Automatic Facial Recognition Based on Facial Feature Analysis.

Kenneth Gavin Neil Sutherland

A thesis submitted for the degree of
Doctor of Philosophy

University of Edinburgh

1992

Abstract

As computerised storage and control of information is now a reality, it is increasingly necessary that personal identity verification be used as the automated method of access control to this information. Automatic facial recognition is now being viewed as an ideal solution to the problem of unobtrusive, high security, personal identity verification. However, few researchers have yet managed to produce a face recognition algorithm capable of performing successful recognition, without requiring substantial data storage for the personal information. This thesis reports the development of a feature and measurement based system of facial recognition, capable of storing the intrinsics of a facial image in a very small amount of data.

The parameterisation of the face into its component characteristics is essential to both human and automated face recognition. Psychological and behavioural research has been reviewed in this thesis in an attempted to establish any key pointers, in human recognition, which can be exploited for use in an automated system. A number of different methods of automated facial recognition which perform facial parameterisation in a variety of different ways are discussed.

In order to store the relevant characteristics and measurements about the face, the pertinent facial features must be precisely located from within the image data. A novel technique of Limited Feature Embedding, which locates the primary facial features with a minimum of computational load, has been successfully designed and implemented. The location process has been extended to isolate a number of other facial features.

With regard to the earlier review, a new method of facial parameterisation has been devised. Incorporated in this *feature set* are local feature data and structural measurement information about the face. A probabilistic method of inter-person comparisons which facilitates recognition even in the presence of expressional and temporal changes, has been successfully implemented. Comprehensive results of this novel recognition technique are presented for a variety of different operating conditions.

In identifying possible avenues of future research to come from this work, the hardware implementation of this algorithm has been considered. By detailed analysis of the proposed algorithm the first tentative steps have been taken in this direction.

Declaration

I declare that this thesis has been completed by myself and that, except where indicated to the contrary, the research documented in this thesis is entirely my own.

Kenneth G N Sutherland

Acknowledgements

There are a number of people whose aid has been invaluable over the last three years and I am very much indebted to them all.

- I must thank my supervisor David Renshaw without whose aid and encouragement I would certain not have got here. I must also thank David for having the patience to read and comment on everything I've put in front of him to over the last three years.
- I am also grateful for the continued encouragement and support of Peter B Denyer.
- I must thank all those students and staff who kindly agreed to take part in my trial. I am particularly grateful to Donald, Gordon and Kevin for agreeing to have their pictures in this thesis.
- I must extend my thanks to Colin Ramsay for his help in performing the codebook generation experiments and Oliver Vellacott for his advice on hardware implementations. My other colleagues in the ISG have all been of some assistance, if only in providing many stimulating political debates.
- I am also indebted to Alan Murray for agreeing to let me use his neural network software and for his hints and tips about neural learning.
- Finally, and most sincerely, I must thank my family for their continuing support. I am particularly grateful to my father for proof reading this thesis and to my brother Andrew for his help and advice at various times.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction.

The futuristist dream of the early science fiction writers, that information could be stored, and controlled, by a machine, is now a reality. An overwhelming amount of personal and sensitive information is now stored in an automated manner. A highly secure method of access control to this information is required to safeguard the interests of the individual and the corporate entity.

The most common means of access control is still the key. For automated systems, the key has been largely superseded by a password or PIN (Personal Identity Number) sometimes in conjunction with a computer readable card. However, possession of a key does not necessarily represent authorisation of access, as the key, if stolen, can be successfully exploited by the purlioner. This unauthorised access is possible because there is no intrinsic relationship between these artifacts (key or PIN) and the identity of the individual attempting to gain access.

In order to guarantee the authenticity of an access claim, it is necessary to establish that the claimant is of a *bona fide* nature. Thus, it is necessary that the identity of the claimant be validated. Anatomical and physiological differences between individuals may well provide sufficient information for an automatic system to identify different individuals.

The study of physical measurements and characteristics which uniquely identify people, has been termed *Biometrics.* A biometric system, is a device which exploits a particular measurement or characteristic, in order to differentiate between several individuals. Such a system can be highly secure, as a physical characteristic cannot be lost, stolen or forgotten. If chosen correctly, a particular physical characteristic cannot be easily mimicked by an imposter.

There is now substantial commercial interest in the development of biometric systems for automated access control[1-3]. The perceived future reduction in fraud is providing considerable impetus for the development of high security systems, particularly as the assumption that PIN and card systems are fail-safe, has come under greater scrutiny.

1.1 Biometric Systems

Independent studies have been regularly undertaken by the Sandia National Laboratories in order to evaluate the performance of various commercially available biometric systems[4, 5]. Such tests require the participation of a controlled *population* containing a number of individuals representing a specified cross-section of the real population. Performance evaluation is then achieved by measuring the ability of the biometric system under test to identify the different members of this population.

To perform this evaluation, it is necessary that the biometric system be *trained* on that given population. This training process allows the device to produce an internal representation for each member of the population. One individual's internal representation in the device, is the parameterisation of that individual, in terms of the biometric being exploited. These internal representations are often referred to as *personal signatures.*

The verification of an individual's identity is most commonly used as the test function. An individual claims to be a particular population member, and the system then accepts, or rejects, that claim, by comparing the claimant with the stored representation of who they claim to be. A more rigorous test of system performance is to require the system to *recognise* an individual. This process involves the comparison of the new individual with all of the population members. The system then suggests the most likely *match* for that individual.

A number of different biometrics systems have been developed, either commercially or academically. The relative merits of a number of different approaches are discussed in this section. Full performance figures for these different biometric systems are not given here, as specific devices are not being considered. However, performance figures for certain commercially available systems can be found in the independent research reports previously cited[4, 5].

It is well known that fingerprint patterns are unique to each individual. The exploitation of fingerprint information for identity verification has been under investigation for a considerable time. Indeed, a number of commercial systems are already available which perform this task. Such systems can fail, when the finger is cut or wrinkled, however, the performance of fingerprint recognition systems is, in general, very good. The amount of data space required to store each fingerprint is, typically, in the region of 512 bytes[6].

A similar, but less well known, biometric is the hand geometry. In this technique measurements and angles are recorded which capture the shape of the human hand, when viewed in silhouette. This approach has the advantage of requiring only a very small number of bytes to be able to store the necessary measurements. The performance figures for devices using this approach are very impressive, however, only one commercial system is available which uses hand geometry analysis to perform automated access control[7].

The recognition of voice is one of the biometric techniques which attempts to mimic a human method of personal identification. In the strictest sense, the voice

patterns used by such systems are not biometrics, however, the processes involved, during the production of speech, are related to the anatomical construction of the vocal tract and the behaviour of the subject. Thus, the **process** of speech can be considered to be biometric in nature. The classification of biometric systems, into physiological and behavioural based systems, is given in Figure 1–1.

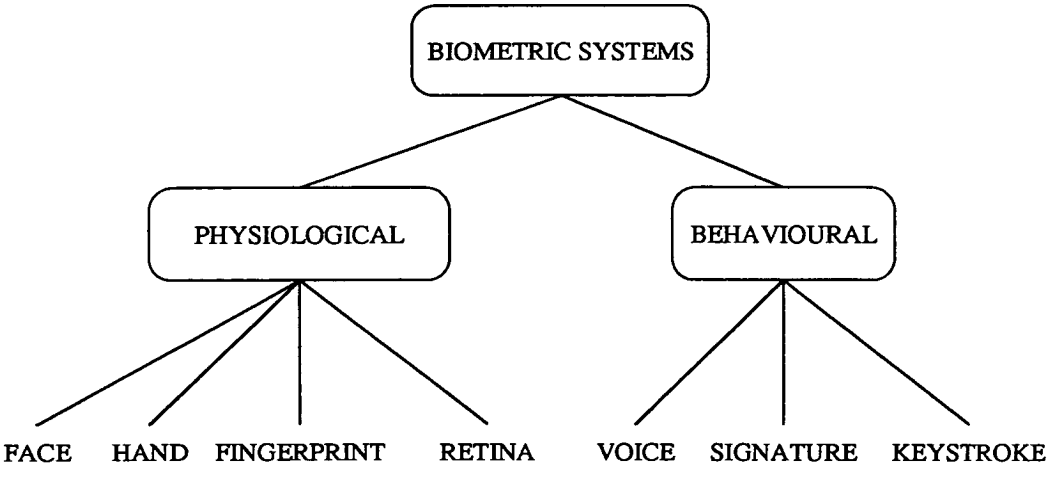

Figure 1–1: Classification of biometric systems.

