, yes

Figure 1.4 *Psychological parameters for PITO coding scheme.*

1.1.2 The motivation and scope of automation.

Coding each subject in terms of their physical and psychological feature would appear to be a tedious and labour intensive process. Our sponsors would like to code faces from 43 police forces in the UK. They estimate 20,000 pictures to be presented each year with a suggested coding time of 2 minutes per face. Due to the volume of pictures and speed conversion rate required, it is obvious why our sponsors are keen to automate the process. Our sponsors are interested to be able to determine:-

- What features might be automatically extracted and at what cost.
- What degree of automation might be possible. User intervention is admissible but not preferred.
- What quality of picture is required.

They also advise that algorithms which facilitate:

- The location of the head, eyes, nose etc, determination of features such as complexion.
- Tidiness of hair etc.

are of particular interest.

Essentially, the main areas of interest lie in the physical points as our sponsors envisage that these parameters will be easier to define and extract than the psychological parameters. However, examination of the psychological features suggests that it may be possible to derive many of their values from the physical points. It is difficult to define an algorithm which robustly identifies hair so we shall focus on the facial features such as eyes, nose, mouth and the skin boundary.

1.1.3 Automatic extraction of physical points

Some of these physical features appear to be quite distinct whereas others would appear less clear, especially for extraction by automatic techniques using computer vision. The eyes are a well defined structure not only because they are the only face feature containing an analytically described shape (the circle in the iris), but also that there is considerable reflected structure. For example, the shape of the eyebrow can follow closely the eye socket which surrounds the eyeball. As such, it is not surprising that there have been computer vision approaches targeted primarily at automatic eye extraction. The mouth would appear to be quite distinct spatially, but it is subject to change in shape with facial expression and when talking. The nose would appear less distinct for automatic extraction since the PITO coding points' accuracy depends primarily on available contrast of the sides of the nose with facial skin. Other features are much less distinct. This is in part because they may depend on the absence of hair for precise definition. For example, the points on the inner hair line just above the eyebrow naturally depend on the arrangement of a subject's hair. Other points depend more on illumination, such as the chin. It is possible that the contrast in the jaw might be poor unless special illumination is used. Of these measures, the eyes appear to be the best defined, justifying their inclusion in model-based recognition scenarios. In a similar manner, and to conform with the ambitions of this project's sponsors, this research has developed new ways to find the eyes, with reliability and precision in large databases and without priming. Then, the other face features marked as of interest to the PITO scheme are extracted. Some of these are exactly those points used within the PITO scheme, such as extrema of the eyes and mouth. Due to the nature of the face data, the nose points are extracted via the minima which represent the positions of the nostrils. Given difficulty in precise location of some of the points defining the face outline, especially in the region of the jaw, these points are determined with a contour which follows the upper face region. In this way, a set of measures is derived which follows closely the original PITO coding scheme, but with selection consistent on the nature and abilities of automatic computer vision algorithms.

1.2 Combining the Face Features for Recognition

After extracting facial features (especially those which a witness may use to describe a suspect) we attempt to combine the features in a manner which reinforces the differences between faces in a database, and thereby minimise the probability of mis-identification. Using a database of 1000+ faces, we aim to show that the variance of a discriminatory measure can be reduced by using measurements which are statistically uncorrelated. It appears that the original application to faces was made by Jia and Nixon [42]. It has been noted that, "The importance of using multisource data.....lies in the fact that it is generally correct to assume that improvements in terms of classification accuracy can be achieved by employing additional independent features provided by separate sensors", [46]. We wish to extend this idea to multiple orthogonal face features and also expect improvements in classification accuracy. Furthermore, by identifying the features containing the highest variation we can minimise effort in extracting redundant features. This is of importance to vision-based biometrics, where the computational effort in feature extraction can be high.

1.3 Thesis Organisation

The remainder of this chapter presents an overview of some of the techniques available in face-based and feature-based recognition systems. Our work for locating face features will require feature-based tools. However, to justify our generic method of feature selection and combination we have adopted principal component analysis (PCA) from face-based techniques. Accordingly, we describe PCA and the eigenface method of face recognition before exploring some feature-based techniques. In chapter 2 we review earlier eye extraction techniques and present a new and robust method of eye centre location. In chapter 3 we extract the remaining features using the eyes as an initialiser. Section 3.4 provides an appraisal of the automatic feature extraction technique used in this thesis. The techniques were chosen to be best suited to system requirements indicated in section 1.1. Chapter 4 compares Fourier descriptors (FDs) representations proposed by Zahn and Roskies [89] and Kuhl and Giardina [41]. Zahn and Roskies descriptors were selected for comparing the forehead boundary, extracted by skin segmentation. A general framework for selecting and combining extracted features is presented in chapter 5 with overall conclusions and further work in chapter 6.

1.4 Face-Based Recognition systems

Turk and Pentland's [78] eigenfaces based on Principal Component Analysis (PCA), is one of the most well known and popular holistic approaches to face recognition. In order to appreciate an eigenface based recognition system the salient points on PCA are summarised from Manly's primer on Multivariate Statistical Methods [56]. Given a data set of n individuals expressed as p vectors, $X_1, X_2, \ldots X_p$, PCA attempts to find p ordered, uncorrelated indices $Z_1, Z_2, \ldots Z_p$, which describe the variance in the original data. The indices Z_i are known as the *principal components* of the data, and are ordered in decreasing magnitude, i.e. $\sigma^2(Z_1) \geq \sigma^2(Z_2) \geq \ldots \sigma^2(Z_p)$ where σ^2 denotes variance. Expressed mathematically,

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \ldots . a_{1p}X_p \tag{1.3}$$

and the variance of the first principal component, $\sigma^2(Z_1)$ is a maximum subject to the constraint

$$a_{11}^2 + a_{21}^2 + \ldots a_{1p}^2 = 1 \tag{1.4}$$

The constraint is introduced to prevent an increase in $\sigma^2(Z_1)$ simply by increasing the j-th coefficient a_{1j}. Similarly,

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \ldots . a_{2p}X_p \tag{1.5}$$

and the variance of the second principal component, $\sigma^2(Z_2)$ is a maximum subject to

$$a_{21}^2 + a_{22}^2 + \ldots a_{2p}^2 = 1 \tag{1.6}$$

with the additional constraint that Z_1 and Z_2 are uncorrelated. In general, the i-th principal component Z_i is given by

$$Z_i = \sum_{j=1}^{p} a_{ij}X_j \tag{1.7}$$

subject to

$$\sum_{j=1}^{p} a_{ij}^2 = 1 \tag{1.8}$$

and the Z_i being uncorrelated. The variances of the principal components, $\sigma^2(Z_i) = \lambda_i$ are the eigenvalues of the sample covariance matrix $\boldsymbol{C}$, where

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & & c_{2p} \\ \vdots & \vdots & & \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} \tag{1.9}$$

in which the diagonal element c_{ii} is the variance of X_i and c_{ij} is the covariance between vectors X_i and X_j. The set of constants $a_{i1}, a_{i2}, \ldots a_{ip}$ (scaled such that the sum of their squared values equals unity as in equation (1.8)) is an eigenvector calculated from λ_i, its corresponding eigenvalue. The sum of the eigenvalues equals the trace of $\boldsymbol{C}$ i.e.

$$\lambda_1 + \lambda_2 + \ldots \lambda_p = c_{11} + c_{22} + \ldots c_{pp} \tag{1.10}$$

The sum of the variance of the principal components is equal to the total variance of original variables. For maximum data compression, the original data should be highly correlated, so that PCA can represent a large number of variables by a smaller set of principal components, each component representing a different dimension in the data. If the original data was already uncorrelated, then there is little to advantage be gained from PCA. If the first few principal components are deemed to contain a sufficiently high percentage of the total variance, these principal components may be used to represent the data. However, at this point it is worth noting that the form of the principal components is still a linear combination of *all* the original variables in equation (1.7) and does not offer any information regarding the variance of the individual variables.

Turk and Pentland [78] used a database of M faces images, $\Gamma_1, \Gamma_2, .. \Gamma_M$, to create the image of an average face, Ψ, defined by

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \tag{1.11}$$

where each face image of size NxN was converted to an N^2 vector. The difference ϕ_i, between each face Γ_i, and the average face Ψ is given by

$$\Phi_i = \Gamma_i - \Psi \tag{1.12}$$

The difference in the database of faces can be characterised using PCA to find M uncorrelated vectors, $\mathbf{u}_k$ which best describe the data chosen such that

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(\mathbf{u}_k^T\Phi_n)^2 \tag{1.13}$$

is a maximum subject to

$$\mathbf{u}_l^T\mathbf{u}_k = \delta_{lk} = \begin{cases} 1 & \text{if } l = k \\ 0 & \text{otherwise} \end{cases} \tag{1.14}$$

where the scalars λ_k and vector $\mathbf{u}_k$ are the eigenvalues and corresponding eigenvectors of the covariance matrix

$$\mathbf{C} = \sum_{n=1}^{M}\Phi_n\Phi_n^T = \mathbf{A}\mathbf{A}^T \tag{1.15}$$

where the N^2xM matrix of face image differences constructed by packing the training vectors in a matrix **A**, given by,

$$\mathbf{A} = \left[\Phi_1, \Phi_2 \cdots \Phi_M\right] \tag{1.16}$$

Applying PCA to the covariance matrix C, with face images of size 128×128 pixels, would require finding the eigenvalues and eigenvectors of **C**, which is a $N^2 \times N^2 = 16384 \times 16384$ matrix requiring approximately 2.7×10^8 bytes of memory. These excessive memory requirements were avoided by first finding the eigenvalues μ_i, and eigenvectors $\mathbf{v}_i$, of $\mathbf{A}^T\mathbf{A}$ such that

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mu_i \mathbf{v}_i \tag{1.17}$$

Pre-multiplying by **A** gives

$$\mathbf{A}\mathbf{A}^T A\mathbf{v}_i = \mu_i \mathbf{A}\mathbf{v}_i \tag{1.18}$$

By definition, if **C** has eigenvalues and eigenvectors λ and **x**, then

$$\mathbf{C}\mathbf{x} = \lambda\, \mathbf{x} \tag{1.19}$$

Substituting $\mathbf{C} = \mathbf{A}\mathbf{A}^\mathrm{T}$, $\lambda = \mu_i$ and $\mathbf{x} = \mathbf{A}\mathbf{v}_i$, shows that $\mathbf{A}\mathbf{v}_i$ are the eigenvectors of *C*. The matrix $L = \mathbf{A}^T\mathbf{A}$ is an $M \times M$ matrix whose elements are given by

$$L_{mn} = \Phi_m^T \Phi_n \tag{1.20}$$

so finding the eigenvalues and eigenvectors of *L* requires less computational effort than finding the eigenvalues and eigenvectors of C from equation (1.15). The eigenvectors $\mathbf{u}_l$ are now found using

$$\mathbf{u}_l = \sum_{k=1}^{M} \mathbf{v}_{lk} \Phi_k \tag{1.21}$$

In practice it may be possible to use a reduced number of eigenfaces, *M'* instead of *M,* for faces to be adequately classified. The characteristics of a class of face images, Ω, can be assigned to a vector of weights which are proportional to the contribution of each eigenface in the class. Each face class is projected into "face space" using

$$\Omega^T = \left[\omega_1, \omega_2, \ldots \omega_{M'}\right] \tag{1.22}$$

where the weights, ω_k are given by,

$$\omega_k = \mathbf{u}_k (\Gamma - \Psi) \tag{1.23}$$

An image with a characteristic pattern vector Ω may be deemed to belong to the *k*-th face class with vector Ω_k if the Euclidean distance between the vectors is less than a threshold value θ_ε,

$$\varepsilon_k = \left\| \Omega - \Omega_k \right\| < \theta_\varepsilon \tag{1.24}$$

Alternatively, the image may be classified as a face which has not been presented to the system or classified as not being a face at all. Introducing a new face into the system requires retraining, a computationally intensive process but one which can be performed as a background task. Turk and Pentland also show that it is possible to use face space to locate a face in an image. A face map is constructed from an input image $I(x, y)$ by calculating the distance ε, between the local sub-image and face space. After some algebra and correlation (denoted by ⊗) they show that

$$\varepsilon^2(x,y) = \Gamma^T(x,y)\Gamma(x,y) - 2\Gamma(x,y) \otimes \Psi + \Psi^T\Psi \sum_{i=1}^{M} \left[\Gamma(x,y) \otimes \mathbf{u}_i - \Psi \otimes \mathbf{u}_i \right] \tag{1.25}$$

where regions of low distance ε indicate the likely location of a face. Their database comprised 2500 faces of 16 different subjects. Although they achieved a peak classification rate of 96%, their system was sensitive to variations in face orientation, illumination and very sensitive to head size. Their system also required the face images to be cropped to prevent intensity variations in the background from contributing to the eigenfaces.

Purnell *et al* [65] have noticed that most face recognition results currently in the literature have been generated from databases which have used Caucasians to test their algorithms, with little or no reference to other population groups. They evaluated an eigenface based recognition system and concluded that the system performance did not depend on the different population groups. The method of face recognition used in this thesis is based on extracting geometric features, as opposed to Eigenface based. However, based on the work of Purnell *et al* it seems reasonable to assume that a face recognition system based on extraction of geometric features would also be invariant to the population group. Robertson and Craw [70] have noted that many researcher have published results on their own, small sized databases with images which may be suited to their own particular application. Testing on different databases makes comparison of techniques difficult, hence the introduction of the Face Recognition Technology (FERET) Test in 1994. This database contains multiple images of a few thousand face images which can be used as a common database for evaluating face recognition systems. Some of the face images were reserved for training while others were reserved for testing. Phillips *et al* [64] evaluated ten face recognition systems in 1996 and improved versions of these systems in 1997. These systems used holistic methods such as, elastic bunch graph matching, template matching, neural networks and variations on eigenfaces. The recognition rates varied from 30% to over 80% if the eye co-ordinates were not given, rising to in excess of 90% where the eye co-ordinate were supplied.

Zong [32] uses singular value decomposition (SVD) for face recognition. The system used 45 training samples of 9 subjects. An error rate of 42.6% was attributed to statistical limitations of the small number of training samples. Isodensity lines (curves of constant grey-level intensity) have been used by Nakamura *et al* [60] on a database consisting ten pairs of pictures; three pairs of men wore spectacles, two pairs of men had thin beards and the other two pairs of face images were of women. On this small data set 100% accuracy was reported. Currently, there appears to be great interest in neural network based face recognition systems. Starkey and Aleksander [75] present an overview of a neural network based face recognition system. In this study, the net was trained to recognise 100% of their 96 face database.

Lanitis *et al* [52] used active shape models, which could account for variances in facial expression, individual appearance, 3D pose, and lighting. PCA was used to characterise the above variances from 690 training images of 30 individuals. Instead of searching for a key feature and using this to locate the remaining features, they attempt to fit a complete model to a face image. Their model used three main components:- (a) A shape model consisting of a manually acquired 152 Points Distribution Model (PDM), with the landmarks distributed around the eyes, nose, mouth, chin and ears. (b) a "shape-free" grey-level model of the face obtained by deforming each face PDM to the mean face PDM in the training set (c) A local grey-level model which uses the grey-level profile perpendicular to the 152 landmark points. The models can then be used in a multi-resolution Active Shape Model (ASM) search for the corresponding points on a new face image. To test the technique a new model was trained using 40 test images from the database and fitted to another 40 faces. Landmarks were successfully located when then the model was initialised ± 20 pixels from the correct position, ± 12 degrees from the correct orientation and 70% of the mean scale. A fast processing time of approximately 2 seconds on a Sun Sparc 20 work station was mentioned to achieve ± 3 pixels from the measured landmarks. A minimum Mahalanobis distance classifier was used on their normal test set consisting of 10 images of 30 people and their difficult database, consisting of 3 images of 30 people. On the normal database they reported recognition rates of 50.3% for the using the shape model up to 92.0% when using shape combined with local grey-level and shape free models. For the difficult test set the corresponding recognition rates were 15.6% and 48.9% respectively. Their results on gender classification suggests that the shape free and grey-level models were best suited for this task, achieving 94% correct classification using 10 images of 20 individuals for training and 10 images of 10 people for testing. Results for variations in illumination were not presented but can be accommodated by training the shape free and grey-level models under varying illumination conditions. The results appear encouraging and robust also offering a certain degree of insensitivity to occlusion. However, Lanitis *et al* point out that although the models are flexible, they are still specific and can therefore only vary in ways encountered in the training set. The training stage for this approach is quite tedious since it requires 152 manually coded landmarks per example face. The amount of manual work required is amplified ten fold since they have used a 10:1 ratio of training images to individuals. Training of this type of system is less attractive than training purely grey-level intensity variation required in Turk and Pentland's scheme [77]. It would be most interesting to use Lanitis *et al's* shape models on a larger database of individuals to evaluate the effectiveness of a training set. It may then be possible to build libraries of shape modes, from which shape-free and grey-level models may be derived.

Chellappa *et al* [13] present an extensive, critical literature review which outlines many of the techniques and applications for face recognition systems. This review compares the results obtained by various numerous authors employing techniques ranging from face-based to feature-based. The task of direct comparison of the techniques was made more difficult because the different authors used different databases of different sizes, which usually contained less than 50 people. Zhang *et al* [90] redressed these difficulties by comparing three topical techniques on a common database of 100 people. The techniques chosen for comparison were eigenfaces [78], elastic matching [49] and auto-association and back-propagation neural networks [15]. Their images were taken from the Massachusetts Institute of Technology, the Olivetti Research Lab, Weizmann Institute of Science and Bern University. A range of illumination conditions were included for each person and the images were cropped and scaled to have roughly the same size. A nearest neighbour classifier was employed for the eigenface approach. For the auto-association neural network, matching was performed using a second neural network operating in classification mode. The elastic matching scheme used Gabor wavelets [19] to generate a set of face features, which often co-incidentally but not necessarily, correspond to features that may be extracted in model-based such as eye, nose, mouth etc. Matching was again performed using a nearest neighbour classifier. Their results, summarised in Figure 1.5, indicate that the elastic matching scheme performed consistently well across the set of databases with recognition rate at least equal to the eigenface scheme. The neural network approach consistently provided the poorest recognition rate. When tested on the database composed of the four databases, recognition rates for the eigenface and elastic matching systems were 66% and 93% respectively. The drop in performance in the eigenface system on the large database was due to variations in illumination between the smaller databases which introduce biases in distance calculations, through the "average face". Given the relatively poor performance of the neural network system on the individual databases, recognition was not attempted on the large databases. However, in the overview of the three techniques, Zong *et al* referenced Bourlard and Kamp [5] who show that the best performance achievable by the neural network is limited to that obtainable by the eigenface system. Using Gabor wavelets to extract key feature points in the elastic matching method made matching algorithm less sensitive to lighting variations, since the key points rather than the whole image is used for matching. In addition, the elastic matching algorithm supports deformation which made the system more tolerant to changes in facial expression. Elastic matching is computationally more expensive than nearest neighbour classification used by eigenface systems. However its benefits outweigh its costs, since unlike eigenfaces, a retraining stage is not required to include a new face into the system for recognition. Finally, deformable matching offers itself to multi-resolution schemes which may be less readily usable in an eigenface environment.

Database	*Eigenface*	*Elastic Matching*	*Neural Network*
MIT	97%	97%	72%
Olivetti	80%	80%	20%
Weizmann	84%	100%	41%
Bern	87%	93%	43%

Figure 1.5 *Recognition rates achieved using popular face-based methods.*

1.5 Feature-Based Recognition Systems

In feature-based recognition we seek to locate, measure and compare constituents and the contour of the face. Kaya and Kobayashi [38] developed face recognition systems based on 9 manually extracted face features and a nearest neighbour classifiers. Their data was from photographs of 62 faces using special equipment to ensure consistent face orientation. They developed an information theory based model of a recognition system in which ideal feature extraction signal sources, corresponded to signal. Noise in the recognition system was modelled as the sum of measurement errors from hardware equipment and intrinsic within-class differences. They extrapolated a 90% recognition rate for 15,000 faces using constants derived from their 62 face database.

Kanade [37] used distance and angle measurements on the eyes, nose and mouth regions to characterise a database of 40 images. The face features were located by searching for local minima in the vertical and horizontal projections of edge maps. First the eyes were located using the vertical projection of the edge map. The vertical projection also enabled the height of the head, the locations of the mouth and nostrils to be determined. The horizontal projection enabled them to determine the width of the face and the bridge of the nose. Their database consisted of 20 training images and 20 test images which were acquired one month after the training images. They compensated for variations in picture size by considering ratios of their 16 face measurements and achieved a peak recognition rate of 75% using a simple distance based similarity measure.

Kelly [44] used many heuristics in a multi-resolution scheme to extract face and body measurements. This work is more significant than those mentioned above because Kelly's scheme used computer vision techniques, rather than manual extraction of features. The head was located by searching the edge map for an oval shape using template matching. The locations of the eyes, nose and mouth were then found using knowledge of the likely position of these features within the head. The outline of the body was segmented by generating an image which was the difference between an image that contained a person, and one that did not. They extracted 10 body measurements and distance measures between various part of the face, e.g. between the eyes, eyes to nose, eyes to the top of the head. Nearest neighbour classification using the leave-one-out rule was used on their database of 72 images of 10 people.

Craw *et al* [16] also used a multi-resolution scheme to locate face features. They reduced their images from 128×128 to 8×8 using local averaging. The head outline on the reduced image is traced using a line follower, which included some heuristic rules, guided by the template, to identify edge pixels belonging to the head outline. They found the process was more successful by iterating from the low resolution of 8×8 through 16×16, 32×32, 64×64 before reaching the full image size. They did not attempt to classify faces from their 20 face database. Instead their subjective appraisal of the extracted features indicated 50% correct eye location, 67% correct eyebrow location, 95% correct mouth location and 60% correct head outlines. Craw *et al* [17] used another template matching scheme to locate face and its features in an mage. They attempted to locate 40 landmark points using a template composed of 1462 feature points. They used simulated annealing to optimise the template which was represented by a polygonal random transformation as described by Grenander *et al* [30] solution than their line following algorithm, but only 1292 were reliably located.

Brunelli and Poggio [8] compared geometric feature extraction with template matching and concluded that the optimum strategy for face recognition is based on holistic methods, specifically template matching. They used feature extraction techniques similar to those used by Kanade [37] to extract face features such as eyes, nose, mouth and eyebrows. Dynamic programming was used to follow the roughly elliptical line of the chin. From these facial features they constructed a 35 dimensional feature vector. The suitability of the features for recognition was examined by considering the Min/Max ratio, which is the minimum distance to a wrong correspondence divided by the maximum distance to the correct correspondence. High values of Min/Max ratios are desirable since it implies high separations of the face classes. For their data set, the Min/Max ratio varied from 1.6 to 1.3 as the class size varied from 5 to 47. A Bayes classifier provided recognition rates which varied from 90% to 50% over the range of the class size. Recognition experiments using correlation were performed to determine the effect of scale and pre-processing. Their results indicated that intensity gradient, computed with an L_1 norm on a Gaussian Regularised image was better than on an unprocessed image, normalisation by the local intensity or the Laplacian of the image intensity. The authors mention a 100% success rate using their correlation based recognition scheme. They also note that the recognition rate achievable using PCA, for a given set of images, will be at best equal to that attainable using correlation, since principal components are linear combinations of the data.

1.6 Conclusion on Current Face Recognition Methods.

Neural Networks, eigenfaces and other holistic face recognition systems may use a large database, but multiple cues of the same face are typically required for training. They do not typically provide individual face features, in which our sponsors are most interested. It is possible to train holistic methods to locate individual features. For example, the idea behind "face space" to locate a face may also be extended to locate local face features. Our sponsors have expressed a preference to minimise human intervention. Model based methods do not generally require a potentially long training phase are therefore better suited to our sponsor's requirement of minimal human intervention. Generally, model based feature extraction is slower than a trained holistic system. However, if required, speed issues may be resolved by using hardware implementations or relying on the ever increasing speed of processors. From the previous sections, it is clear that there appear to be few face recognition systems with the ability to recognise a face within a large number of distinct faces using model based techniques. The feature extraction approach is readily analogous to our application since the witness describes the constituent face features. Although there are quantitative accuracy measures for locating individual features, systems which provide such measures for a combination individual features, are rare.

1.7 Databases used for face recognition in this thesis

In this thesis we have used faces from four different databases which are

- An existing database from Jia and Nixon's work [42] with some additions from the Southampton University World Wide Web (WWW).
- A large database of single shots of police officers supplied by our sponsors, the PITO database.
- The Aberdeen 1000 database which is a large database of 1000 faces also, supplied by our sponsors.
- The XM2VTS database [54] is a low cost commercially available system which contains multiple front-view shots of each person.

Publishing restrictions apply to our sponsor supplied databases and the resolution of existing databases from Jia and Nixon was not as high as that in the XM2VTS database. Consequently, towards the end of the research programme the XM2VTS database was chosen to show the results of our new feature extraction algorithms, but we shall describe the other database used throughout our research.

Jia and Nixon's images were of relatively low resolution (e.g. 174 x 250 x 8 bit grey scale) and were captured in controlled lighting conditions to minimise shadows. It is significant to notice that these images have a light face on a dark background whereas the background is approximately the same grey-level intensity as the face for the Web images. The face images in Jia and Nixon's database captured the subject's head to just below the chin, excluding clothing. Thus by using careful lighting Jia and Nixon started with face images which were readily segmented from background. Consequently, their idea for eye location of searching the whole image for pair of local intensity minima under the eyebrows seems reasonable on their images. The images from other databases are not subject to such favourable lighting conditions.

The face images from the PITO database are usually of uniformed police officers. The vast majority of the faces in this database were of Caucasian males in the age range 25 to 45, with short hair and without beards. These images contained additional items which were useful in the task of manual extraction and subsequent coding of extracted face features. These items included colour charts so that the human expert, but could subjectively classify the subject's skin colour, rulers to provide a reference of size of the face features and text to label the file currently being coded. Such additional facilities may have proved useful to face coding by a human expert could hinder automatic segmentation of the face. A typical image from this database, shown in Figure 1.6(c), used 509 x 634 x 8 bit colour. The grey-level intensity of the background is approximately the same as that for the skin, so the face can be segmented using colour information without producing very strong gradients at the boundary between the background and the skin. This is a benefit for extracting the chin which usually has weak edge strength.

The face images from the Aberdeen 1000 database are coded to 24 bit colour and the face could be segmented from the background by using colour information. Revealing any faces from this database is prohibited. Even with this restriction, Aberdeen 1000 database has already been used in some studies [16]. Although these images did not contain extraneous items as in the PITO database, there was often a loop on the subject's clothing for example on the bottom right of Figure 1.6 (d), which made eye extraction more difficult. We cannot display any of the faces in this database, but it is important to note that these pictures would appear to have been taken in the mid 1970s as opposed to the early 1990s for the PITO images. The Aberdeen 1000 faces were not constrained to have short hair and minimal facial hair as in the PITO database. Long hair, sideburns and other facial hair present in the Aberdeen 1000 gallery of faces also caused problems for feature extraction.

Images from the XM2VTS database (24 bit colour) have a dark background resulting in strong gradients between the skin and the background, which hinders the process of extracting the chin using snakes. However the main advantage of using this database over all the aforementioned databases is that there were up to four copies of each face. For a given similarity measure, we require the within class similarities to be better than between class similarities. The other databases contained only one instance of each face, i.e. a within class size of one which renders these databases less suitable for face recognition. The XM2VTS database is a large inexpensive database which is widely available. We believe that it will become a popular database providing a common set of research quality images, enabling direct comparison of new algorithms.

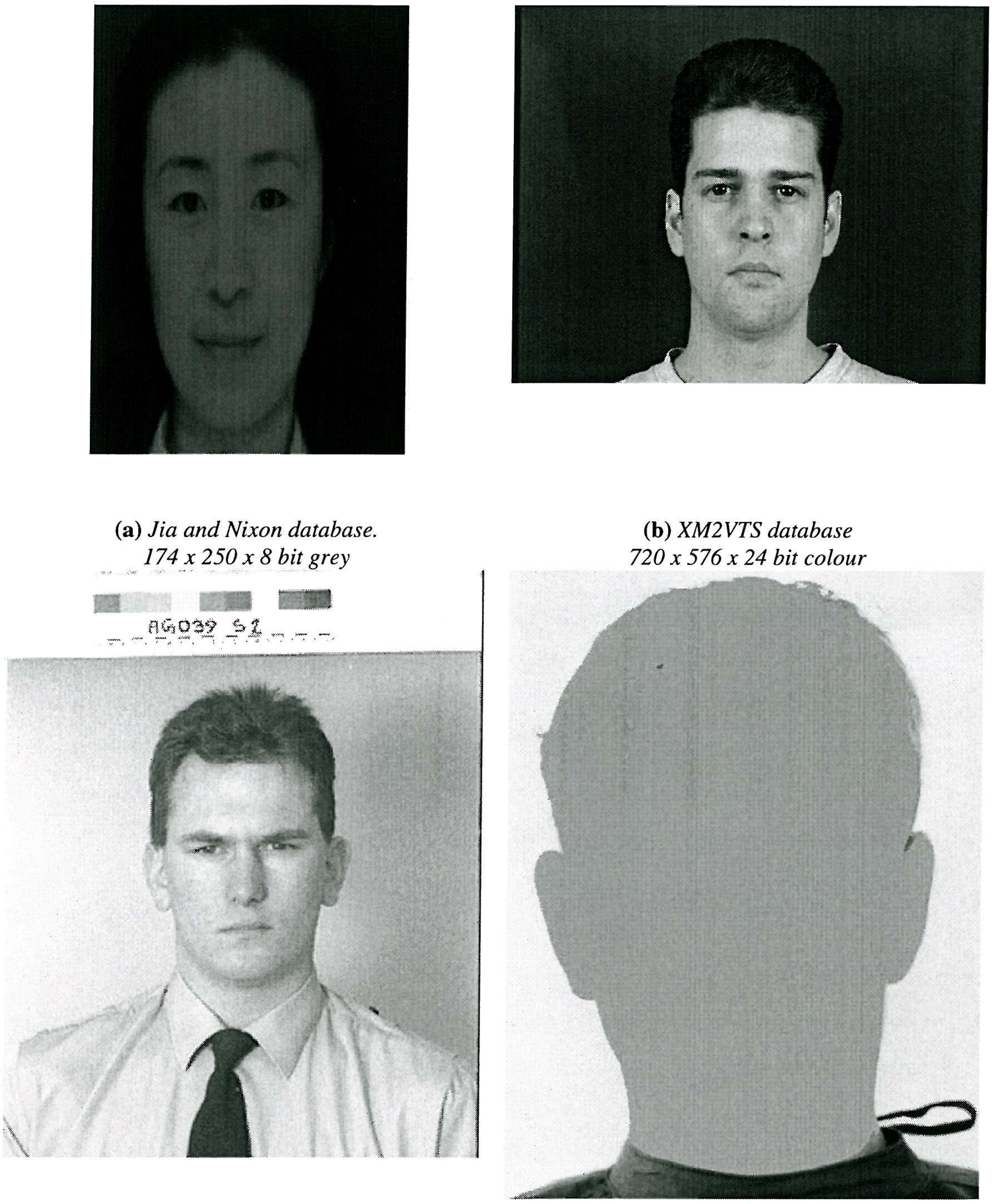
1.8 Contributions of this thesis

This thesis describes a programme of research aimed at improving the extraction, selection and combination of features for model-based face recognition systems. New algorithms for face feature extraction are developed which are both of academic interest and practical interest to our sponsors. The main contributions contained in this thesis are indicated below.

- We begin our search for features with new method for finding eyes in an image. An evidence gathering process is reformulated to implement a new concentricity operator. The concentricity operator is applied to the whole image and eye candidates are identified by peaks in the concentricity map. Our method offers invariance to image scale and rotation, yet requires few parameters.
- The location and parameters of the best candidate eyes are determined by applying an improved deformable eye template initialised at the peaks in the concentricity map. The internal energy of the template is optimised when the template contracts to a point; a state which is in conflict with the template's image energies. Existing templates attempt to resolve these conflicting energy requirements, by introducing artificial internal energy terms designed to prevent the template from contracting. By improved modelling, we alleviate this problem and avoid the problem of determining the associated coefficients of the balancing energy terms.
- Using our concentricity operator and improved eye template, we achieved 91% and 93% successful location rates on face databases of size 1000 and 88 respectively. The size of our database is significantly larger than many of the databases used for eye location and combined with the high levels of success provides a useful alternative to other methods used in the literature.
- The skin boundary can be extracted by assuming the region under the eyes is skin and applying standard region growing techniques. Features such as eyebrows, lips and nostrils may then be extracted by noting that they will be enclosed by the skin boundary, but have different statistical properties. We tested this method on a database of 44 faces so the techniques do not command the level of confidence that may be afforded to our eye location work. Nonetheless, it provides an alternative method of face organ location which opens further avenues for research.
- Next the discriminating power of the components of a 90 dimensional feature vector was examined. By normalising a set of features we derived a measure of the intrinsic variance contained within each feature set. The intrinsic variance of each feature provided the

coefficients for a distance based similarity measure which in turn provides the basis for a general method for combining features of disparate magnitude and independent sources.

- Our method embraces some of the concepts used in PCA in as much as it is able to account for the major components of the total variance. PCA requires that *all* the original variables (or features) are used to represent the variations in the data. Using our formulation, the variance of each *individual* feature is extracted which enables us to select the most discriminating feature. The new approach produced an increased variance in the classification matrix compared to a system of equal coefficients.
- The feature vector was composed of 29 geometric measures, 6 colour measures and 55 Fourier descriptors. Surprisingly, although the eyes were essential for locating the other face features, their deformable template parameters were amongst the least discriminating set of features, while the Fourier descriptors of the forehead contour boundary contained much higher discriminatory power. Classification tests on a database of 44 faces from the XM2VTS database yielded modest recognition rates of 72% and 77% using automatic and manual feature extraction. However, the XM2VTS images for each person were captured over a number of months which makes the database more realistic than other databases used in the literature and more difficult to achieve high classification rates. A simulation on the effect of measurement noise on system performance revealed that errors of only 6 pixels on the geometric feature set could result in a 43% classification error. However when we combined the geometric measures with perfectly extracted contour measures from its first eight Fourier descriptors we achieved 100% classification.
- In addition to presenting new methods for model-based feature extraction, we have presented a new method for identifying the features which offer the most discriminating power for use in a model-based recognition system. Given that model-based feature extraction is computationally expensive, these contributions are important system considerations, since they can be used to guide feature selection for model-based automatic face recognition.



(a) *Jia and Nixon database.*
174 x 250 x 8 bit grey

(b) *XM2VTS database*
720 x 576 x 24 bit colour

(*c*) *PITO database*
509 x 634 x 8 bit colour

(*d*) *Aberdeen 1000 database.*
384 x 512 x 24 bit colour

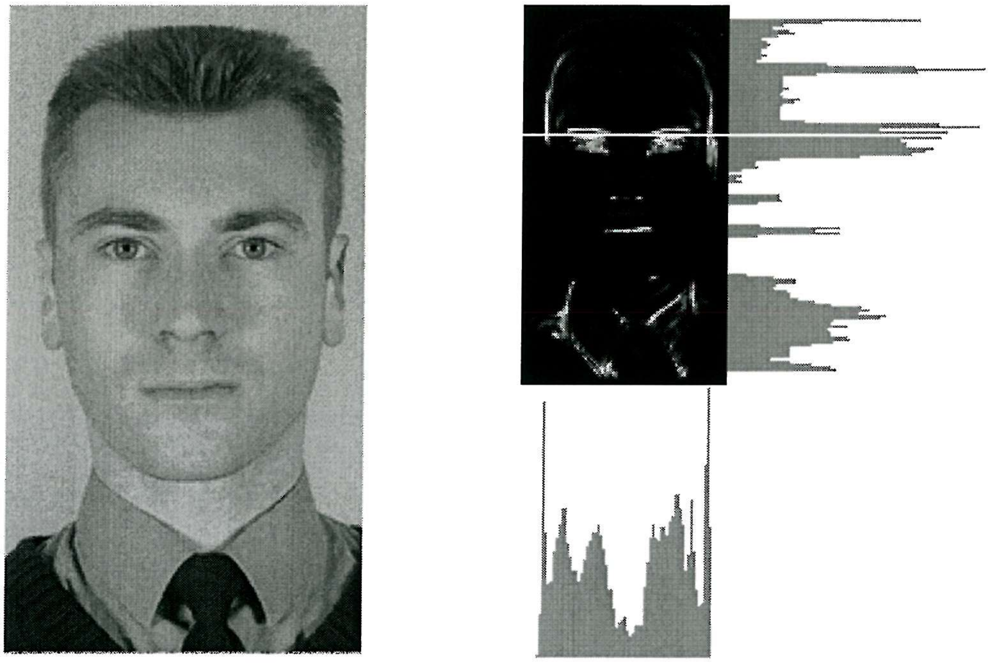
Figure 1.6 *Sample images from the databases used for feature extraction*

2. Locating the Eyes

2.1 Introduction

Finding the eyes is an important stage of feature extraction in automatic face recognition. They are spatially well-defined round structures on the face and the distance between them varies little with different facial expressions. Thus, measurements of eye spacing have been used to provide invariance to distance from the camera and to normalise other face measurements [42] [75]. Having located the eyes, we can use them combined with other anthropometric measures to locate the position of other face features.

Reisfield and Yeshurun [68] [69] used a generalised symmetry accumulator to locate the eyes. A face usually exhibits a high degree of symmetry about a line through the vertical axis of the nose. For example, the eyes are either side of this axis, as well as being locally symmetric. Thus the symmetry accumulator can be used to locate the position of the eyes. One apparent advantage of the approach over a correlation-based method such as [42] is that it is independent of scale or orientation. However since no *a priori* knowledge of the face is used in the search for symmetry, the process is computationally intensive, requiring several hours to produce a symmetry map. The authors mention a success rate of 95% (on a very small database) providing the faces occupy between 15 and 60% of the image. The main difficulty encountered when applying this symmetry formulation to our database was the lack of feature selectivity. In Figure 2.1 we show a face image in (a) and the corresponding symmetry map in (b). In (b) we also show the integrals of the vertical and horizontal projections of the symmetry map. Considering the vertical projection, it can be seen that there are local peaks which correspond to the location of the nose and mouth. The eye region produced the largest peak as shown by the white line which has been extended into the symmetry map. Note however, that there are large peaks due to the subject's shirt collar and hair, which suggests that it would be difficult to select the desired features using symmetry alone. Parsons and Nixon [62], refined the symmetry operator to improve local feature sensitivity, enabling large features to be distinguished from small ones. By adding just one parameter, the position and value of the peak in the symmetry accumulator space can vary with the choice of the size of a target feature. However, the full effectiveness of this improvement was not confirmed by *backmapping* to the image space and identifying the extracted feature points. Overall, the general isotropic symmetry concept does not use the circular shape of the iris; a better result may have been achieved using a rotational symmetry accumulator as described by Yip *et al* [86].



(a) *Original image* **(b)** *Symmetry map.*

Figure 2.1 *Hair and clothing as well as eye have high symmetry.*

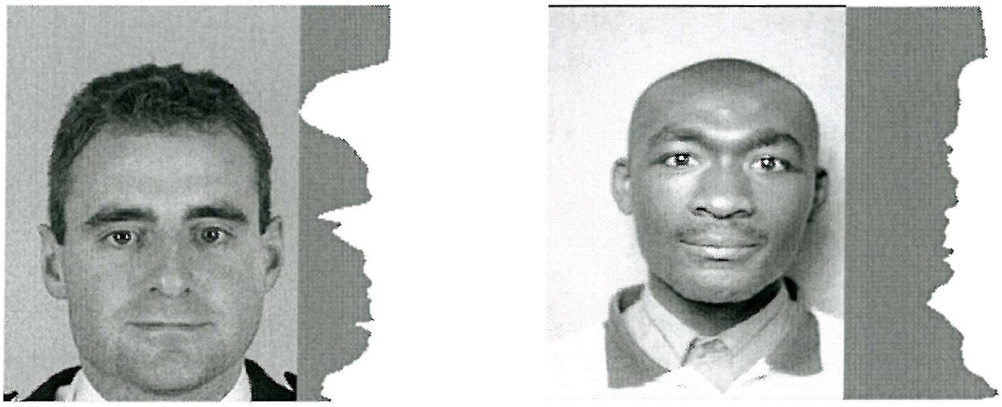
A deformable template was used by Yuille *et al* [88] and Xie *et al* [84] to locate the eyes. The eyes can be modelled mathematically as a circular iris which is enclosed (or occluded) by two parabolae, the eyelids, containing two white regions, the sclera [85] [88]. In order to determine the parameters of the iris, sclera and eyelids, appropriate energy functions are defined using the information about the valleys, edges, peaks and intensity of the face image. The eye template interacts with the face image by adjusting its parameters to minimise a composite energy functional. Thus, finding the eyes reduces to numerical optimisation of minimising the energy functional. The template is sufficiently flexible to locate the eyes despite variations in size, orientation and lighting conditions. More shape information is used, but the algorithm sometimes does not converge to the desired result. The algorithm was computationally costly, involving sequential change in the values of up to 11 parameters, followed by numerical optimisation of the energy term. The success of the process was very sensitive to initial starting conditions. When the template was initially positioned above the eyebrows the algorithm failed to distinguish between the eyebrows and the eyes. It was claimed that by using a better choice of parameters and optimisation techniques, the problems resulting in non-convergence suffered in [88] were alleviated, however the size of their database is not reported.