The performance figures for voice pattern recognisers are not as good as fingerprint or hand geometry systems. This is partly because the voice can change as a result of spurious factors, notably health and mood. Voice recognisers can also be fooled by good mimics. There is one significant advantage of voice recognition over most of the other biometrics, in that voice recognition can be performed over the telephone[1] and personal identity verified remotely. Despite a large number of potential problems, voice pattern recognisers were easily the best sellers in the biometric market in 1989[7].

Another biometric process which has been used to validate identity is the process of writing. The signature, or mark, has long been accepted as proof of identity, however, the automation of this process, has required a substantial amount of research effort. As a result of this research, two very different solutions

[1] Assuming a low noise connection.

to the problem have evolved[8]. Static signature matching is performed using image processing pattern recognition algorithms on two images of a signature and evaluating the similarity between them. Dynamic signature matching uses information regarding the way the signature was written, *eg* pen speed and angle *etc*, to discriminate between different writers. Static techniques are more easily misled by conventional forgery and thus dynamic approaches perform better. However, the dynamic approach, requires the installation of expensive equipment at each site at which identification is required. The intrinsic variability of a subject's signature presents problems for both types of signature matching devices. Another disadvantage of the dynamic approach, is that the user may well feel intimidated by the device and be unable, or unwilling, to provide a good example of their signature.

A similar problem could arise for another system of biometric research – keystroke dynamics. This technique uses the frequency and pressures with which keys on a computer keyboard are struck to perform identification [9]. However, the viability of this approach, as a highly secure access control system, has not yet been proven.

An example of a passive system, which requires no active role from the user, is the retinal scan technique. Using a low-level infrared scanning technique, this approach records the vein pattern within the retina and uses it for comparison between individuals. The performance of retinal scan devices is very good, however, public mistrust of the intrusive nature of the device, must surely limit its future popularity. Vein patterns in the hand have also been examined as possible biometrics. The information is obtained by placing the hand over a bright light with the vein pattern being detected automatically. The performance of this type of system has not been evaluated in independent research.

All of the above solutions to the problem of biometric personal identification are, at present, at differing stages of commercial development as automated access control systems. One other biometric, which has not yet reached this stage of

development, is automatic facial image recognition. An individual's face is their most distinct attribute and has, for many years, been used as the most secure method of identification (*eg* Passport photographs, personal identity cards and driver's licences, in certain countries).

The attraction of video based facial image recognition has been enhanced by the recent developments in imaging technology[10]. The cost of video camera equipment is likely to continue to fall for the foreseeable future, as these developments continue. It is also likely that initial image processing functions may be included in future video camera equipment[11]. Thus, video based biometric analysis will become more competitive with the other biometric approaches.

The amount of information contained in the face is substantial, not only does the face contain identity information, but also clues regarding age, race, gender and emotional state. The human demonstrates an astounding ability to extract this type of information, from only a brief exposure, to a particular facial image. When performing recognition, the human appears to be able to isolate some salient characteristics of the face, and retrieve a name to match that face, with great rapidity and high accuracy. This recognition ability, appears to be largely unaffected by changes in expression and general appearance.

The clear advantage of an automated facial recognition system, over the other biometric systems available, is that facial recognition is a very much more natural process. The public acceptance of the system, which is so important if any proposed biometric system is to be successfully exploited commercially, is more likely to be extended to automatic facial recognition systems because of this perceived naturalness. The implementation of automatic facial recognition is likely to be much less obtrusive than the other biometrics, and should require only a passive role from the subject. For these reasons, automatic facial recognition must be considered as a real contender in the biometric identification market. However, the challenge for facial recognition research is the implicit requirement that the human ability to recognise faces, be mimicked in an automatic way.

1.2 Research Goal

The goal of the research reported in this thesis is to establish an algorithm for automatic facial recognition capable of future realisation as an automated access control device. To this end the thesis will address a number of key points.

- The salient facial features, which can be used for recognition, must be isolated and their extraction performed automatically.
- The comparison of faces must be performed in such a way as to facilitate recognition even when faces are changed by expression and other factors.
- Only minor constraints should be placed on the practical operation of the proposed system.
- The performance figures of the system must be sufficiently positive, so that they suggest that facial recognition is a true competitor in biometrics.
- The algorithm devised, must be suitable for future realisation in a real-time environment.
- A substantial amount of data reduction must be achieved to allow for cost effective storage and rapid comparisons of different faces.

The final point is of particular importance with the advent of *smart-card* technology. It may be that in the near future, everyone will be expected to carry a large amount of personal data on their own smart-card. If this is so, then low storage requirement, facial data, could be included for use in secure transactions. A realistic target, would be to reduce the storage requirement of the facial data, down to a few hundred bytes.

1.3 Thesis Layout

Substantial research effort has been expended in the pursuit of understanding the human visual perception system. In particular, the perceptual and cognitive processes involved in the function of human facial recognition have been widely discussed. Chapter 2 reviews the present understanding of how humans perform facial recognition, with reference to the theoretical processes involved in early vision and perception.

In addition to this theoretical approach, the process of human facial recognition in action – as analysed by experimental research – is also reviewed. The goal of this review is to identify, and summarise, the key pointers as to how humans perform facial recognition and to establish whether it is possible to use these pointers as a guide to the construction of an automatic system of facial recognition.

Chapter 2 also reviews the present state of the art regarding automatic facial recognition systems, emphasising the importance of the facial parameterisation process and illustrating where parallels exist between the present understanding of human facial recognition processes, and their automated counterparts. It is not suggested by this review that an automatic facial recognition system should attempt to mimic the functional performance of the human perception, it is, however, suggested that some key observations of the human performance can be used as a starting point for the construction of an automatic facial recognition device.

Moving away from the facial recognition process for a moment, it is essential that the facial features and characteristics can be located and identified using automated techniques, if they are to be used in any comparative analysis. Chapter 3 discusses the methods whereby it is possible to isolate a face from an image. From this analysis, two different approaches are apparent. Firstly, the face can

be located as a single object or, secondly, the face can be located as a network of smaller objects which **together** represent a face. Chapter 3 illustrates the way in which a number of standard image processing functions have been utilised to perform both full facial location, and the location of the component facial features.

A novel technique of facial detection, based on limited feature embedding, is introduced in chapter 3. The implementational details of this algorithm are presented. Finally, full performance figures of the new algorithm are reported and analysed.

Chapter 4 returns to the specific problem of facial parameterisation. It identifies the factors which must be considered when selecting certain facial features for use in facial comparisons. A novel set of facial features which can be used to perform facial parameterisation is introduced. Initial experiments are reported which demonstrate the potential of this novel feature set for facial recognition. The signal processing algorithm of vector quantization is then introduced as a possible means of providing substantial data reduction of the chosen facial features. The importance of vector quantization codebook generation is identified, and a number of different possible methods of performing this task are described.

The incorporation of facial measurements, relating to the inter-relationships of the different facial features, is suggested as a means of improving the facial parameterisation process. A number of multivariate analysis techniques are used to assess the relative significance of each of the different facial measurements. A compact set of discriminative and uncorrelated measurements are then produced using canonical analysis. The complete algorithm for facial comparison, incorporating facial feature information and facial measurement information, is then described.

A number of practical problems associated with automatic facial recognition are identified in chapter 5. These problems include facial variations caused by temporal and expressional changes, and the use of spectacles. The importance of

generalisation in the training of the system, coupled to a probabilistic approach to personal comparison, is established as the key to solving some of these apparent problems of automatic facial recognition. A number of different methods of performing comparisons, using both the facial features and their positions, are discussed. An overall method of facial training and subsequent recognition, which incorporates the facial location and parameterisation functions discussed in chapters 3 and 4, is presented.

Chapter 5 describes the best known facial recognition device, namely WISARD. A number of differences, between WISARD and the novel facial recognition technique introduced in this thesis, are described. The different methods of performance analysis associated with biometric systems are discussed. Comprehensive results are then presented for a number of different variants of the new system of facial recognition. As a benchmark, comparative results are presented for the WISARD system functioning on the same test data.

In addition to these recognition results, experimental results are also presented for verification analysis, performed using both systems. A full discussion and explanation of the experimental results reported, is included in chapter 5.

Finally, chapter 6 discusses some of the practical problems associated with the implementation of the proposed algorithm, in a real-time environment. A number of potential avenues of future research are identified in this section. The philosophical questions regarding the future public acceptance of biometric recognition systems are also discussed. The final conclusions of this thesis are presented in chapter 6.

Chapter 2

Automatic Facial Recognition : A Review.

2.1 Introduction

The underlying process of *facial parameterisation* is crucial to both human and automatic facial recognition. Parameterisation is the process whereby an object is broken-up into its salient features[1] or characteristics. In image analysis, this function involves the selection of salient patterns which must be present for a particular image to contain a specific object[12]. For example, if a face is present within an image, one would expect to find a broadly oval object containing eye, nose and mouth shapes. The process of determining this *short-hand* representation of the face, is termed facial parameterisation. By using this parameterised facial model, it should be possible to determine whether any new, previously unseen, image contains a face, or not.

Thus, to be able to recognise a face, two functions must be performed. Firstly, the salient characteristics which are fundamental to the face must be identified (*ie* a short-hand parameterisation of the face must be derived[2]). Secondly, a method

---

[1]The term *features* is used here to refer to the statistical parameters or characteristics which can be used to identify different patterns. However, in this thesis, the term *feature* is also used to refer to the facial features like eyes, nose and mouth *etc.*.

[2]The identification of the salient characteristics of an object, in the manner described here, can be termed *feature extraction*.

of extracting these characteristics from any given image must be established. To perform the first task it may be helpful to gain some knowledge of the psychological processes in the brain, which allow humans to identify each other. The second task is the domain of digital signal processing, or more specifically, image processing.

As well as being able to recognise the fundamental elements of the face, an automatic face recognition system must also be able to isolate the factors that enable humans to **differentiate** between faces of different people.

This chapter will discuss a number of the key pointers to the facial recognition process suggested from psychological research. It will then review the ways in which these theories have been exploited to produce prototype automatic facial recognition devices. Incorporated in this review, is a discussion of a number of other image processing algorithms, which have no significant links to human perception, but which have been exploited to perform face recognition.

2.2 Human Facial Recognition Theory

From the spatial light patterns falling on the human retina, to the cognitive retrieval of a name, to match a given stimulus face, an enormous amount of visual and perceptual processing has occurred. Inherent to this process, is the abstraction of information from the visual scene to form a logical pathway to the *recognition* of a complex object (in this case, a face). Perceptual primitives could form levels in a hierarchy isolating *features* from the spatial scene, *en route* to the full analysis of the visual data. Unfortunately, no unified theory of visual perception exists to describe the different processing stages involved here. However, various different aspects of visual perception have been identified by research into physiology and psychology, much of this work has been summarised in [13], some of the most important points are discussed below.

Early visual processing is thought to involve the abstraction of lines, shapes and surfaces, as a first stage towards the construction of a full internal representation of the visual scene. Physiological studies have attempted to identify individual cells, within the visual pathway, dedicated to extracting certain features from input spatial patterns. Such *feature detectors* would form the ideal building blocks for the construction of the lowest level of a hierarchical visual system. Indeed, some animal experimentation, has identified feature detectors tuned to the recognition of particular patterns of light. This research revealed, that when a given pattern of light falls on such a cell, an excitatory pulse is transmitted up the visual pathway. This pulse indicates the presence of that chosen pattern, in the visual scene.

In the hierarchical model of visual processing, such feature detectors would transmit information to higher level detectors. The next stage in the hierarchy would be capable of performing further data abstraction given certain inputs from the lower levels. This hierarchy would culminate in a dedicated cell for each visible object. Termed the *grandmother cell* approach, this theory hypothesises that a large number of cells would be allocated to, for example, face recognition; each of which would be tuned to recognise a particular known face. A number of weakness of this approach are apparent, not least of which is the requirement to dedicate such a large number of cells to recognising objects. However, the main objection to this theory, is the high likelihood that ambiguities would occur at all the component levels. For example, confusion can occur within the simplest form of these cells, as a result of variations in scale and visual intensity.

Marr's research[14] moves away from the cellular approach, to consider the computational algorithms required to extract visual information. The initial process involved in Marr's model of the visual pathway, was the parallel extraction of *edge* information at a number of different spatial frequencies (or resolutions). These edges represented rapid transitions in intensity from dark to light (or *vice versa*) within the spatial pattern. Marr suggested, that at each of these spatial

frequencies, different sets of edges were located, relating to the scale of the objects visible in the image.

In the computational implementation of this theory, Marr and Hildreth[15], performed edge location using a Laplacian operator on Gaussian filtered images. A Gaussian filter was used to preprocess the images in order to produce a certain, measured, amount of blurring. By using several different *sizes* of filter, the different spatial frequencies required were obtained. The Laplacian operator was chosen for its ability to locate edges regardless of their orientation[16, Page 340]. However, the Laplacian operator does introduce false edges (or noise). The next stage of Marr's theory involved the identification of, as he termed, *blobs*, *bars* and *edge segments*. This information was obtained by combining the data contained in the different frequency views. Marr termed this representation the *raw primal sketch*; and considered it to be the result of the computation undertaken in the visual cortex. The raw primal sketch still contained vital information regarding motion, and position, inherent to the visual scene.

The *full primal sketch* resulted from further processing the raw sketch, to obtain contour, texture and shading information. The integration of this information, with motion and stereoscopic data, produced the $2\frac{1}{2}$D sketch. Marr suggested that this $2\frac{1}{2}$D sketch was a *viewer-centred* representation of the surrounding scene, as taken from a particular vantage point. He thought of this sketch, as the culmination of early visual processing; with true object recognition, being dependent on full 3D (position-independent) analysis of the visual scene.

While describing a plausible system of visual perception, Marr's theory, is unfortunately lacking in substantiating physiological evidence. However, the hierarchical surface/object based approach to visual analysis is very attractive for any implementation of an image processing function.

Addressing the specific area of human facial recognition there is one question that arises most frequently[17] :

Is facial recognition special ?

That is to say: Does the process of human facial recognition involve visual and perceptual processes not used in other object, or pattern, recognition? This question relates to a further step in image understanding from the visual perception considered by Marr. The process of human facial recognition, deals with a much higher level of data abstraction and perceptual organisation. In order to answer this question, it is necessary to move away from the consideration of low level vision, and investigate observations of human facial recognition in action.

Newborn infants appear to have an inherent preference for face like shapes. Morton and Johnson[18] reported research in which they provided a number of infants with visual stimuli representing faces, scrambled faces and blank faces. By monitoring eye and head movements of the infants, they observed that significantly more attention was paid to the correct face shape.

Similar research, reviewed by Walk[19], has shown that at 2-3 months of age, infants look directly at the eyes of the facial stimulus; and by 5 months notice the mouth. Around 6 months of age, the infant can differentiate between male and female faces and around 6-7 months can start to recognise familiar faces from unfamiliar ones. These findings are consistent with the suggestion that some rudimentary facial recognition is innate.

Other evidence suggesting the special nature of facial recognition, has been obtained from the identification of the medical condition termed *prosopagnosia.* With this condition, the sufferer loses the ability to recognise faces, while similar brain functions remain unimpaired[20]. However, some other research[21], has suggested that prosopagnosia patients still have some *covert* abilities to recognise faces, which can only be revealed by using suitable prompting. Thus, no clear conclusions can be drawn from the observation of prosopagnosia sufferers.