Lam and Yan [50] reported that they found the eyes by searching for the eye corners. They suggest that these corners would be located where the two parabolae of the eyelids intersect and where an eyelid parabola intersects with the circular arch of the iris. This approach cannot be applied across the whole face because there are several places on the face where corners may be detected, notably the mouth. On our databases, we also found that there was often poor contrast between the lower eye lid and the skin as illustrated by the edge map of Figure 2.21. Our remedy to the lack of corner information was to develop Yuille's eye template as detailed in section 2.7.2. As with the symmetry operator, we suggest that searching for corners alone would be insufficient to locate eyes but a second stage of examining the candidate eye locations for the existence of an eye should improve the process.

In [76] Stringa uses the bright spot of light reflected from an illuminating light source to locate the pupil. Even though it is possible for the light spot be absent, Stringa relies on this and then postulates that the horizontal intensity variation across the eye will be symmetric about this bright spot, with a change in intensity between sclera and iris and another sharp change in intensity between bright spot and iris. The shape of the eye is used only indirectly, via the intensity variations. Stringa reported 100% success on a database of 333 faces with a fast processing time. However, the technique is intrusive and may not be so successful in cases where an artificial light source is not be available.

A standard Hough transform (SHT) has been used [58] to detect the instance of a circular shape and an ellipsoidal shape which approximate the perimeter of the iris and sclera respectively. Both the measured iris and sclera centre were used to provide eye spacing measurements. The iris centre measurements were within ± 2 pixels of the subjective estimate of the eye centre. A recent approach has employed a non explicit form of concentricity [48] and achieved good results. However, the databases used in [58] and [48] were relatively small, comprising only 6 pairs of eye measurements.

Jia and Nixon [42] and Craw *et al* [16] share the premise than the eyes are a pair of local intensity minima under the eyebrows. Jia attempted to locate the iris using template matching to define the search window for the eyes. Within this window the intensity either side of a circular or round edge was considered. If the intensity inside the round edge was at least 1.5 times the intensity outside then the edge pixel was deemed to be a candidate iris pixel. The final location for the eye centre determined as the location which had the highest ratio of candidate iris pixels per unit radius of the round edge. This idea of count per unit radius can work well if the eyes are wide open, however in the not uncommon case where the iris is occluded by the eyelids, the technique tends to favour small circles since they appear more completely formed. Numerous small circles yielding a high count per unit radius can be found within the eye window which are not near the eye centres. The eye window is likely to be less useful if the image is that of a person with a dark skin tone. Clearly, dark skin combined with light intensity eyes would exacerbate the search for the pair of "local intensity minima" as illustrated by Figure 2.2. In this figure, we project the mean intensity in grey to the right of each face. For Figure 2.2(b) the local minima are less pronounced because his skin tone and face features are dark and does not offer regions of high variance in intensity. In Figure 2.2(a) there are more pronounced local minima due to a combination of lighter skin tone and dark facial features. However, there is still a problem to decide which local minimum corresponds to the eyes. The global minimum actually corresponds to the head, midway down his hair. The 2nd largest local minimum corresponds to the eyebrows while the eyes are only the 3rd most significant local minimum. The width of the eyebrows is usually approximately twice that of the eyes. This implies that the eyebrows correspond to a more significant local minimum that the eyes, assuming the eyes and the eyebrows exhibit the same grey-level intensity. Locating the eyes by means of local minima appears to introduce uncertainty regarding which local minimum to select for an eye window.



(a) *Adequate eye window localisation* **(b)** *Local intensity minimum less pronounced*

Figure 2.2 *Defining eye windows based on local intensity minima.*

Shackleton and Welsh [71] combined Yuille *et al*'s deformable eye template [86] with Turk and Pentland's use of PCA [76] as a basis for classifying and recognising facial features. They envisaged a complete recognition system based on recognition of local features such as the eyes, nose, mouth etc which are classified and recognised using PCA. Their scheme used a deformable eye template as an example local feature and applied PCA to geometrically normalised eye images. A training set of sixty faces was used before recognition tests were performed on a further 24 faces. They included a term to improve segmentation between the sclera and skin. The results produced 31 fits comparable to a manual fit, 8 fits which could have been improved by hand, 2 fits with some error noticeable, 12 obviously poor fits and 7 instances where the eye template failed to locate the iris. In cases where the deformable template results were unsatisfactory, manually extracted eye images were used in order to verify the value of the PCA approach for recognition. The left eye of each face was used as the cue while the right eye was used as the target. They were able to match 16 left eyes to the corresponding right eye of the same face. Of the remaining eight eyes, five were amongst the best five matches. PCA was used to classify the extracted eye features, however the crucial stage of locating the eyes was not performed automatically. Specifically, they manually initialised the deformable template approximately 20 pixels from the iris.

In [3] we implemented and compared the SHT versus the candidate count per unit radius approach used in [42]. Our results indicated that attempts to locate the iris based on a count per unit radius tend to favour small circles in the corner of the sclera. Using 23 faces of the original database (half the original database) it was found that the SHT performed better than the method used in [42]. In order to circumvent the problem of defining an eye window and improve the performance obtained by the SHT a new method was developed which exploits the inherent concentricity of the eye region. This method, based on the HT uses edge gradient information instead of intensity.

None of the approaches hitherto (except Kothari *et al* [48] where concentricity was not explicit) have used one particularly strong feature of the eye, namely that it is the centre of a set of concentric shapes. This allows extension of a proven technique (the HT provides a result equivalent to matched filtering) in a non-heuristic manner. The resulting technique is invariant to scale and rotation, requires few parameters, and is not specific to the eyes. Also, it offers the potential for finding the eyes without priming by prior knowledge. Since the eyes are the only feature with this concentricity, a technique formulated to use it will deliver the eyes automatically, when applied to a whole face image. The eyebrows enclose the eyelids which enclose the iris which contains the pupil. We suggest that the eyes are the epicentre of two holes in the head which emanate concentricity. This premise is less prone to exceptional circumstances, which result in failure, than other methods use in the literature. Thus, concentricity affords the basis for a new method of eye location, which needs little prior knowledge concerning the face or its location. We formulate our concentricity operator in terms of an evidence gathering process, typified by the Hough transform.

Before the subtleties of our new concentricity formulation is employed on real face images we discuss the Hough transform.

2.2 Origins of the Hough transform (HT).

The Hough Transform [33] was developed by Paul Hough while studying particle tracks in a bubble chamber. To detect these complex patterns of points in a binary image, Hough interchanged the role of the parameters and the independent variable in the image. A discrete version of the parameter space, known as an accumulator, was constructed to accumulate evidence for the existence of a specified shape within the bounds of the parameter space. Each image point votes for the existence a specified shape for a range of parameter values. A peak in the accumulator indicated that the image points fit a shape corresponding to the parameter values at the peak. Although much work has been done in reducing the dimensional requirements, the memory and computation requirements increase exponentially with the number of parameters. However the HT has been shown to deliver the same results as template matching [83] but faster, since the HT requires only the feature points not the whole image space. This simple parametric transformation mapping was shown by Deans [21] to be merely a special case of the Radon transform [67]. The Hough transform is currently a popular technique for finding parametric shapes in images with low dimensionality such as lines, circles and ellipses [53].

2.3 The Standard Hough transform for Lines

A straight line connecting a sequence of pixels can be expressed in the form

$$y = mx + c \tag{2.1}$$

where x, y are points on a line of gradient m and intercept c. Re-arranging Equation 2.1 such that m and c are the variables and x and y are the parameters, then

$$c = -xm + y \tag{2.2}$$

which is the equation of a line with gradient $-x$ and intercept y passing through the point (m, c). Thus a point (x, y) in the image space can be mapped to a line in the parameter space and each line in the parameter space corresponds to a point in the image space. The intersection of lines in the parameter space indicate that corresponding image points are co-linear. Furthermore, the point of intersection in the parameter space give the parameters of the co-linear points in the image space and the number of intersecting lines equals the number of co-linear points. In practise, in order to search for a line in an image, we first define a parameter space in which to accumulate evidence of a line in the image. The parameter space is therefore usually simply referred to as an *accumulator*. For each point in the image, votes are cast for a range of one of the parameters, say *m*, which generates corresponding values of *c* as described by equation 2.2. A line of votes in the accumulator is implemented by incrementing the cells in an array, whose vector corresponds to the parameters, in this case the values obtained for *c* and *m*. If the same cell is incremented by votes from another image point, then this represents an intersection of a lines of votes. The cell with the highest count corresponds to the largest number of co-linear points.

There are two fundamental practical constraints to consider before the accumulator can be used. These are the quantization and the range of the parameters for the accumulator. Any point in an image has an infinite number of lines that may pass through it. In Figure 2.3, point C is shown with five lines passing through it ranging from $m = 2$ to $m = 0$. If the parameter quantisation is too coarse the desired line may not be detected. If the parameter space quantisation is too fine, the computation cost of evaluating equation 2.2 may become significant. The Cartesian parameterisation of a vertical straight line is difficult to realise on computer since vertical lines require a representation corresponding to an infinite slope. This problem can be circumvented by using the polar representation of a point (x, y) on a line given by

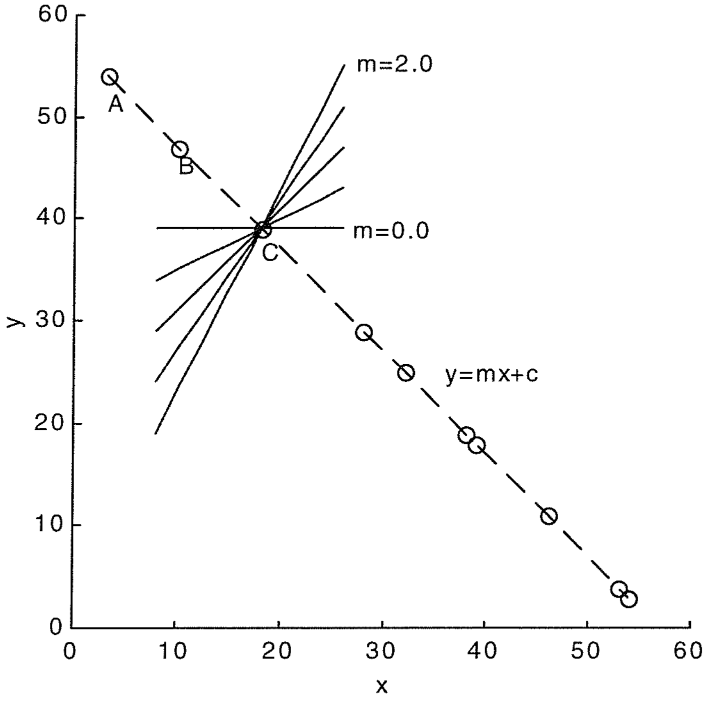


Figure 2.3 *Cartesian Representation of line y=mx+c*

$$\rho = x\cos\theta + y\sin\theta \qquad (2.3)$$

where ρ is the perpendicular distance from the origin to the line and θ is the angle subtended by the perpendicular and the x axis as illustrated in Figure 2.4. Figure 2.5 shows 10 image points corresponding to the line in Figure 2.3. Using the polar representation, the range of the parameters are θ from 0 to 360 degrees and ρ from 0 to $\sqrt{\left(nr^2 + nc^2\right)}$, where nr and nc are the number of rows and columns in the image. The parameters line in Figure 2.3 are m = -1 and c = 57 using Cartesian representation, which translates to θ = 45 degrees and ρ = 40 pixels using polar representation. The corresponding parameter space, shown in Figure 2.6, was quantised to an accuracy of 1 degree for θ and one pixel for ρ. The full range of the parameters were used to find the peak, but to provide clarity and detail, the accumulator is displayed only in the region of the peak. The peak in the accumulator has value 10 and is located at the co-ordinates θ = 45, ρ = 40, which is in agreement with the actual parameters of the line in the image of Figure 2.5.

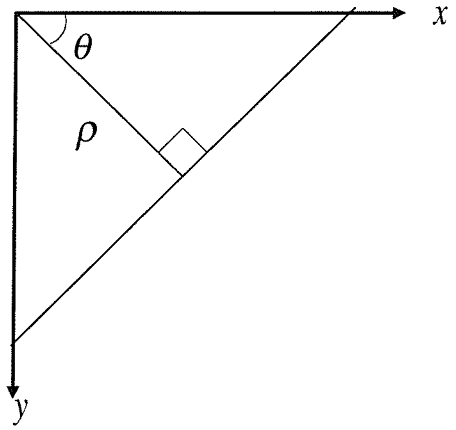


Figure 2.4 *Polar representation of a line*

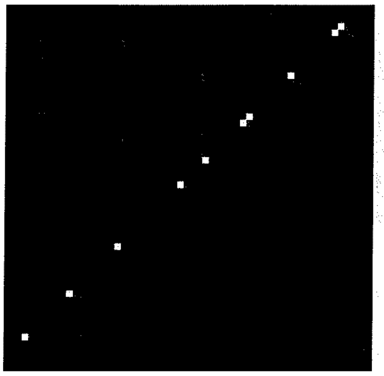


Figure 2.5 *Image of points in a line*

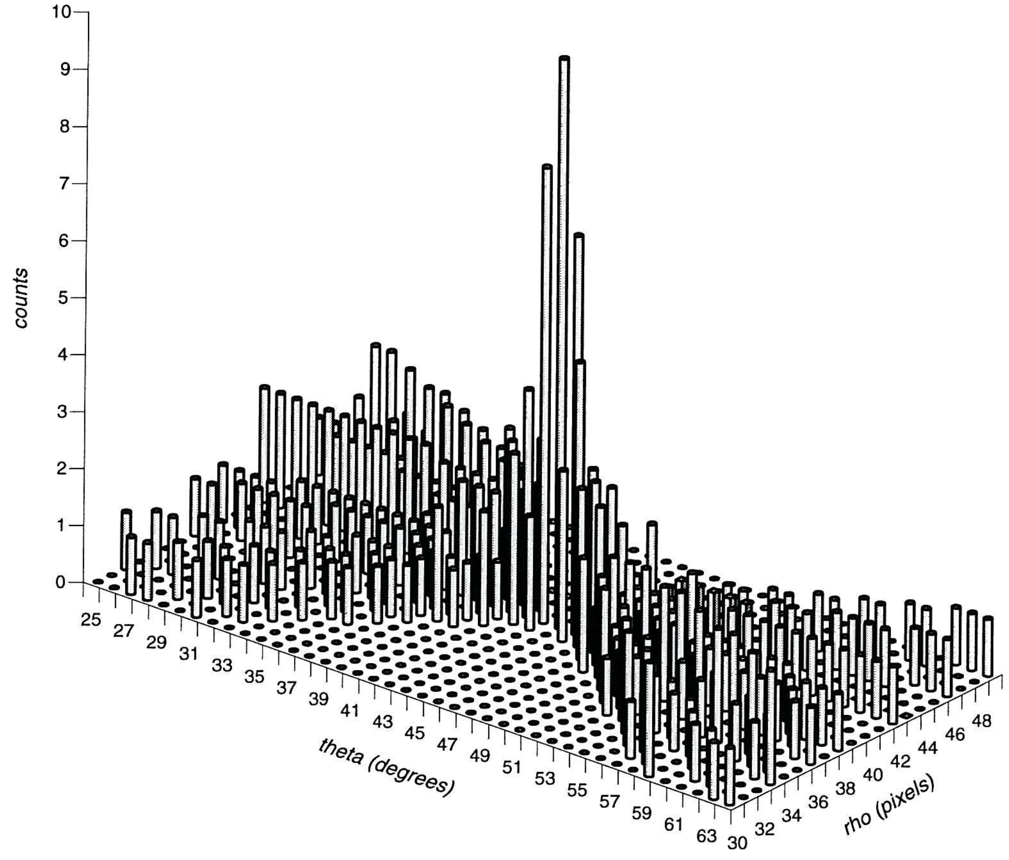


Figure 2.6 *Accumulator for image of a line of points.*

2.4 The standard Hough transform for circles

The equation of a circle can be expressed as

$$(x-x_o)^2+(y-y_o)^2=r^2 \tag{2.4}$$

where (x, y) is a point on the circle with centre (x_o, y_o) and radius r. In the previous section, a point in the image mapped to a line of votes in the two-dimensional parameter space. In the case of detecting circles, a point in the image space maps to a circle of votes of a given radius. Thus, each point in the image space generates a cone of votes as shown in Figure 2.7 The location of the peak count of intersecting votes in the accumulator space gives the parameters of the circle in the image.

Before discussing how gradient information can be used to implement a concentricity accumulator, it is useful to note that this may not be readily achieved using the SHT. Figure 2.8 shows a circle of radius 11 pixels, the accumulator space for radii ranging from 5 to 17 pixels and the resultant accumulator achieved by summing these counts over the range of radii. For the accumulator which corresponds to the radius of the circle in the image, $r = 11$, there is a large peak at the centre of the circle. For all other radii there are local minima of intersecting votes at these locations. The result of summing these counts over all radii indicates that there is not a peak at the desired locations and thus a concentricity accumulator using the SHT fails.

Brown [10] applies traditional signal detection terminology to the Hough transform. The target peak may loosely be referred to as "signal" while the counts elsewhere that do not contribute to the signal may be considered to be "noise". The target may be aliased by a circle of another size which is displaced from the true target location. Kiryati and Bruckstien [45] show that the SHT implies sampling of a non-bandlimited signal and also propose an effectively alias-free Hough transform.

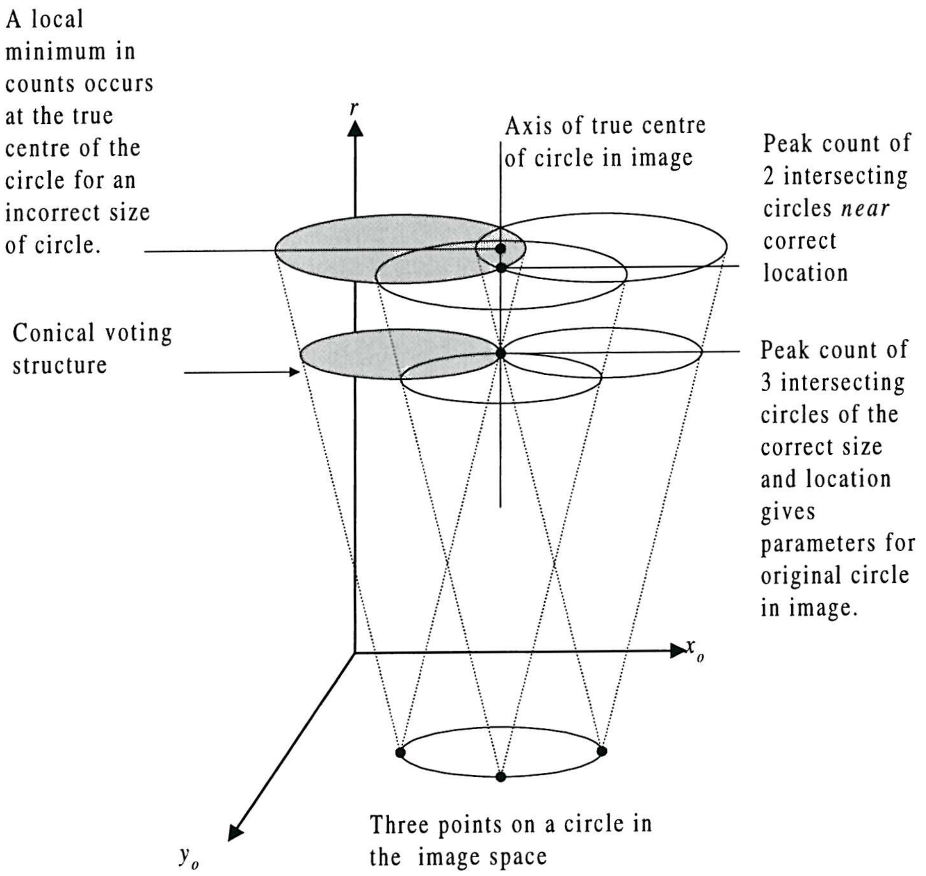

A local minimum in counts occurs at the true centre of the circle for an incorrect size of circle.
r
Axis of true centre of circle in image
Peak count of 2 intersecting circles *near* correct location
Conical voting structure
Peak count of 3 intersecting circles of the correct size and location gives parameters for original circle in image.
x_o
y_o
Three points on a circle in the image space

Figure 2.7 *Accumulator space for circles.*

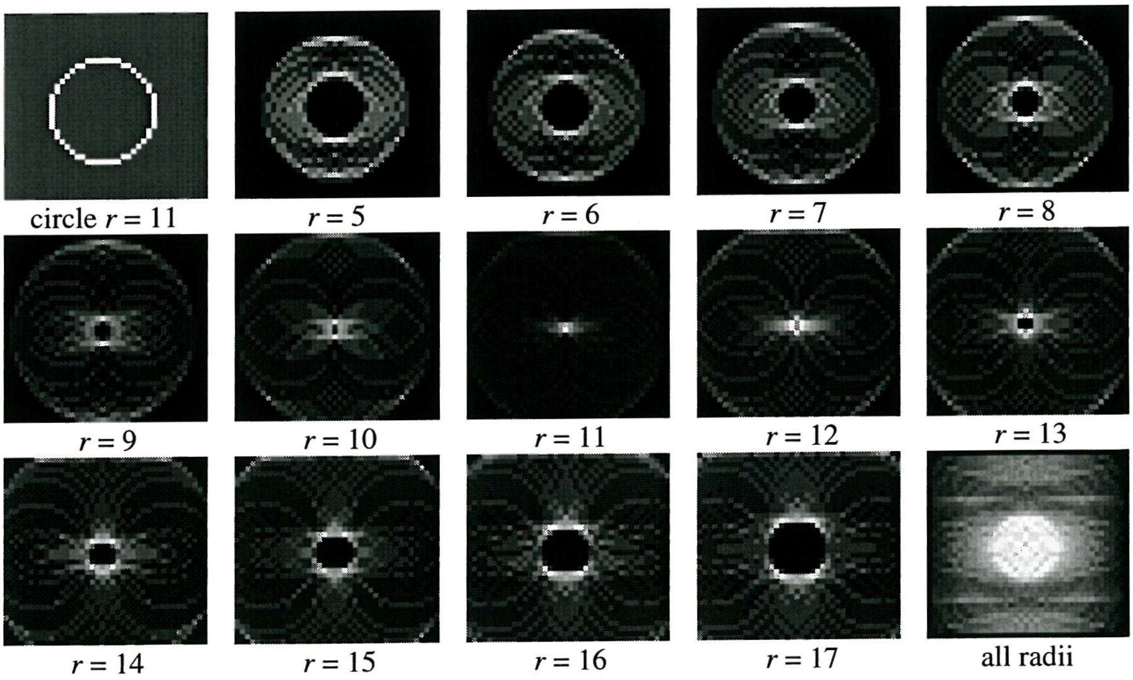


Figure 2.8 *Concentricity by SHT fails.*

2.5 Concentricity using gradient decomposed HT

The gradients of concentric circles can be arranged to combine constructively to define the centre of concentricity as shown in Figure 2.9. If point 2 is on the iris and point 1 is on an eyebrow, then the process of locating the eye centre may actually be enhanced by the eyebrows. Differentiating equation 2.4 with respect to x yields

$$dy / dx = -(x - x_o) / (y - y_o) \tag{2.5}$$

where dy and dx are the vertical and horizontal components of the gradient intensity at the point (x, y). By substitution,

$$x_o = x \pm \frac{r}{\sqrt{1+(dx/dy)^2}} \tag{2.6}$$

$$y_o = y \pm \frac{r}{\sqrt{1+(dy/dx)^2}} \tag{2.7}$$

Using equations 2.6 and 2.7, the centre of a circle can now be found using a pair of two-dimensional accumulators (x_o, r), (y_o, r) or a single two-dimensional accumulator (x_o, y_o). Using gradient information, we gain reduced memory requirements (two-dimensional space, *c.f.* three-dimensional space for SHT) at the expense of accurate measurement of the gradient at a given point.

In our first concentricity formulation [3], we defined the co-ordinates of the centre of concentricity as

$$(\max(X_o), \max(Y_o)) \tag{2.8}$$

where

$$X_o = \sum_{rmin}^{rmax} x_o \tag{2.9}$$

and

$$Y_o = \sum_{rmin}^{rmax} y_o \tag{2.10}$$

where *rmax* and *rmin* are the minimum and maximum values likely for the radii of the iris.

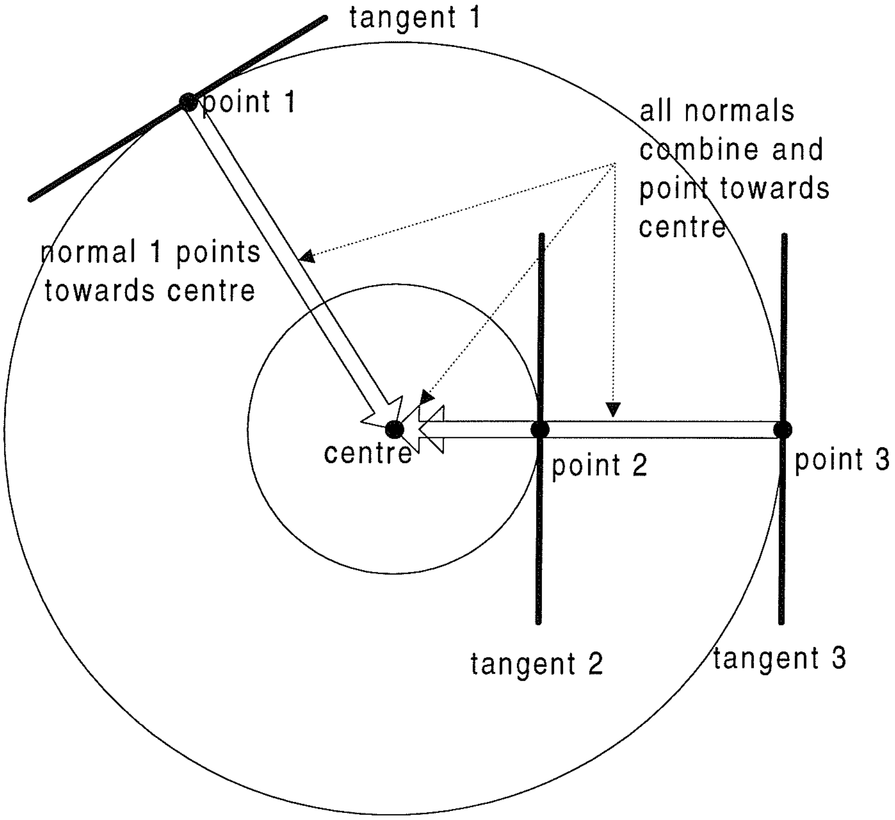
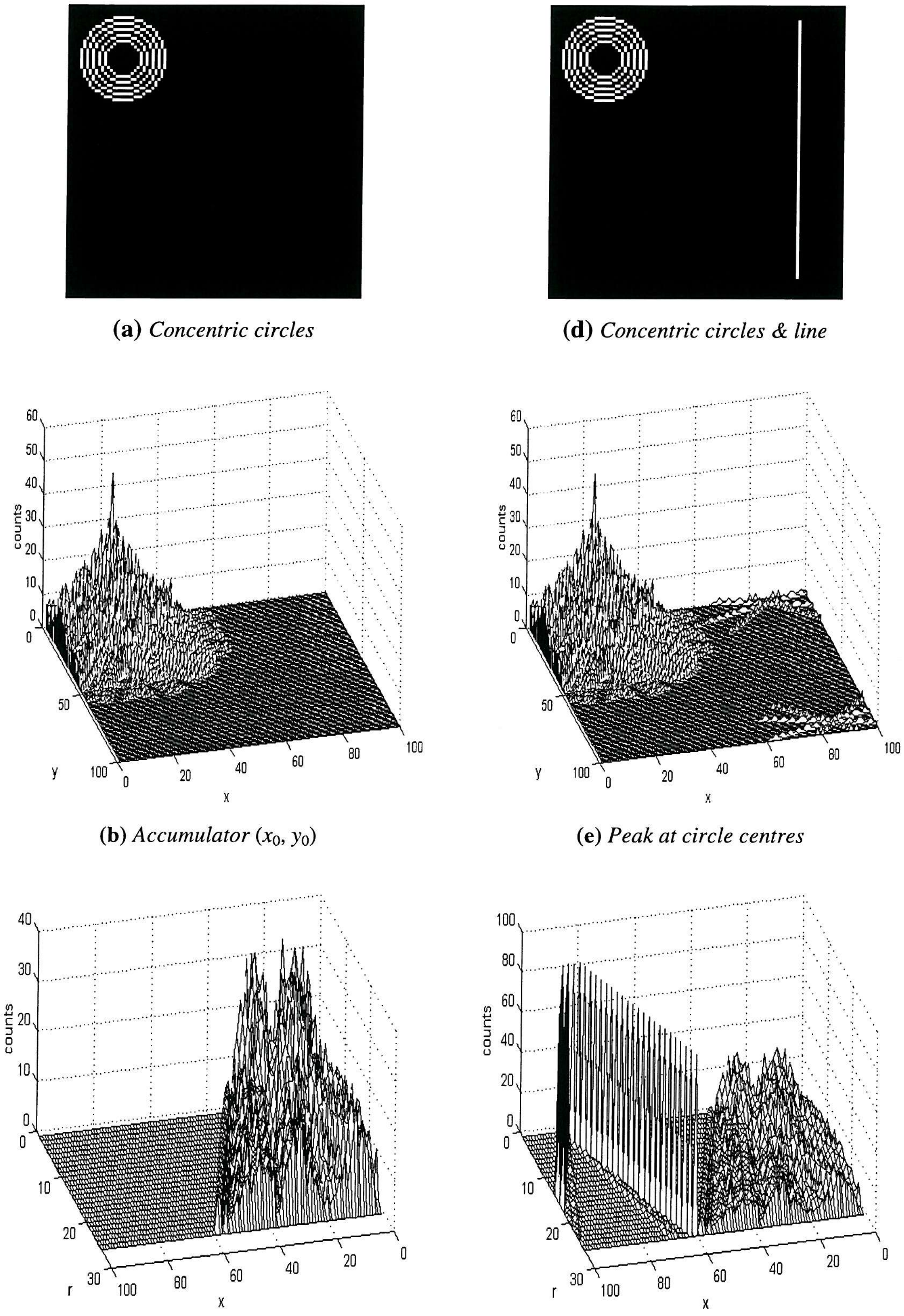

tangent 1
point 1
all normals combine and point towards centre
normal 1 points towards centre
centre
point 2
point 3
tangent 2
tangent 3

Figure 2.9 *Gradients of concentric circles intersect at the circle's centre.*

In the dual accumulator arrangement, horizontal and vertical lines can introduce valid but undesirable solutions to equations (2.6) and (2.7). In the case of a vertical line, equation (2.6) reduces to $x_o = x \pm r$. Each point on the vertical line results in the same cell, (x_o, r) being incremented and, the longer the line, the larger the undesirable peak. Equivalent difficulty can be experienced with horizontal lines in accumulator (y_o, r). In the single accumulator case, equation (2.6) again reduces to $x_o = x \pm r$ for a vertical line but equation (2.7) reduces to $y_o = y$. Now, cells given by (x_o, y_o) are only incremented once for each point on a vertical line. Figure 2.10 illustrates how the single accumulator is less susceptible to a vertical line than the (x_o, r) accumulator. Figure 2.10(a) show a synthetic image of concentric circles while Figure 2.10(b) and Figure 2.10(c) show the resultant (x_o, y_o) and (x_o, r) accumulators. The co-ordinates at the peak in the (x_o, y_o) accumulator correspond to the co-ordinates of the centre of the concentric circles in Figure 2.10(a). The x co-ordinate of the concentric circles can be found by summing the counts over the radii of interest and locating the peak of this distribution as described in [4]. Figure 2.10(d) shows a set of concentric circles in the presence of a vertical line. Using the (x_o, y_o) accumulator, Figure 2.10(e), it



(a) *Concentric circles*

(d) *Concentric circles & line*

(b) *Accumulator* (x_0, y_0)

(e) *Peak at circle centres*

(c) *Accumulator* (x_0, r)

(f) *Peaks due to vertical line*

Figure 2.10 *Locating concentricity peaks using accumulators.*

can be seen that the detected centre of concentricity is unchanged. However, in the case of the (x_o, r) arrangement. Figure 2.10(f), the accumulator counts are dominated by the presence of the vertical line in the image and the desired co-ordinates are less readily extracted.

2.6 Finding eye candidates.

In this section we present details of experiments for eye extraction techniques discussed in the previous section. The first experiment suggests that the SHT performed better than the correlation method of Jia and Nixon [42]. We then show that the dual two-dimensional concentricity accumulator performed better than the SHT, despite relying on some heuristics to filter straight lines. From the discussion in the previous section, we expect the single concentricity accumulator to provide the same advantages over the SHT as the dual concentricity accumulator, but without the need to filter straight lines from the image. The single concentricity accumulator was applied to the whole image without the need for ambiguous search windows. When this accumulator was applied to the bulk of the Aberdeen 1000 database and XM2VTS database we achieved success rates of 50% and 84% respectively.

2.6.1 Circular correlation vs SHT for eye extraction.

In this experiment we compare the performance of the SHT and Jia and Nixon's circular correlation method [42] when applied to real images for eye extraction. Their technique is similar to a Hough transform for circles but also attempts to incorporate grey-level information present in the eye region. The search space for points inside and outside the circle i_{in} and i_{out} are determined using

$$(i_{in}, j_{in}) = (i_0 - 2\cos(\theta), j_0 - 2\sin(\theta)) \qquad (2.11)$$

$$(i_{out}, j_{out}) = (i_0 + 2\cos(\theta), j_0 + 2\sin(\theta)) \qquad (2.12)$$

where (i_0, j_0) is the centre of a circle of variable radius representing the iris, (i, j) is a point on the circle representing an iris pixel and θ is the angle subtended by the x axis, the centre of the circle (i_0, j_0) and a point on its perimeter (i, j). The circular correlation method considers a pixel as a suitable iris pixel if the average intensity inside the circle representing the iris is 1.5 times less than that outside the circle, otherwise it is ignored. Candidate iris pixels are accepted if

$$e(i_0, j_0) = \begin{cases} 1 \quad \text{if} \quad \sum_{(i,j)=(i_{in}-1,j_{in}-1)}^{(i_{in}+1,j_{in}+1)} f(i,j) < \sum_{(i,j)=(i_{out}-1,j_{out}-1)}^{(i_{out}+1,j_{out}+1)} 1.5 f(i,j) \\ \\ 0 \quad \text{otherwise} \end{cases} \tag{2.13}$$

The circle with the largest ratio of pixels on an acceptable iris edge to those on a circle with variable radius, $c(i, j)$ gives the final estimate of the iris position. This ratio, R, is given by

$$R = \frac{\sum_{(i,j)\in c_2} e(i,j)}{\sum_{(i,j)\in c_5} c(i,j)} \tag{2.14}$$

where c_2 and c_5 are circular ring, two and five pixels wide, respectively. Figure 2.11 shows some sample faces from the database of faces used for this experiment. The sample face images bmc94 and ajgh were from the Southampton University World Wide Web gallery, whereas images Face10 and Face11 were from Jia and Nixon's original database. The position of the extracted eye centres, within a manually selected window, is shown in by a cross "+" for both the SHT and circular correlation. The result was classified as good if an algorithm produced a result which was less than 3 pixels from manually estimated location of the eye centre. If the result error was greater than 3 pixels but within the iris, the result was classified as marginal and a result outside the iris was classified as bad. Using a database of 23 faces, including 19 from Jia and Nixon's database, the SHT performed better than the correlation method. The SHT achieved 37 good, 6 marginal and 3 bad results, whereas the correlation method achieved 32 good, 5 marginal and 9 bad results.

2.6.2 SHT vs dual concentricity accumulators for eye extraction.

In theory, the new approach can be applied to the whole face and the two peaks in the X_o, histogram, equation 2.9, provide the x co-ordinates of the eyes centres. In practice it is necessary to pre-filter the edge map to remove edge points that lie on vertical or horizontal straight lines which can introduce valid but, undesirable solutions to equations 2.6 and 2.7 resulting in a peak in the accumulator which corresponds to a straight line, instead of a circle. The sides of the face can contribute to this undesirable effect, Consequently, straight lines consisting of more than 40 pixels and ± 10° of vertical or horizontal were removed. In order to achieve this one might consider rejecting pixels ± 10° of vertical or horizontal straight from the output of the Sobel edge detector, however this may also reject the pixels in the eye region. It is necessary to examine the whole

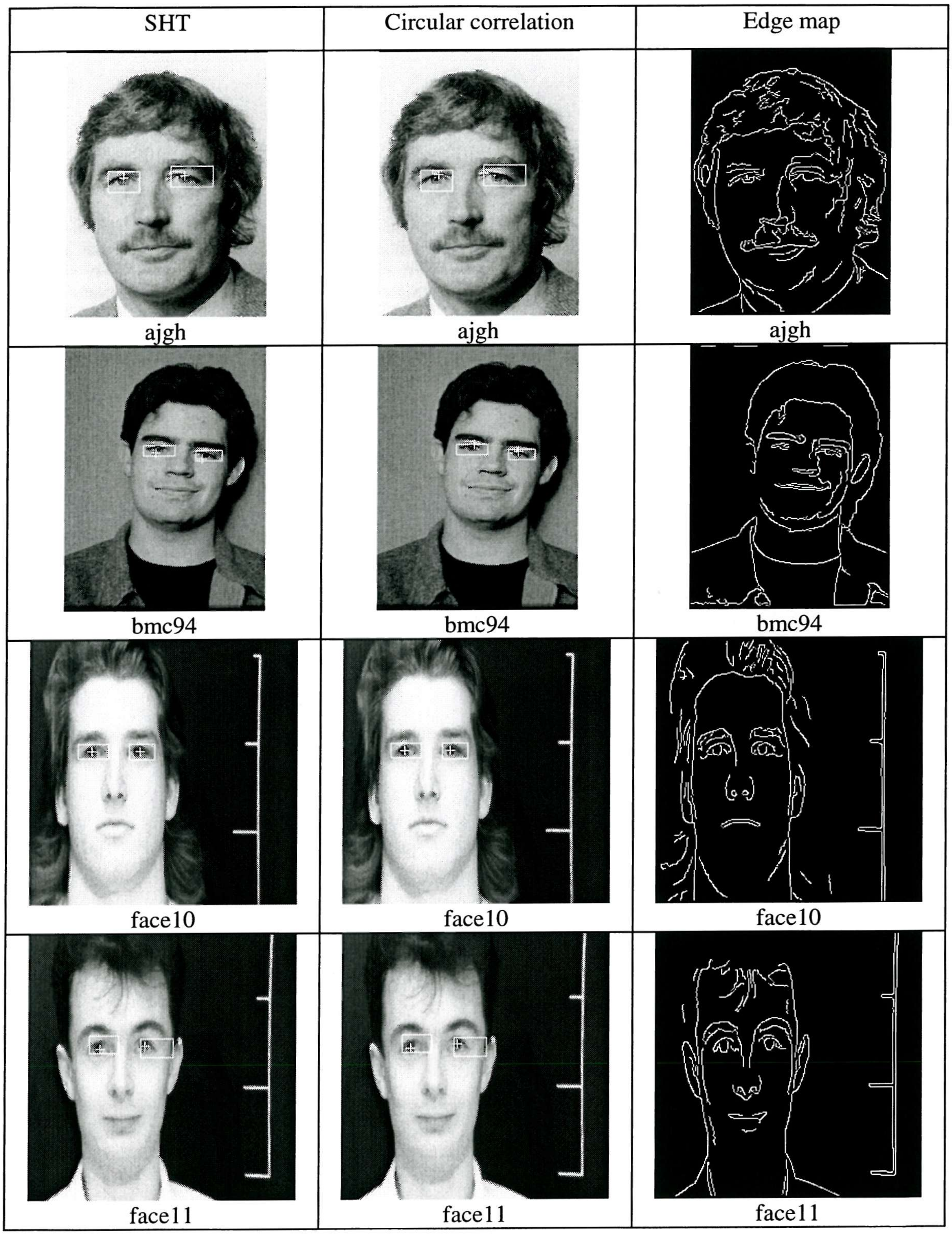
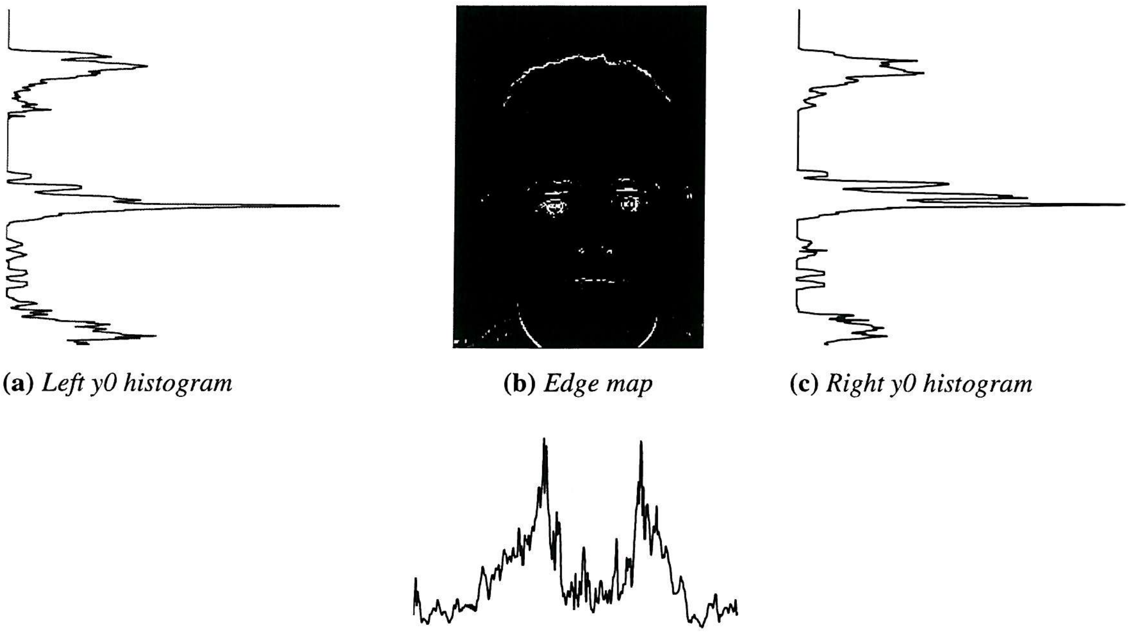


Figure 2.11 *Sample faces used for eye extraction with manually defined eye windows.*

picture after edge detection to decide whether a pixel is part of the side of the face. To this end, a crude heuristic was employed, namely using a Hough transform for lines to find and remove the sides of the face. The face images measured approximately 300×400 pixels. The pre-processing stage consisted of intensity normalisation, edge detection via a 3×3 Sobel edge detector implementing *dx* and *dy* from equation 2.5 followed by uniform thresholding above 160 to remove the sides of the face as described above, see Figure 2.14(b). This simple pre-processing was preferred to Canny edge detection which may require more parameter tuning to achieve the desired edge map. In addition, our new method's strength lies in its ability to detect the highest density of concentric edge points, but the Canny edge detector, with its non-maximum suppression may filter out some of these necessary edge points. Figure 2.14 shows a processed face image with a cross to mark the detected eyes. First *x* co-ordinates of the eyes corresponding to the two prominent peaks in the X_o histogram are found using the whole face (see Figure 2.12(a)). The face can then be split into two halves, the centre line being midway between the eyes found using the X_o histogram. The *y* co-ordinates of the eyes are found using separate accumulators for the left and right side of the face, thus compensating for any non vertical orientation of the face (see Figure 2.12(b) and (c)). The SHT requires a 3-dimensional accumulator in *r, a, b.* In order to minimise memory requirements of the SHT, both methods were re-applied in a window measuring 110×110 pixels centred on the extracted eye centres.

Figure 2.15 clearly illustrates the advantage of the new method. The extracted eye centre using the new method is marked with a "+" while a "×" is used to mark the eye centre extracted via the SHT. Using the concentricity property of the eye a cluster of pixels, centred around the pupil is extracted. In contrast, the SHT located a large arch of an eyelid which is not centred on the pupil. Figure 2.16 shows the statistics of the difference between the estimates of the eye co-ordinates provided by the standard HT and the new formulation, both from the values of manually-obtained estimates of the eye co-ordinates. Figure 2.16 gives the statistics for a large radius variation (as to be expected in application) and for a smaller radius variation. (A small radius variation can be justified in applications where face contour extraction [27] precedes eye location and primes a small expected radius variation.) The difference between the eye co-ordinates is less between the new formulation and the manual estimates, compared with the standard HT. For large variations, the mean difference for the new formulation on the left eye is 1.8 pixels, whereas for the same eye the mean difference for the SHT is 6.8 pixels. For the right eye, the mean difference is 2.5 and 6.4 pixels for the new and the SHT methods respectively. Reducing the potential radius variation improves the mean difference for the new concentricity formulation which is still considerably lower than for the SHT, being nearly half its value. Given that the average iris radius is approximately 8 pixels we see that the new formulation will locate eye centres within the iris.



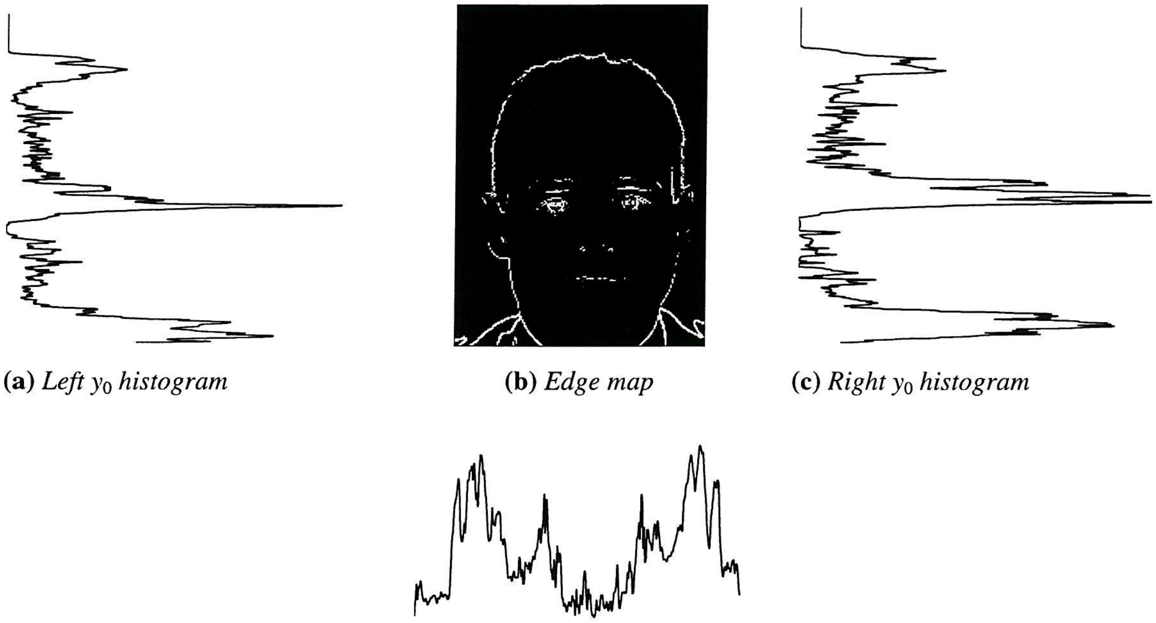
(a) *Left y0 histogram*

(b) *Edge map*

(c) *Right y0 histogram*

(d) *x_0 histogram. Highest peaks locate eyes*

Figure 2.12 *x_0 and y_0 histograms after removing vertical and horizontal lines*



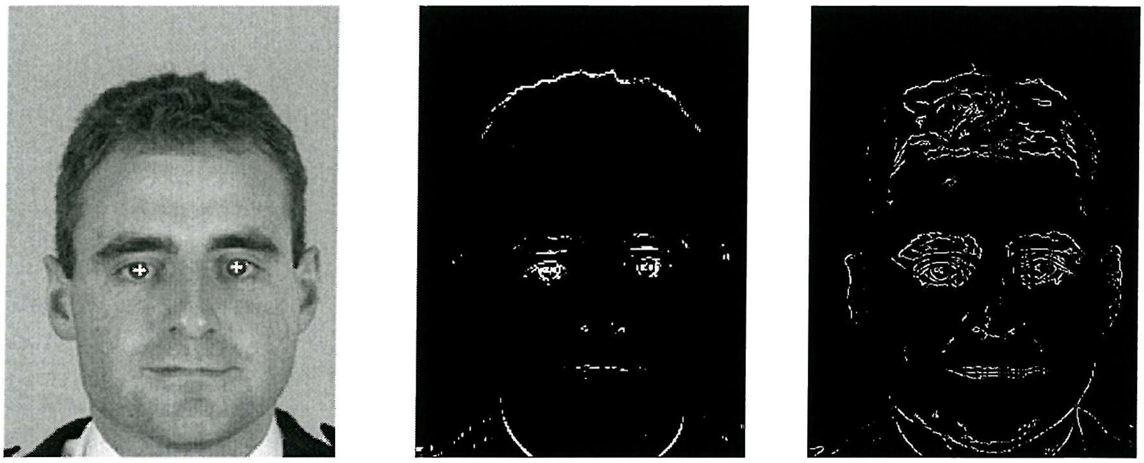
(a) *Left y_0 histogram*

(b) *Edge map*

(c) *Right y_0 histogram*

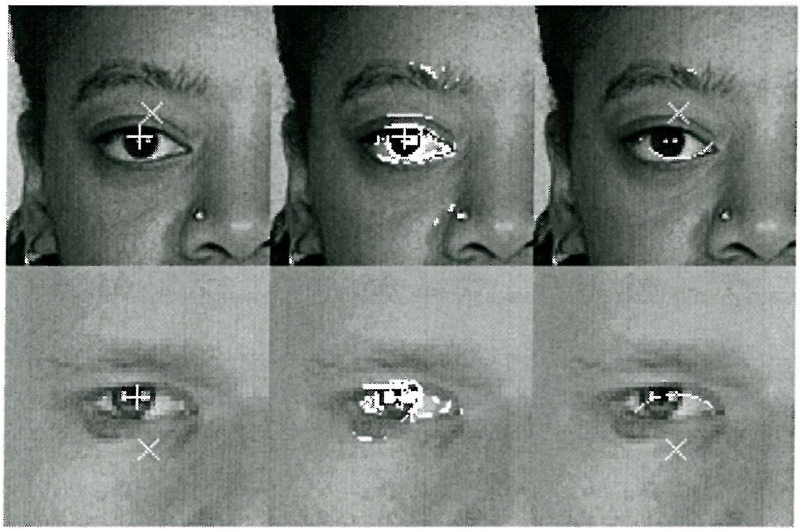
(d) *x_0 histogram. Highest peaks locate sides of face*

Figure 2.13 *x_0 and y_0 histograms without removing vertical and horizontal lines.*



(**a**) *Extracted eye centres* (**b**) *Filtered Sobel edge map* (**c**) *Filtered Canny edge map*

Figure 2.14 *Extracted eye centres*



(a) SHT = x *concentricity* = +

(b) *concentricity edge data*

(c) SHT *edge data*

Figure 2.15 *Eye pixels for new method vs SHT*

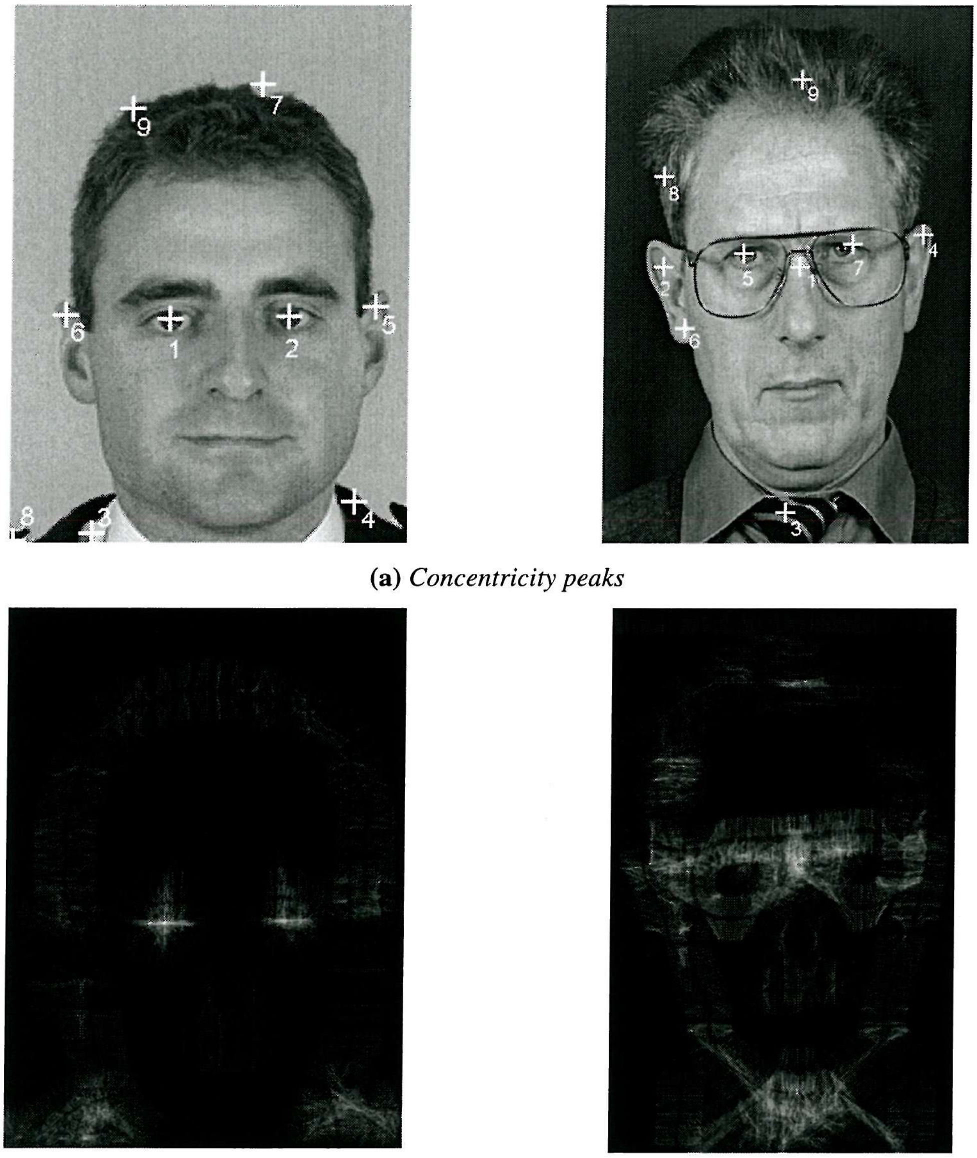
method	SHT (left)			new (left)			SHT (right)			new (right)		
actual - measured	x	y	d	x	y	d	x	y	d	x	y	d
Mean error radius=4-25	3.6	10	11	2.4	1.3	3.1	3.8	9	10	3.3	1.7	4.1
Sdev error radius=4-25	2.5	5.3	5.4	3.6	2	3.9	3.1	4.9	5.2	2.6	2.5	3
Mean error radius=4-10	2.5	2.7	4.3	1.4	0.7	1.9	1.7	1.9	3.1	1.8	1.7	2.8
Sdev error radius=4-10	2.8	3.8	4.1	1.6	1	1.6	3.5	3.2	4.4	2.4	3.2	3.7
(Note: average iris radius = 8 pixels, all measurements in pixels) x=measured error in x_0 y=measured error in y_0 d= Euclidean error												

Figure 2.16 *Comparing SHT and new method for finding eye centres*

Overall, for the 54 eye database, the use of concentricity in the new technique resulted in a mean difference of 3 pixels, whereas the mean difference for the SHT was 10 pixels, a three-fold improvement in accuracy.

2.6.3 Eye extraction using a single concentricity accumulator.

In the previous section we showed that the dual concentricity accumulator performed better than the SHT for circles when applied to locating eyes in a face image. The problem of vertical and horizontal lines dominating the dual concentricity accumulator was circumvented by removing the straight lines. This solution is unsatisfactory for two reasons. Firstly, pixels in the eye region may be removed if they are on the path which includes the straight lines being removed. Secondly, it is difficult to determine the length of the line of pixels to be removed. The single concentricity accumulator (using the same equations as for the dual accumulators) elegantly solves the problems posed by vertical and horizontal lines in the image. Figure 2.17(b) shows the resultant concentricity maps for two face images. The locations of peaks in concentricity are marked by cross "+" in Figure 2.17(a), the peak labelled 1 corresponding to the highest peak count down to 9 corresponding to the lowest 9th highest peak count. For face 1, the two highest peaks in concentricity correspond to the eyes. However, for face 2, although the peaks in concentricity are close to the eye centres, they are only the 5th and 7th highest peaks in concentricity. The subject's glasses, ears and tie provided higher concentricity peaks.



(a) *Concentricity peaks*

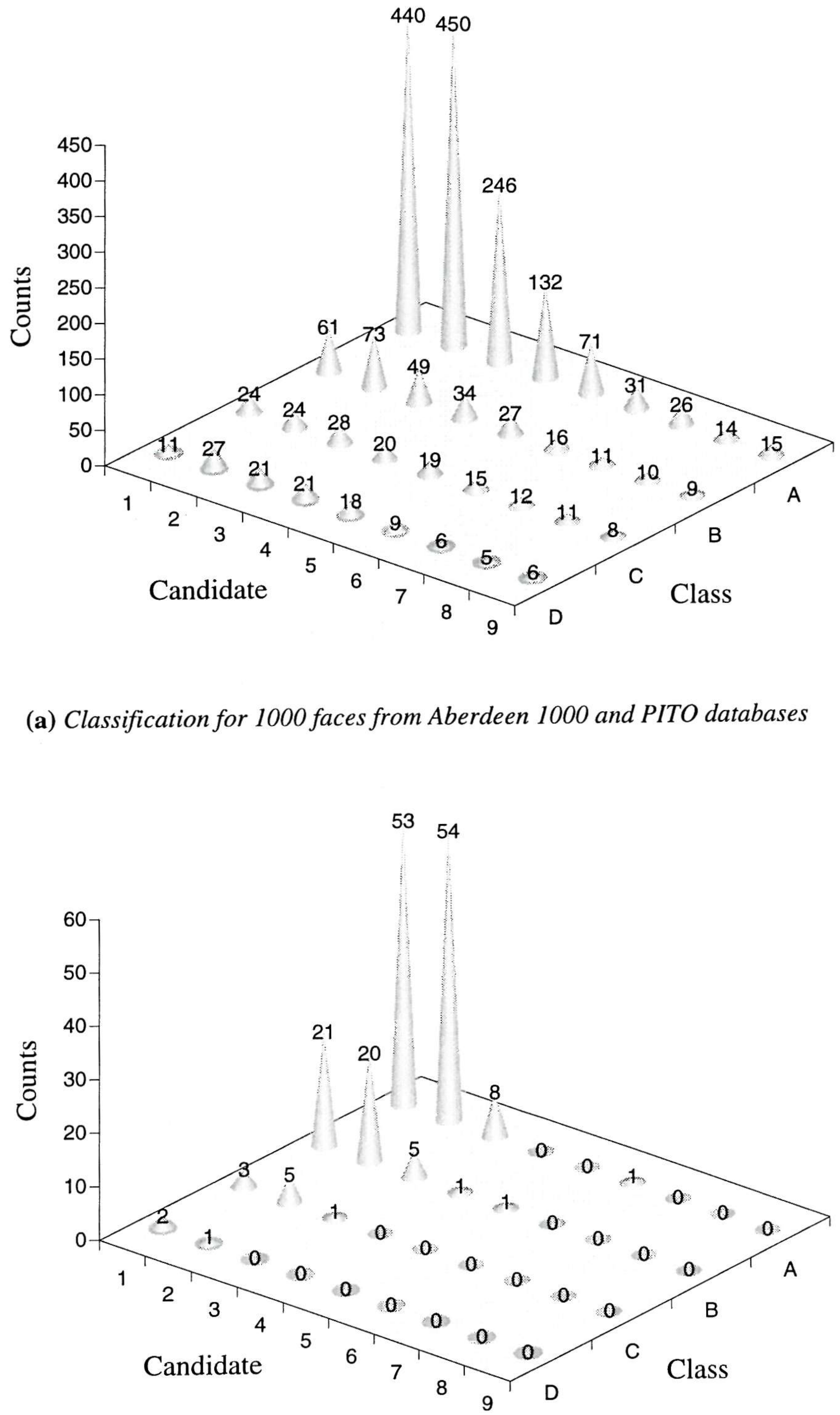
(b) *Concentricity map*

face 1 *face 2*

Figure 2.17 *Face processing for eye candidates using single concentricity accumulator*

To determine whether the results for face 1 or face 2 were more representative, the improved technique was tested on two databases. The first database contained 1000 faces images, of which approximately 800 came from the Aberdeen 1000 database while 200 PITO faces were included to round the numbers up to 1000. The second database comprised of 88 face images form the XMVTS database. The PITO images were cropped to remove extraneous artefacts such as rulers and text labels which could yield high values of concentricity and confuse the search for the eyes. Similarly, the XMVTS face images were selected to exclude images in which clothing was excessively striped or checked.

Figure 2.18 summarises the performance for these two databases. In the ideal case, the extracted eye centres would correspond to the two highest peaks in a concentricity accumulator. We consider the nine highest peaks and their Euclidean distance from the manually measured eye centres, classified into four classes as follows. Class A represents an actual distance of between 0 to 4 pixels, class B between 5 to 8 pixels, class C between 9 to 15 pixels and class D greater than 15 pixels. Each peak in concentricity is a candidate for an eye centre, so the ideal distribution for eye location would be all counts in the cells given by class A and candidates one and two. Class A represents a very good automatically located eye position. Class B is also good because this approximates to a result which is still within the iris. If Classes A and B are considered, our new concentricity algorithm achieved 84% for a selection of the XM2VTS database faces, which is better than a 74% success rate achieved by Sobottka and Pitas using similar images on a smaller database of 38 images [73]. For the one thousand face database only approximately 50% of the extracted eye centres were inside the approximate radius of the iris. However, it is worth noting that there are a large number of extracted eye centres which are in class A but not amongst the two highest peaks in concentricity, for example class A candidate 3 has 246 counts. These high concentricity counts may be attributed to hair or clothing. The images in the one thousand face database had a high edge gradient around the head, usually caused by dark hair against a light background.



(a) *Classification for 1000 faces from Aberdeen 1000 and PITO databases*

(b) *Classification for 88 faces from the XM2VTS database*

Figure 2.18 *Performance classification using concentricity analysis.*

2.7 Deformable templates for eye extraction

2.7.1 Standard deformable eye templates

In Yuille *et al's* deformable eye template model [88], the eye consists of two parabolas positioned at $\vec{x}_e$ for the eyelid and a circle of radius r centred at $\vec{x}_c$ for the iris. The points p_1 and p_2 represent the centres of the whites of the eyes. The upper eyelid has height a, the lower eyelid has depth b and both have width c all measured with respect to the parabola co-ordinate axis. The parabola co-ordinates axis is rotated by an angle of θ, with respect to the iris co-ordinate axis. In order to determine the parameters of the iris, sclera and eyelids, appropriate energy cost functions are defined using the information about the valleys, edges, peaks and intensity of the face image. The eye template interacts with the face image by adjusting its parameters to optimise a composite energy functional. If a good match between the template and the image has been obtained; the perimeter of the iris will correlate with the edge data for the iris, the area within the iris will correspond to a valley in the image intensity and the sclera area corresponds to peaks in the image intensity. In general, finding the eyes reduces to the construction of a suitable model and numerical optimisation of the energy cost function.

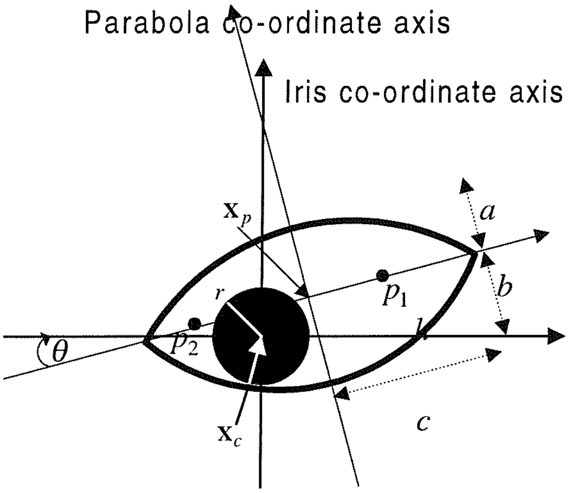


Figure 2.19 *Yuille et al Deformable Eye Template*

To provide explicit representation for the model boundaries, Yuille *et al* define two unit vectors

$$\vec{e}_1 = (\cos\theta, \sin\theta) \tag{2.15}$$

$$\vec{e}_2 = (-\sin\theta, \cos\theta) \tag{2.16}$$

which are used to represent variations in the orientation of the model. A point $\vec{x}$ can be represented by the co-ordinate (x_1, x_2) where

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 \tag{2.17}$$

The parabola for the upper eyelid can be represented by

$$x_2 = a - \frac{a}{b^2}x_1^2 \tag{2.18}$$

and the parabola for the lower eyelid can be represented by

$$x_2 = -c + \frac{c}{b^2}x_1^2 \tag{2.19}$$

where $x_1 \in [-b, b]$. The energy function of the template, $E_c(\vec{x}_e, \vec{x}_c, p_1, p_2, a, b, c, r, \theta)$, which is a function of eleven variables, can be expressed in terms of the valley, edge, peak and internal energies,

$$E_c = E_v + E_e + E_i + E_p + E_{in} \tag{2.20}$$

The valley, edge and peak fields are defined in terms of the image intensity at point **x**, $I(\mathbf{x})$ as

$$\Phi_v(\mathbf{x}) = \text{-I}(\mathbf{x}) \tag{2.21}$$

$$\Phi_e(\mathbf{x}) = \nabla\text{I}(\mathbf{x}).\nabla\text{I}(\mathbf{x}) \tag{2.22}$$

$$\Phi_p(\mathbf{x}) = \text{I}(\mathbf{x}) \tag{2.23}$$

The valley energy is given by the surface integral of the valley field defined by area of the iris, normalised by the area of the iris,

$$E_v = -\frac{c_1}{Iris\ area} \iint_{Iris\ area} \Phi_v(\vec{x})dA \tag{2.24}$$

The edge energies for the eyelids and iris are given by

$$E_e = -\frac{c_2}{Iris\ length} \int_{Iris\ bounday} \Phi_e(\vec{x})ds - \frac{c_3}{Eyelid\ length} \int_{Eyelid\ boundary} \Phi_e(\vec{x})ds \tag{2.25}$$

The energy which due to intensity in the iris and sclera is given by

$$E_i = \frac{c_4}{Iris\ area} \iint_{Iris\ area} \Phi_i(\vec{x})dA - \frac{c_5}{Sclera\ area} \iint_{Sclera\ area} \Phi_i(\vec{x})dA \tag{2.26}$$

The energy due to the sclera is given by

$$E_i = -\frac{c_5}{Area} \iint_{Sclera} \Phi_i(\vec{x})dA \tag{2.27}$$

The energy due to the peak points p_1 and p_2 is given by

$$E_p = c_6\left\{\Phi_i(\vec{x}_e + p_1\vec{e}_1) + \Phi_i(\vec{x}_e + p_2\vec{e}_1)\right\} \tag{2.28}$$

and the internal energy is given by

$$E_{in} = \frac{k_1}{2}(\vec{x}_e - \vec{x}_c)^2 + \frac{k_2}{2}(p_1 - \frac{1}{2}\{r+b\})^2 + \frac{k_2}{2}(p_2 + \frac{1}{2}\{r+b\})^2 + k_3(b-2r)^2 \tag{2.29}$$

Yuille *et al* used steepest descent to minimise equation (2.20). They defined six sequential time epochs in which the $\{c_i\}$ and $\{k_i\}$ coefficients were allowed to vary during the matching process in order to exploit the salient features of the eye as follows:-

1. The coefficients for the valley energy are designed to dominate the template energy. If the initial position of the template is displaced from the desired position, only $\vec{x}_c$ and $\vec{x}_e$ are allowed to vary by steepest descent which should pull the whole template towards the iris.
2. The intensity coefficients are increased to match the size of the iris.
3. The edge coefficients are increased to fine tune the iris boundary.
4. The peak coefficients are increased to rotate the template to the correct orientation.
5. The intensity coefficients for the whites of the eyes are increased to adjust the outer boundary of the template
6. Fine tuning of the edge boundaries are achieved by increasing the edge coefficients

Yuille *et al* [88] noted that their template failed to converge to the required solution if the template was initialised above the eyebrows. This failure could be attributed to strategy employed in the first epoch. If the template was initialised above an eyebrow, the valley energy would attract it to an eyebrow, which was the first local minimum encountered by the template. Xie *et al* [84] proposed simultaneous rather than sequential optimisation of the eye parameters using the Levenberg-Marquardt (L-M) method. They defined an energy function as the weighted sum of ten non-linear functions,

$$F(\vec{X}) = \sum_{i=1}^{10} W_i E_i^{\,2}(\vec{X}) \tag{2.30}$$

where, $\vec{X}$ is the vector of template parameters, W_i is a vector of weights and the E_i^2 correspond energy terms similar to those used by Yuille *et al* [88]. This energy function was optimised with all template parameters allowed to vary simultaneously using an iterative process such that

$$\vec{X}_{k+1} = \vec{X}_k + d_k \tag{2.31}$$

$$(J_k^{\,T} W J_k + \lambda I) d_k = -J_k^T W F_k \tag{2.32}$$

where J_k is the Jacobian matrix for the non-linear functions $E_i(\vec{X})$ and λ is an adjustable constant. If $\lambda \rightarrow \infty$ in equation 2.32 then the optimisation process approximates the steepest descent used by Yuille *et al* [88]. Xie *et al* used a small value for λ, in which case the optimisation process is similar to Newton's method. However, both steepest descent and the L-M method require gradient information which may be difficult to obtain accurately for all parameters. Figure 2.20 shows an optimisation sequence where all the parameters were allowed to vary simultaneously. Initially, the template was pulled towards the iris as required. However due to the difficulty in establishing the correct weights for the internal energies over a large range of eyes, the template began to shrink towards a point.

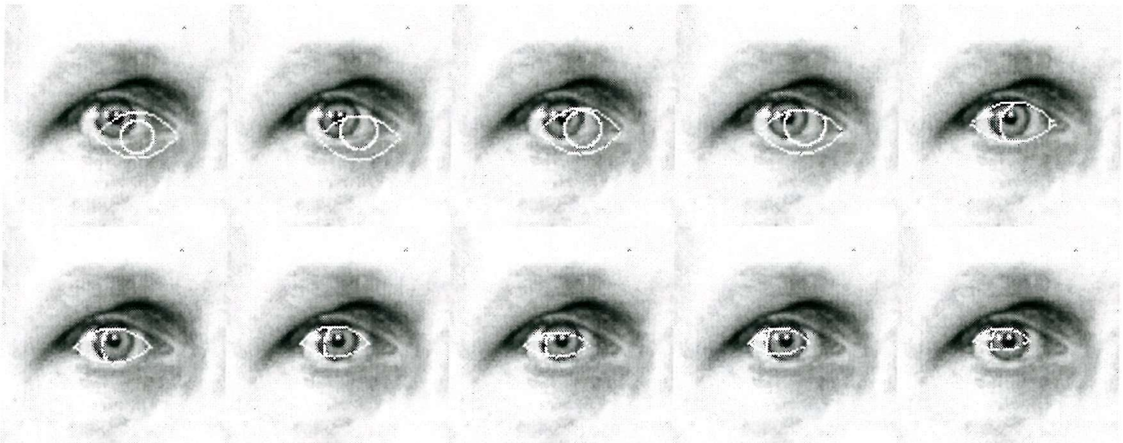


Figure 2.20. *The template shrinks to a point if its internal energy weights are wrong.*

Another potential problem lies in the implementation of equation 2.25. Even if we assume that there were edge pixels in the eye corner, then using a count per unit length basis, the parabola which has a smaller width will yield the best edge energy. This results in the eyelid parabolas collapsing onto the iris as shown in Figure 2.21. The internal template energies used throughout the literature attempt to compensate for this drawback. Xie *et al* [84] also included a number of internal energies to their template which were designed to control the expected final shape of the template. They attempted to force the intersection of the parabolas into the eye corners using internal energy force,

$$E_{i1} = k_{i1}|b\text{-}2r|. \tag{2.33}$$

A further internal force

$$E_{i2} = k_{i2}|b\text{-}4c| \tag{2.34}$$

is applied to keep the eye lids open. However such terms are somewhat artificial and require some extra weighting constants (k_{i2} and k_{i1}) in the template energy function.

In summary, using gradient methods, the energy function needs to be differentiable and preferably smooth for the template parameters to proceed in the correct direction. These goals may be partially satisfied by Gaussian filtering the peak, edge and valley fields. However the pre-processing can play a significant role on the steady state template parameters. If a parameter is far from the desired solution, pre-processing should effect rapid convergence. Over filtering may prevent the parameters moving from their initial values. Alternatively, excess filtering may cause the template to pass over and subsequently oscillate about the desired solution. Ill-chosen weighting factors could cause the contours to attempt to shrink to a point. In the next section we detail our next contribution to the literature in the form of an improved eye model which does not require the use of internal energies. The improved model is then optimised using an improved search strategy which does not require the use of gradient information. We used a genetic algorithm (GA) although the simplex method [125] used by Chow and Xi [122] maybe a reasonable alternative.

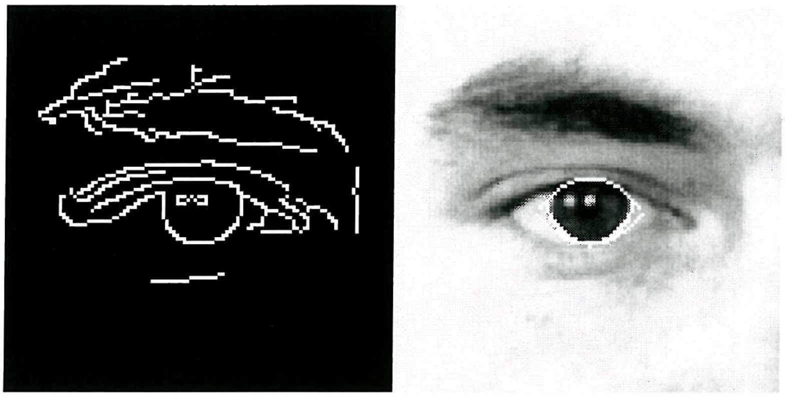


Figure 2.21. *Poor edge data results in undesirably small eyelid width being selected.*

2.7.2 Improved deformable eye template techniques

Instead of searching for the small areas defined by p_1 and p_2, we use the whole of the region inside the parabolas (minus the iris) to represent the sclera and remove these two parameters from our eye model. We define an energy function E, composed of edge energy E_e, peak energy E_p, valley energy E_v and internal energy E_i.

$$E = k_p E_p + k_v E_v + k_e E_e + k_i E_i \tag{2.35}$$

where k_p, k_v, k_e and k_i are tuning coefficients. The image energies (E_p, E_v and E_e) represent line or surface integrals of the image intensity, normalised by the length or area of the bounding contour. The problem of the parabolas intersections collapsing onto the iris can be alleviated by not allowing pixels in the eyelid edge boundary to be shared with those of the iris edge boundary. Applying this rule to Figure 2.22, yields a low edge count per unit boundary length because many pixel on the parabola arch are already part of the iris edge boundary. The remaining parabola boundary lies in the sclera where there are few edge pixels. At this stage of processing, the Canny edge detector [12] is preferred to the Sobel as it provides arcs which are one pixel wide, minimising the likelihood of pixel being shared by iris and parabola boundaries. Since pixels in the iris cannot be shared with those of the eyelid, a high count per unit length can best be achieved by using the pixels in the eye corners, as required.

Accordingly, the required surface and edge boundaries are represented in terms of sets of points. We denote the set of points in the areas enclosed by the circle and parabolas as A_{circle} and A_{parabola} respectively; similarly the boundaries as B_{circle} and B_{parabola}. The iris and sclera, areas and boundaries are now given by

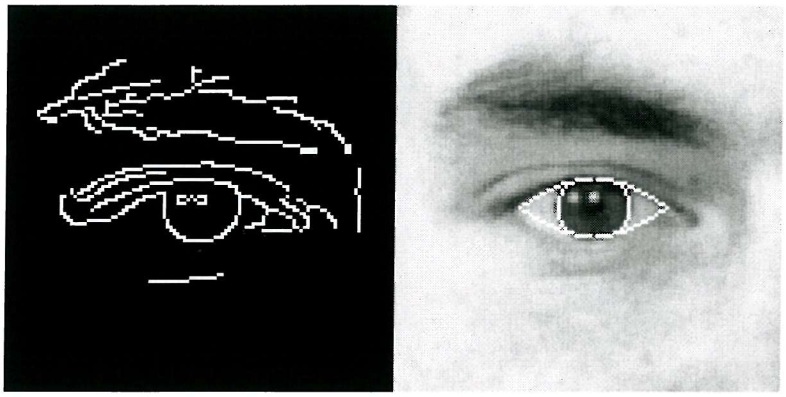


Figure 2.22. *Correct eyelid extraction even in presence of poor edge data.*

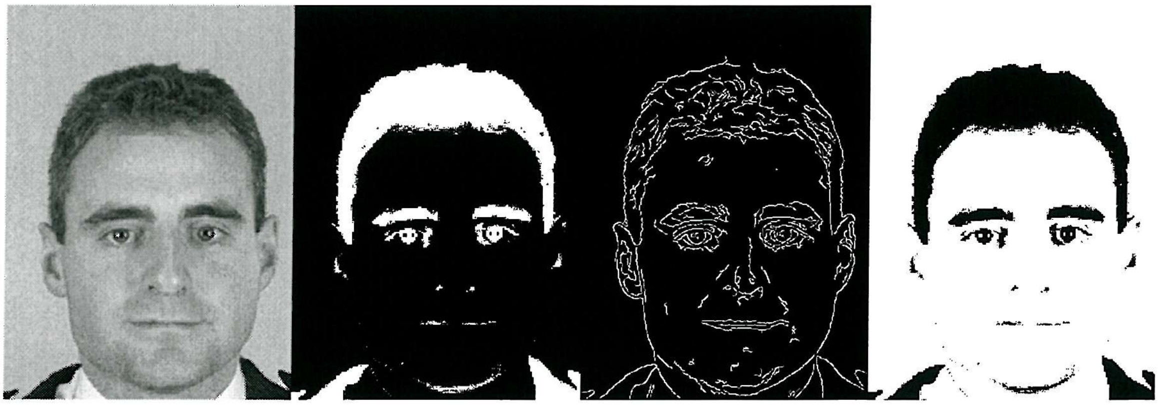
$$A_{iris} = A_{parabola} \cap A_{circle} \tag{2.36}$$

$$B_{iris} = A_{parabola} \cap B_{circle} \tag{2.37}$$

$$A_{sclera} = A_{parabola} \notin A_{iris} \tag{2.38}$$

$$B_{sclera} = B_{parabola} \notin B_{iris} \tag{2.39}$$

The image, intensity normalised and represented in 256 grey levels, is pre-processed to obtain binarized valley edge, and peak fields, Φ_v ,Φ_e and Φ_p as shown in Figure 2.23. The valley field was inverted by evaluating the maximum grey-level minus the image intensity, then uniformly thresholded. The feature pixels, in white are set to 1 and the non-feature pixel, in black are set to 0. Uniform thresholding is also used for the peak field, while a Canny edge detector [12] is used for the edge field.



(a) *original face* **(b)** *valley field* **(c)** *edge field* **(d)** *peak field*

Figure 2.23 *Field processing for valley, edge and peak fields.*

In many deformable eye template applications e.g.[84][88] the peak field processing is designed to segment the sclera using morphological operators [82]. The processing is applied to a small window in the image which is assumed to already contain an eye. Referring to Figure 2.17 we note that there are many eye candidates near the background of the image. A morphological operator may group large areas of a light background as a candidates for the sclera. Worse still, the background intensity may be lighter than the true sclera intensity, in which case it is safer to use simple thresholding but note that a valid sclera candidate must be bounded by edge pixels belonging to the eye template. This can be achieved by including the edge energy, E_e, in the definition of the peak energy E_p, as indicated in equation (2.41). The values for tuning constants k_p, k_v, k_e were set to one for simplicity, although they may be optimised by discriminant function analysis and k_i is not required thus; the extended energy is

$$E_e = \frac{1}{|B_{sclera}|}\sum_{B_{sclera}}\phi_e + \frac{1}{B_{iris}}\sum_{B_{iris}}\phi_e \qquad (2.40)$$

$$E_p = E_e\frac{1}{|A_{sclera}|}\sum_{A_{sclera}}\phi_p \qquad (2.41)$$

$$E_v = \frac{1}{N(A_{iris})}\sum_{A_{iris}}\phi_v \qquad (2.42)$$

2.7.3 Genetic Algorithm (GA) optimisation methods.

GAs perform a stochastic search for an optimum solution and can be applied to ill-behaved functions of high dimensional spaces [123]. They work in a manner analogous to the survival of the fittest in natural evolution. For a given species, the strongest (or *fittest*) tend to live longer than the weaker individuals and therefore bear more offspring than their weaker counterparts. If the attributes which provide the advantage over the weaker individuals are inherited by their offspring, then eventually the weak individuals will die, leaving a population of strong individuals.