Ellis[22] addressed the question of facial recognition being a special case, when analys–ing facial recognition, under the influence of a number of other factors. He concluded then that :

> *There does not seem to be any unambiguous evidence that faces are handled by a special and specific recognition system.*

However, in a later study, Ellis and Young[23] identified the role of facial recognition as one factor, in personal identification. They concluded then, that facial recognition was a special process, but not unique. Facial recognition thus forms a key part of personal identification, considered in parallel with other identifiers like gait, voice, location and appearance. It would appear to be correct to say that facial recognition is not a unique process and, as such, it must rely on the same cognitive mechanisms underlying human perception as a whole.

Hay and Young[24] formalised the process of facial recognition, when linked to the other personal identifiers discussed above. They suggested that the retrieval of a matching name, to a given stimulus face, was achieved by performing facial recognition, in parallel, with the other identification processes. They suggested that *Facial Recognition Units (FRUs)* contained structural information of each familiar face and, that if these FRUs were suitably stimulated by an input face, they would produce a positive response. This piece of information is then incorporated with the other personal information, in order to retrieve the name for that individual. A schematic of this model is presented in Figure 2–1.

Bruce and Young[25] refined this concept further, to produce a dedicated facial recognition model. This new model incorporates expression, *facial speech*[1] and *semantic information*, Figure 2–2. The semantic information can be partitioned into two types; firstly, visually derived information, like age, sex *etc.*, and, secondly, identity specific information such as location and occupation. During

[1] Facial speech is the manner in which the face deforms under the process of speech.

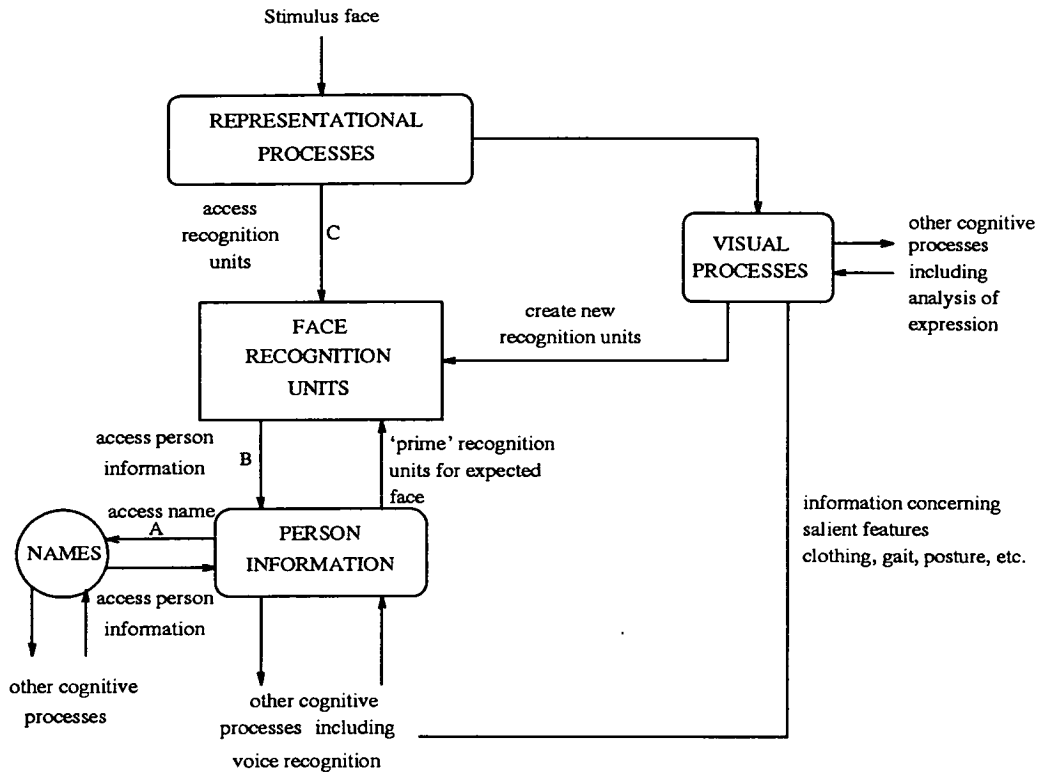

Figure 2–1: Hay and Young's model of face recognition.

recognition, the FRUs have access to the identity specific information contained at the *Personal Identity Nodes (PINs)*. By identifying the link between FRUs and PINs, this model explains a number of common observations about facial recognition; for example, why it is easier to recognise a familiar person in their home than it is to recognise a known person in an unlikely place. It is also suggested by this model, and by other research[26], that while analysis of the facial expression occurs in parallel with facial recognition, the two processes are independent.

While such models identify plausible perceptual processes used to perform human facial recognition, they do not attempt to specify the computational processes involved in extracting the actual facial intrinsics, a point acknowledged by Bruce *et al* [27]. These models deal more with the hypothesised cognitive processes than the low-level visual analysis, required by the *front-end* of any facial recognition system. Thus, such models can yield few pointers to aid the implementation of an automatic system. However, much psychological human facial

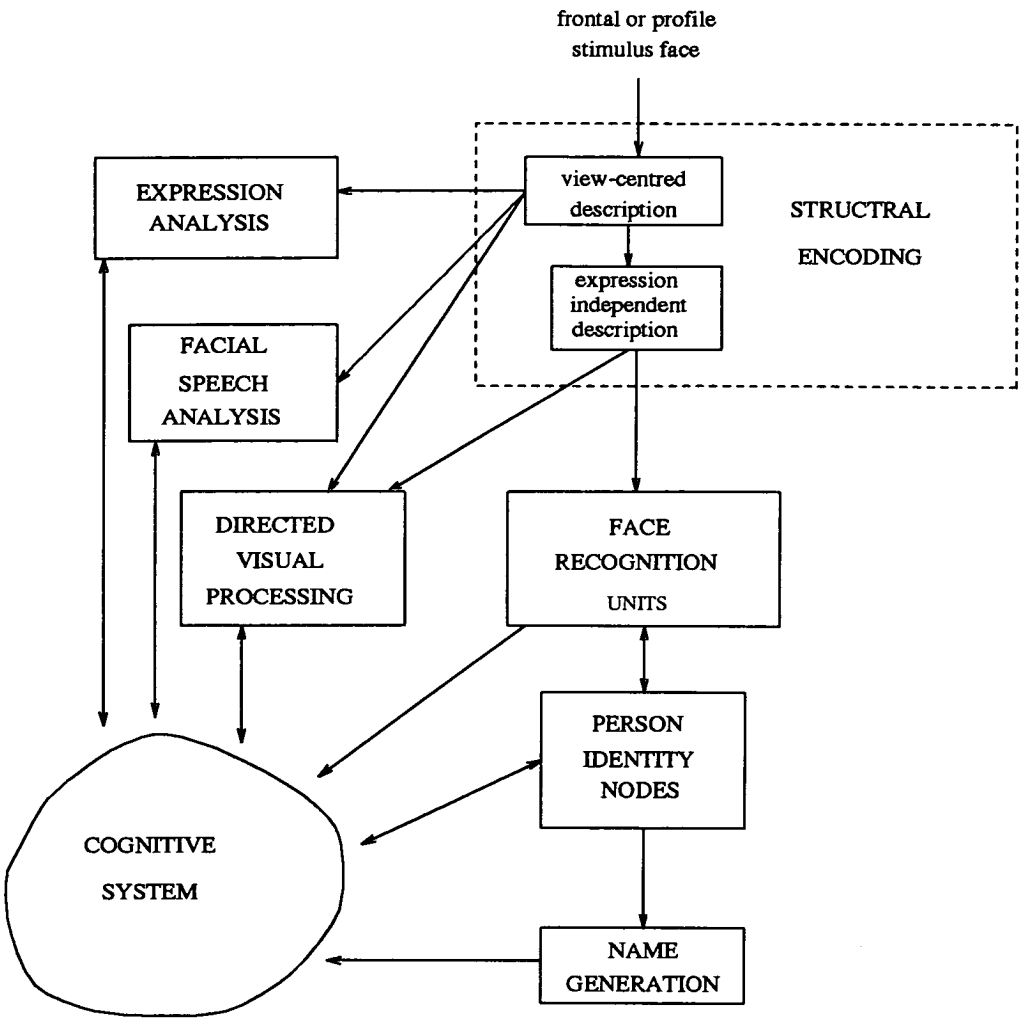

Figure 2–2: Bruce and Young's model of face recognition.

recognition experimentation has focused on analysing the key facial characteristics required for facial recognition, this research will now be reviewed.