The terminology used in genetic algorithm (GA) implementations for optimisation is itself inherited from the field of biological genetics. Each parameter to be optimised is coded as a gene. A typical coding scheme would map the range and resolution of each parameter to a binary Gray scale [26]. A complete set of parameters, or set of genes, comprise a chromosome for an individual. The *objective function* is simply the function to be optimised enabling the fitness of an individual to be evaluated in terms of the objective function. The optimisation process starts with a random population of N individuals or chromosomes. The parameters are then decoded and applied to the objective function to evaluate the fitness of the *i*-th chromosome, f_i. Parent individuals are selected for reproduction with a probability of selection which is proportional to the their fitness. The *cross-over* operator is used to implement the mating of two parent chromosomes to produce two offspring. A randomly chosen point is chosen to spilt both parent chromosomes. Two offspring chromosome are produced by concatenating half the genes from one side of the spilt point of one parent, to half genes on the other side of the split point of the second parent. The *mutation* operator is then applied to the children produced from the cross-over stage in order to introduce new genetic material into the population pool. Randomly changing a gene, may yield a particularly fit chromosome which may flourish in successive generation. Alternatively, it may be particularly weak, in which case in may die quickly in successive generations. A very low probability of mutation was used to prevent the pool of chromosomes loosing the attributes inherited from their parents. It should be noted that if a small mutation or perturbation of a chromosome is combined with a high probability of mutation, then the process is approximating simulated annealing instead of a GA. Successive generations of parent selection, cross-over and mutation are applied until a predefined termination condition is met. For example the evolution process may be deemed complete if the mean population fitness is within 5% of the max population fitness or simply after a fixed number of generations.

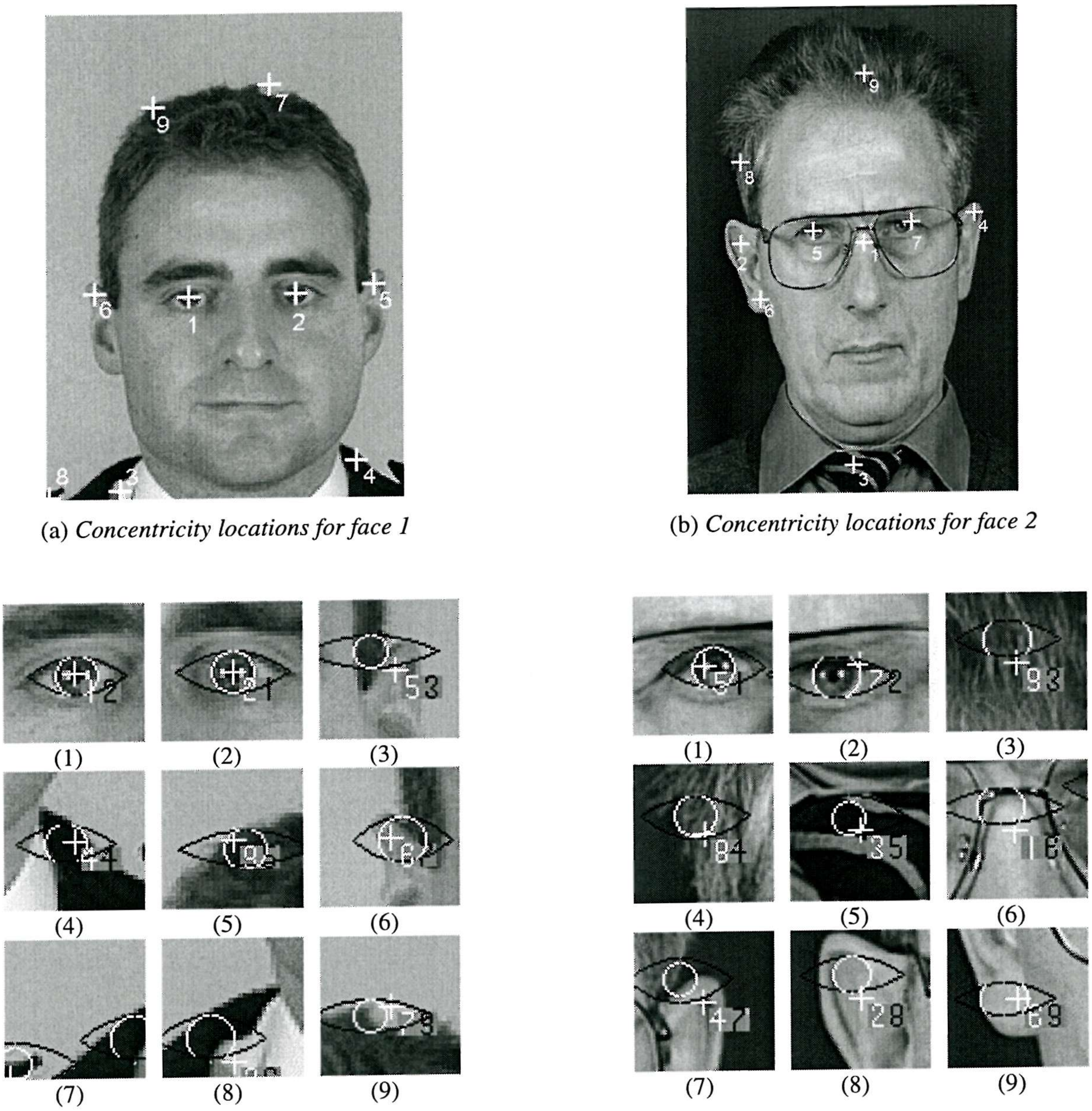
2.7.4 Eye Template Energy Optimisation

During the evolution process a population of 100 genes with a mutation probability of 0.001 was used. The evolution was deemed to be complete after 50 generations or if the maximum fitness was within 5% of the average population fitness. An initial population of 1000 genes (ten times the evolution population) were randomly selected for mating. By selecting a large initial population the process is given a wide selection of genetic material from which to evolve which also implies a wide sampling of the parameter space. By using a large initial population we increase the probability of selecting a parameter set which locates the template at the correct position. In the event that we also include a parameter set which corresponds to the location of an eyebrow, we expect that the parameter set corresponding to the eyebrow will eventually be evolved from the pool of eye candidates. In the unlikely event that the initial pool of 1000 chromosomes does not contain a parameter set which is sufficiently close to the desired location of the iris, then there is still a chance that the mutation process will generate the chromosome which corresponds to the template being positioned near the iris. Thus for our application, the GA optimisation appears to offer significant advantage over gradient methods in eluding local minima such as the eyebrows, from which Yuille *et al's* method [88] could not recover.

In order to obtain a good template solution in a reasonable time, it is necessary to be able to evaluate the energy function as quickly as possible. The energy function, was pre-evaluated for all the parameters of interest and stored in a lookup table. To reduce the size of the lookup table we set the parabola centres equal to the iris centre ($x_c = x_p$) and $\theta = 0$. Subsequent evaluations of the energy function can then be obtained by indexing the lookup table by x_c, r, a, b, c. Using the lookup table enable us to process all nine candidate sites very quickly (in approximately 20-30 seconds on a Pentium 233Mhz).

2.8 Deformable eye template results

Figure 2.24 illustrates two examples of the eye location method. Figure 2.24(a) shows the result of the first stage of processing, in which the nine best sites are ordered (1 through 9) in term of peak concentricity magnitude. The white cross shows the location of a local concentricity peaks. The results of the final processing stage for face 1 are shown in (b) ordered (1 through 9) in term of best eye match. For face 1 the best two concentricity locations are excellent and can hardly be improved by the deformable template. However, it is important that it does not degrade the result obtained by concentricity. In fact, the best deformable template matches at these location are the top two and the templates are very good fits to the eyes. The extracted iris edge boundary, B_{iris}, is shown in white, while the extracted sclera edge boundary, B_{sclera} is shown in black. Note also, that since sclera pixels cannot be shared with iris pixels, by equation (2.39), it is unlikely that the sclera will collapse onto the iris and there is no need for a term in the energy function to prevent this. Considering face 2 shows the benefits our two stage process even move clearly. The eyes are included in the list of candidate sites delivered by concentricity but are not the highest ranked (at 5th and 7th). After deformable template analysis, the eyes have moved to be the 1st and 2nd candidates, an excellent result. Referring to the right eye of face 2, note how the template has moved its initial location, marked by the cross, to find the centre of the iris. This example is particularly challenging as the subject is wearing spectacles. However, we found that providing the eyes were not obscured by reflected light, spectacles can contribute to the concentricity counts in the eye region.



(a) *Concentricity locations for face 1*

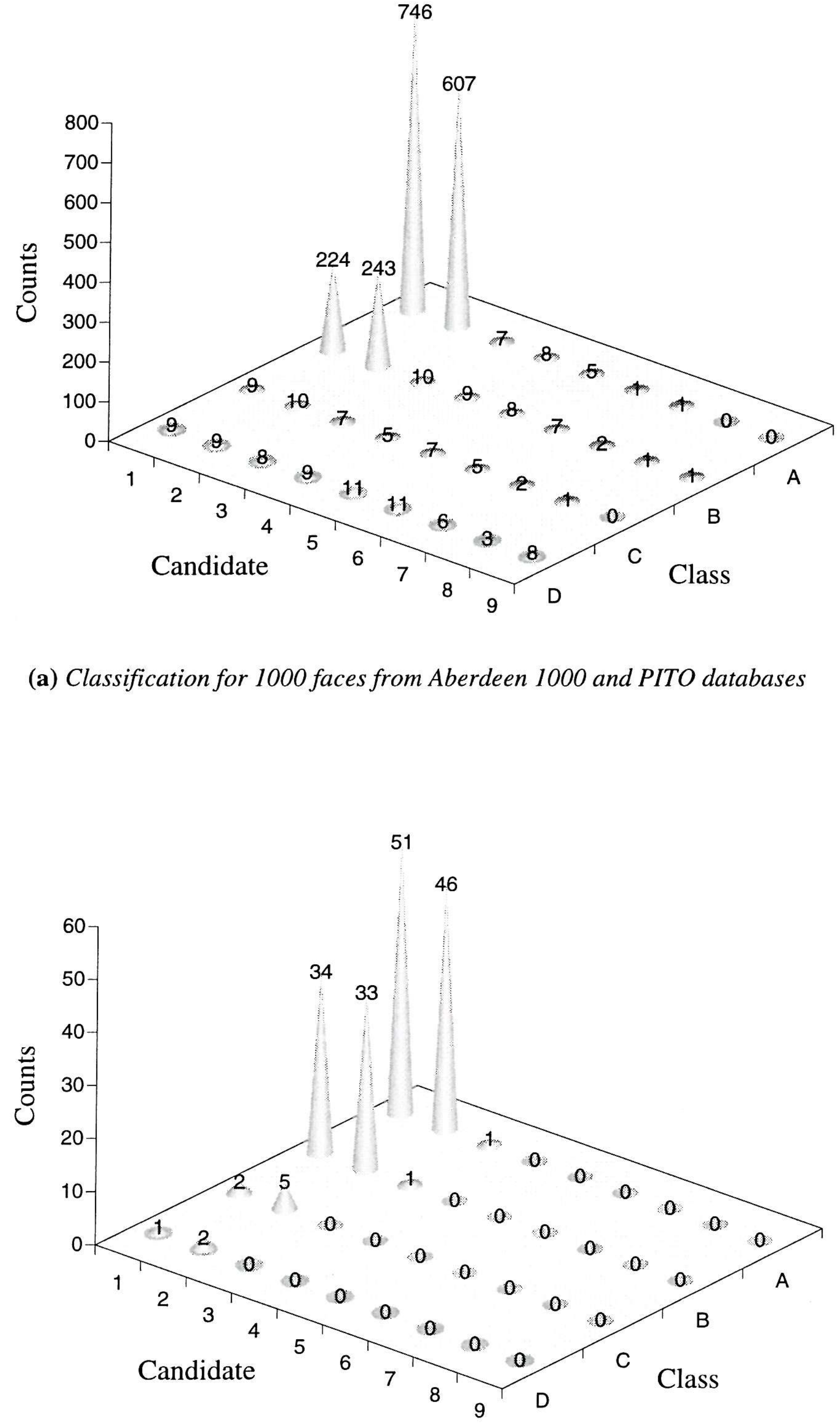
(b) *Concentricity locations for face 2*

(c) *Best eye locations for face 1*

(d) *Best eye locations for face 2*

Figure 2.24 *Improved eye location using enhanced deformable template*

Figure 2.25 shows the classification for the PITO and 1000 face database using concentricity combined with improved deformable template analysis. Referring back to Figure 2.18 which shows classification on the same databases using only concentricity, it can be seen that the deformable template has improved the eye location system. Using only concentricity analysis on the 1000 face database, just over 50% of the counts are in class A with rating 1 and 2. As such, the majority of the candidate eye centres are determined correctly by concentricity analysis. However, the majority of the remainder are in the right place (class A) but not ranked correctly as the most likely to be eye centres. After applying the deformable template, initialised at the location of peak concentricity, candidates. 91% of the counts are class A, B candidates 1 and 2. Similarly, the results of the improve deformable template combined with concentricity achieved 93% successful eye location using the XM2VTS database. The performance improvement gained by using the deformable template is not as pronounced on this database as the 1000 face database, since the output of the concentricity analysis was arguably sufficiently high at 84%. The average radius of the irises was approximately 10 pixels, which means that most of the extracted eye centres were within the iris. The benefit of the deformable template analysis can been seen by noting the shift in the data distribution when using concentricity analysis alone, compared to concentricity and deformable template analysis.



(a) *Classification for 1000 faces from Aberdeen 1000 and PITO databases*

(b) *Classification for 88 faces from the XM2VTS database*

Figure 2.25 *Classification using concentricity and improved deformable template analysis*

2.9 Conclusions and further work

Concentricity affords a new basis to extend the HT for eye extraction in a manner which is well suited to the normal appearance of eyes in a face image. The standard HT (SHT) can provide good results when a good prior estimate of iris radius is known. However, the results obtained with our model of the eye region being the centre of concentricity, has surpassed the SHT in terms of accuracy, providing a factor of three improvement over the SHT. The approach to finding eye centres using the SHT was to search for a single circle which represents the iris. A maximum count approach could produce poorer results because the iris pixels may not yield a higher count than the pixels from an eyelid or an eyebrow. A maximum count per unit radius may also not extract the iris pixels but may instead find the eye corners since small circles in this area would have a higher count per unit radius. The problem of attempting to select a specific radius for the iris was desensitised by using all the radii of interest to effect a measure of concentricity. If the SHT is used as the basis of a concentricity operator, the centres of concentric circles combine destructively. However, using gradient information, the concentric circles in the eye region combine constructively to reinforce the location of the eyes. Thus the new approach can be applied to a wide range of radii over the *whole* face without the need for an intermediate eye window location stage or *a priori* knowledge of the face. Our concentricity operator has been successfully tested on images, which are similar to passport type photos which include the shoulders as opposed to images often used in research which have to be cropped to only include the head. For 88 faces on from the M2VTS database, 84% of the extracted eye centres were within the approximate region of the iris. Using the same definition of success on the 1000 face database, only 50% of the extracted eye centres were within the iris and amongst the two highest peaks in concentricity. The low success rate on this database was due to the light colour of the background which produced strong gradient information near the head which in turn attracted the algorithm to high concentricity found in the hair. We have improved Yuille *et al's* deformable template [88] which when initialised at the peaks of concentricity, improve the classification of extracted eye centres. The improved template model does not require internal energies or tuning constants which may be difficult to specify. Local minima in the template energy space can be avoided using a genetic algorithm which inherently obviates the requirement for accurate gradient information. Our improved eye template model and search strategy was verified on a database of over one thousand faces yielding in excess of 90% successful eye location.

3. Locating features using the Eyes.

The next set of features that we attempt to extract are those related to the head boundary. Robust extraction of the skin should help in locating points 6, 8 and 9 from Figure 1.1. We expect that any region that is enclosed by facial skin but exhibiting different statistical properties is likely to be of interest. This is the basis on which we locate the eyebrows. If the eyebrows are occluded by hair, they will not be correctly detected. The eyebrows will also be occluded and be less likely to be detected if the face is not positioned normally to the camera. A database consisting of 44 images of 28 different people were pre-selected at random to prevent eyebrow occlusion.

We investigated the feasibility of the dynamic programming technique for chin extraction [7], [27] to find points on the chin 33, 3, 37, 4 and 36. We did not attempt to locate the bridge of the nose, points 23 and 23 or the width of the nose, points 26 and 27. However, as an alternative to finding the base of the nose, we used the local intensity minima to locate the nostrils. The points on the mouth 30, 31, 32, 34 and 35 were located using a closed-mouth model described by Yuillie *et al* [88], but using a GA instead of gradient methods.

Good initialisation for the search for the above features can be obtained if the head boundary can be extracted. Turk and Pentland [78] showed that an eigenface based system could be trained to recognise a face by pre-determining the expected distance to face space, c.f. equation (1.25). This method may require some operator interaction at the end of the learning phase before being able to locate face space. The effect of size and the number of images required for training would also require specification. Although a limited amount of human intervention is acceptable, it is not desirable.

Jia and Nixon [42] used a quadratic curve fit to the chin region. The search space for the chin was identified by first locating the sides of the neck which were assumed to approximate to a pair of roughly parallel vertical straight lines. In a large database, head boundary extraction using this heuristic method is likely to suffer from two drawbacks (a) initialisation of the quadratic curve fitting process by searching for the neck may fail if the neck is obscured by clothing (b) the quadratic fit may not be a good match to the chin.

Brunelli and Poggio [7] integrated the horizontal projection of the edge map to locate the sides of the head. The difficulty in applying this method for our sponsors' purposes would be in differentiating between points 7 and 8 the outer and inner boundaries of the hair.

The task of fitting a set of points to the head boundary, which includes the chin, can be more accurately achieved using an active contour or snake. Snakes have the ability to track a target feature, yielding a flexible implicit boundary description instead of an explicit parametric description. Although the human head is roughly oval in shape, biometric differences in head size and shape can be used to distinguish between face images. Thus, accurate head boundary determination can be used as part a face feature vector as well as enhancing the performance of another feature vector, the eye spacing measurement. Fortunately, we have been able to locate the eyes on a large database with reasonable accuracy and without using heuristic methods. Concentricity combined with our improved deformable eye template work gives the locations of the points 19, through 25 on Figure 1.1. In addition the eyes can be used as an initialiser for all the other face features.

3.1 Active contours

An active contour will attempt to vary the spatial distribution of its contour points from some initial distribution and location to align itself with the target feature. The extracted feature results from a minimum of a combination of internal energy, derived from local spatial constraints, and image energy from the image edge data. The original snake model was introduced by Kass [43]. A contour can be described parametrically by $\mathbf{v}(s)=(x(s), y(s))$ where $x(s)$, $y(s)$ are the x and y co-ordinates along the contour and $s \in [0, 1]$ is the normalised arc length. The snake model defines the energy of a contour $\mathbf{v}(s)$ to be

$$E_{\text{snake}}(\mathbf{v}(s)) = \int_{s=0}^{1} \lambda E_{\text{int}}(\mathbf{v}(s)) + (1-\lambda) E_{\text{image}}(\mathbf{v}(s)) ds \tag{3.1}$$

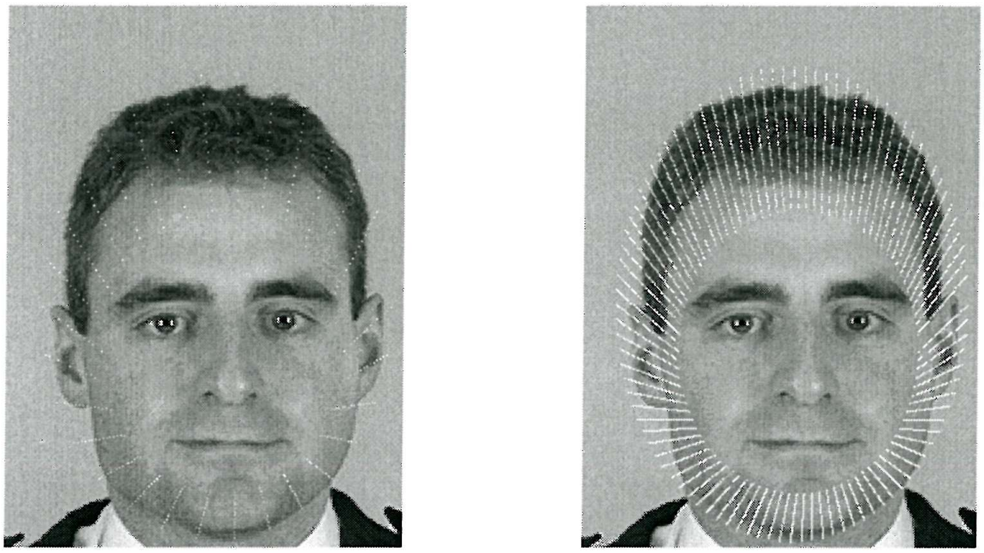
where E_{int} is the internal energy of the contour E_{image} is the image energy and $\lambda \in [0, 1]$ is the regularisation parameter used to bias the solution either to the internal energy or to the image energy. The energy integral is called a functional since its independent variable is a function. In the original formulation, minimisation of equation 3.1 was achieved using an evolutionary approach which used local energy variations to locate the target feature. A weakness of this local minimisation approach was the difficulty in determining suitable parameters and sensitivity to initialisation. The performance of the local minimum criterion can suffer if the target feature is obstructed by extraneous features or noise. Furthermore, the technique can exhibit a tendency for the contour to contract into a point in the absence of image energy. The problems caused by extraneous obstructions can be alleviated using global minimisation techniques. Global minimisation employ an exhaustive search for the minimum contour energy within a region containing the target feature. Gunn and Nixon [27], Brunelli and Poggio [7] and Amini [1] have successfully applied dynamic programming to find a globally minimised solution to equation 3.1

3.2 Application of Dynamic Programming

The search based technique uses two initial contours to define the head boundary; one outside the head boundary and the other inside the head boundary. In contrast to the evolutionary technique, the search based technique does not employ "snake-like" behaviour but considers all possible solutions in the search space to find the optimum contour. Dynamic programming uses the principle of optimality [74] which states: *Whatever the path to a node X, there exists an optimal path between X and the end point. In other words if the optimal path (start point to end point) goes through X then both its parts start point to X and X to end point are also optimal.*

3.2.1 Head Boundary Search Space

Figure 3.1 shows an image and the search space used for the dynamic programming search. The inner and outer contours are open contours consisting of N points or *stages* of the dynamic programming search. The corresponding point on each contour is joined by a line of M points or *nodes*. M must be sufficiently large to ensure that local intensity minima can be sampled while N governs the number of points on the extracted head boundary. The aim is to find the minimum cost path from the first stage through to the end stage. For this mode of dynamic programming there should be no circular path from the first stage to the last therefore, the initial contours are open rather than closed contours.



(a) *nodes* = 15, *stages* = 30

(b) *nodes* = 70, *stages* = 100

Figure 3.1 *Search Space for Dynamic Programming.*

3.2.2 Dynamic Programming Search

A closed discrete contour can be described as

$$\mathbf{v}_i = (x_i\ ,\ \ y_i)\ \ i\text{=0..}N\text{-1} \tag{3.2}$$

where N is the number of points and the subscript arithmetic is modulo N. The energy of an open contour is given by

$$E_{\text{snake}}(\mathbf{v}) = E_0(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2) + E_1(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) + \ldots + E_N(\mathbf{v}_{N-3}, \mathbf{v}_{N-2}, \mathbf{v}_{N-1}) \tag{3.3}$$

and the energy at each snake point or node is given by

$$E_i(\mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}) = \lambda_i E_{\text{int}}(\mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}) + (1-\lambda_i) E_{\text{ext}}(\mathbf{v}_i) \tag{3.4}$$

where $\lambda \in [0, 1]$ is a regularisation parameter. In order to apply dynamic programming to equation 3.4, a two element vector of state variables $(\mathbf{v}_{i+1}, \mathbf{v}_i)$, is calculated at each stage. The optimal value function, S_i, is a function of two adjacent points on the contour and is calculated as

$$S_i(\mathbf{v}_{i+1}, \mathbf{v}_i) = \min_{v_{i-1}}[S_{i-1}(\mathbf{v}_i, \mathbf{v}_{i-1}) + \lambda_i E_{\text{int}}(\mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}) + (1-\lambda_i) E_{ext}(\mathbf{v}_i)] \tag{3.5}$$

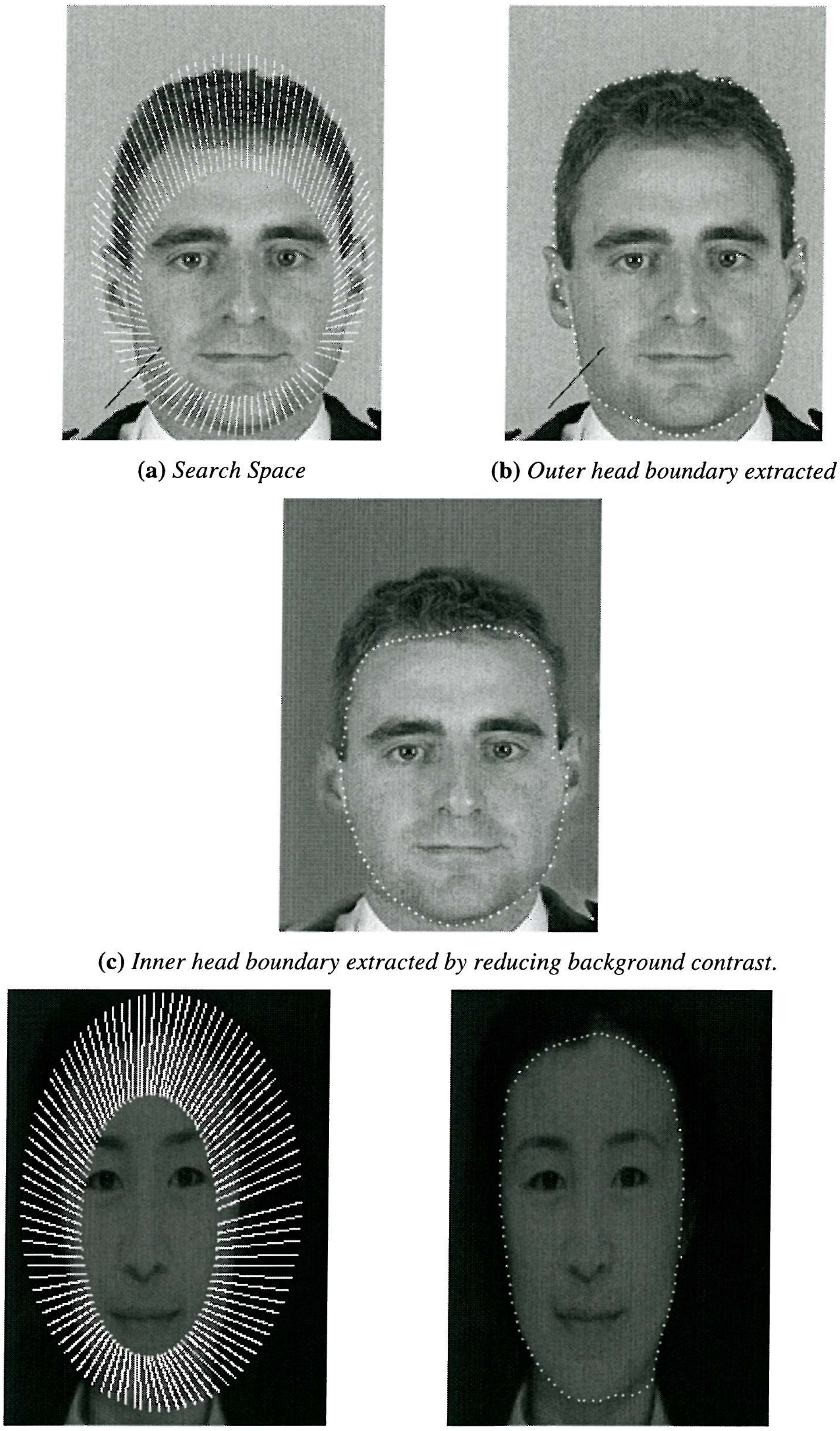
given the initial conditions $S_0(\mathbf{v}_1, \mathbf{v}_0) = 0$

where, $$E_{\text{int}}(\mathbf{v}_{i-1},\mathbf{v}_i,\mathbf{v}_{i+1}) = \left(\frac{\left|\mathbf{v}_{i+1} - 2\mathbf{v}_i + \mathbf{v}_{i-1}\right|}{\left|\mathbf{v}_{i+1} - \mathbf{v}_{i-1}\right|} \right)^2 \tag{3.6}$$

In addition to the energy matrix corresponding to the optimal value function, a position matrix is also required. Each entry of the position matrix at stage *i* stores the value of $\mathbf{v}_{i-1}$ that minimises the optimality function Equation 3.5 This is evaluated for *i*=1..*N*-2. The result is obtained by back-tracking through the position matrix.

3.2.3 Manually Initialised Snakes.

Gunn and Nixon [27] automatically initialise the search space using a number of heuristics. This resulted in subjective success rates of 95% and 73% for the outer and inner contours respectively on a database of 75 faces. In order to demonstrate the effectiveness of the dynamic programming approach to active contours we have initialised the search space manually. Figure 3.2 shows examples of a search space and the corresponding extracted head boundary. Figure 3.2 (a) shows a search space with a black line on the subject's left cheek which is in the search space. Figure 3.2 (b) shows that despite being in the search space, the dynamic programming algorithm (unlike the greedy algorithm) is sufficiently robust to avoid the local minimum offered by the black line in the search space. However, it is important to note that by reducing the background contrast of image (b), the inner head boundary was extracted instead of the outer head boundary, see image Figure 3.2(c). This will cause problems in a large database if it is desirable to compare all inner or all outer head boundaries.



(a) *Search Space*

(b) *Outer head boundary extracted*

(c) *Inner head boundary extracted by reducing background contrast.*

(d) *Search space*

(e) *Extracted contour*

Figure 3.2 *Type of head boundary (inner or outer) extracted dependant on background.*

3.2.4 Automatically Initialised Snakes.

In the previous chapter the eyes were located to a high degree of accuracy. The distance between them can be used as a guide to where nose, mouth and eye brows may be located. Referring to Figure 3.3, the eyes are located at co-ordinates $e1$, $e2$ and the distance between them is d pixels. An ellipse with semi-major axis d and semi-minor axis $0.75d$, centred at (x, y) can be used as a search space which should enclose the eyes, nose and mouth. This ellipse was used to define the inner contour. The top of the head and the width of head measured at height of the eyes can be used to define the ellipse for the outer contour.

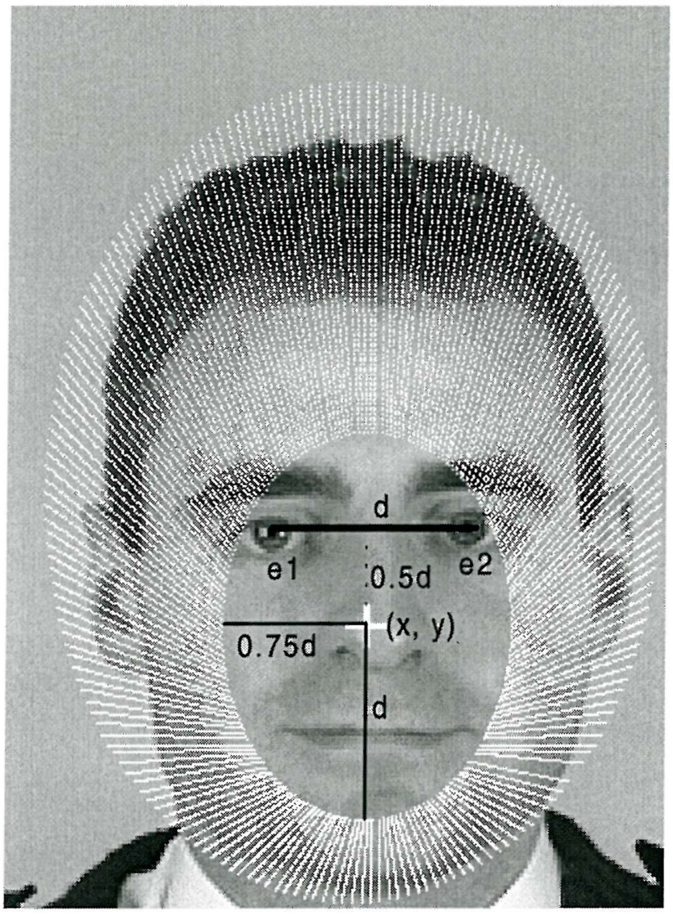


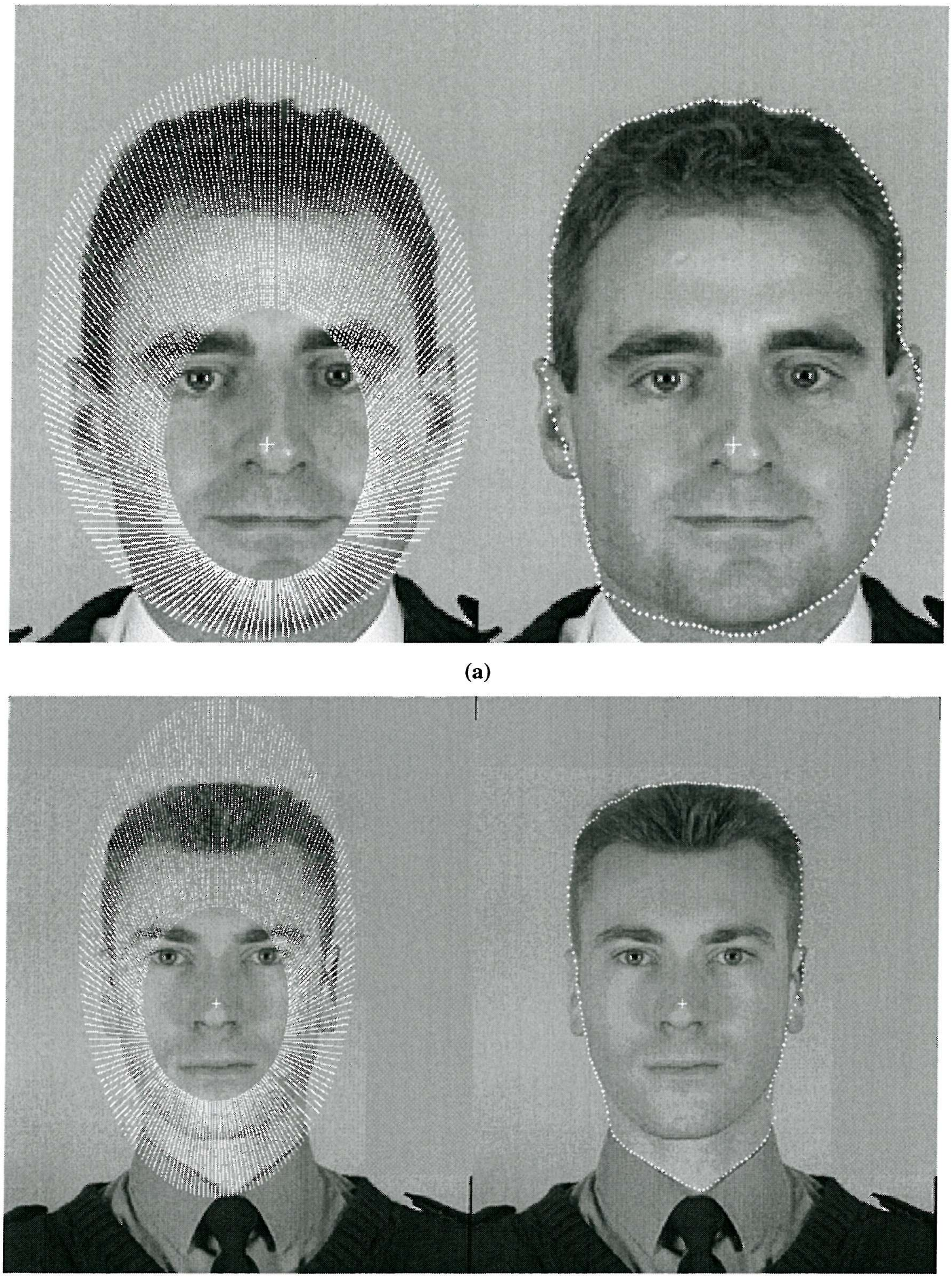
Figure 3.3 *Search space for face organs based on eye spacing.*

Figure 3.4 shows four examples head boundary extraction based on the eye locations. In example (a) the final contour has been attracted to the strong edge data provided by the shirt collar. Recall, that the final solution is the minimum energy around the contour which is evaluated at each stage in terms of curvature and normalised edge strength. The extracted contour in Figure 3.4(a) is just as valid as the solution of Figure 3.4(b) for the corresponding search spaces. The main difference between the two search spaces is that in Figure 3.4(a) a section of clothing with high gradient information was included. This problem is further highlighted in Figure 3.4(b)/(c) where the shirt collar is selected in preference to the chin by virtue of its stronger edge data. Example (d) shows an extreme case where the lower portion of the extracted contour is quite poor. Using the eyes alone, we are unable to satisfactorily control the size of the search space. Ideally, we need to restrict the search space to areas of skin. In the next section we describe a chromatic clustering algorithm which can be used to segment the skin and could act as a pre-filter to head boundary extraction stage.

3.3 Skin segmentation

Active contours need to be near the vicinity of desired contour to provide the required extracted contour. In the absence of *a priori* knowledge which defines the boundary of the skin, the quality of the head boundaries extracted by active contour methods will depend on the edge gradients produced by clothing. If the skin boundary can be located, this may offer good initialisation for the active contour.

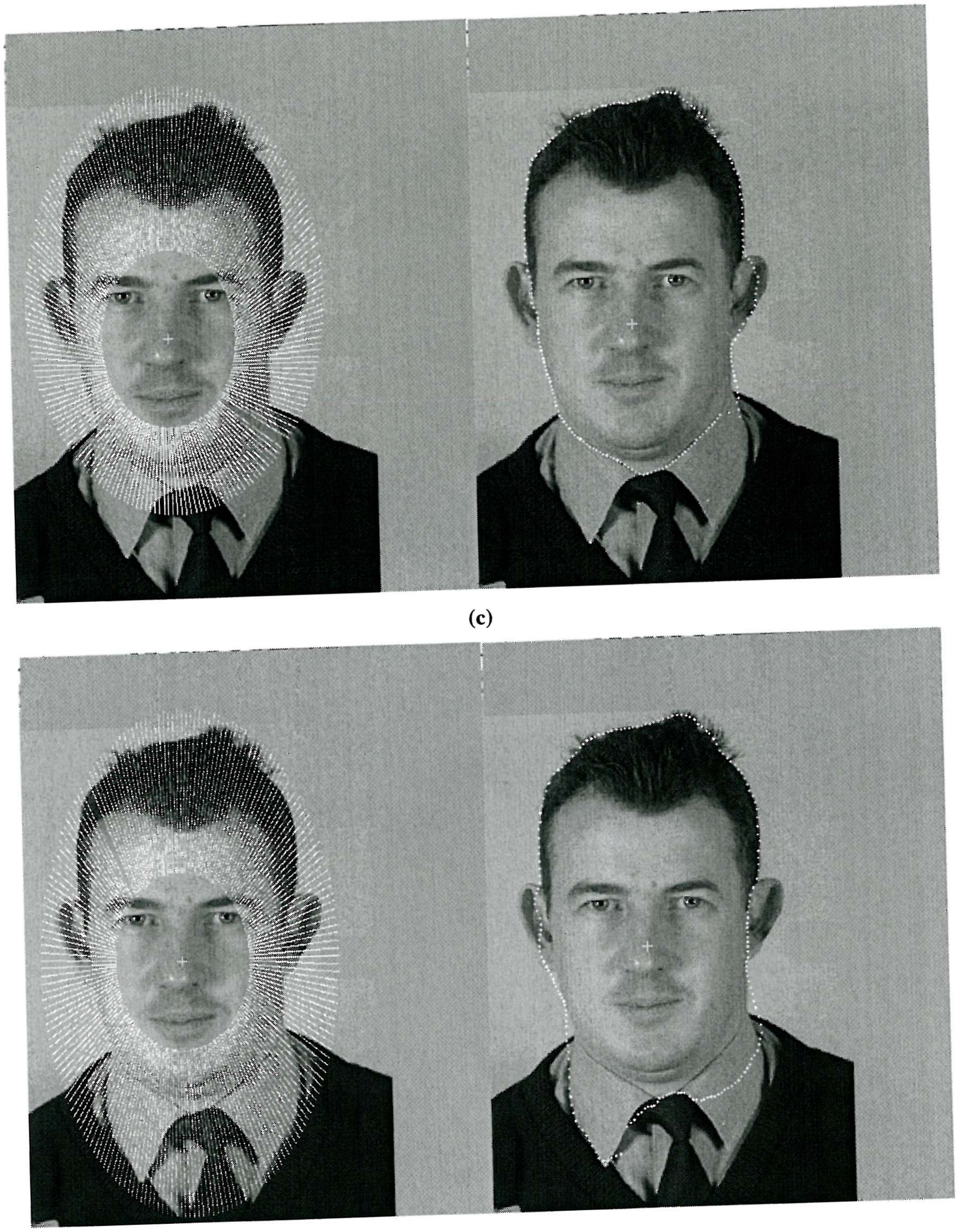
From our work in eye location, we have to located the eyes with reasonable frequency. It is also reasonable to assume that the region below the eyes is skin. If we further assume that the skin is a homogeneous region, then a region growing algorithm [61], seeded from the region below the eyes, can be used to extract the required skin boundary. Immediately the assumption of the skin being a homogeneous region means that moustaches and beards which do not match sample of skin under the eyes in terms of homogeneity may extract incorrect outer skin boundaries. However, if there are relatively few such faces in the database then this approach is acceptable. Given a sample region, s, below the eyes containing N pixels, if the colour of a pixel $I(i, j)$ at co-ordinates (i, j) is split into its red, green and blue components, I_r, I_g, I_b then the mean colour of a sample of skin below the eyes is given by



(a)

(b)

Figure 3.4 *Poor head boundary extraction due to strong edge data.*



(c)

(d)

Figure 3.5(cont.) *Poor head boundary extraction due to strong edge data.*

$$\mu_s(i,j) = (\mu_r, \mu_g, \mu_b) \tag{3.7}$$

where μ_r, μ_g, μ_b are the mean of the red, green and blue components of a pixel given by

$$\mu_r(i,j) = \frac{1}{N} \sum_{(i,j)\in s} I_r(i,j) \tag{3.8}$$

$$\mu_g(i,j) = \frac{1}{N} \sum_{(i,j)\in s} I_g(i,j) \tag{3.9}$$

$$\mu_b(i,j) = \frac{1}{N} \sum_{(i,j)\in s} I_b(i,j) \tag{3.10}$$

A pixel may be considered to belong to the skin region if its Euclidean distance to the sample skin pixel is less than a threshold value, D_t, and it is region-4 connected to a pixel in the skin sample. The Euclidean distance between two pixels p_1 and p_2 is given with colour components (r_1, g_1, b_1) and (r_2, g_2, b_2) is given by

$$d = ((r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2)^{\frac{1}{2}} \tag{3.11}$$

A region which is completely surrounded by skin but does not satisfy the criteria for skin is also of interest, since this region is likely to be the location of a face organ such as a mouth or nostril. Furthermore, we observed that the lips and nostrils usually have a high proportion of red compared to green and blue which can be used to refine the skin location process. A candidate for a non-skin pixel was identified if the pixel the red component was significantly higher than other colours, viz.

$$\frac{I_r}{\left(I_g + I_b\right)} > S_t . \frac{\mu_r}{\left(\mu_g + \mu_b\right)} \tag{3.12}$$

where S_t is a empirically determined threshold. Figure 3.6 shows an example of the skin segmentation process. The segmented skin is shown in grey, holes in the skin which correspond to face organs are shown in black and the background is represented as dark grey. This example shows that the holes in the skin can be used to initialise the search for the face organs such as the nose and mouth. The nose and mouth windows can now be initialised using the distance between the extracted eye centres.

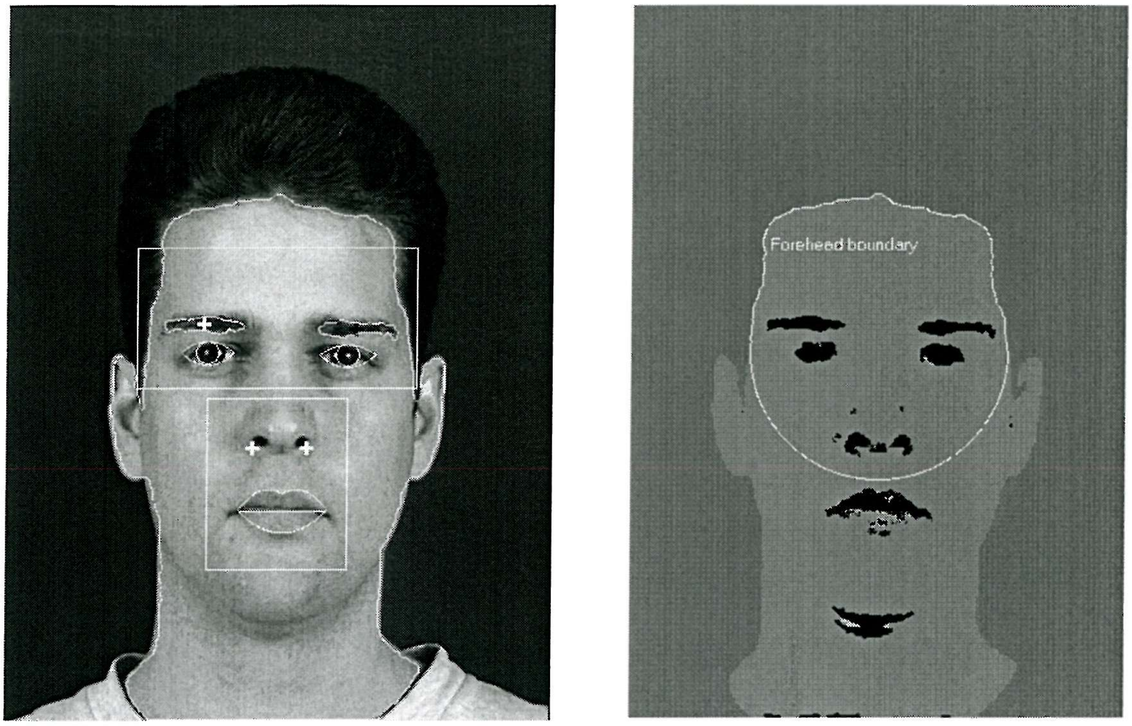


Figure 3.6 *Skin segmentation using region growing, seeded from skin under the eyes.*

The full set of extracted features are shown in the appendix. From Figure 3.6 we can extract automatic distance measures listed in terms of co-ordinates of landmark points from Figure 1.1. The names of these measures are listed in Figure 3.7. In addition, we can extract the red, green and blue components (skinColourRed, skinColourGreen, skinColourBlue) from the skin. Similarly, the colour components for the iris can also be extracted.

Name	x_1	y_1	x_2	y_2
headHeight	X	X	37	37
foreheadright	18	18	A	A
foreheadleft	19	19	B	B
headWidth	33	33	36	36
noselength	X	X	Y	Y
nosewidth	E	E	F	F
chinwidth	33	33	36	36
eyeaxistomouth	X	X	31	31
intereyebrowdist	15	15	16	16

mouthwidth	34	34	35	35
mouthdepth	31	31	32	32
mouthheight	30	30	31	31
scerlawidthl	22	22	23	23
scerlawidthr	24	24	25	25
scleraheightl	18	18	21	21
scleradepthl	18	18	20	20
eyetohairl	18	18	C	C
eyetohairr	19	19	D	D
eyebrowwidthl	13	13	15	15
eyebrowwidthr	16	16	17	17
eyebrowdepthl	38	38	14	14
eyetocheekl	18	18	C	C
eyetocheekr	19	19	D	D
mouthtojawl	31	31	33	33

Figure 3.7 *Features by automatic or manual extraction.*

3.4 Assessment of Extracted Features.

In this section we appraise the overall automatic feature extraction process with particular reference to issues that may be of interest to our sponsors. The appendix shows features extracted for 44 faces of 18 different people from the XM2VTS database. The eye location techniques have been tested on a large database (more than one thousand faces) compared to the size of many databases used in the literature. The success of the eye location techniques indicates that we have made a good start to production volume feature extraction.

However, it was not possible to extract the chin with any degree of accuracy due to inadequate active contour initialisation. We note that a different contour may be extracted from the same face image simply by changing the background intensity. It may be possible to determine an optimum regularisation parameter, but this may require consistent edge strength at the skin-background boundary. This consistency would be required throughout our sponsor's 20,000 face images. If correctly illuminated, the active contour initialisation for the chin may be obtained from the skin extraction algorithms. However, it does appear that inconsistent illumination has caused shadows under the chin on a number of occasions. We would also suggest that poor illumination was responsible for poor within-class feature segmentation, in particular, the eyebrows. Ignoring the illumination effects under the chin, facial skin was successfully segmented from the background using empirically derived thresholding limits. These limits would need to be evaluated over a range of skin colours or an alternative approach employed. We have used a simple distance measure to a sample of skin using its red, green and blue components. Perhaps more research into illumination invariant colour representations [55] could be useful, although from the introduction, it is clear that illumination is a difficult problem for face-based as well as feature-based recognition.

The hair is of particular interest to our sponsors but its extraction was not attempted. It would be easy to segment the hair from the background in the XM2VTS because the background is dark blue in colour and probably not easily confused with hair. However our sponsors expressed an interest in being able to obtain a subjective description of hair which might include tidiness, waviness, spikiness etc. This would require very high resolution images so that texture information could be extracted. Such high resolution images would also be useful for extracting the eyebrows.

From the extracted skin boundary, skin colour can certainly be determined from the resolution of image supplied. However, to determine more detail such as gender, age and complexion much higher resolution images may be required.

The mouth extraction results could possibly be improved by an increased population size in the genetic algorithm used for the mouth template optimisation. However it is difficult to suggest methods to reliably measure the nose parameters over a range of skin colours.

Clearly, there is still much research to be done before acceptable results are likely to be obtained on our sponsors' database of 20,000 images. The main proposals for our sponsors' consideration are to use high resolution images and pay much greater attention to illumination, background and to subjects' pose. However, this remains a visual assessment of the results of extracting natural features using eyes as a primer. Visual analysis suggests that within class variation in some features by these extractions might reduce their use in an automatic recognition system. Prior to confirming this we shall consider descriptions of the contour for use within a recognition system.

4. Fourier Descriptors for contour comparison.

We are ultimately interested in using the contours, which have been extracted from the face, as part of a face feature vector. Any properties of a contour representation which enhance automatic face recognition are important considerations. There are many techniques available for describing contours such as, chain codes [24] and B-splines [31]. However, Fourier descriptors (FDs) offer a method for shape comparison which can be made invariant to rotation, scaling, translation (RST) and start point. They also enable the use of numerous analysis and manipulation techniques after the contour has been converted to the frequency domain. In addition, the wealth of experimental and theoretical research suggests that Fourier descriptors are a suitable choice for our application. In this chapter we compare the methods proposed by Zahn & Roskies [89] and Kuhl & Giardina [41] for comparing contour boundaries which may be obtained from head boundary extraction.

4.1 Elliptic Fourier Descriptors

Kuhl and Giardina [41] represent a continuous closed contour in two dimensions, as a parametric function $v(t)$ of time t. The projections of $v(t)$ on the x and y axes are $x(t)$ and $y(t)$. Since the contour is closed, both $x(t)$ and $y(t)$ are periodic with period T, where T is the time taken to traverse the contour at a constant rate. The functions $x(t)$ and $y(t)$ are eligible for Fourier analysis and can be written as

$$x(t) = A_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{2n\pi}{T}t + b_n \sin\frac{2n\pi}{T}t\right) \tag{4.1}$$

$$y(t) = C_0 + \sum_{n=1}^{\infty}\left(c_n \cos\frac{2n\pi}{T}t + d_n \sin\frac{2n\pi}{T}t\right) \tag{4.2}$$

For the x projection,

$$A_0 = \frac{1}{T}\int_0^T x(t)dt \tag{4.3}$$

$$a_n = \frac{2}{T}\int_0^T x(t)\cos\frac{2n\pi}{T}t \quad dt \tag{4.4}$$

$$b_n = \frac{2}{T}\int_0^T x(t)\sin\frac{2n\pi}{T}t \quad dt \tag{4.5}$$

and similar term can be evaluated for the coefficients C_0, c_n and d_n for the y projection. Kuhl and Giradina [41] showed that for a piece-wise linear contour (e.g. a Freeman chain coded contour) the elliptic Fourier descriptors a_n, b_n, c_n and d_n can be evaluated as

$$a_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{K}\frac{\Delta x_p}{\Delta t_p}\left[\cos\frac{2n\pi}{T}t_p - \cos\frac{2n\pi}{T}t_{p-1}\right] \tag{4.6}$$

$$b_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{K}\frac{\Delta x_p}{\Delta t_p}\left[\sin\frac{2n\pi}{T}t_p - \sin\frac{2n\pi}{T}t_{p-1}\right] \tag{4.7}$$

$$c_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{K}\frac{\Delta y_p}{\Delta t_p}\left[\cos\frac{2n\pi}{T}t_p - \cos\frac{2n\pi}{T}t_{p-1}\right] \tag{4.8}$$

$$d_n = \frac{T}{2n^2\pi^2}\sum_{p=1}^{K}\frac{\Delta y_p}{\Delta t_p}\left[\sin\frac{2n\pi}{T}t_p - \sin\frac{2n\pi}{T}t_{p-1}\right] \tag{4.9}$$

where K is the number of piecewise linear sections in the boundary, Δx_p, Δy_p are the lengths of the projection of the p-th link on the x and y axis and

$$\Delta t_p = t_p - t_{p-1} \tag{4.10}$$

is the time to traverse link *p*. The terms A_0 and C_0 represent a contour positional offset and are given by

$$A_0 = \frac{1}{T}\sum_{p=1}^{K} \frac{\Delta x_p}{2\Delta t_p}(t_p^2 - t_{p-1}^2) + \xi_p(t_p - t_{p-1}) \tag{4.11}$$

$$C_0 = \frac{1}{T}\sum_{p=1}^{K} \frac{\Delta y_p}{2\Delta t_p}(t_p^2 - t_{p-1}^2) + \delta_p(t_p - t_{p-1}) \tag{4.12}$$

where

$$\xi_p = \sum_{j=1}^{p-1} \Delta x_j - \frac{\Delta x_p}{\Delta t_p}\sum_{j=1}^{p-1} \Delta t_j \tag{4.13}$$

$$\delta_p = \sum_{j=1}^{p-1} \Delta y_j - \frac{\Delta y_p}{\Delta t_p}\sum_{j=1}^{p-1} \Delta t_j \tag{4.14}$$

and

$$\xi_1 = \delta_1 = 0 \tag{4.15}$$

An approximation to the original contour can be reconstructed using

$$x(t) = A_0 + \sum_{n=1}^{N} X_n \tag{4.16}$$

$$y(t) = C_0 + \sum_{n=1}^{N} Y_n \tag{4.17}$$

where the projections for X_n and Y_n are given by,

$$X_n(t) = a_n \cos\frac{2n\pi}{T}t + b_n \sin\frac{2n\pi}{T}t \tag{4.18}$$

$$Y_n(t) = c_n \cos\frac{2n\pi}{T}t + d_n \sin\frac{2n\pi}{T}t \tag{4.19}$$

Giardina and Kuhl [25] show that the points (X_n, Y_n) describe elliptic loci. Thus the approximation of a closed contour using can be represented as the vectorial sum of rotating phasors, which are defined by the above projections. The frequency of each phasor, relative to the fundamental frequency of the first harmonic is determined by its harmonic number, n. They also show that the error between a point on the original contour and the reconstructed contour, ε, can controlled to any degree of accuracy by the number of harmonics, N. If the reconstruction error, ε, is defined as

$$\varepsilon = \max\left[\sup_{t}\left|x(t) - x_N(t)\right|, \quad \sup\left|y(t) - y_N(t)\right|\right] \tag{4.20}$$

then the error is bounded as

$$\varepsilon \leq \frac{T}{2\pi^2 N}\max\left[\overset{T}{\underset{0}{V}}(\overset{\circ}{x}(t)), \quad \overset{T}{\underset{0}{V}}(\overset{\circ}{y}(t))\right] \tag{4.21}$$

where $\overset{T}{\underset{0}{V}}$ represents the sum of the variation and $\overset{\circ}{x}(t)$ and $\overset{\circ}{y}(t)$ are the derivatives of $x(t)$ approximated by

$$\overset{\circ}{x}(t) = \frac{\Delta x_i}{\Delta t_i} \tag{4.22}$$

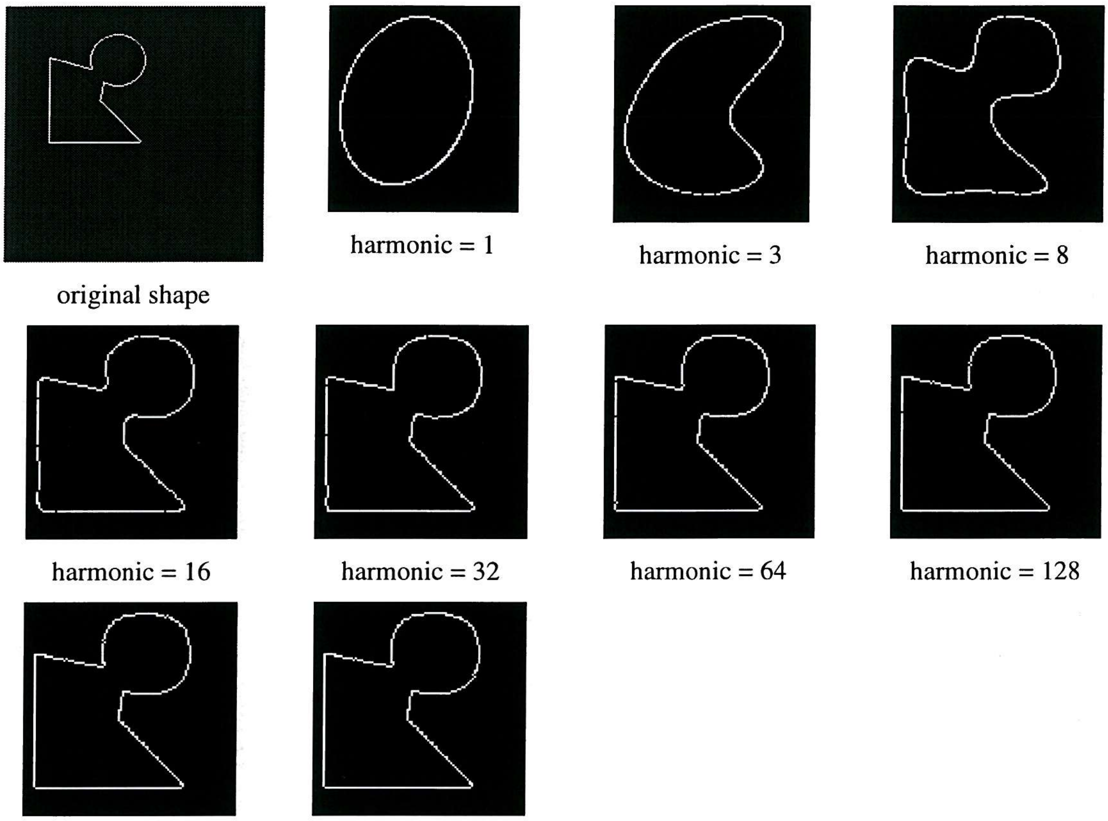
and

$$\overset{\circ}{y}(t) = \frac{\Delta y_i}{\Delta t_i} \tag{4.23}$$

The accuracy of the original representation can be increased by increasing N. Translation invariance can be achieved by ignoring the positional offset terms A_0 and C_0. Kuhl and Giardina note that the locus for the first harmonic can be either circular or elliptic. In the case of an elliptical first harmonic, scaling invariance can be achieved by dividing each of the coefficients by the magnitude of the semi-major axis or the radius of the circle. A nearest neighbour classifier was used to classify the shape of class m minimising the distance D of the unknown shape r to the known shape p.

$$D_m^2 = \min_{r=1,2,....P}\left[\min_{p=1,2,....P}\sum_{n=1}^{N} D_n(r,p,m)\right] \tag{4.24}$$

$$D_n^2(r,p,m) = \left({}_r a_n^{**} - {}_r a_{nm}^{**}\right)^2 + \left({}_r b_n^{**} - {}_r b_{nm}^{**}\right)^2 + \left({}_r c_n^{**} - {}_r c_{nm}^{**}\right)^2 + \left({}_r d_n^{**} - {}_r d_{nm}^{**}\right)^2 \qquad (4.25)$$



original shape

harmonic = 1

harmonic = 3

harmonic = 8

harmonic = 16

harmonic = 32

harmonic = 64

harmonic = 128

harmonic = 160

harmonic = 200

Figure 4.1 *Shape reconstruction using elliptic Fourier descriptors.*

4.2 Zahn and Roskies Fourier Descriptors

Zahn and Roskies [89] describe a simple closed curve, γ, with parametric representation $u(l) = x(l) + jy(l)$ where $0 \leq l \leq L$ is arc length and L is the total length of the curve. The accumulated angular change in direction of the curve since the start point is defined $\phi(l)$. For a closed curve, $\phi(0)$=0 and $\phi(L)$= 2π. To apply Fourier analysis, ϕ is mapped to a periodic function, ϕ^* where

$$\phi^*(t) = \phi\left(\frac{Lt}{2\pi}\right) + t \qquad (4.26)$$

is periodic with period 2π and $0 \leq t \leq 2\pi$. Expressed mathematically, the boundary may be described in term of its Fourier series as,

$$\phi^{*}(t) = \mu_0 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt) \tag{4.27}$$

Or in polar form the expansion is

$$\phi^{*}(t) = \mu_0 + \sum_{k=1}^{\infty} A_k \cos(kt - \alpha_k) \tag{4.28}$$

where the $\{ A_k , \alpha_k \}$ and $\{a_k, b_k\}$ are the Fourier descriptors in polar and Cartesian form respectively and μ_0 represents a positional offset. The Cartesian form of the Fourier coefficients for the n th harmonic evaluated at the m th vertex is given by,

$$\alpha_n = -\frac{1}{n\pi} \sum_{k=1}^{m} \Delta\phi_k \sin \frac{2\pi n l_k}{L}, \tag{4.29}$$

$$b_n = \frac{1}{n\pi} \sum_{k=1}^{m} \Delta\phi_k \cos \frac{2\pi n \Delta l_k}{L} \tag{4.30}$$

and the positional offset is given by

$$\mu_0 = -\pi - \frac{1}{L} \sum_{k=1}^{m} \Delta l_k \Delta\phi_k \tag{4.31}$$

where the terms Δl_k and $\Delta\phi$ represent the length of the polygon and change in angle at vertex k, see Figure 4.2. The polar representation of Zahn and Roskies' Fourier descriptors can be evaluated using,

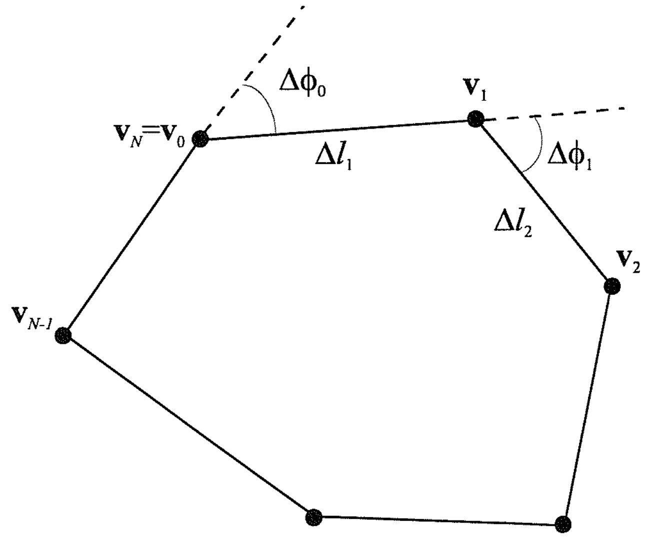


Figure 4.2 *Closed Planar polygon in terms of Edge Lengths Δl_i and Vertex Bends $\Delta\phi_k$*

$$A_n = \sqrt{\left(a_n^2 + b_n^2\right)} \tag{4.33}$$

$$\pi < \alpha_n \leq \pi \tag{4.34}$$

Reconstruction of an approximation to the spatial domain curve given its Fourier descriptors can be achieved via numerical integration of

$$Z(l) = Z(0) + \frac{L}{2\pi} \int_0^{\frac{2\pi l}{L}} \exp\left\{ i \left[-t + \delta_0 + \mu_0 + \sum_{k=1}^{N} A_k \cos\left(kt - \alpha_k\right) \right] \right\} dt \tag{4.35}$$

where Z(0) represents a positional start point. Furthermore Zahn and Roskies noted that given an originally closed contour, the reconstructed contour is not necessarily closed. The requirement for numerical integration for contour reconstruction and the appearance of the reconstructed are unattractive characteristics. However these drawbacks can be overlooked when the ease of shape comparison is considered. Essentially, shapes can be compared by measuring the amount by which the shapes differ from a circular shape. Fourier descriptors of a shape can be made invariant to scale by normalising the A_n by the magnitude of the first harmonic, A_1. The expressions for Zahn and Roskies Fourier descriptors are invariant to the start point and rotation. The α_n terms can be used to detect mirror images if necessary. Thus, the magnitude of the Fourier descriptors offer a simple means of comparing two shapes in term of rotation, scale and translation (RST). Given two curves p and q with normalised Fourier descriptors $\{A_{pn}, \alpha_{pn}\}$ and $\{A_{qn}, \alpha_{qn}\}$ a distance measure between the two contours can be evaluated as

$$D = \sum_{k=1}^{n} \left| A_{pk} - A_{qk} \right| \tag{4.36}$$

A distance measure of $D = 0$ indicates that curves γ and γ' are identical.

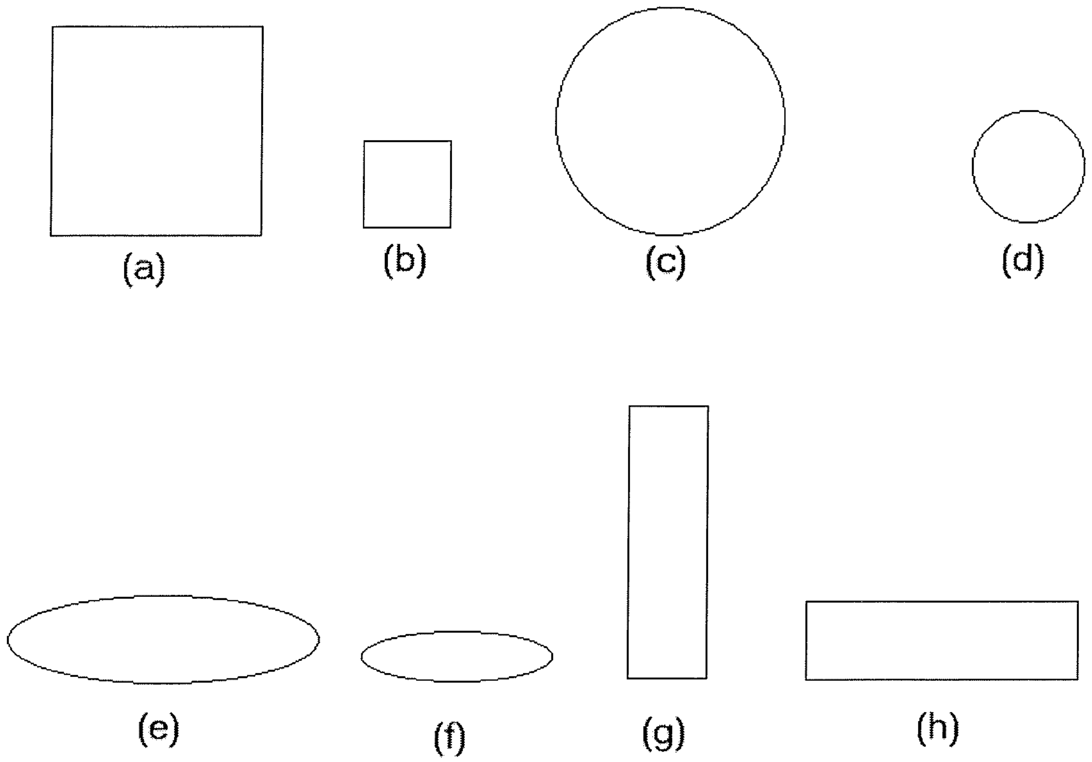


Figure 4.3 *Sample shapes for recognition.*

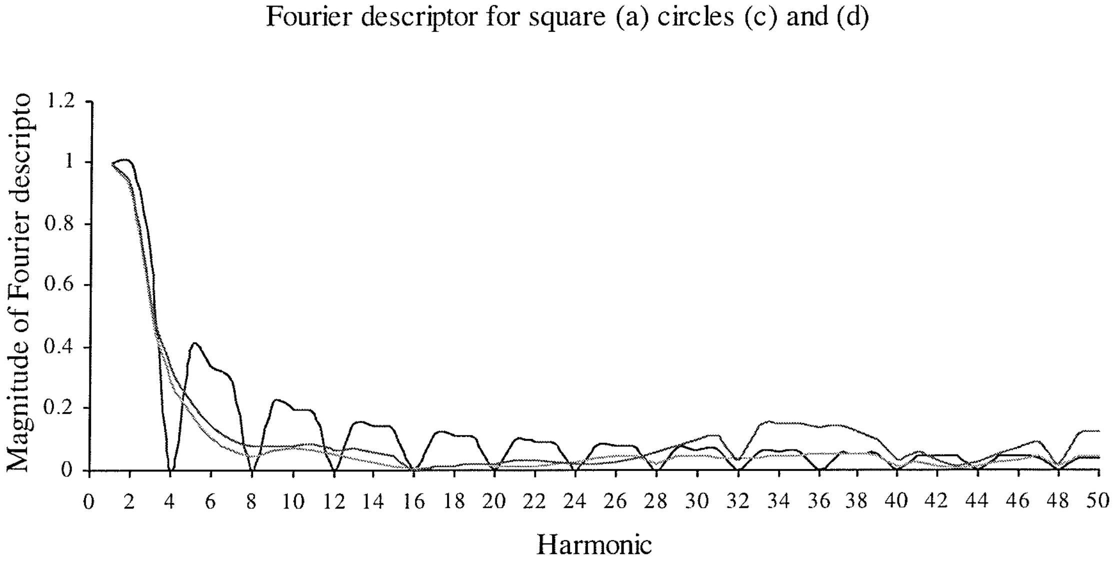


Figure 4.4 *Fourier descriptors for three shapes using Zahn and Roskies method.*

4.3 Elliptic vs Angular Fourier Descriptors

Van Otterloo [79] compares the properties of many of the popular forms of Fourier descriptors and shows that the convergence rate for Zahn and Roskies' method is much slower than other methods, for example elliptic loci [41] or complex number [29] based Fourier descriptors (FDs). The reconstruction algorithm for Zahn and Roskies' FDs, described by equation (4.35) requires a numerical integration solution. This often results in a contour, which was originally closed, being represented as an open contour after reconstruction. If our major concern was to achieve a faithful reproduction of the original contour and to control the number of harmonics required to achieve the desired reproduction quality, as perhaps in the case of digital transmission, then elliptic FDs seem most appropriate. Using elliptic FDs, the error between the original contour and the reconstructed contour can be controlled using equation (4.21). Furthermore, the loci of reconstructed contour always describes a closed contour. However we are most interested in shape comparison, rather than shape representation or reconstruction and Zahn and Roskies' method results in a very simple expression for shape comparison as expressed by equation (4.37). Two shapes are the same if their Fourier descriptors are identical. In Zahn and Roskies presentation, the amplitude of the FDs offers direct shape comparison by means of a distance measure between the two sets of FDs. In Kuhl and Giardina's presentation [41] shapes are classified using a nearest neighbour classifier. The nearest neighbour classifier requires a distance measure to be minimised over the set of shapes. Shapes with the smallest minimised distance measures are considered to be most similar. We have adopted Zahn and Roskies' method for further experimentation rather than Kuhl and Giardina's method since we can avoid minimisation over the set of shapes.

4.4 Enhanced shape discrimination for FDs

We note that although the first harmonic is required for a shape's reconstruction, it does not vary across the set shapes since it is usually normalised to unity for shape comparison. Considering the example of shape reconstruction in Figure 4.1 it may be necessary to use many Fourier descriptors to achieve the desired degree of accuracy. However, only a few descriptors may be required to distinguish different shapes. Accordingly, we prescribe that the lack of variation in the value of the first harmonic, should be incorporated into the similarity measure for shape comparison. On the other hand, a harmonic which contains a large amount of variation across the sample set should be used to amplify the difference in the shapes. Thus, we introduce a new set of weights into the distance measure for comparing Zahn and Roskies' FDs, equation (4.37). The distance measure between shapes p and q can now be written,

$$D = \sum_{k=1}^{n} w_k \left| A_{pk} - A_{qk} \right| \tag{4.37}$$

where the coefficient weights to be determined w_k are proportional to the variation of the k-th harmonic, across the set of n shapes. The form for the weights can be expressed as

$$w_k = \overset{n}{\underset{i=1}{V}}(F_{ik}) \tag{4.38}$$

where $\overset{n}{\underset{i=1}{V}}$ represents the variation over the set of shapes and F_{ik} represents the k-th harmonic (or feature) for the i-th shape (or sample). For the case of Zahn and Roskies' FDs, we can directly use the variation in the amplitude of the FDs to amplify important harmonics by setting set F_{ik}=A_{ik}. On this basis, the coefficient weighting for the first harmonic, w_1 = 0, since there is no variation in amplitude of the first harmonic A_{1k}. The nearest neighbour classifier for the elliptic FDs, equation (4.9) can also be enhanced by variation weighting coefficients by applying the variation weights to distance measures that require minimisation. We note however, that for the elliptic FDs we would need to quantify the variation of the set of vectors (a_{ik}, b_{ik}, c_{ik}, d_{ik}) whereas with Zahn and Roskies' FDs we only need to quantify the variation of A_{ik} which is a scalar quantity.

The terms "features" and "samples" are included in parenthesis to emphasise that weights defined in this manner could also apply to a general set of features, F. The form of the distance measure for the Zahn and Roskies FDs is conveniently similar to distance measures which may be used compare feature vectors. A method for determining the amount of variation contained in a set of measures, V, is the topic of the next chapter.

5. Feature Selection & Combination.

We have described some of the techniques available to extract face features, both in our studies and elsewhere, but much less attention has been paid to how to select or combine features to achieve a useful result. Even if features can be robustly extracted over a large database of faces, we still need to quantify the usefulness of a set of extracted features. The question posed by feature selection is; *which features, and in what proportions, maximise the systems recognition rate, yet optimise computational costs* ? An exhaustive examination of the feature set rapidly becomes impractical as the number of features increases. One approach to feature selection could be to formulate and optimise a cost function which uses the features as its input. We have already discussed the use of techniques such as generic algorithms, dynamic programming, or gradient methods which might be use for optimisation. Alternatively, one might attempt to identify a subset of features that provide a good acceptable performance using domain knowledge of the feature space.

Kittler *et al* [40] provide a theoretical framework for combining multi-modal data sources. One face recognition study compared the relative merits of features versus templates [7] summing the performance from the individual features as a means of feature combination and then extended the approach to fuse speaker recognition and face data derived from templates around the eyes, nose and mouth [6]. Jain and Zonker [35] illustrate the value of feature selection in combining features from different data models and demonstrate the potential difficulties of performing feature selection in small sample size situation. This study compared fifteen search techniques and emphasised the power of the sequential forward floating selection method. These tests used, twenty dimensional, two-class data sets. Fisher's Linear Discriminant Analysis (LDA) is designed to maximise the ratio of between-class to within-class scatter [23]. However in a face retrieval system we do not typically have such *a priori* class information.

Roeder and Li [71] qualitatively analyse accuracy requirements for face-based face recognition. Twelve face-based features including eyes, nose, mouth and contour measurement were used. Individual and groups of feature measurement were perturbed in order to examine the effects on recognition success. They used a database of 333 faces with a recognition system based on a nearest neighbour classifier. Their results indicated that the eyes were least sensitive to the recognition process whereas the measurements involving the cheeks and chin were amongst the most sensitive.

Nixon *et al* [59] present theoretical arguments which further support the combination of features from different sources. The theory was applied to the complete (112 images) Brodatz texture set and provided an improved classification from 82% and 76% up to 88%. Jia and Nixon [42] used four feature sets; geometric measurements of the major face organs, Fourier descriptors on the head boundary contour, moments of the eye area and the Walsh power spectrum of a face profile. They found that an extended feature vector composed of ¼ weighting of the original features proved a better discriminator than any single feature. This result was achieved on one face in their database but suggested that discriminatory power may be increased by combining orthogonal feature sets.

Kaya and Kobayashi [38] performed manual measurements on 62 enlarged photographs of 8 different faces and developed an information theory based approach to face recognition. They modelled the extracted face features as a signal

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{D} \tag{5.1}$$

where $\boldsymbol{D}$ represents noise corrupting an otherwise perfectly extracted one dimensional feature vector $\boldsymbol{X}$**.** They note that the maximum number of classifiable patterns is limited by the average mutual information between $\boldsymbol{X}$ and $\mathbf{Y}$ defined as

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) \tag{5.3}$$

where $H(\boldsymbol{X})$ is the entropy of $\boldsymbol{X}$ and H($\boldsymbol{X}|\boldsymbol{Y}$) is the entropy of $\boldsymbol{X}$ given $\boldsymbol{Y}$. When the noise $\boldsymbol{D}$ and the feature vector are normally distributed the mutual information between $\boldsymbol{X}$ and $\mathbf{Y}$ is given by

$$I(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2}\left(\log_2\left|M_X + M_D\right| - \log_2\left|M_D\right|\right) \tag{5.3}$$

Where M_X and M_D are the covariance matrices of $\boldsymbol{X}$ and $\boldsymbol{D}$ respectively. Two sources of noise were identified: (a) measurement noise $\boldsymbol{D}_m$, introduced by from the hardware equipment etc. which would cause different results to be obtained from the same photograph and (b) intrinsic noise $\boldsymbol{D}_i$, which accounts for differences in multiple versions of the same face. Sources of intrinsic noise would include variations facial expression and orientation to the camera, changes of hair styles etc. PCA was applied to $\boldsymbol{X}$ and $\boldsymbol{D}$ to produce orthogonal linear transformations of the features and noise $\boldsymbol{LX}$ and $\boldsymbol{LD}$. Assuming $\boldsymbol{D}_i$ rather than $\boldsymbol{D}_m$ to be the dominant source of noise,

$$I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{X};\boldsymbol{LY}) \tag{5.4}$$

$$= \frac{1}{2}\log_2 \frac{|M_{LX} + M_{LD}|}{|M_{LD}|} \tag{5.5}$$

$$= \sum_i \frac{1}{2}\log_2 \left|1 + \frac{\sigma_i^2}{\gamma_i^2}\right| \tag{5.6}$$

where M_{LX} and M_{LD} are the covariance matrices of ***LX*** and ***LD*** respectively and σ_i^2 and γ_i^2 are the variances of the diagonalised matrices ***LX*** and ***LD.*** Kaya and Kobayashi then continued by developing a nearest neighbour classifier relying on PCA as a method of data reduction.

The problem with PCA from a feature selection point of view, is that although it can result in considerable data reduction for highly correlated data, it is still a transformation of *all* of the original variables. As such, it does not tell us which variable or feature contains the most information. However, if our feature selection strategy is based on selecting uncorrelated features, whose covariance matrix is already a diagonal matrix, then PCA will merely order the variables in terms of their variances without an interdependence on all the original variables. A heuristic answer to the rhetorical question poised by feature extraction at the beginning of this chapter is to select statistically orthogonal features, since all the variance and information is captured in the leading diagonal of its covariance matrix.

Brunelli and Poggio [8] found that template matching produced better recognition rates than geometric feature extraction. They found that the discriminating power of their templates, arranged in decreasing performance, were the eyes, nose, mouth and whole face. They considered approaches to combine the similarity score of different features to produce a global score. The following strategies were identified:

1. Choose the score of the most similar feature.
2. Add the feature scores.
3. Add the feature scores, but include a different weight for each feature. The weight for each feature being the same for each person in the database.
4. Add the features including a different feature weighting for each person in the database.
5. Feature score are optimised using a nearest neighbour classifier or a hyper basis function neural network.

Brunelli and Poggio found that even the simple approach of adding the feature scores improved overall recognition. Jia and Nixon in effect implemented this approach since they simply used equal weighing of ¼ for each of their 4 statistically independent features. In the next section we define weights for each feature and applied them to each person database. Ignoring the inevitable noise introduced by feature extraction, these weights quantify the amount of variation present in each extracted feature and can be used to weight the components of a distance measure.