2.3 Human Facial Recognition Performance

Before considering actual facial recognition research results, the manner under which such research is performed, must be discussed. Many studies have used exactly the same photographs for human learning, as for testing, and it has been suggested[28] that this method of experimentation may be flawed. The argument

is, that by using the same image twice, some of the recognition performance recorded for human trials may be due to the viewer's *pictorial* recognition abilities, and not necessarily their ability to recognise faces. However, Sergent[29] has argued, that while this effect may be a factor, it is more likely that when a human views a face it is perceived as a face and not just a pictorial pattern. An additional factor to consider here, is that in many of these studies, the human subjects have been told that the purposes of their exposure to certain facial images was to facilitate their future recognition. This may well have caused the subjects to process the faces presented to them, in a different manner from normal. For these two reasons, some measure of uncertainty must be attached to the experimental results described in this section.

A number of studies have shown that the ability of the human to recognise faces is conditional on a number of practical presentational details. One such study[30] compared the recognisability of the same facial images when presented in different manners. Using a panel of human judges, this research compared the ability of the panel to recognise faces when presented to them as photographs, accurate line drawings and outline drawings (featuring only the outlines of the face and the major facial features).

The reported experimentation suggested that the photographic images were significantly better recognised than the drawings; with the full line drawings more easily recognised than the outline drawings. Assuming that the full line drawings contained all the relevant feature and structural information, the implication of this research, is that there is more information in the photograph which is of relevance to facial recognition. Davies *et al* suggested that the other factors present which could influence recognition, included shading, texture and depth cues. This research suggests that an automatic facial recognition system should, at least, attempt to incorporate this kind of information.

Further research[31] considered viewing angle at which target faces were pre-

sented. The study used four different facial poses; the frontal view, the profile view and left and right $\frac{3}{4}$ views (or portrait views). The experimental results were slightly unexpected. The results obtained contradicted the researchers' initial assumption that the portrait views of the face, which contain both frontal and profile information, would be much more recognisable than the other views. In fact, the results showed that neither of the portrait views of the face, were significantly better at conveying the subject's identity than the frontal pose. The profile view was shown to be significantly more difficult to recognise than any of the other three views. This study suggests that the relative ease with which a frontal view of the face can be posed and captured is sufficient justification for its use in an automatic system.

Acknowledging the greater information content of colour photographs, the same research work[31] analysed the relative recognisability of both colour and black and white facial photographs. The experimentation performed, revealed that there was no significant difference between the ease of recognition using colour or black and white stimuli. This suggests that the inclusion of colour information in a facial recognition system may not improve recognition performance.

Moving away from the arguments regarding presentation of facial stimuli and returning to the underlying facial characterisation, several questions arise. Fundamentally, there is considerable debate over the overall mechanism of recognition, whether the human adopts a *holistic* or *feature-based* strategy to perform facial analysis. Some human experimentation[32] has concluded that humans may perform analysis of the face, **either** holistically, or *featurally*, depending upon the purpose of storing the information. Other research[33] has attempted to assess the roles of the left and right brain hemispheres, in performing either holistic or *featural* analysis. Unfortunately, as such work remains inconclusive in its findings, consideration of both holistic and feature based facial recognition is required.

A large number of different studies have analysed the relative importance of

the various facial features. As a very comprehensive review of this literature is included in[34], it has been decided here only to discuss the major research works in this area and update that review.

Ellis *et al* [35] attempted to assess the relative recognisability of the inner and outer facial feature configurations. This was performed by presenting a panel of human judges with facial images with either the inner features (the eyes, nose and mouth) or the outer features (the hair, chin and facial outline) removed. Two experiments were performed, firstly, using familiar faces (from well known celebrities) and, secondly, using unknown faces. For the familiar faces, the inner facial features proved to be more valuable for recognition, however, the unknown faces were recognised with equal frequency, from either the inner, or the outer, feature configurations. The authors of this work suggested that this apparent discrepancy, could be attributed to the fact that a greater amount of attention is paid to the more expressional, inner parts of the face for known (and frequently seen) faces.

A completely different source of experimental information also appears to point to the importance of these inner facial features[36]. By monitoring the eye movements of a human, while studying an image of a face, it can be noted that significantly more attention is paid to the eyes and mouth, than that given to the outer facial appearance.

The importance of these two pieces of evidence should not be overstated. However, the relative significance of the inner facial features has been noted for further consideration.

Rather than dividing the face on a feature basis, Haig[37, 38] used a spatial grid to partition the facial image into 38, equal sized, sub-parts. Again using human jurors, Haig recorded the recognisability of a number of facial images, on the basis of each of the different image blocks. He concluded, that there was no general set of features used to recognise faces, and, that when presented with a particular facial stimulus, the human viewer selects certain distinctive features,

to use for recognition, on an *ad hoc* basis. Haig further suggested, that there was little consistency between different jurors, when viewing the same facial image.

The relative difficulty with which humans perform facial recognition when the stimulus face is up-side down has been widely noted[39]. This difficulty has been exploited by Endo[40] in order to assess the relative importance of each of the facial features. Endo presented a number of human viewers with two upside-down faces, one of these images was the original photograph, and the other had some of the facial features altered. The viewer was then requested to record whether or not they detected a difference between the two images. The significance of each facial feature was derived from how easily an alternation was noticed in each of the features analysed. This research identified the eyes and the hair as being of primary importance; with less attention paid to the mouth and the chin; and least significance given to the nose. This study was broadly in agreement with a number of other studies which have established the top half of the face (*ie* the hair, forehead and eyes) as being of significantly greater importance than the lower features (the mouth, nose and chin)[34]. In a similar vein, Fraser and Parker[41] reported a study in which they separated the face into four items, namely, the eyes, the nose, the mouth and the facial outline. Of these features, the eyes and the facial outline proved to be of most use in facial recognition.

Taking a more holistic approach to facial recognition, researchers have attempted to evaluate the recognition of facial images as a function of spatial frequency. In this representation, the high frequency components of an image are likely to convey the fine edge detail within the image, whereas, low frequency image parts preserve more of the overall shapes and less of the detailed information. Two such studies[42, 29] have identified the fundamental importance of the low frequency (coarsely sub-sampled) views of the face.

By identifying the significance of the low frequency information in facial recognition, these studies suggest a high level of importance is attributable to the un-

derlying structures present within facial images. However, this result has not been entirely incorporated into the established models of vision and perception[28].

To summarise, the above discussion has identified the various pieces of evidence suggesting the likely importance of certain aspects of the face, in the facial recognition process. Notably, the inner facial features have been widely identified as being of primary importance, however, positive information for recognition, is also contained in some of the other features. Likewise, the role of facial structure, in recognition, cannot be denied. The best conclusion to draw from this information, is that any facial recognition system must incorporate elements of both feature and structural information.

Rhodes[43] performed a study using the concept of first and second order facial features. The first order features took the form of verbal descriptions of the local facial features; the eyes, mouth, hair *etc.* The second order information related to the measurements between these features. In this manner, both the local feature information, and the global configurational information, were captured for use in recognition. In a human facial recognition trial, Rhodes proved the importance of both types of facial information.

2.4 Facial Image Retrieval

Recognition of criminals has provided the impetus for the first steps to be taken towards developing automatic facial recognition systems. In the devices which have been produced, automation has been restricted to the comparative stages of the system. The analysis of the facial data is still performed by a human judge (or a witness).

Goldstein *et al* [44] presented a questionnaire to a panel of human judges. They were then asked to provided a qualitative assessment of a particular facial stimulus, in terms of a number of facial features. The characteristic information,

contained in this questionnaire, was then used as an identifier to that particular person.