5.1 Classification via variance weightings

We have extracted a set of face measures which need to be compared in order to realise a recognition system. Some of the popular methods for measuring the similarity or difference between two vectors in earlier chapters. These include the Euclidean distance, the Mahalanobis distance and the nearest neighbour classifier. The relative merits of these and other similarity metrics are discussed by Webb [81]. Although these measure can be used to implement a recognition system, they do not provide information which indicates which individual components of the feature vector are efficient discriminators. To identify these components we first define the similarity measure of a feature vector, of length M, for two faces a and b as,

$$S(a,b) = \sum_{i=1}^{M} w_i S_i(a,b) \tag{5.10}$$

where $S_i(a, b)$ is a distance based similarity measure and w_i are weighting coefficients to be determined. The form of $S_i(a, b)$ that we have chosen is shown in equation 5.12 but first we determine the terms w_i. The set of feature vector measurements extracted from a database of faces can be expressed in matrix form,

$$\boldsymbol{F} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{ij} & & x_{2N} \\ \vdots & \vdots & & \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{bmatrix} \tag{5.10}$$

where, element x_{ij} represents the i-th feature measurement for the j-th face and N is the number of faces in the database. Assuming that the between class variation is greater than the within class variation, we seek to choose values for the weights which are proportional to a feature's population variation. The unprocessed sample variance may unfairly bias the coefficient weighting in favour of features of large magnitude. This is because the sample variance of large measures, may be greater than for a set of measures of small magnitude. To compensate for this, each feature is normalised by the maximum feature measurement in the vector of N faces before calculating its sample variance. Normalising the data in this way before calculating the sample variance provides a more useful measure of a feature's intrinsic variance. For example, the head width for each person in the database is normalised by the largest head width within the database. The normalised feature vector is given by

$$x_i^* = \frac{x_i}{\max\limits_j(x_i)} \tag{5.10}$$

By normalising each feature by its maximum value, we obtain a new set of feature measures, $\boldsymbol{F}^*$ which are bounded in the range [0, 1]. A feature that does not vary will have zero variance and correspondingly any weighting term used in a similarity measure should also be zero. The weights for improved classification are now given by the ratio of the variance of a row of a particular feature measure, to the total normalised variance,

$$w_i = \frac{\sigma^2_{x_i^*}}{\sigma^2_{F^*}} \tag{5.10}$$

subject to the constraint

$$\sum_{i=1}^{M} w_i = 1 \tag{5.11}$$

where σ denotes variance. A similarity measure, based on the Canberra distance measure was chosen as

$$S_i(a,b) = \sum_{i=1}^{N}\left(1 - \left|\frac{x_{ia}^* - x_{ib}^*}{x_{ia}^* + x_{ib}^*}\right|\right) \tag{5.12}$$

where a and b are two feature vectors. If the extracted features are to be used as part of an effective recognition system, the within class similarities should be greater than between class similarities. In this case a face may be considered as being recognised. As discussed in the introduction, there are many criteria used in the literature to provide benchmark system performance. We shall define a recognition rate to be the ratio of recognised faces to the number of faces in the database. This definition of success is more stringent that used by Kamel *et al* [39] who achieved a 95% recognition rate by considering the best four matches on a database of 84 faces. Our definition of recognition is not only simpler, it is also closer to the requirements of a commercial system or a human recollection of a specific face. We also define success in terms of a system classification error, *ce*, which is related to the recognition rate *rr* by,

$$rr = 1 - ce \tag{5.13}$$

5.1.1 Variance weights applied to Fourier descriptors

The method of calculating the variance in each feature is exemplified in Figure 5.1. The table in Figure 5.1 (a) shows the normalised Fourier descriptors for the sample shapes of Figure 4.3. Note that first harmonic is normalised to unity and therefore does not contain any discriminatory power. To prevent the coefficient weights from being dominated by the variance of the larger lower order harmonics, we first normalise the row of the feature vector matrix by the largest harmonic in the row as shown in Figure 5.1 (b). The sample variance of the row of normalised Fourier descriptors are the coefficient weights are also shown in Figure 5.1 (b). The resulting weights profile is shown in Figure 5.2. The cumulative weight shows what percentage of the total variance is accounted for by using up to the *n*-th harmonic. For example the 43rd harmonic, corresponds to a coefficient weighting of approximately 3 and 83% of the total variance if all harmonics up to the 43rd are used. If the cumulative weight was a sharp exponential, it may be possible to capture the majority of the variance with the first few harmonics. However, the cumulative distribution is approximately linear and we cannot truncate the series of harmonics without losing significant amounts of variance. Nonetheless, it is clear that some harmonics have greater variance and accordingly are assigned greater weighting. The benefit of applying variance weights can be seen in the system performance summarised in Figure 5.3. The variance of the confusion matrix is generally higher when variance weights are applied. This is particularly apparent for a small number of harmonics because in the case of equal weights the first harmonic is given the same weighting as the other harmonics when it actually does not contain any discriminatory power. We have calculated the Fourier descriptors so in the absence of noise it would be desirable to be able to be able to recognise the full set of sample shapes. The desired 100% recognition rate is achieved using variance weighted similarity coefficients. However, in the case of using equal weighted similarity coefficients, the recognition rate falls because the weighting (1/harmonics) become too small to act on the important terms in the similarity measure. It is important to note that in this experiment the variance weights have improved recognition in an environment where there is no measurement noise, since we have calculated the Fourier descriptors.

Harmonic	shape a	shape b	shape c	shape d	shape e	shape f	shape h	shape g
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	1.000	0.896	0.919	0.253	0.222	0.645	0.645
3	0.667	0.667	0.515	0.546	0.662	0.667	0.667	0.667
4	0.000	0.000	0.288	0.335	0.204	0.177	0.493	0.493
5	0.400	0.400	0.184	0.225	0.392	0.401	0.400	0.400
6	0.333	0.333	0.104	0.144	0.184	0.161	0.287	0.287
7	0.286	0.286	0.064	0.096	0.275	0.287	0.286	0.286
8	0.000	0.000	0.046	0.077	0.164	0.144	0.083	0.083
9	0.222	0.222	0.066	0.077	0.208	0.223	0.222	0.222
10	0.200	0.200	0.069	0.081	0.149	0.135	0.071	0.071

(a) *Zahn and Roskies Fourier descriptors for sample shapes.*

Harmonic	shape a	shape b	shape c	shape d	shape e	shape f	shape h	shape g	Weights
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.00
2	1.000	1.000	0.896	0.919	0.253	0.222	0.645	0.645	1.90
3	0.999	0.999	0.772	0.819	0.992	1.000	0.999	0.999	0.17
4	0.000	0.000	0.585	0.68	0.414	0.359	1.000	1.000	2.89
5	0.998	0.998	0.459	0.562	0.979	1.000	0.998	0.998	0.97
6	1.000	1.000	0.311	0.432	0.551	0.483	0.861	0.861	1.42
7	0.997	0.997	0.222	0.334	0.959	1.000	0.997	0.997	2.09
8	0.000	0.000	0.28	0.468	1.000	0.882	0.505	0.505	2.52
9	0.996	0.996	0.296	0.343	0.931	1.000	0.996	0.996	1.82
10	1.000	1.000	0.346	0.406	0.747	0.676	0.357	0.357	1.54

(b) *Variance weights from normalised feature vector matrix.*

Figure 5.1 *Calculating variance weights for sample shapes using first 10 harmonics*

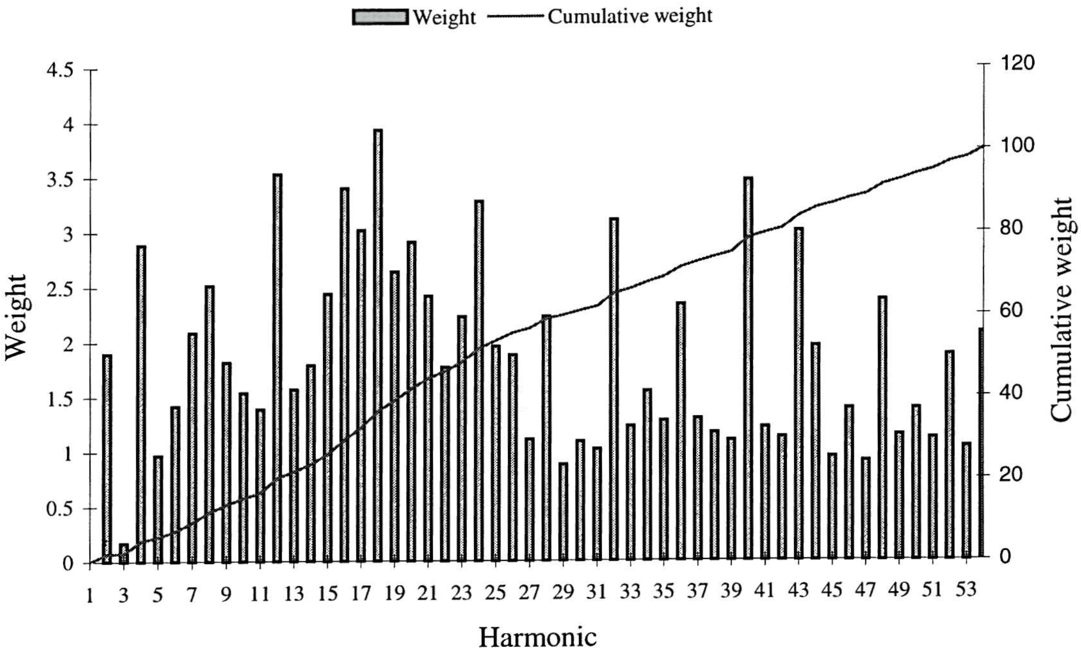


Figure 5.2 *Variance weights and cumulative variance weight.*

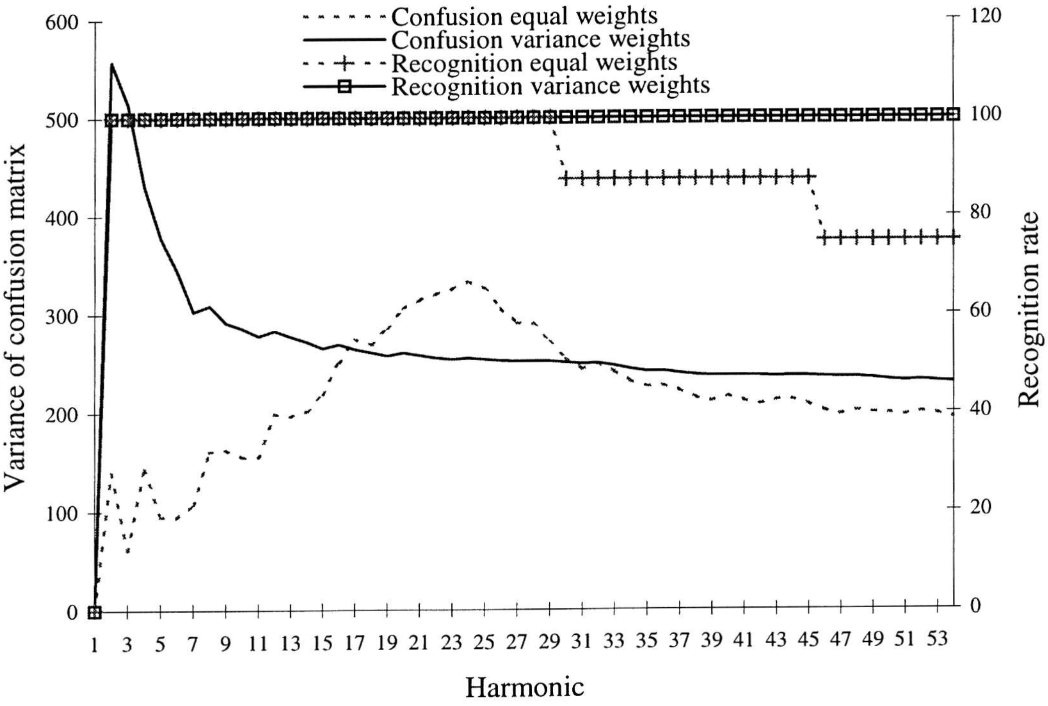


Figure 5.3 *Variance of confusion matrix and recognition rate.*

5.1.2 Variance weights applied to geometric face measures

The face measurements of 44 faces from the M2VTS database were manually extracted by two expert coders. The database contained multiple copies of each of the 18 different faces so that recognition experiments could be performed. Our sponsors kindly provided a database of 200 faces, one face per subject with face features coded in accordance to the PITO coding scheme. Even using expert coders, we must expect some measurement noise on each feature. We assume that the noise due to data acquisition and sampling is small compared to the intrinsic noise introduced by variations in illumination and facial expressions. Ideally, the intrinsic noise should be small compared to the actual feature being measured, e.g. head height, otherwise the weights are reflection of the noise and the recognition rate will suffer. Figure 5.4 shows the variance weighting coefficients obtained by applying equation (5.10) to 44 faces from the XM2VTS database.

Figure 5.5 compares the results obtained by applying variation weightings and constant weighting to 200 faces from the PITO database and 44 faces from the M2VYS database. It can be seen that the variance of the confusion matrices using variation weightings are always higher then those obtained without variation weights. However, the recognition rates obtained using variance weights were lower than those obtained by equal weightings, which indicates a high level of measurement noise was present.

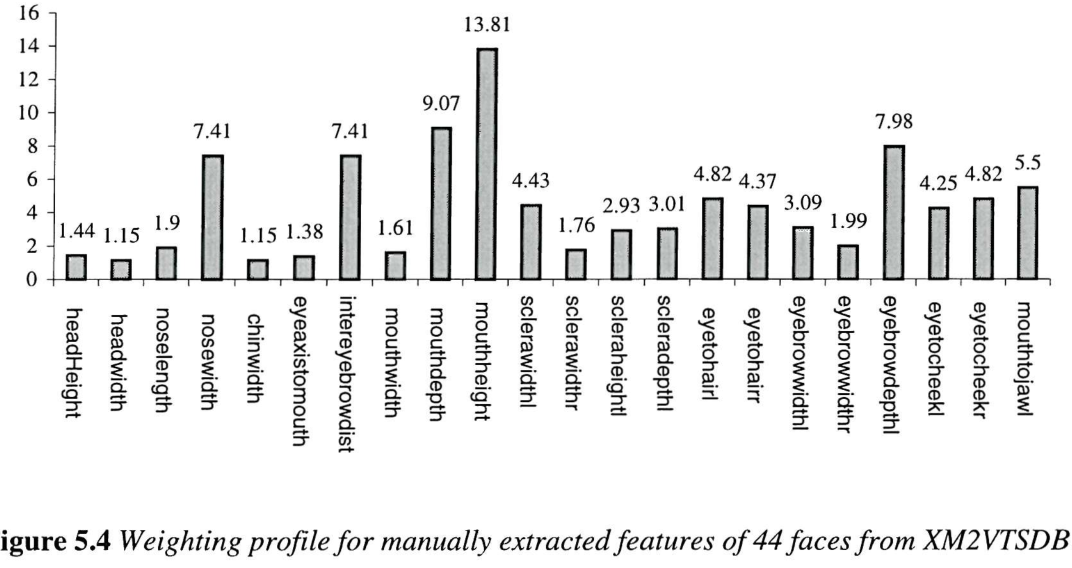
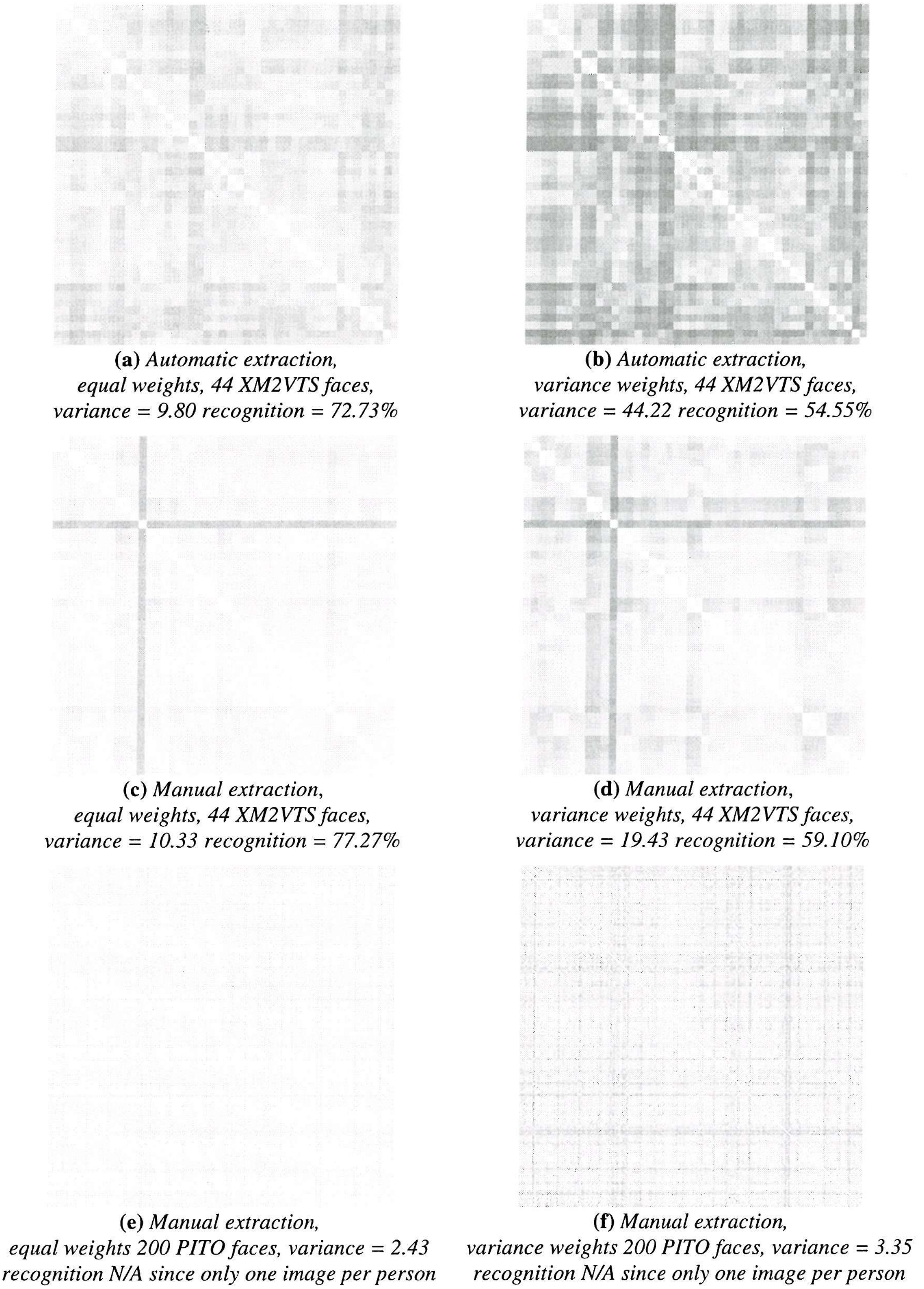


Figure 5.4 *Weighting profile for manually extracted features of 44 faces from XM2VTSDB.*



(a) *Automatic extraction, equal weights, 44 XM2VTS faces, variance = 9.80 recognition = 72.73%*

(b) *Automatic extraction, variance weights, 44 XM2VTS faces, variance = 44.22 recognition = 54.55%*

(c) *Manual extraction, equal weights, 44 XM2VTS faces, variance = 10.33 recognition = 77.27%*

(d) *Manual extraction, variance weights, 44 XM2VTS faces, variance = 19.43 recognition = 59.10%*

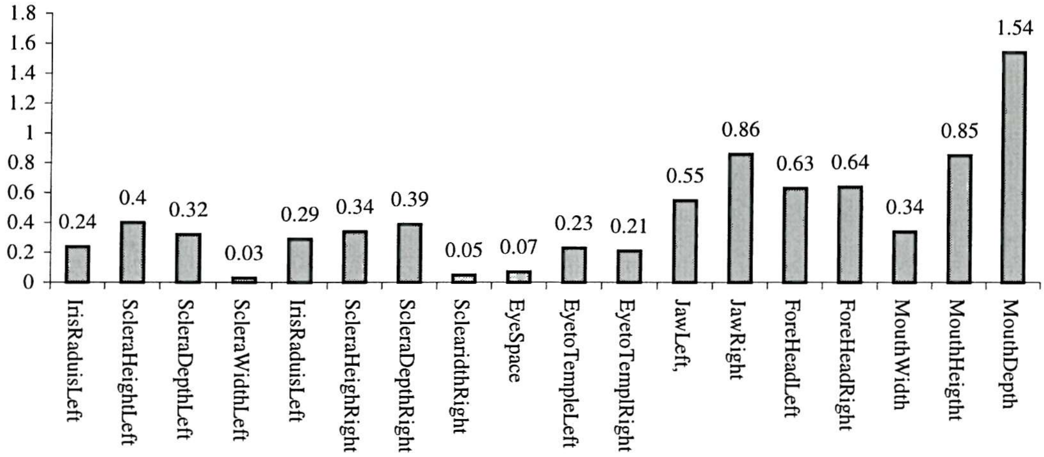
(e) *Manual extraction, equal weights 200 PITO faces, variance = 2.43 recognition N/A since only one image per person*

(f) *Manual extraction, variance weights 200 PITO faces, variance = 3.35 recognition N/A since only one image per person*

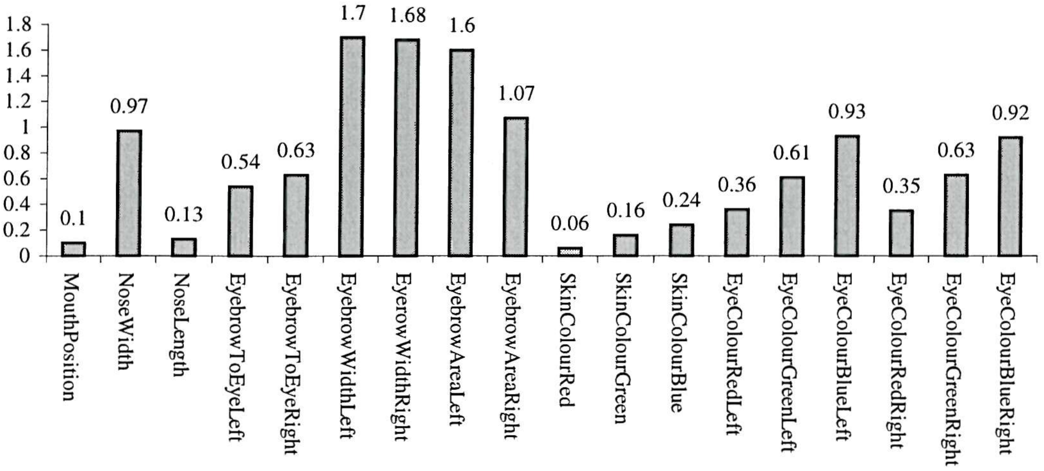
Figure 5.5 *Confusion matrices for manual and automatically extracted features.*

5.1.3 Variance weights applied to composite feature vector

Within the extracted feature set, there are a number of different types of features. For example, we have distance measures, such as the eye parameters obtained from the deformable eye template, area measures for the eyebrows, colour information from the skin and iris boundaries and contour information from the Fourier descriptors of the forehead boundary.

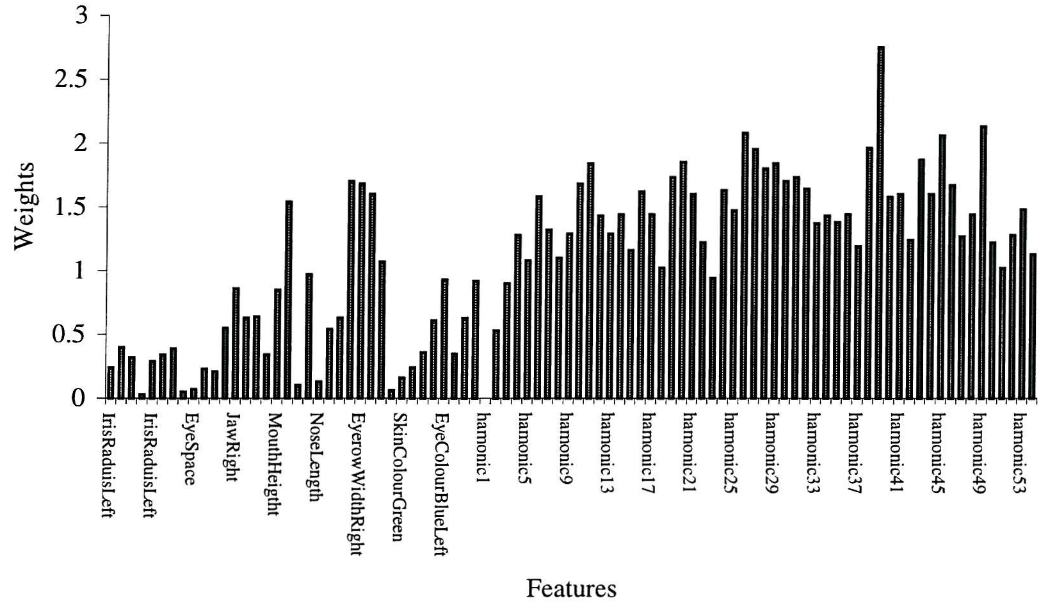


(a) weightings for first 18 features

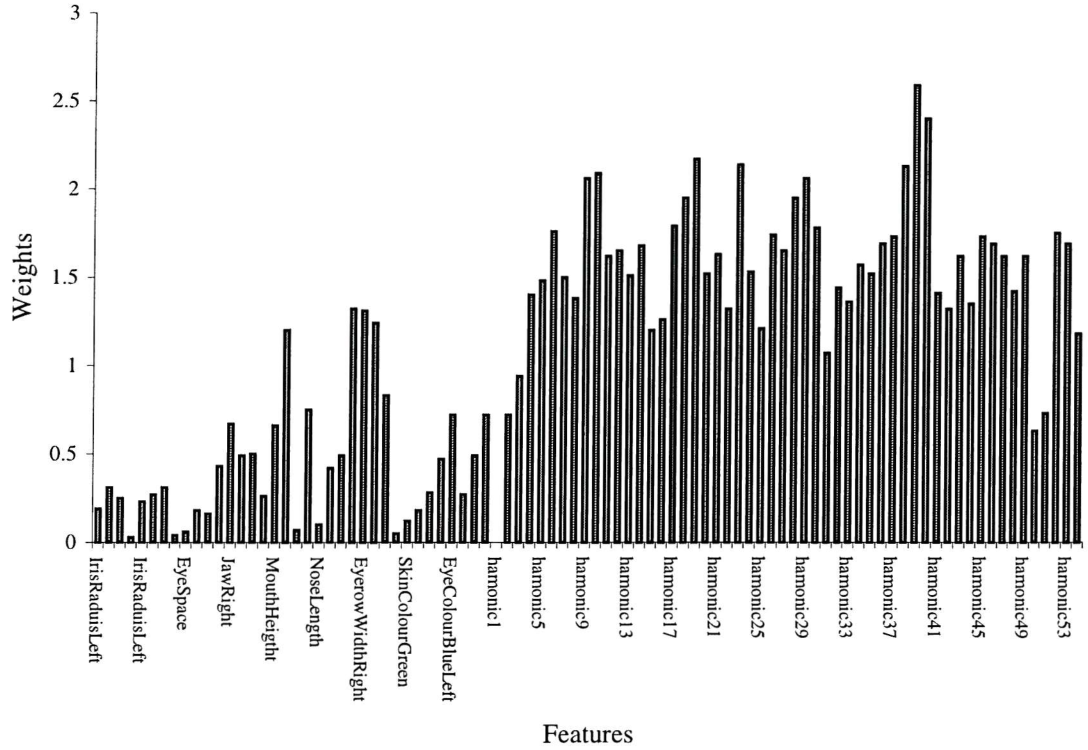


(b) weightings for next 17 features

Figure 5.6 *Weightings for geometric, colour and contour features*



(c) *Weights for extracted features.*



(d) *Weights for extracted features using perfect class information for Fourier descriptors*

Figure 5.6 *continued from previous page*

Figure 5.6(c) shows the variance weights distribution using automatically extracted feature for the database of 44 faces. The distribution in Figure 5.6(d) also uses the full set of extracted features, but in this case we have assumed perfectly robust within class contour extraction by assigning within classes to the same contour.

The results are interesting because they show that despite being the most robustly extracted, the geometry of the deformable eye template offers little in terms of discriminatory power. The most discriminating geometric features are the area of the eyebrows and those measurements relating to the parameters of the deformable mouth template. Examination of the extracted features in the appendix suggests that extraction of the eyebrow boundary is amongst the least robustly geometric features. This distribution is in contrast to the results obtained by Brunelli and Poggio using template matching [7]. They found that the most discriminating features in decreasing order of importance were the eyes, nose, mouth and whole face template. They suggested that the whole face templates were the least successful because of difficulties in normalising the scale of their pictures and sensitivity to head rotations. We suggest that discriminating order for the eye, nose and mouth templates may be reasoned in terms of their likely variance content.

The associated high weighting given to the eyebrow areas in the similarity measure will have an adverse effect on the recognition rate or classification error. One might be tempted to assume that the variance weightings are proportional to the magnitude of the original measurements which might explain why the eye parameters have less discriminatory power than the eyebrows or mouth parameters. This argument can be easily refuted when we consider the complete feature vector Figure 5.6(c) and (d) which include the Fourier descriptors whose initial magnitude are the smallest in the feature vector set. It is most interesting to note that the weightings for the Fourier descriptors are generally higher than those for the geometric and colour information features. Comparing the weights for Figure 5.6(c) and (d) it can be seen that in the case where we have assumed perfect forehead boundary extraction (d), the weights are even greater in favour of the set of Fourier descriptors.

The effect of noise and the benefit of robust feature extraction can be seen by comparing the system performance reported in Figure 5.9 (a) and (b). Both cases start with a classification error of approximately 30% and 47% for equal and variance weights respectively using geometric and colour features, plus the first Fourier descriptor from the forehead boundary. It can also be seen that in both cases, the variance of the confusion matrix is higher if variance weights are used instead of equal weights. Examining Figure 5.9 (c), the classification error using variance weights increases up to a peak of 81% when at the 40th harmonic. This is in contrast to the result achieve when perfect feature extraction is used for Figure 5.9 (d). The classification error now rapidly decreases to zero as the similarity measure becomes dominated by the high weights for the Fourier descriptors. Indeed, it would appear that our variance weightings has enables us to identify the principal features for discrimination. The original feature vector was 90 dimensional, 55 harmonics for the Fourier descriptors and 35 for the remanding features, but the same result could have been obtained using the just the first eight Fourier descriptors. Our method thus serves as a powerful principal feature selector.

5.1.4 Effect of noisy feature extraction on recognition

Perfectly robust feature extraction on real face images is desirable but usually too difficult to achieve in practice. The effect of noisy feature extraction is to degrade the system recognition rate. To avoid system degradation, features should extracted such that the within-class distances are smaller than the between-class distances. In this section we simulate the effect of noise on system recognition by perturbing the set of feature values

The geometric face measurements of 18 different people were automatically extracted using our feature extraction methods. To simulate the effect of noise extra within classes were generated with increasing amounts of noise. A portion of the feature vector matrix constructed by adding noise to create a class size of five within class measures is shown for individual 000_1 with noise level equal up to ± 5 pixels of the extracted measure.

	000_1_1	000_1_2	000_1_3	000_1_4	000_1_5	009_2_1	009_2_5
headHeight	211	214	216	219	221	204	214
headwidth	192	195	197	200	202	192	202
noselength	65	68	70	73	75	72	82
nosewidth	29	32	34	37	39	35	45
chinwidth	192	195	197	200	202	192	202
eyeaxistomouth	119	122	124	127	129	119	129
intereyebrowdist	34	37	39	42	44	19	29
mouthwidth	72	75	77	80	82	89	99
mouthdepth	5	8	10	13	15	0	10
mouthheight	8	11	13	16	18	17	27
sclerawidthl	12	15	17	20	22	14	24

sclerawidthr	42	45	47	50	52	41	51
scleraheightl	11	14	16	19	21	13	23
scleradepthl	23	26	28	31	33	23	33
eyetohairl	114	117	119	122	124	188	198
eyetohairr	111	114	116	119	121	175	185
eyebrowwidthl	73	76	78	81	83	86	96
eyebrowwidthr	66	69	71	74	76	73	83
eyebrowdepthl	7	10	12	15	17	13	23
eyetocheekl	47	50	52	55	57	58	68
eyetocheekr	54	57	59	62	64	50	60
mouthtojawl	58	61	63	66	68	54	64

Figure 5.7 *Noise added extracted to measures to create a class size of five.*

For a given set of measures, the within class distribution will overlap the between class distribution with increasing noise, resulting in impaired system performance. Figure 5.8 shows the relationship between additive noise and classification error. With up to six pixels error the equally weighted coefficient similarity measures are unable to detect any classification errors. The variance weighted coefficient similarity measures method is able to detect the effect of noise at a lower level of about four pixels. Thus variance weighting has enabled the system to be more sensitive to differences in the distance measures which may be due to a different face being presented to a recognition system. The important factor is that again the variance of the confusion matrix is again greater when variance base coefficient weightings are employed. This result strongly suggests that in the absence of noise, or with sufficiently robust and accurate feature extraction, the variation based coefficient similarity measures are the most useful.

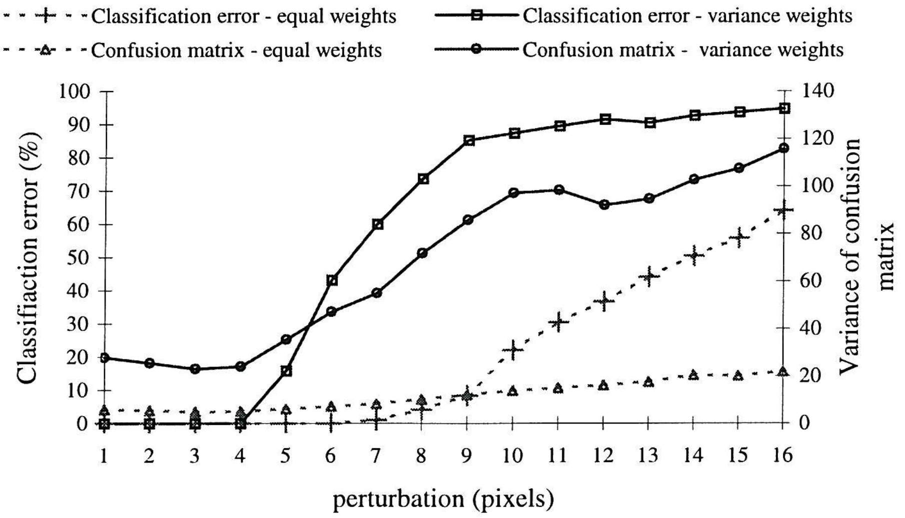


Figure 5.8 *Effect of perturbing face distance measurements on system performance.*

The simulated effect of feature perturbation was useful not only to determine how much measurement can be tolerated, but also to serve as a benchmark for classification results that may be attained using automatic feature extraction system.

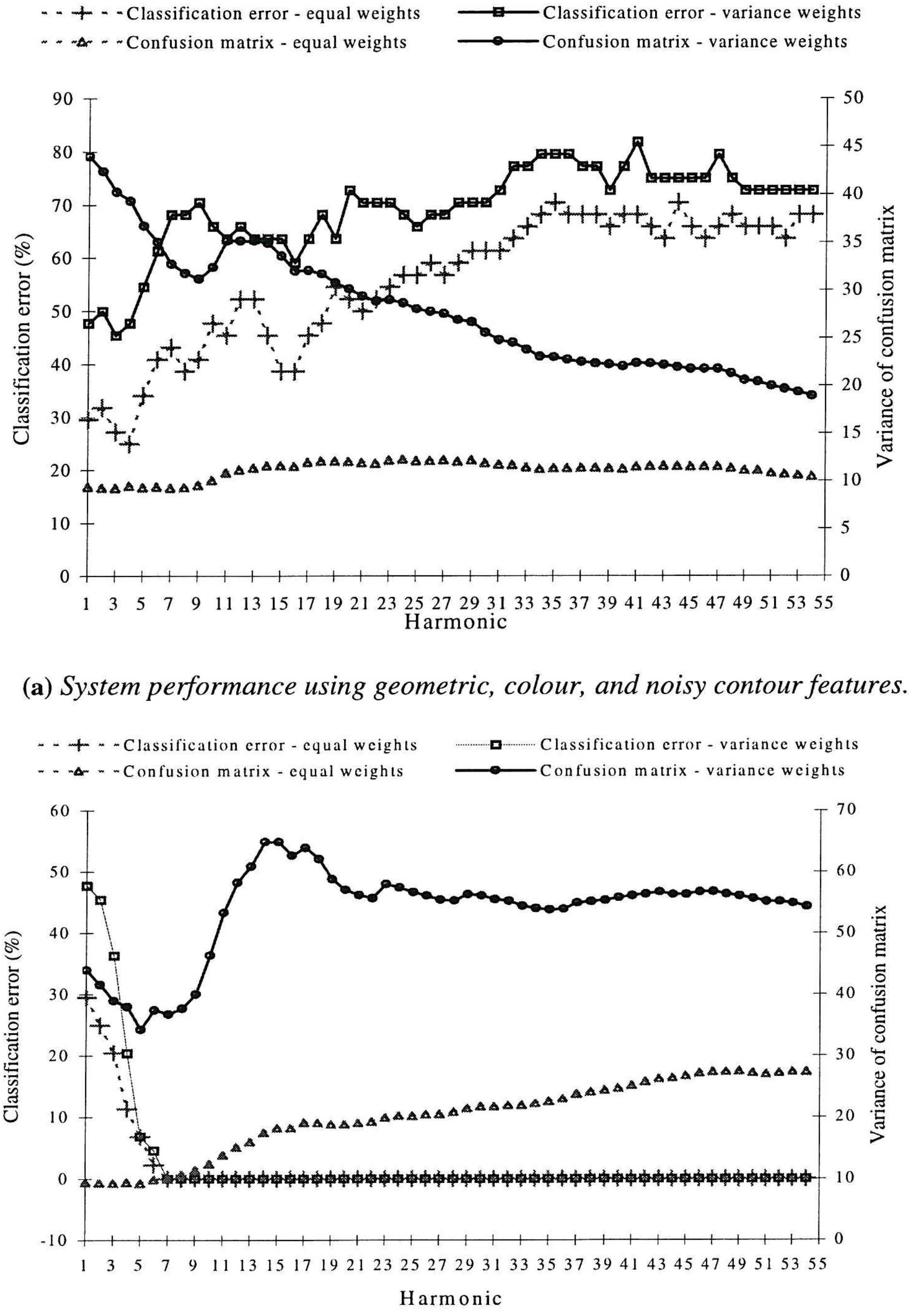
5.2 Conclusions.

We have presented a useful method of combining features of various magnitudes from different sources. The approach consists of determining coefficient weights which optimise a distance based similarity measure. The weights are computed from the natural variance in the feature vector matrix, yet are insensitive to the relative scale of the components of the feature vector matrix. We have used scalar quantities such as the Euclidean distance between landmark face measures to show the increase variance obtained using the weighted coefficient. The principle can be extended for use on vector quantities, if the vector can be converted to a scalar quantity.

We have simulated the effect of extraction noise on system performance and determined that an error of 5 pixels on geometric measure may reduce the system recognition rate to 84%. Using automatically extracted features we achieve a recognition rate of 72% which is lower than the success rates achieve by Brunelli and Poggio [7] and Lam and Yan [51] who achieved 90% and 96% recognition rate respectively. Our reduced success rate may be due to the more realistic image capture conditions in which the images were captured over a period of several days. Even using manual coding of the database the success rates of Brunelli and Poggio and Lam and Yam could not be achieved. If it is assumed that manual extraction could be acquired to within 5 pixels, then the reduced recognition rate may be attributed changes of facial expression, hair styles or a facial hair.

The approach has been tested on features sets originating from face geometric facial measurement combined with the Fourier descriptors from forehead contours. The results showed an in increase in the variance of the confusion matrix if our method for calculating the coefficients were used compared to using equally weighted coefficients. From the distribution of calculated coefficients we note that the model of the eye [88] is amongst the least discriminatory feature sets, whereas the eyebrow's area were amongst the most discriminatory geometric features. The coefficients for Fourier descriptors were the most discriminatory feature within the 90 dimensional feature vector. They were significantly higher than the geometric and colour features, despite being smaller in magnitude. Actually, 100% recognition could be achieved using just eight robustly extracted Fourier descriptors. Our method thus identifies the principal features for face recognition which is a useful information considering the expensive computational costs of feature extraction.

The investigation into the variance coding of face feature measures is at an introductory stage only, rather than an established one. It would certainly appear that using the variance of face features can be used to control their contribution to a recognition metric. However, some areas require further study, such as the effect of noise on the measures, the accuracy of the estimates of variance and the accuracy of the measures themselves. These effects combine, in a practical face recognition system, to control the maximum achievable recognition rate. As such however, the full extent of the contribution of these factors awaits further research.



(a) *System performance using geometric, colour, and noisy contour features.*

(b) *System performance using geometric, colour, and perfect contour features.*

Figure 5.9 *System performance using noisy and perfect feature extraction.*

6. Conclusions and Further Work

This thesis has developed new ways for face feature extraction within a model-based recognition scenario. More detailed conclusions can be found at the end of each chapter, however in broad summary we observed that the eyes represent regions of high concentricity. This has enable us to locate them using an evidence gathering process, designed to identify such regions. The success rate using concentricity was quite variable ranging from 50% on a database of a thousand faces to 84% using 88 faces from the M2VTS database. We have improved the model of the standard eye template by imposing rules on its construction, and thereby avoiding the use of explicit internal energy terms. The original template optimisation used gradient information, which could result in the template being trapped in a local energy minima at the eyebrows. We also experienced difficulty in the determining suitable termination criteria for the iterative gradient based approach. The problem of eluding local minima and specifying termination criteria for gradient based optimisation were both solved using genetic algorithms. Although more computationally expensive, they proved more suited to optimisation of the eye template, using populations of stochastic solutions to evade local minima. Applying our improved deformable eye template to the top ten peaks of concentricity, increased the success rates to 91% and 93% on the one thousand face and 88 faces from the XM2VTS database.

We have presented a new method for combining features which may be different in terms of their magnitude and/or source. The method involves using the feature vector matrix to derive coefficients for an Euclidean based similarity measure. These coefficients, calculated from a normalised feature vector represent the intrinsic variance of a feature. The benefits of this relatively inexpensive pre-processing was an net increased variance in the classification matrix and a variance profile indicating the discriminating capability of each feature. The variance profile may be used to determine which features should be extracted make the most efficient use of computational effort. In this respect, the parameters of the deformable eye temple were very expensive for their return in terms of discriminatory power. The concentric algorithm could achieve a success rate of 84% success in seconds using 88 faces from the XM2VTS database. Application of our improved deformable template improved the eye location rate from 84% to 93% but added several minutes to the eye extraction process for a mere total 2% weighting in the Similarity measure. Clearly, robust feature extraction is crucial for a face recognition system based of features is to be practicable. The features with the highest coefficients should be extracted with the highest confidence and robustness to avoid classification rates dominated by noisy feature extraction.

6.1 Further Work

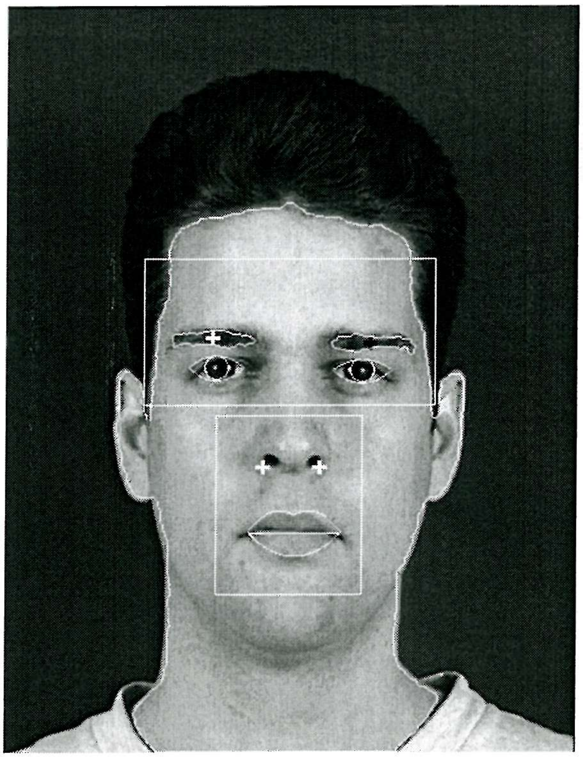
Although the eye template parameters offer little as a discriminator, they are a key initialiser in the search for other facial features. We have shown that, without strong initialisation, the results from the dynamic programming solution to head boundary extraction may vary with the background intensity, which is a most undesirable effect. We were able to extract the skin boundary, using the eyes as an initialiser, but even so, it was very difficult to robustly extract the chin boundary and hence measure the head height. The variance profile for manually extracted features suggests that the computational overhead for dynamic programming may not merit the extraction of this feature. On the other hand, the chin contour is potentially a very useful feature because it is a part of the facial bone structure. As such, it is less easily altered, unlike the forehead contour. In addition we have seen that Fourier descriptors of the contours provide more discriminating capacity that geometric measures, so perhaps perseverance should be extended to extract this elusive feature.

Images for the XM2VTS database were obtained over a number of months. The time interval between successive images of the same person allowed changes in illumination, clothing, hairstyles, facial expression, facial hair, orientation to the camera, any combination of which could impair classification results. Where possible, these effect may need to be incorporated into the recognition process.

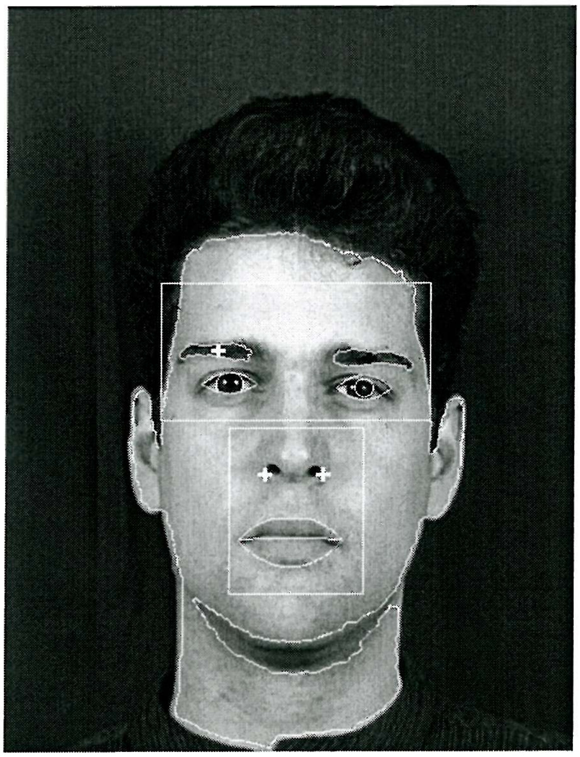
We have suggested that the face features could be found by assuming that they are located at holes in the skin boundary. We have applied simple thresholding with some empirically derived constants to segment these features. If we assume the number of features are known, there may be some benefit in applying clustering techniques rather than simple thresholding to locate face organs. The eyebrows nose mouth etc. were extracted by looking at holes in skin rather than intensity variations. If a model such as a deformable eye template is used there will be an error between the a point on the contour and the closest point on a simple geometric parameter model. Our variance profile shows that there is substantial discriminatory capability in the Fourier descriptors of boundaries. It would be interesting to compare the classification performance of a system of face features extracted as contour models rather than their approximation by geometric models.