In the questionnaire, the jurors had to provide ratings for a number of facial features deemed (by Goldstein *et al*) to be of importance for facial recognition. For example, the facial shape could be classified as *square*, *round*, *oval* or *long*. For certain features a gradation of description was allowed, *eg* the eye opening was given a value of between one and five, where 1 was *slit* and 5 was *wide*.

In this research, Goldstein *et al* used a set of thirty-four facial characteristics to encapsulate the intrinsic facial information of any one individual. However, twelve of these features w–ere removed as they were either irrelevant, or statistically highly correlated with other facial features. For example, the upper and lower lip thicknesses, were very rarely different from each other, and therefore these two features could be replaced by one with little, or no, loss of information. Thus, a reduced set of twenty-two facial characteristics were maintained for facial comparison.

The automated stage of the system, performed the comparison between the characteristics recorded on one questionnaire, with all the other characteristic information stored for each member of the test population. The comparison was performed by representing each *face* as a point in 22-dimensional space. Each axis in that space referred to one of the 22 selected facial characteristics. The *difference* between two different faces was measured as the straight line distance between their respective points in 22-dimensional space. This straight-line distance measurement is termed the Euclidean distance.

On the basis of the population feature statistics, it was possible for the system to identify a particular feature as being very discriminative, for a given population. The system could then automatically prompt the human viewer to provide information on this particular feature. In this way, the system was able to perform limited automatic feature selection.

An important aspect of this research work was to analyse the way in which different human judges subjectively described the same facial images. In general, a high level of consistency was recorded between different judges, however, *good* and *bad* judges were identified. For interest, the system was able to produce the most and least similar pairings from the population.

A similar system of identification of individuals was presented by Della Riccia and Iserles[45]. Motivated by criminal identification requirements, their system matched a sketched identi-kit image, with a database of known criminals (in *mug shot* form). Thus, given an identi-kit sketch, the system was required to retrieve the most similar mug shot from its database.

Again this technique used subjective verbal descriptions of the fundamental facial characteristics. However, this information was supplemented with a set of actual facial measurements, drawn from the identi-kit image. To select the most discriminative characteristics, a rigorous feature selection algorithm was employed.

The experimentation involved a database of 506 mug shots and twelve identi-kit images. For each stimulus, the system produced names of the six most similar mug shots in the database. In nine out of the twelve test cases, the correct mug shot was contained in the six produced. The other three cases failed. However, Della Riccia and Iserles attributed these failures to badly drawn identi-kit images, and not to the failure of the comparative analysis performed automatically. A number of other facial retrieval algorithms, operating on similar principles, have been reviewed by Laughery *et al* [46].

In the narrow field of computer guided retrieval of facial images, the use of syntactic descriptions of the face is a viable process. However, the subjective nature of the descriptions used, makes full automation of this process very difficult. The key aspect of the various research programmes in this area is that they have attempted to identify the distinctive facial characteristics and perform multiple

comparisons using these features. Thus, indications of the most likely features to exploit in an automatic system have been provided.

2.5 Automatic Facial Recognition Systems

A number of prototype automatic facial recognition systems have been designed and implemented. These systems have been based on a number of the different theories of human approaches to facial recognition. However, some of the prototype systems, considered in the following sections, are much more loosely based on human vision and perception and relate little to the present theories of actual human facial recognition.

The following sections will link these two areas together, to reorganise the research of automatic facial recognition into a number of general approaches. The review will start with local, feature-based, techniques and then move on to discuss the more holistic systems. In this way, the review given here, reveals a clearer picture of the present state of the art in automatic face recognition, than has previously been published[47, 48].

2.5.1 Facial Feature Based Systems

Video images capture light intensity changes, across the scene, in a manner analogous to light falling on the retina. However, a video image represents an arbitrary partitioning of this scene, into a number of picture elements (or pixels). Each of these elements, is assigned a numerical value signifying the light intensity at that image point. Baron[49] described a system in which certain facial features were extracted from the entire image, in terms of their pixel patterns, and compared with previously stored features to assess facial similarity.

Baron imposed a hierarchy on facial recognition, by storing each face as a series of image sub-parts. Typically, the stored data included image parts (or

templates) representing; a view of the whole face, the right eye, the mouth, the chin and the hair. Each of these templates was selected at a suitable image resolution so that all five parts where the same size.

To form a database of faces, Baron used a sample image from each member of the test population and partitioned it into five facial sub-parts. Recognition was performed by taking a new image, partitioning it into parts, and comparing each of these templates with all the stored examples for the population.

Recognition was only established if, firstly, the whole face template was matched to a particular member of the population, and, secondly, three out of four of the other templates were also matched to the same population member. A *correlation* algorithm was used to measure the similarities between corresponding templates from different people.

The weakness of Baron's system lay in the fact that the facial sub-parts, used in recognition, were manually extracted from the image data, by a human operator. Also, the facial parts chosen for each facial image, were judged to be the most distinctive features for that particular face. Thus, much of the vital feature selection stage of the device, was being performed manually.

On a trial population of forty-two people, and a test set of one hundred and fifty images, the system performed perfectly; correctly matching each test image with its owner, and correctly rejecting those faces not in the database. This study suggests the possible viability of using local feature data for comparative analysis. However, this test experiment is not sufficiently large to be viewed as significant.

Another research team[50] has attempted to use a biologically plausible process to underpin a similar system of feature based facial recognition. Unlike Baron's approach, this system did not use a hierarchy of recognition, instead, it used all the available features to perform recognition, and assumed an equal importance for each different feature. The actual feature comparison involved an

additional level of data extraction, through the use of *Cortical Thought Theory (CTT).*

The process of CTT involves the image data being displayed on the cortex surface of the brain, the cortex then extracts a two dimensional vector to characterise that image. The researchers termed this two dimensional vector the *gestalt* of the image. To perform facial recognition, the system calculated six gestalt values for a chosen set of six facial sub-parts. The system required manual location of these features only if the automatic location algorithm failed. The facial partitioning used in this study is shown in Figure 2–3.

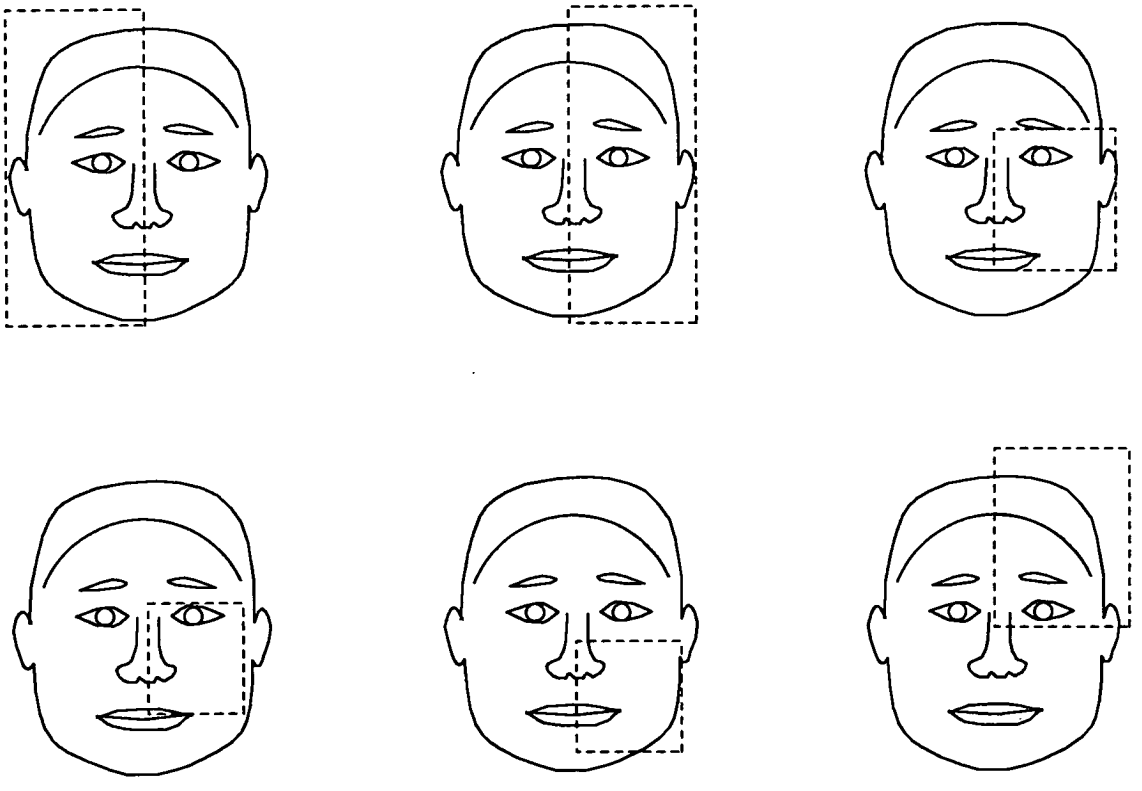

Figure 2–3: The facial partitioning used for cortical thought theory.