7. Appendix

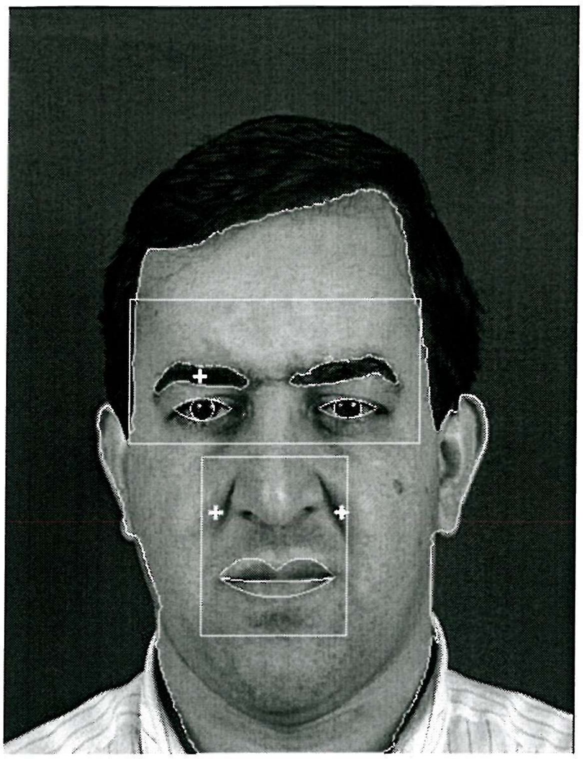
7.1 Extracted features



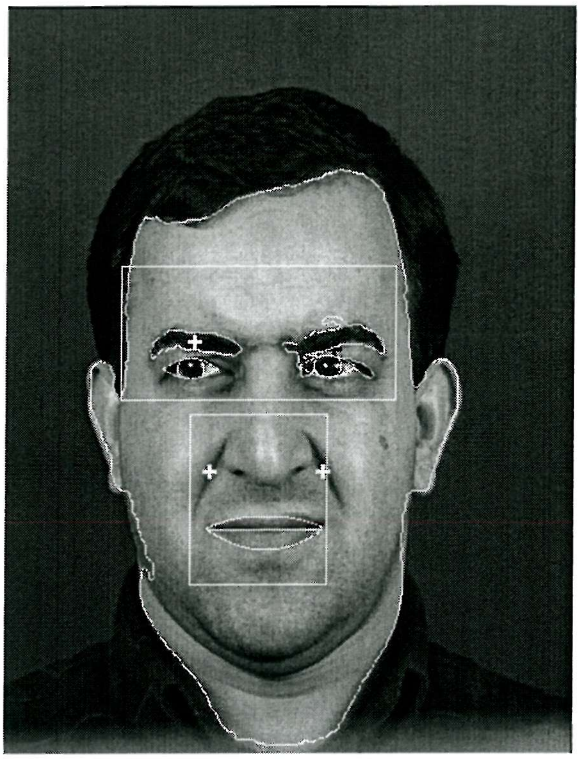
000_1_1



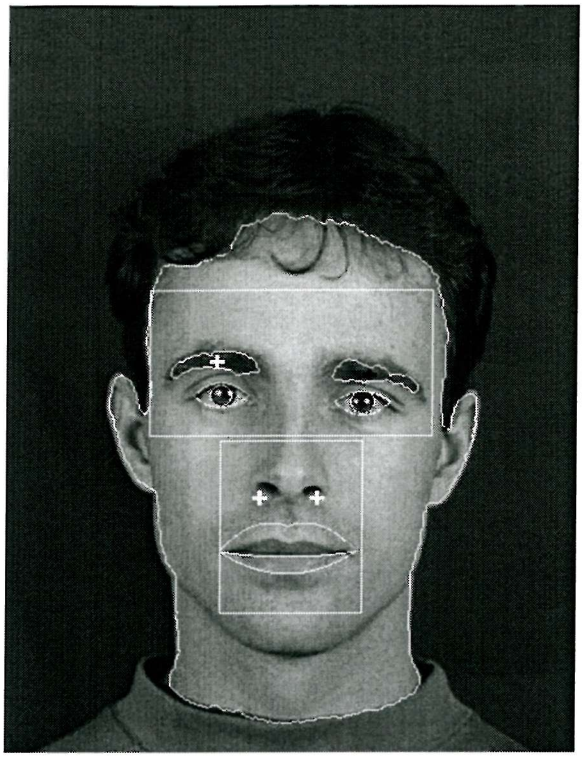
000_3_1



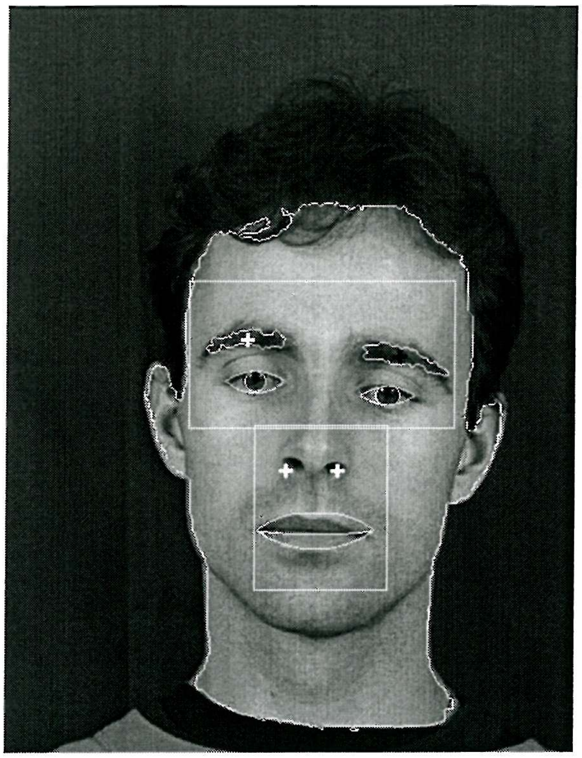
003_1_1



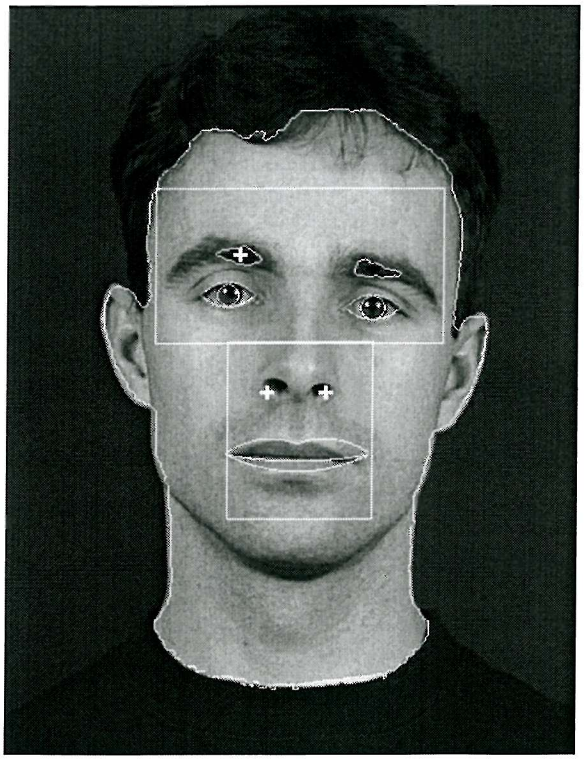
003_4_1



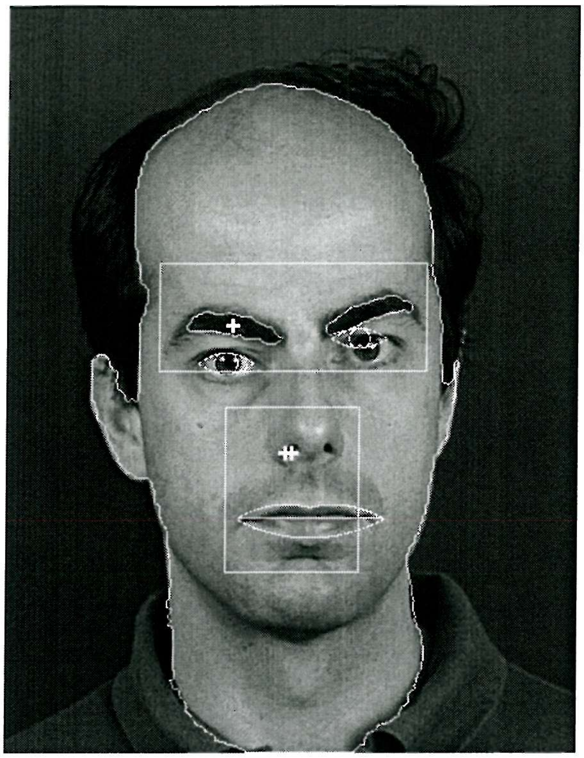
004_2_1



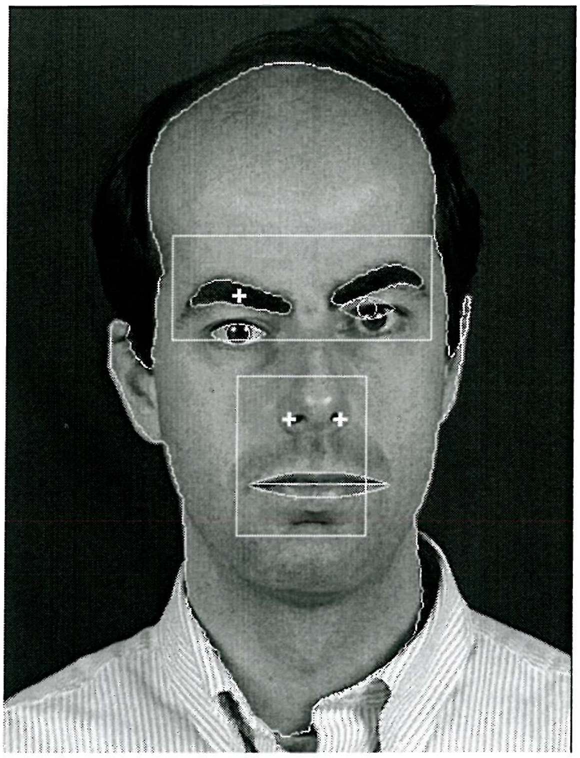
004_3_1



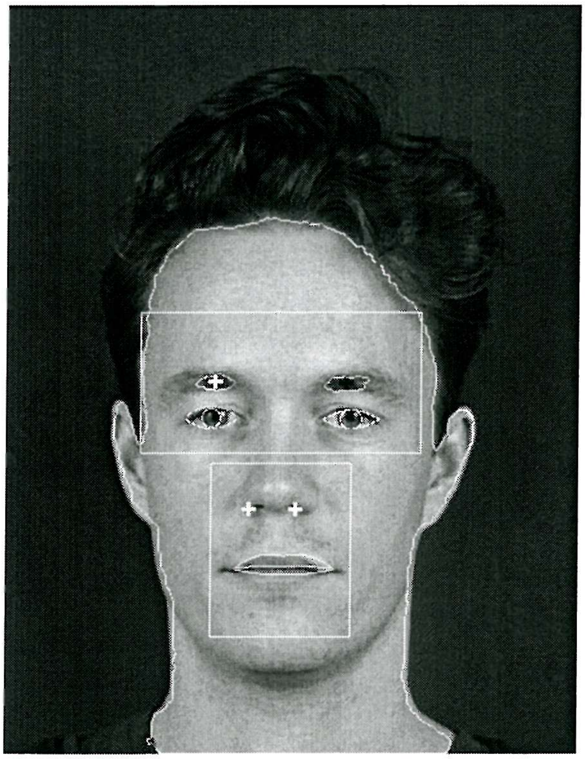
004_4_1



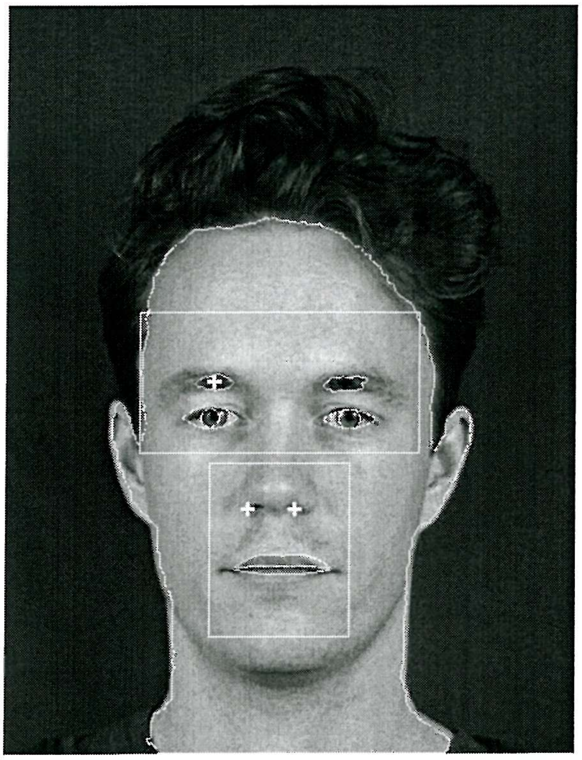
009_4_1



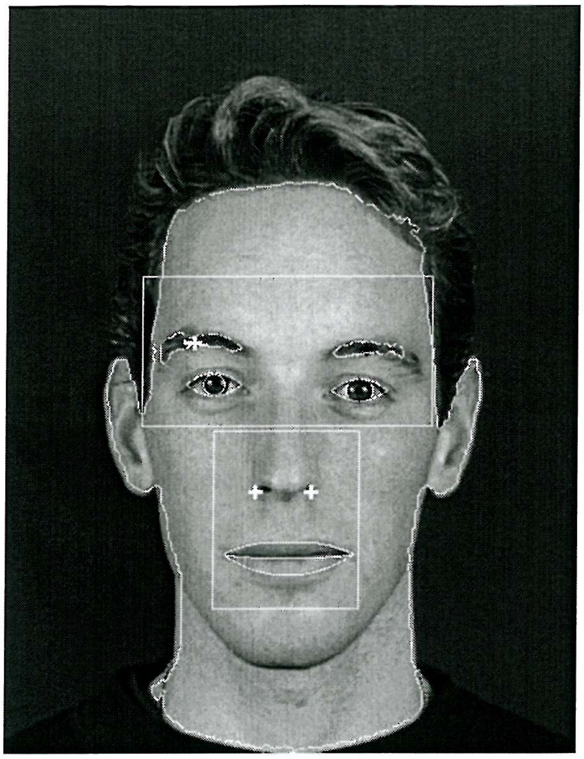
009_2_1



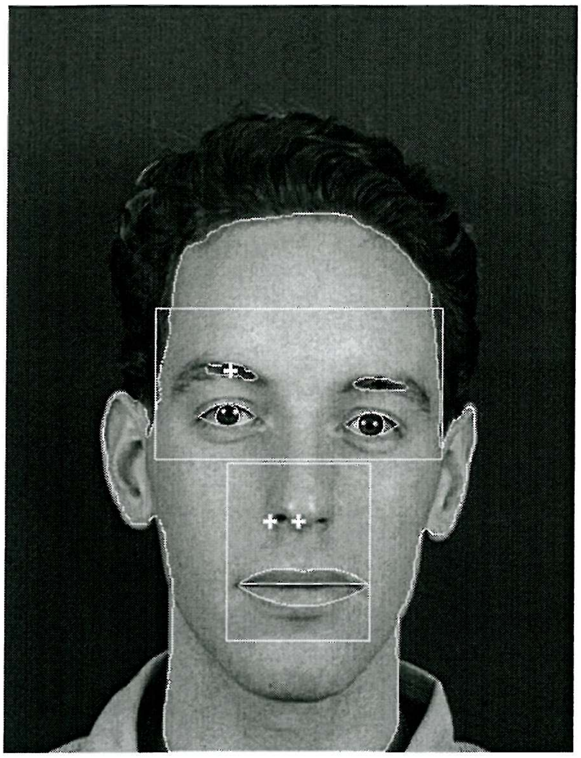
017_4_1



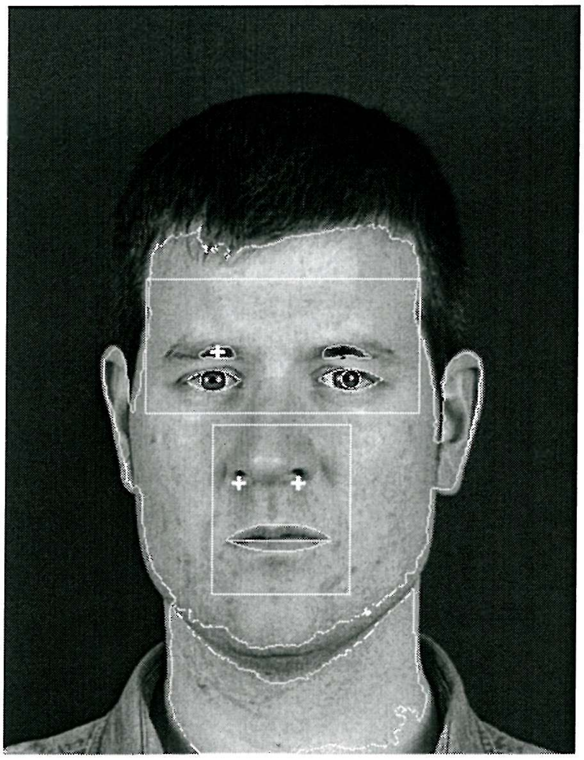
017_1_1



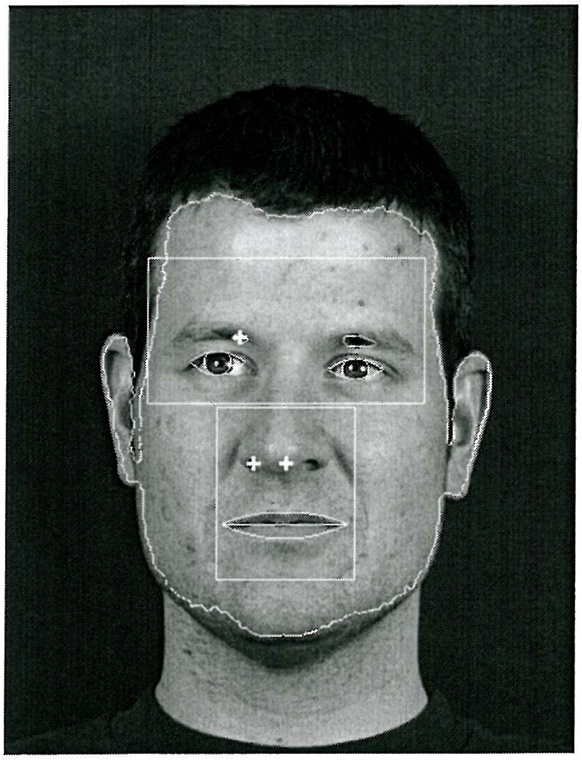
021_1_1



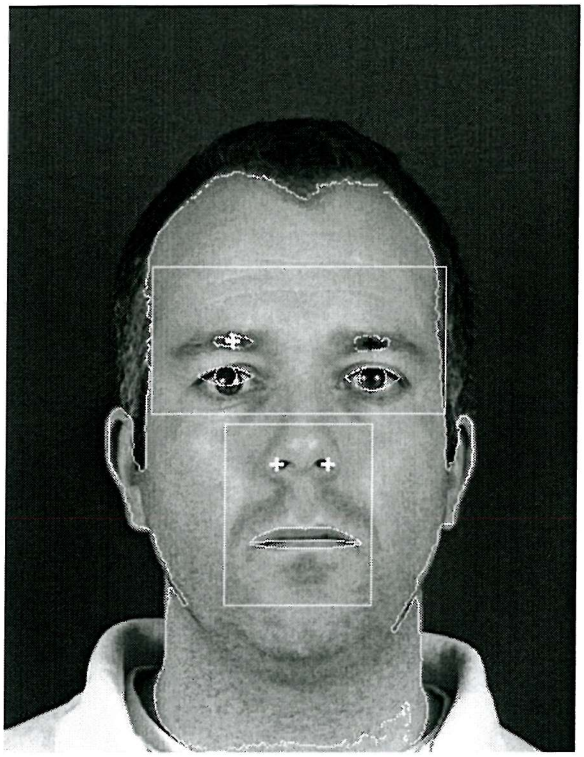
021_4_1



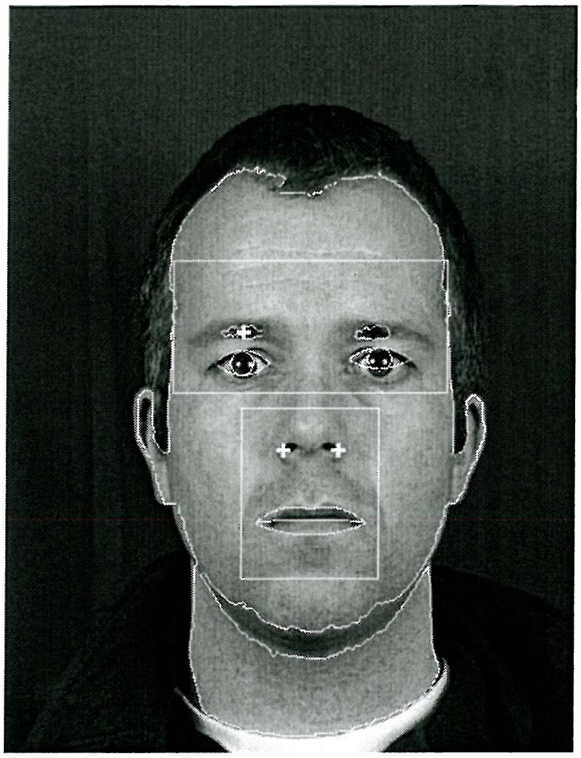
030_1_1



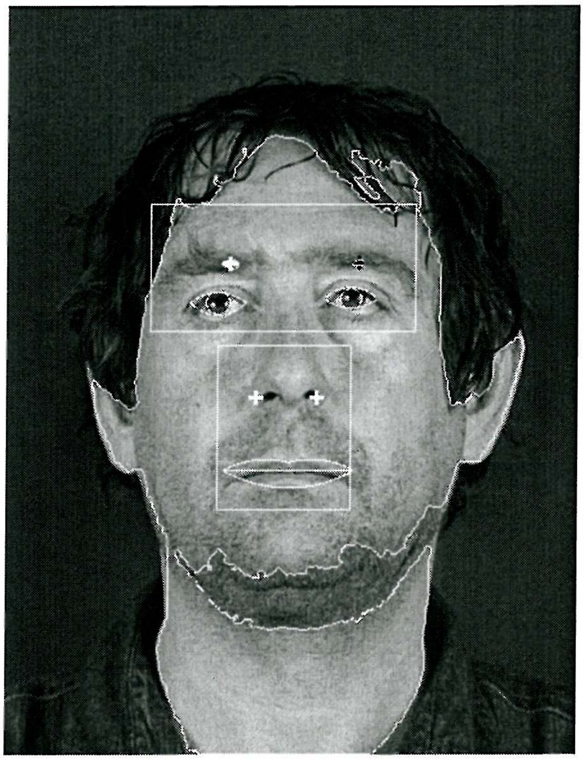
030_4_1



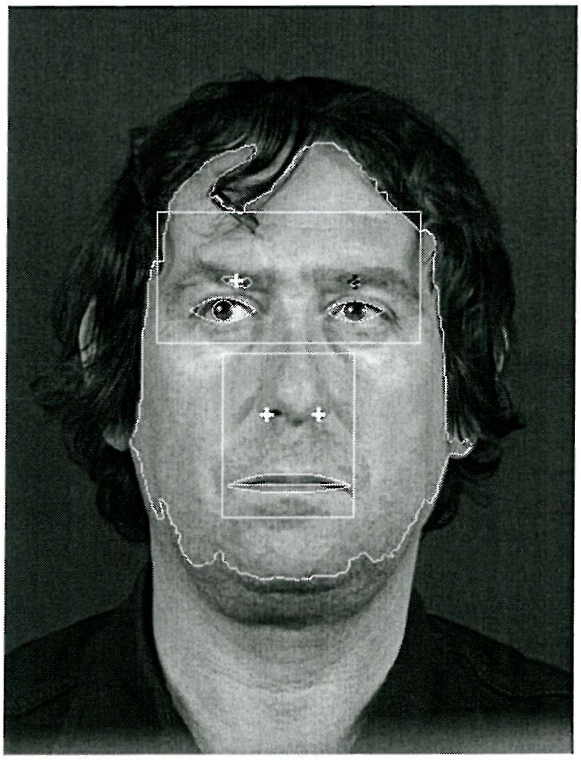
032_1_1



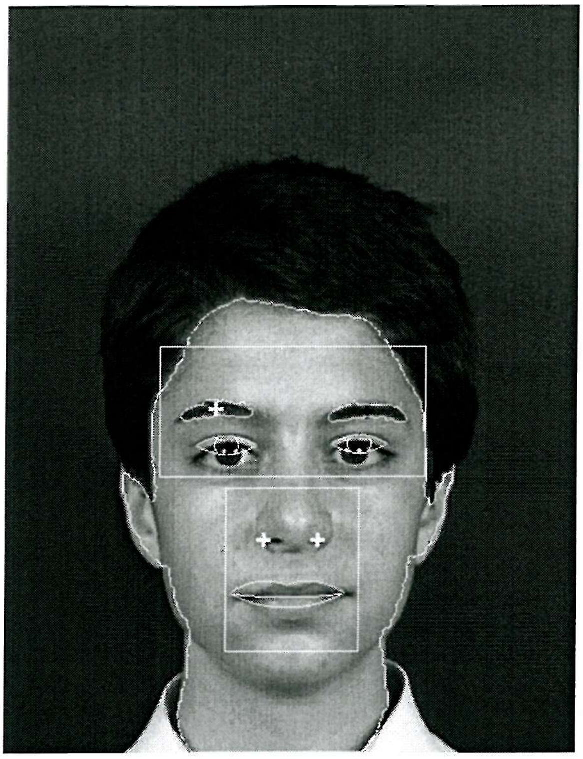
032_3_1



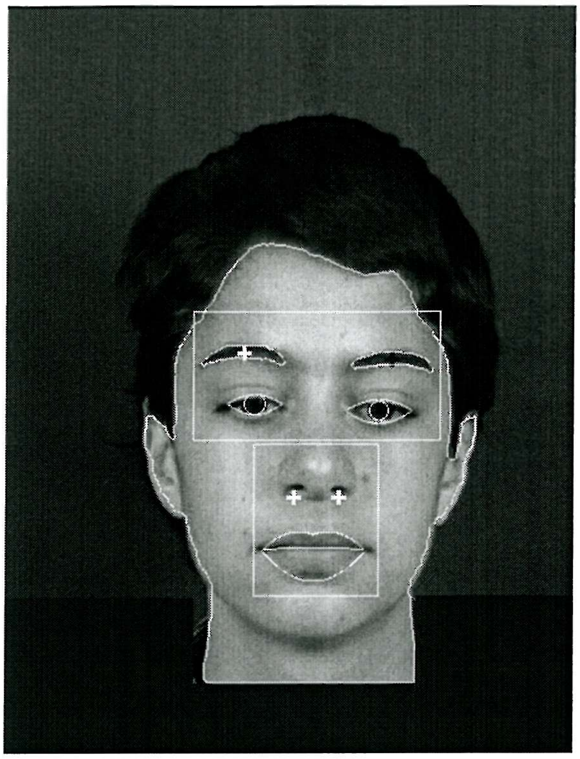
051_1_1



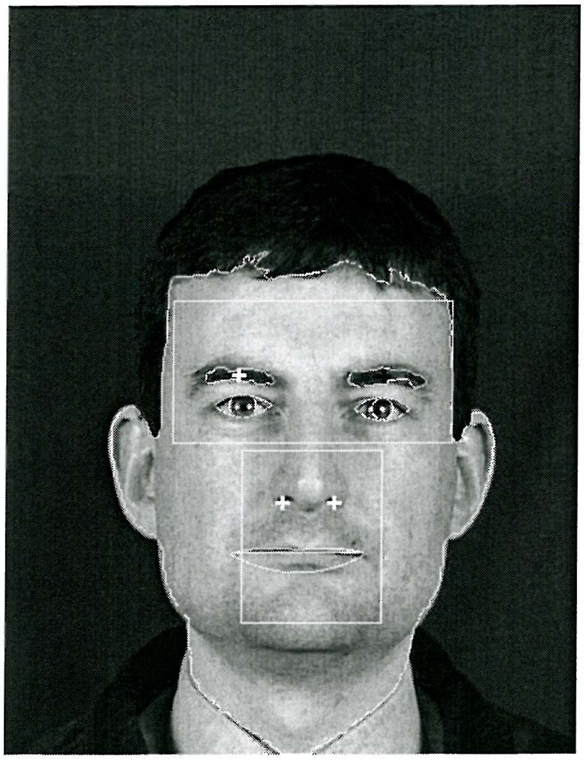
051_4_1



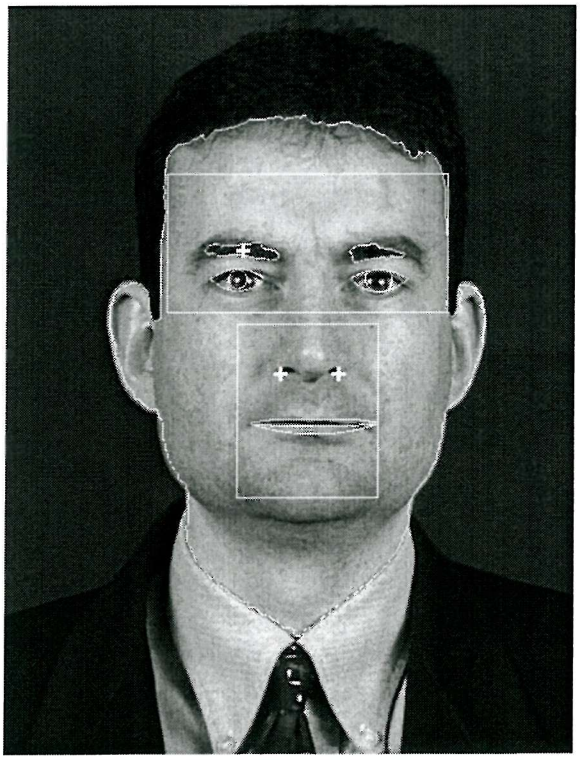
068_3_1



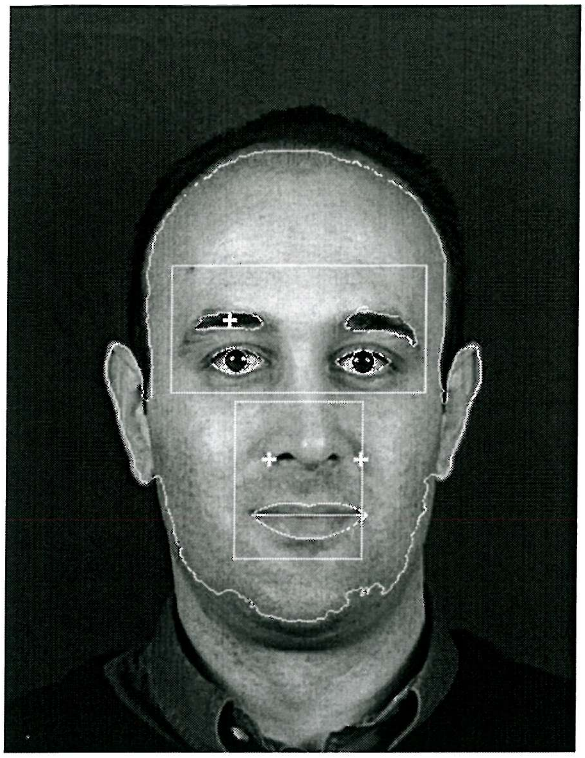
068_4_1



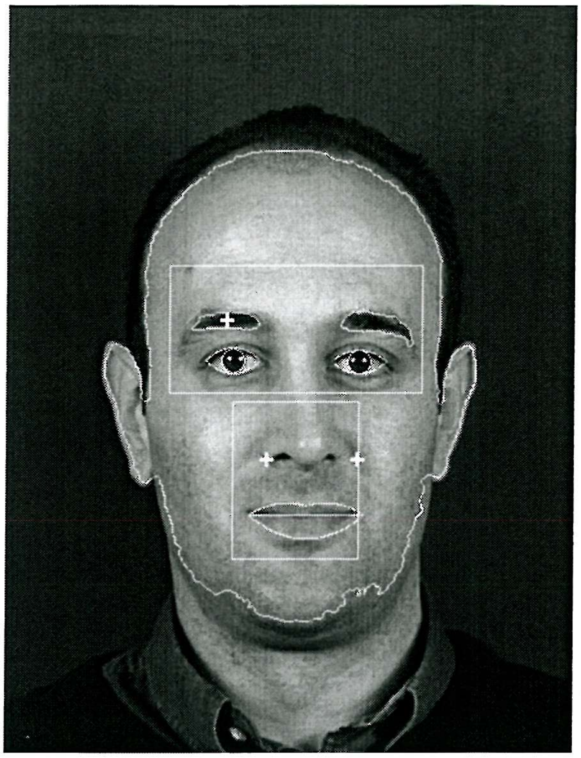
074_1_1



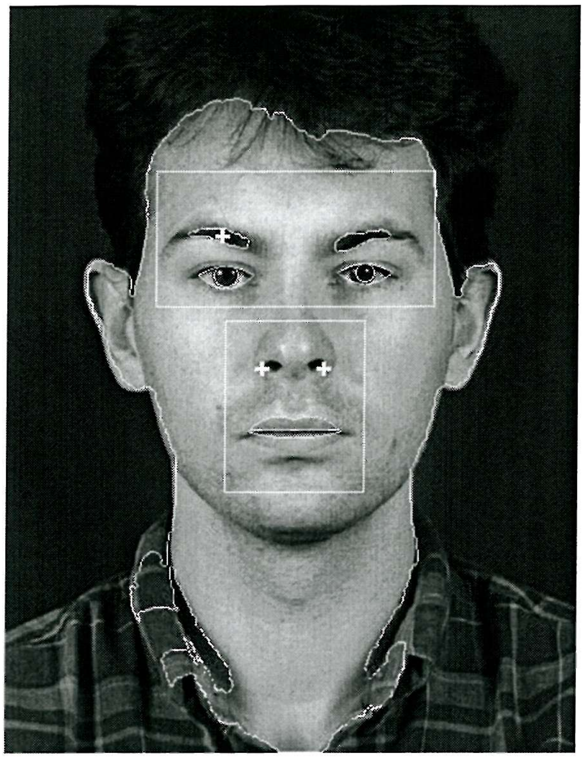
074_4_1



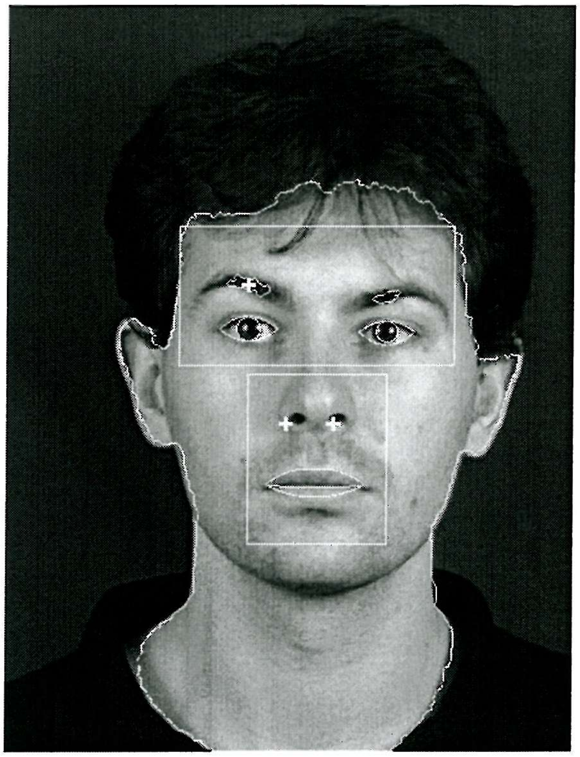
105_1_1



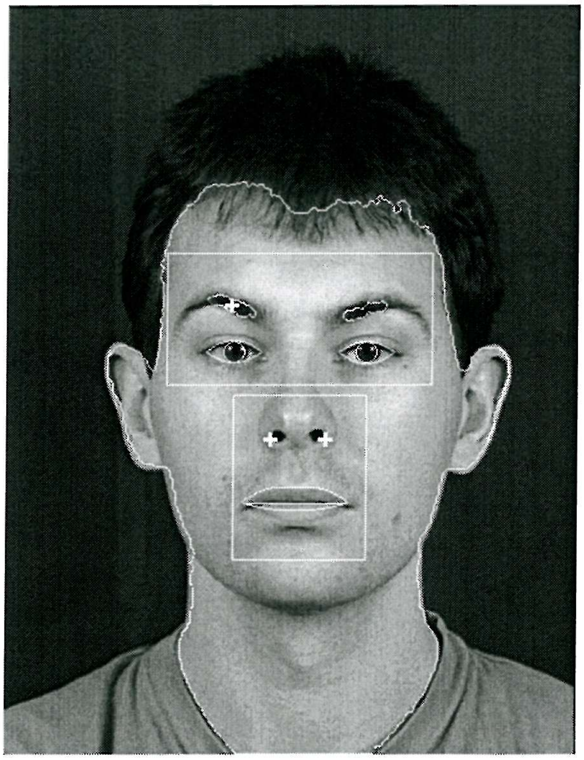
105_4_1



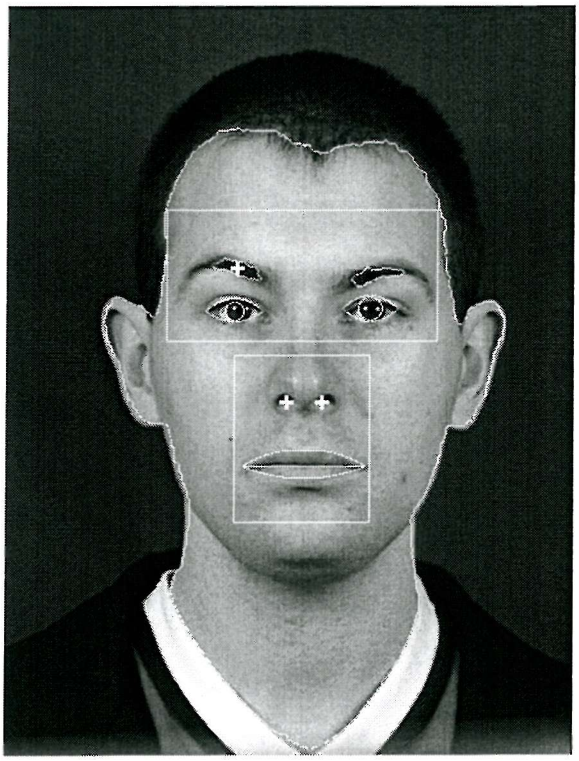
131_1_1



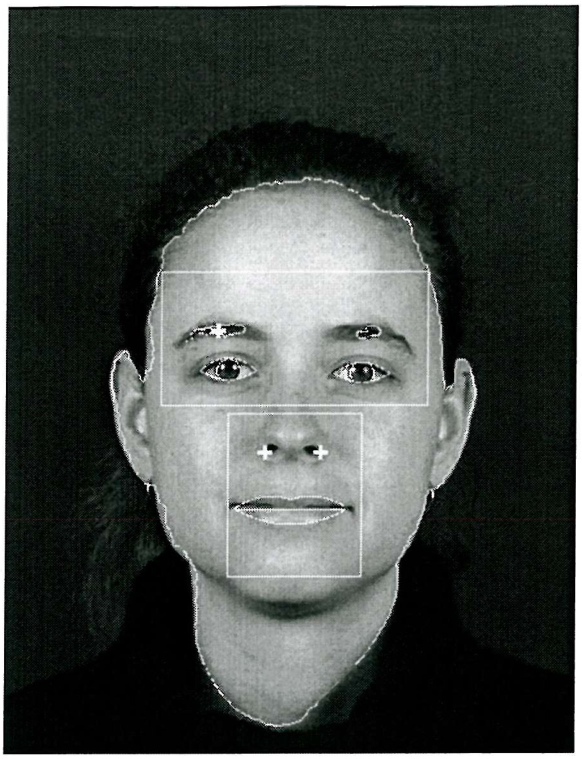
131_2_1



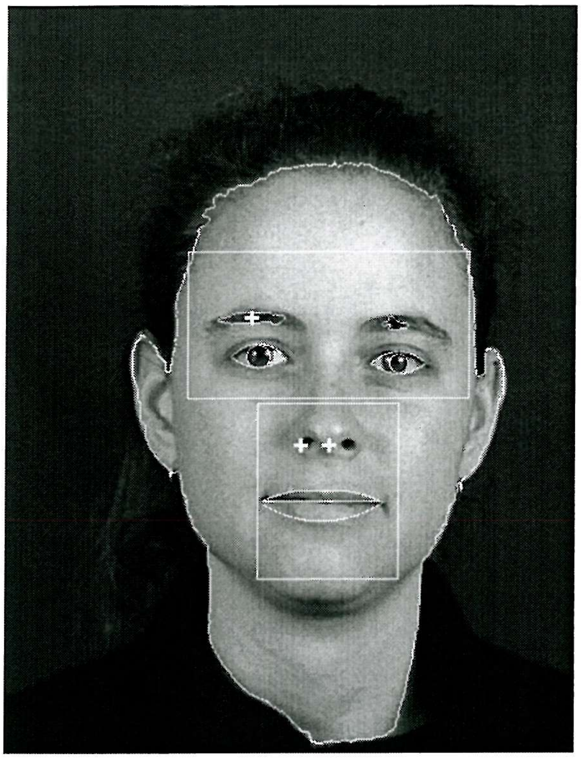
131_3_1



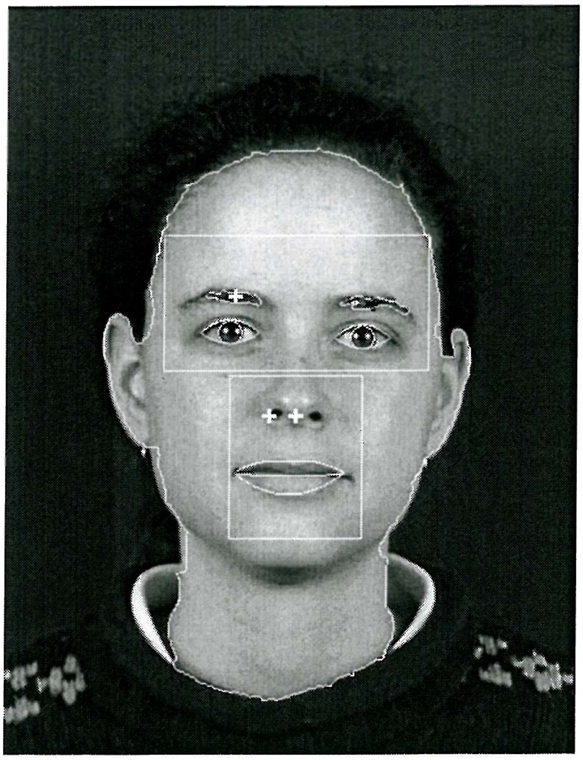
131_4_1



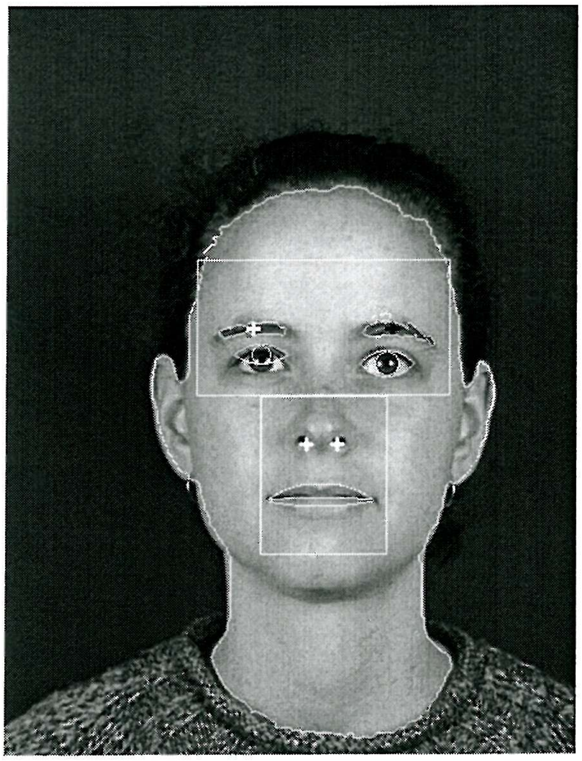
188_1_1



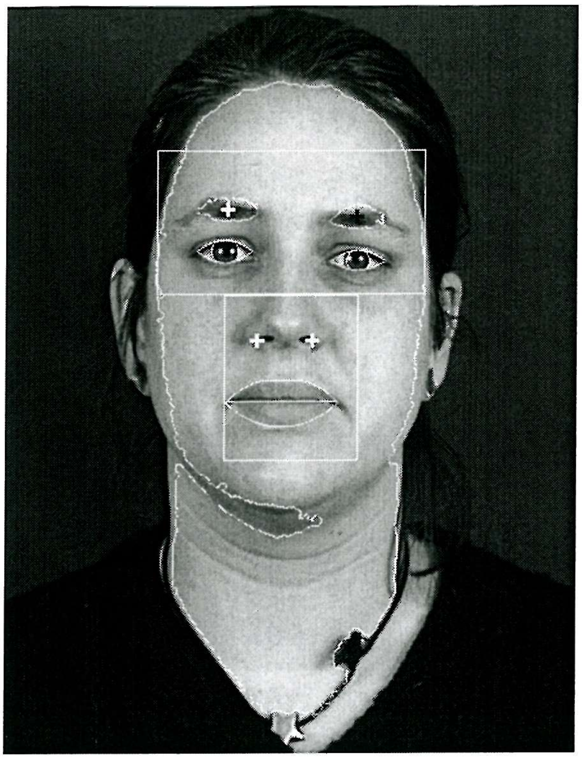
188_2_1



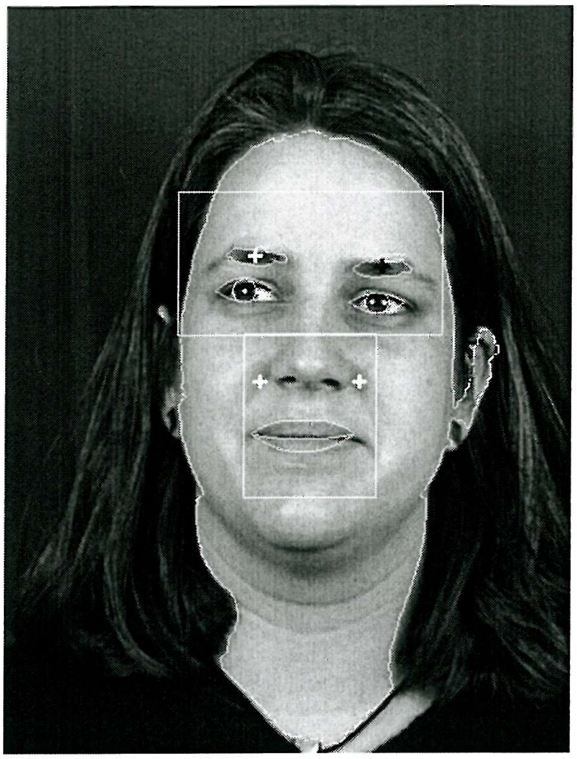
188_3_1



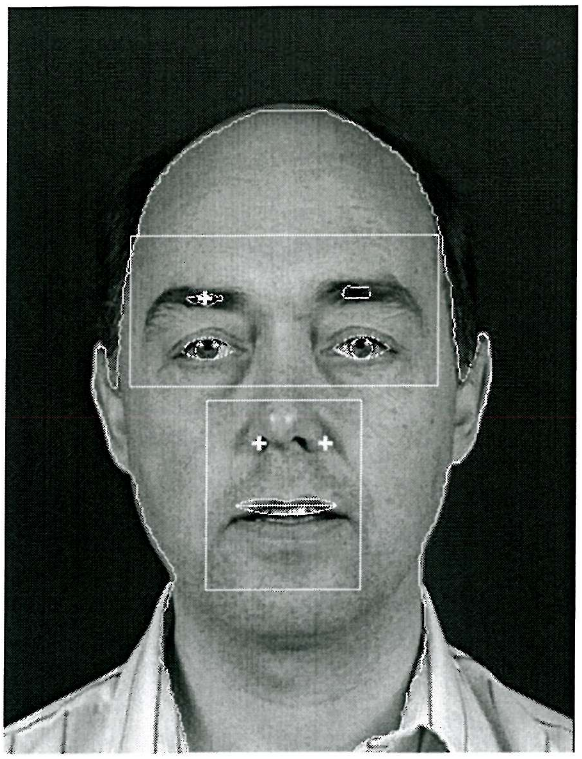
188_4_1



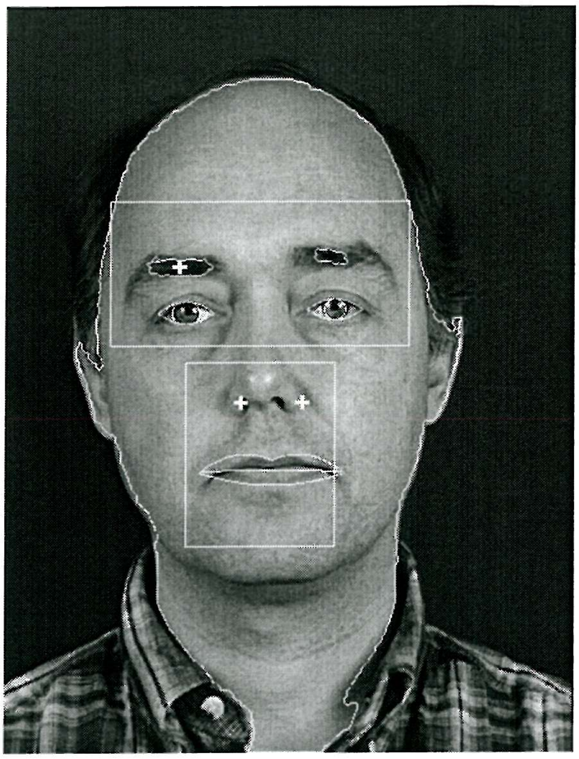
202_2_1



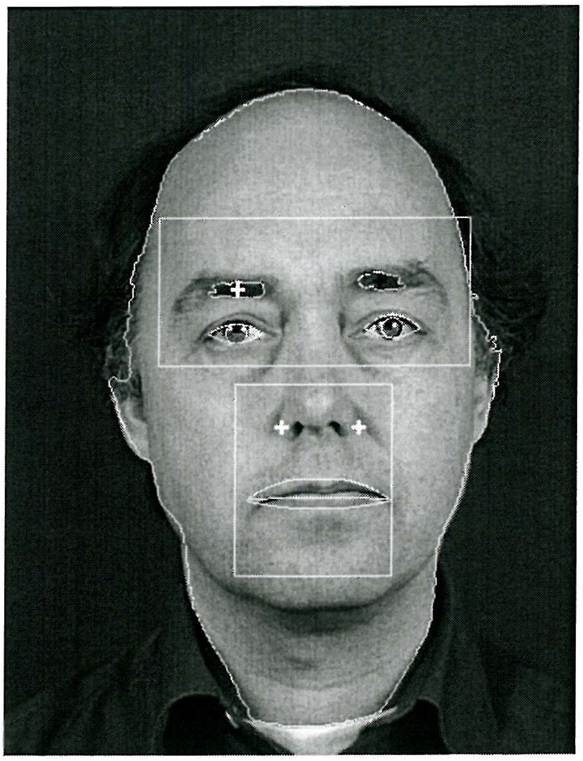
202_3_1



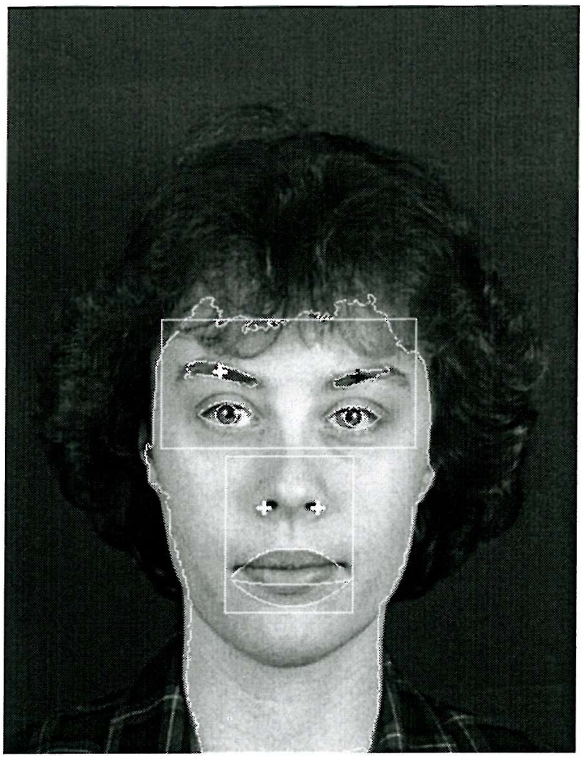
212_1_1



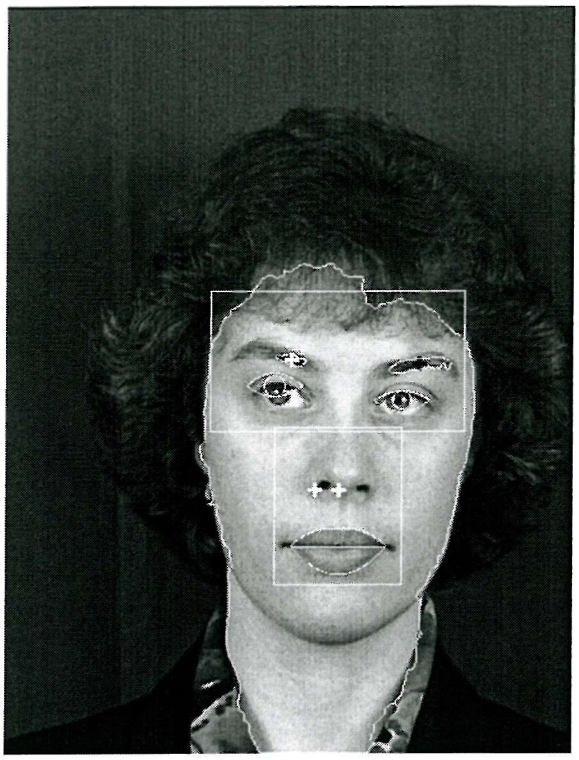
212_2_1



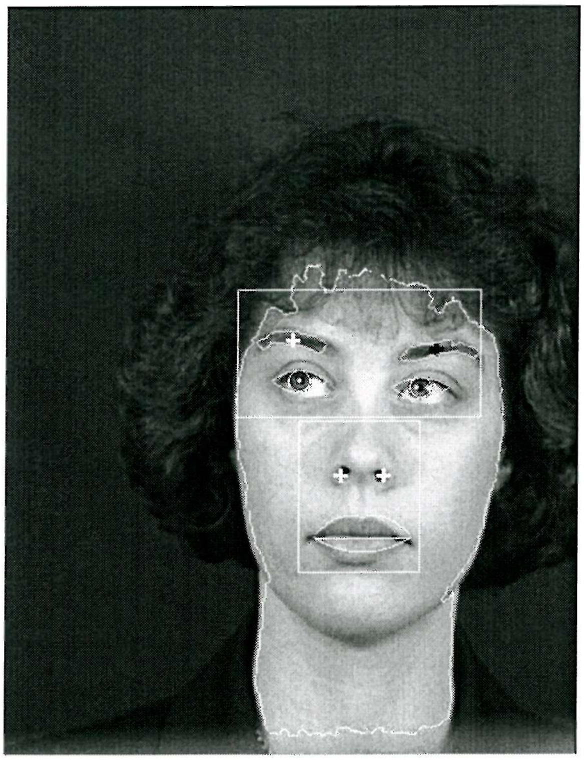
212_4_1



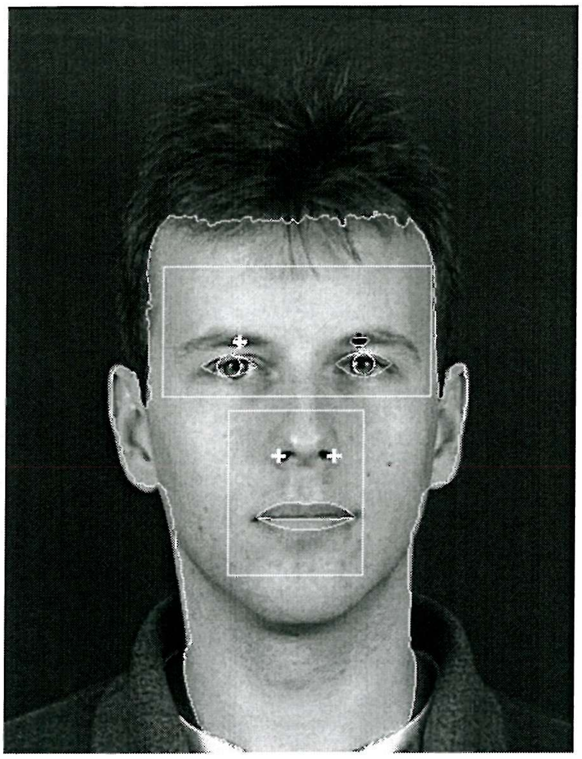
279_2_1



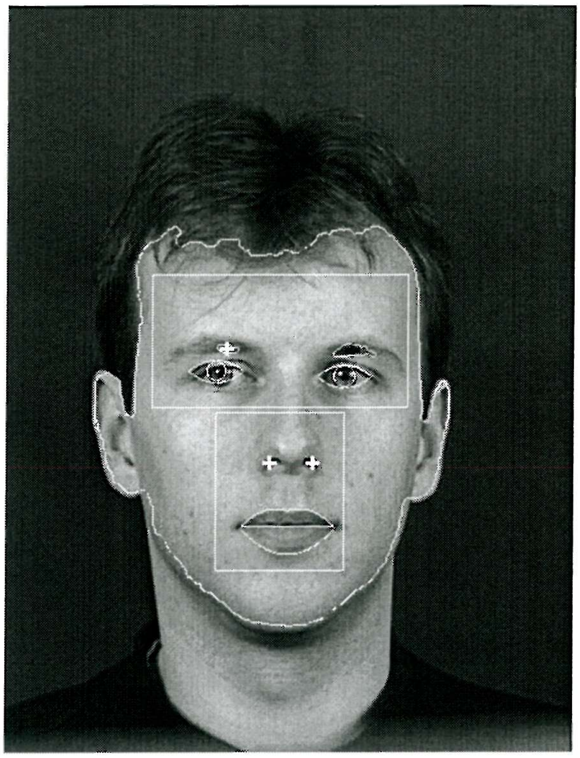
279_3_1



279_4_1



285_1_1



285_4_1

8. References

[1] A. A. Amini, T. E. Weymouth and R. C. Jain, Using Dynamic Programming for Solving Variational Problems in Vision. *IEEE Trans. PAMI.*, **12**(9) pp.855-867, (1990).

[2] A. S. Aguado, M. E. Montiel and M. S. Nixon, On Using Directional Information for Parameter Space Decomposition in Ellipse Detection. *Patt. Recog.*, **29**(3) pp. 369-381, (1996).

[3] D. E. Benn, Eye Location using the Standard Hough Transform and Circular Correlation. *Southampton University Interim Report* , (March 1996).

[4] D. E. Benn, M. S. Nixon and J. N. Carter, Robust Eye Centre Extraction Using the Hough Transform, *Lecture Notes in Computer Science*, **1206** pp. 3-9, Audio and Video-based Biometric Person Authentication. (AVBPA) (1997).

[5] H. Bourlard and Y. Kamp, Auto-association by multi-layer perceptrons and singular value decomposition, *Biological Cybern.*, **59** pp. 291-294, (1988).

[6] R. Brunelli and D. Falavigna, Person Identification using Multiple Cues. *IEEE Trans. PAMI.*, **17** (10) pp. 955-966, (1995).

[7] R. Brunelli and T. Poggio, Face Recognition: Features Versus Templates. *IEEE Trans. PAMI.*, **15** (10) pp. 1042-1052, (1993).

[8] R. Brunelli and T. Poggio, Template Matching: Spatial Filters and Beyond. *Pattern Recognition*, **30** (5) pp. 751-768, (1997).

[9] J. L. Blue, G. T. Candela, P. R. Grother, R. Chellappa and C. L.Wilson, Evaluation of Patter Classifiers for Fingerprint and OCR applications. *Pattern Recognition*, **27**(4):485-501, (1994).

[10] C. M. Brown, Inherent Bias and Noise in the Hough Transform. *IEEE Trans. PAMI.*, **5**(5) pp. 493-505, (1983).

[11] D. K. Burton, Text-dependent Speaker Verification using Vector Quantization Source Coding. *IEEE Transactions on Acoustics, Speech and Signal Processing* **35**(2) pp. 133, (1987).

[12] J. Canny, A Computational Approach to Edge Detection, *IEEE Tran. on PAMI* **8**(6) pp. 1208-1216, (1993).

[13] R. Chellappa, C. Wilson and S. Sirohey, Human and Machine Recognition of Faces: A survey. *Proc. of the IEEE* **83** (5) pp. 704-741, (1995).

[14] G. Chow and X. Li, Towards a System for Automatic Facial Feature Detection, *Patt. Recog.* **26**(12) pp. 1739-1755, (1993).

[15] G. W. Cottrell and M. Fleming, Face Recognition using unsupervised feature extraction, in *Proc. Int Neural Network Conf.* **1**(3) pp. 322-325, (1993).

[16] I. Craw, H. Ellis and J. R. Lishman, Automatic Extraction of Face-Fetaures. *Pattern Recognition Letters* **5** pp. 183-187, (1987).

[17] I. Craw, D. Tock and A. Bennett, Finding Face Features in *Proc. 2nd Europe Conf. On Computer Vision* pp. 92-96, (1992).

[18] D. Cunado, M. S. Nixon and J. N. Carter, Using Gait as a Biometric, via Phase-Weighted Magnitude, *Lecture Notes in Computer Science*, **1206** pp. 95-102, AVBPA 1997.

[19] I. Daubechies. The Wavelet Transform, Time Frequency Localisation and Signal Analysis, *IEEE Trans. On Information Theory*, **36**(5) pp. 961-1004, (1990).

[20] E. R. Davies, A Modified Hough Scheme for General Circle Location. *Pattern Recognition Letters*, **7** (1) pp. 37-44, (1988).

[21] S. R. Deans, Hough Transform from the Radon Transform, *IEEE Trans. PAMI.-3*, pp.185-188, (1981).

[22] R. O. Duda and P. E. Hart, Use of the Hough Transform to Detect Lines and Circles in Pictures. *Communications of the ACM* **15**(1) pp. 11-15, (1972).

[23] R. A. Fisher, The use of Multiple Measures in Taxonomic Problems, *Ann. Eugenics* **7**, pp.179-188, (1936).

[24] H. Freeman, On the Encoding of Arbitrary Geometric Configurations. *IRE Trans* **EC-10**(2) pp. 260-261, (1961).

[25] C. R. Giardina and F. Kuhl, Accuracy of curve approximation by harmonically related vectors with elliptical loci. *Computer Graphics and Image Processing* **6** pp 277-285, (1977).

[26] D. A. Goldberg, *Genetic Algorithms in Search Optimization and Machine Leaning,* Addison-Wesley, (1989)

[27] S. R. Gunn and M. S. Nixon, Snake Head Boundary Extraction Using Global and Local Energy Minimisation. *Proc. 13 ICPR*, **2**, pp. 581-585, (1996).

[28] G. Gerig and F. Klein, Fast Contour Identification Through Efficient Hough Transform and Simplified Interpretation Strategy. *8th International; Joint Conf Pattern Recognition*, Paris France pp. 498-500, (1986).

[29] G. H. Granlund, Fourier pre-processing for hand print character recognition, *IEEE Trans. Computers* **C21** pp. 195-201, (1972).

[30] U. Grenander, Y. Chow and D. Keenan, *Hands: A Pattern Theoretic Study of Biological Shapes,* New York: Springer-Verlag, pp. 195-201, (1972).

[31] Z. Haung, Affine-Invariant B-Spline Moments for Curve Matching. *IEEE Trans. on Image Processing* **5**(10) pp. 1473-1480, (1996).

[32] Z. Hong, Algebraic Feature Extraction of Images for Recognition. *Pattern Recognition*, pp. 179-187, (1962).

[33] P. V. C. Hough, Method and Means for Recognising Complex Patterns, *U. S. Patent No. 3069654*, (1962).

[34] A. K. Jain, R. Bolle and S. Pankanti, *Biometrics Personal Identification in Networked Society,* Kluwer Academic Publishers (1999).

[35] A. K. Jain, Jia and D. Zongker, Feature Selection: Evaluation, Application, and Small Sample Performance, *IEEE Trans. PAMI.*, **19**(2) pp. 153-158, (1997).

[36] T. Kanade, Picture Processing System by Computer Complex and Recognition of Human Faces, Dept of Information Science Kyoto University, 1973.

[37] T. Kanade, Computer Recognition of Human Faces, Baseland Stuttgart: Birkhauser, (1977).

[38] Y. kaya and K. Kobayashi, *A Basic Study on Human Face Recognition, Frontiers of Pattern Recognition,* Academic Press, pp 265-290, 1972.

[39] M. S. Kamel, H. C. Shen, A. K. C. Wong, T. M. Hong and R. I. Campeanu. Face Recognition using Perspective Invariance Features. *Pattern Recognition Letters*, **15**, pp 877-883, 1994.

[40] J. Kittler Y. P. Li, J. Matas and M. U. Ramos Sanchez, Combining Evidence in Multimodal Personal Identity Recognition Systems, *Lecture Notes in Computer Science*, **1206** pp. 327-334, AVBPA 1997.

[41] F. P. Kuhl and C. R. Giardina, Elliptic Fourier Descriptors of a Closed Contour. *CVGIP* **18** pp. 236-258, (1990).

[42] X. Jia and M. S. Nixon, Extending the Feature Vector for Automatic Face Recognition. *IEEE Trans. PAMI.*, **17** (12) pp. 1167-1176, (1995).

[43] M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active Contour Model, *Int. J. Comp. Vision* **8** pp. 321-331, (1988).

[44] M. D. Kelly, Visual Identification of People by Computer, *Technical Report A1-130, Stanford AI Project, Stanford*, CA, (1970).

[45] N. Kiryati and A. M. Bruckstein, Antialiasing the Hough Transform, *CVGIP: Graphical Models and Image Processing* **53**(3) pp. 231-222, (1991).

[46] H. Kim and P. Swain, Evidential Reasoning Approach to Multisource-Data Classification in Remote Sensing. *IEEE Transactions on Systems, Man and Cybernetics* **25**(8) pp 1257-1265, (1995).

[47] C. Kimme, D. Ballard and J. Sklansky, Finding Circles by an Array of Accumulators. *Commun. ACM* **18**(2) pp 120-122, (1975).

[48] R. Kothari and J. L. Mitchell, Detection of Eye Locations in Unconstrained Visual Images, *Proc. 13 ICPR*, **2**, pp. 519-523, (1996).

[49] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. V. D. Malburg and R. Wurtz, Distortion Invariant Object Recognition in the Dynamic Link Architecture, *IEEE Trans. Comput., vol* **42**(3) pp. 300-311, (1993).

[50] K. M Lam, and H. Yan, Locating and Extracting the Eye in Human Face Images, *Pattern Recognition* **29**(5) pp. 771-779, (1996).

[51] K. M Lam, and H. Yan, An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View, *IEEE Trans. PAMI* **20**(7) pp. 673-686, (1998).

[52] A. Lanitis, C. Taylor and T. F. Cootes, Automatic Interpretation and Coding of Face Images using Flexible Models, *IEEE Trans. PAMI.*, **19**(7) pp. 743-755, (1997).

[53] V. F. Leavers, Which Hough Transform, *CVGIP: Image Understanding* **58**(2) pp. 250-264, (1993).

[54] K. Messer, J. Matas, J. Luettin and G. Maitre, XM2VTSdb: The extended M2VTS Database, *Proceedings 2nd Conference on Audio and Video-base Biometric Personal Verification (AVBP99), Springer Verlag, New York,* (1999). The internet URL can be found at http://www.ee.surrey.ac.uk/Research/VSSPP/xm2vtsdb.

[55] J. Matas, *Colour-base Object Recognition.* PhD thesis, University of Surrey (1995).

[56] B. F. Manly, Multivariate Statistical Methods, A Primer, Second Edition, Chapman and Hall, (1994)

[57] J. A. Nelder and R. Mead, A Simplex Method for Function Minimization, *Comput. J.* **7**(4), pp. 308-313, (1965).

[58] M. Nixon, Eye Spacing Measurement for Facial Recognition. *SPIE Proc.*, **575**, pp. 279-285, (1985).

[59] M. S. Nixon, L. S. Ng, D. E. Benn and S. R. Gunn, *Considerations on Extended Feature Vector in Automatic Face Recognition,* Paper Invited for Special Session on Face Recognition IEEE SMC '97.

[60] O. Nakamura, S. Mathur and T. Minami, Identification of Human Faces Based on Isodensity Maps. *Patt. Recog.*, **24**, pp. 263-272, (1991).

[61] J. R. Parker, *Practical Computer Vision Using C,* John Wiley & Sons, Inc, (1994)

[62] C. J. Parsons and M. S. Nixon, Introducing Focus in The Generalised Symmetry Operator, *IEEE Signal Processing Letters* ***6(3)*** pp. 49-51 (1999) .

[63] E. Persoon and K. S. Fu, Shape Discrimination Using Fourier Descriptors *IEEE Transactions on Systems, Man, and Cybernetics, SMC* **7**(3) pp 170-179, (1977).

[64] P.J. Phillips, H. Moon, S. Rizvi and P. Rauss, The FERET Evaluation *in Face Recognition From Theory to Application,* Edited by H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie and T.S. Huang **7**(3) pp 244-261, (1998).

[65] D.W. Purnell C. Nieuwoudt and E.C. Botha, Automatic Face Recognition in a Heterogeneous Population. *Pattern Recognition Letters 19* pp 1067-1075, (1998).

[66] Y. Y. Qi and B. R. Hunt, Signature Verification using Global and Grid Features, *Pattern Recognition*, **27**(12): pp. 1621-1629, (1994).

[67] J. Radon, Uber die Bestimmung von Funkitionen durch ihre Integralwerte langs gewisser Mannigfatkeiten. *Berichte Sachsische Akademie der Wissenschaften Leipzig, Mat. Phys. Kl.* **69,** pp. 262-267, (1917).

[68] D. Reisfield, H. Wolfson and Y. Yeshuran, Context-free Attentional Operators: The Generalised Symmetry Transform. *Int. J. of Comp. Vision* **14** pp. 119-130, (1995).

[69] D. Riesfield and Y. Yeshurun, Robust Detection of Facial Features by Generalised Symmetry. *Proc. 11th Int Conf. on Patt. Recog.*, pp. 117-120, (1992).

[70] G. Robertson and I Craw, Testing Face Recognition Systems, *Image and Vision Computing*. **12** pp. 609-614, (1994).

[71] N. Roeder and X. Li, Accuracy analysis for Facial Feature Detection. *Pattern Recognition.* **29** (1) , pp. 143-157 (1996).

[72] M. A. Shackleton and W. J. Welsh, Classification of Facial Features for Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, **1206** pp. 573-579, AVBPA 1991.

[73] K. Sobottka and I. Pitas, A Fully Automatic Approach to Facial Feature Detection and Tracking, *Lecture Notes in Computer Science*, **1206** pp. 77-83, AVBPA 1997.

[74] M. Sonka, V. Hlavac and R. Boyle, *Image Processing, Analysis and Machine Vision.* Chapman Hall (1993).

[75] R. B. Starkey and I. Aleksander, Facial Recognition for Police Purposes using Computer Graphics and Neural Networks. *IEE Colloquium on Image Processing in Security and Forensic Science*, pp 2/1-2/2, (1990).

[76] L. Stringa, Eyes Detection For Face Recognition. *Applied Artificial Intelligence* **7** pp. 365-382, (1993).

[77] M. Turk and A. Pentland, Eigenfaces for Recognition. *Journal of Cognition Neuroscience.* **3**(1) pp. 71-86, (1991).

[78] M. Turk and A. Pentland, Face Recognition using Eigenfaces*: Proc. Int. Conf on Patt. Recog.*, pp. 586-591 (1991).

[79] P. J. Van Otterloo, A Contour-Orientated Approach to Shape Analysis. Prentice Hall International (UK) Ltd., Hemel Hempstead (1991).

[80] D. J. Williams and M. Shah, A Fast Algorithm for Active Contours and Curvature Estimation, *CVGIP:Image Understanding,* **55**(1), pp.14-26, (1992).

[81] A. Webb, *Statistical Pattern Recognition*, Arnold Publishing Group ISBN 0340741643 pp. 352-361 (1999).

[82] J. Serra, *Image Analysis and Mathematical Morphology*. Academic Press, New York, (1982).

[83] J. Sklansky, On the Hough Technique for Curve Detection, *IEEE Transactions* **C-27**(10) pp. 923-926, (1978)

[84] X. Xie, R. Sudhakar and H. Zhuang, On Improving Eye Feature Extraction using Deformable Templates. *Patt. Recog.*, **27**(6) pp. 791-799, (1994).

[85] X. Xie, R. Sudhakar and H. Zhuang, Real-Time Eye Feature Tracking from a Video Sequence Using Kalman Filter. *IEEE Trans. on SMC*, **25** (12) pp. 1568-577, (1995).

[86] R.K.K. Yip, W.C.Y. Lam, P.K.S. Tam and D.N.K. Leung, A Hough Transform Technique for the Detection of Rotational Symmetry. *Pattern Recognition Letters*, **15**(9) pp. 919-928, (1994).

[87] H. K. Yuen, J . Princen, J. Illingworth and J. Kittler, Comparative Study of Hough Transform methods for Circle Finding. *Image and Vision Computing* **8**(1) pp. 71-77, (1990).

[88] A. Yuille, D. Cohen and P. Hallinan, Feature Extraction from Faces using Deformable Templates. *Int. J. Comp. Vision* **8**(20) pp. 99-111, (1989).

[89] C.T. Zahn and R. Z Roskies, Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers* **21**(3) pp. 269-281, (1972).

[90] J. Zhang, Y. Yan and M. Lades, Face Recognition: Eigenface, Elastic Matching and Neural Nets. *Proc. IEEE* **85**(9) pp. 1423-1435, (1997).

9. Author's Relevant Publications

D. E. Benn, M. S. Nixon and J. N. Carter, *Robust Eye Centre Extraction Using the Hough Transform*, Lecture Notes in Computer Science, **1206** pp. 3-9, AVBPA 1997.

M. S. Nixon, L. S. Ng, D. E. Benn and S. R. Gunn, *Considerations on Extended Feature Vector in Automatic Face Recognition,* IEEE SMC **5** pp. 4075-4080 (1997).

D. E. Benn, M. S. Nixon and J. N. Carter, *Extending Concentricity Analysis by Deformable Templates for Improved Eye Extraction*, Proc. AVBPA (1999).

D. E. Benn, M. S. Nixon and J. N. Carter, Locating and Measuring Eyes on Large Databases using the Hough Transform and Deformable Templates, to be submitted to Pattern Recognition.