The gestalt value of each facial sub-part was taken as the location of the peak of the Fast Fourier Transform of that image part. To create a population database, a number of images (typically 5) of each member of the population, were presented to the system. The device then recorded the position, and the variance of the position of the six gestalt values, derived for each of the population

members. To recognise a new face, the location of its gestalt points are compared to the locations recorded for each person's face in the population database.

In a trial of the system, a success rate of 90% was achieved on a population of twenty people, using one test image of each person. Again this result cannot be deemed to be significant because of its low test size. The principles of the system are similar to Baron's approach, however, the data storage requirements of each face have been reduced substantially.

2.5.2 Facial Measurements

Moving one step away from the actual facial features; the configurational arrangement of these features can be captured in a simple set of measurements. A set of measurements obtained in this way, could contain not only the feature sizes, but also much of the underlying structural information. Unfortunately, as in common with much of this research, there is no agreement over what these distinctive facial measurements should be. Thus, a number of different measurement strategies have been suggested[51-54] as possible approaches to facial recognition. One possible measurement set is illustrated in Figure 2–4.

Despite this plethora of different schemes, there have been remarkably few significant facial recognition trials performed using a system based purely on facial measurements. This is partly due to the practical problems of automatic facial feature location; the accuracy of which would be crucial to a measurement system. In addition to this factor, the extraction of truly *distinct* and *uncorrelated* facial measurements has also been identified as a possible problem[55].

Kanade[56] performed a measurement based facial recognition trial with a population of twenty people and a test set of twenty images (one from each population member). Using an automatic feature location and measurement technique, he obtained a recognition rate of only 50%, however, using manually derived measurements, the recognition rate rose to 75%. This suggests, that while some level

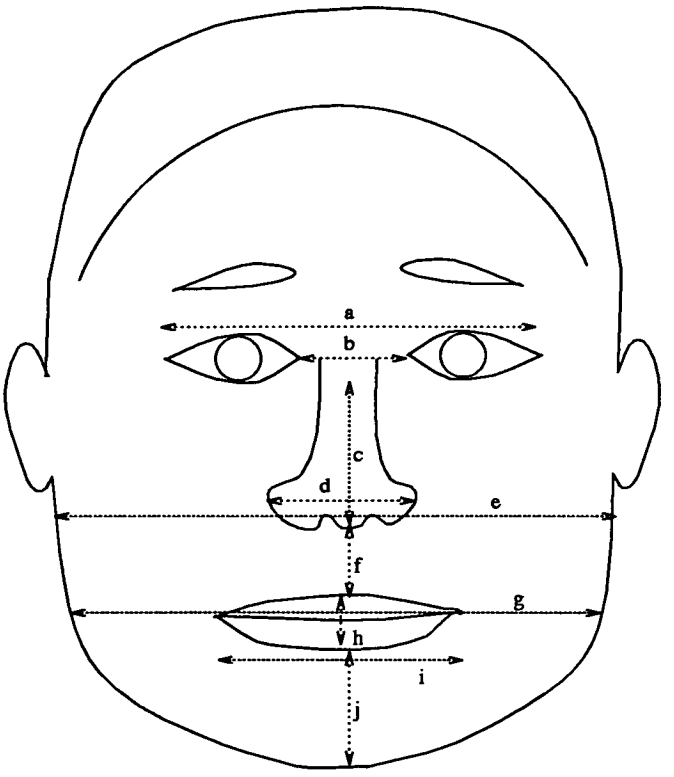

Figure 2–4: An example set of facial measurements.

of facial recognition could be performed in this way, the measurement set Kanade used did not capture enough of the distinctive information to perform complete facial recognition.

In a more recent study Wong *et al* [57] performed successful recognition using a much smaller set of measurements. However, as their results referred to a population size of only six members this is not a very valid study.

2.5.3 Principal Component Analysis

Kohonen[58] made use of full facial images for a demonstration of his *linear associator*. The associator was a memory matrix in which information could be stored regarding a certain class of objects. The training, or storage, involved presenting the matrix with a stimulus vector and the desired output vector for that stimulus. This is a form of *supervised learning*. In recognition mode, the

associator was presented with a stimulus and an output was produced. For facial recognition, the input and output vectors were pixel array images of faces. Kohonen tested the associator using new views, and corrupted data, of individuals the device had previously learnt to recognise. In both cases, the system gave the correct response. The system did require the image data to be pre-processed to contain only the facial information.

O'Toole *et al* [59] used the linear associator to assess the effect of spatial frequency transformations on facial recognisability. The experiments involved training the associator with one type of facial image (either high-pass filtered, low-pass filtered or normal) and measuring the recognition rate recorded for testing on the other two types. The applicability of this research is limited, however, it does provide further evidence for the viability of the linear associator for facial recognition.

In later research, O'Toole and Abdi[60] recognised the link between Kohonen's linear associator and *Principal Component Analysis (PCA)*[1] and illustrated the way in which this process can be used to break down the input face, into a set of component *features.* Sirovich and Kirby[61] used this technique to perform characterisation of a sub-part of the face containing just the eyes and the nose, again breaking the image up into a set of component parts. To understand the processes in action here, it is necessary to consider the way in which PCA operates.

To perform principal component analysis on facial images each face must be considered as a single n-dimensional vector. This transformation is achieved by concatenating all the rows of the image together. The process of PCA then involves representing this vector as a linear combination of a set of *basis* vectors.

[1]The technique of Principal Component Analysis has appeared in many different research areas under a variety of names, including; The Karhunen-Loeve transform; The Hotelling transform and eigenvector analysis.

These basis vectors (termed *eigenpictures* or *eigenfaces*) are formed from a training set of facial images. These eigenpictures can contain the *features* described by O'Toole and Abdi.

These basis vectors define an orthogonal set of axes in which to describe the vectors representing any facial image. It is the optimal nature of this orthogonal transformation, that has led to the consideration of PCA for image data compression[62, 63]. For faces, the image of a face can be represented as a weighted sum of these optimal axes.

Data compression is obtained, because the *energy* within the image information is usually concentrated in the first few of the component axes. Thus, it is possible to truncate this description at a much lower dimensionality than that of the original face. To store any particular face, it is only necessary to store the coefficients relating to the proportions of the different basis vectors required to reconstruct that face. For example, Kirby and Sirovich[64] and Craw and Cameron[65, 66] have both characterised facial images in terms of the first fifty eigenfaces, with remarkably little loss of recognisability. Similar work has also been reported by Turk and Pentland[67, 68].

A complete, automatic, face recognition system involving the use of PCA has not yet been implemented. This is partly due to the fact that the faces must be fully normalised in size and orientation, before they can be encoded using the PCA technique. This point was addressed in a study by Craw *et al* [69], in which they were able to distort a new facial image into a predefined *standard* geometric shape. A similar technique has also been suggested for the encoding of eyes using PCA[70]. However, the computational requirements of such geometric transformation are not inconsiderable.

The use of PCA for faces is just the extension of a well known statistical technique to the process of facial characterisation. There is no physiological, or psychological, evidence to suggest that this process mimics human facial process-

ing in any way. However, PCA as a means of facial recognition remains a very promising research area.

2.5.4 Neural Networks

A number of different neural network architectures have been applied to the problem of automatic facial recognition. All of the research studies in this area adopt a holistic approach to facial recognition. It is assumed that the network will identify the most discriminative information itself (*ie* that the network will perform automatic feature extraction). This feature extraction is able to occur, because the network is presented with a number of training facial images and allowed to *learn* the most distinctive features about them. Thus, the features used by the network, to perform recognition, are implicit in nature. It is hoped that neural networks will be able to mimic human recognition performance because the network is allowed to learn facial representations in a similar manner to that used by humans. The weakness of this argument is that very few of the learning algorithms advocated at present are, in any way, biologically plausible[71]. However, this fact has not diminished the enthusiasm of neural network proponents to apply different neural strategies to facial recognition.

One such approach was reported by Cottrell and Fleming[72] in which they exploited the data compression abilities of neural networks. They used a network to construct a set of internal features suitable for image compression, they then used these features, to perform facial recognition. The similarity between extracting features in this way, and the process of principal component analysis has been identified by Cottrell and Munro[73].

The facial compression was performed by the use of a three layer *Multi Layer Perceptron (MLP)*, consisting of 4096 input neurons, 80 hidden neurons and 4096 output neurons. The network was presented with a number of example faces, until it was able to produce an output sufficiently similar to the input stimulus.

By allowing the MLP to compare its output with its input, the hidden layer was able to adapt to encode the images in an efficient manner. Thus, the hidden layer was able to perform automatic feature extraction, with each neuron dedicated to isolating a particular characteristic composed of parts of the entire face.

The outputs from these 80 feature detecting neurons, termed *holons*, were then used as inputs to a further MLP. This second network was trained to produce three different outputs. Firstly, the system indicated whether the input image was a face or not; secondly, it gave the gender of the face; and finally a name for the stimulus face. Further research[74] demonstrated the same approach was able to determine emotion, as well as gender and identity. The recognition performance claimed for this system was very good when used on a population of twenty subjects. Although, one possible problem for this approach, is the manner in which it performs feature extraction. By training the network to minimise the absolute error for facial image compression, it may not be yielding the correct features for use in recognition.

One of the first commercial prototype automatic facial recognition systems has been developed by SD Scicon Ltd[75]. Their system is based on the development of a standard neural architecture to perform facial recognition. They have claimed a perfect recognition rate on a population of 100 members. However, the company is unwilling to divulge the actual algorithmic details involved in the system.

A number of other neural systems have been proposed for facial recognition, however, no other fully tested systems have been reported. For example, another neural technique has been suggested (by a research team at Brunel University[76]) for facial image retrieval from an image database. However, the system has not yet been applied to full facial recognition.

The use of contextual information in facial recognition has also been examined using neural techniques[77]. Genetic Algorithm based neural networks have also been suggested as a possible future avenue of facial recognition research[78].

In their review Phillips and Smith[79] concluded, that there is a very high likelihood that some form of neural network can yield a functioning facial recognition device. However, they acknowledge that the form in which data is presented to the network may well be crucial in determining the likely system performance. Thus, neural networks may be better employed classifying the output of some other image analysis technique, than attempting to perform complete facial recognition. They suggest that a hybrid, neural and conventional approach may be one solution to automatic facial recognition.

2.5.5 Three-Dimensional Facial Recognition

All of the facial recognition devices described above, perform based on a frontal view of the face. However, it is undeniable that further information is contained in, firstly, the profile view and, secondly, the full 3D map of the face. For example, several researchers[80, 81] have established that subjective identity and expressional recognition can be enhanced by using image *valleys* in addition to edges, in facial transmission. The *valledge* technique, as it is termed, isolates valleys in the face where substantial changes in the facial surface occur; this information provides the viewer with cues to the 3-dimensional characteristics of the face.

A large amount of research work has been dedicated towards extracting characteristic information from the facial profile. In an accurately posed silhouette image of the face, the stark profile information can be easily obtained. It is then possible to extract a number of critical features, from this profile, to facilitate some level of facial recognition. The isolation of these key features is fundamental to the process, however, as with frontal facial analysis, no one strategy exists for performing this task.

A number of researchers[82, 83] have suggested different characterisation methods for profile images. However, one aspect is common to all these profile recognition systems; that is their ability to reduce the facial information into a very

small amount of characteristic data. Indeed, Harmon *et al* [85] demonstrated a profile based system capable of distinguishing between 112 subjects which stored only seventeen key features extracted from the profile.

However, it is this high level of data reduction which may present problems for a profile based system when used on a large population. The amount of information that it is possible to extract from only one profile line is, ultimately, quite small. Thus, it would be very difficult for a profile based system to maintain the distinctiveness between different faces, when they are compressed to such a degree. However, as yet, this point remains unresolved.

Obtaining full three dimensional depth maps of the facial data is a lot more difficult than the frontal, or profile, facial views discussed above. There are a number of techniques available which can be used to obtain a depth map of the front of the face; these include laser scanning techniques[86] and stereo photogrammetry[87]. One research team[88] has chosen to extract 3D information for the whole head, by moving a camera around a stationary subject. Thus, capturing more information than the more standard stereo camera approach.

Having obtained an accurate 3D map of the face, or the head, there still remains the problem of characterisation. One particular approach to 3D facial analysis has attracted substantial recent interest, in this approach[84] the facial surface is described in terms of a number of primitive 3D geometric shapes. It may be, that if the facial surface can be parameterised in these terms, then a characteristic data set of relatively low dimensionality can be obtained. A practical system based on this approach has not yet been implemented.

A slightly less rigorous approach has been adopted by Lapreste *et al* [89] in which they identified the fundamental turning points in the three dimensional face, and stored the distances between these key points. They produced a successful facial recognition result for this approach, but it was only based on two example faces. Similar research results have been reported, by another research group, using a similar technique[90].

A two dimensional technique of facial recognition which is very similar to these 3D approaches, exploits isodensity lines[91]. In this technique, intensity contours (*ie* lines connecting points of equal image intensity) are preserved for facial analysis.

Three dimensional facial information is not only useful for 3D facial recognition, but also for performing facial transformations prior to the use of one of the 2D frontal facial recognition techniques described above. For example, Aitchison and Craw[92] described a technique of mapping any 2D view of the face, onto a *generic* 3D facial model and extrapolating new views of the face from that 3D model. They suggested, that this technique could be used to standardise the facial view as input for their PCA facial recognition system. Shashua reported a technique which used a combination of images to compensate for changes in view, and illumination, [93], however, substantial results for this technique have not been published.

The mapping of 2D data onto a 3D model has also been used for coding of facial images in videophone equipment[94, 95]. In this research, more *realistic* facial images have been obtained by exploiting the fundamental three-dimensional nature of the face. However, the recognisability of such images has not been properly assessed.

Three dimensional information regarding the face, must contain additional clues to the individual identity of that face. However, it may be that the added investment in hardware required, as described above, to obtain the 3D information for an automatic system, cannot be justified by its possible increase in performance. It may be, that while 3D information contains a lot of recognition information, sufficient hints to the underlying three dimensional structure are contained in the shading and depth cues in a frontal 2D photograph. A much larger recognition study using 3D information is required to resolve this argument.

2.5.6 Other Techniques

There are a number of miscellaneous facial recognition studies which do not fall into the categories given above.

One such study by Cannon *et al* [96] performed facial recognition using a hybrid approach. The system performed recognition using a number of different indicators including; the height of the subject; the colour of their eyes; the intensity of their hair; the intensity of their cheeks and the height of their forehead. These five characteristics where recorded as five scalar numbers. In addition to this information, the system also stored two pixel matrices, one for each of the subject's eyes. Recognition was established if all these characteristics were matched to within certain tolerances. In experimental research, using 50 test images, a recognition rate of 96% was recorded. Similar work has been reported by Kelly[97].

Another hybrid approach has been reported recently[98], in which a number of different facial characteristics are extracted in a variety of different ways. This system uses geometric information and intensity information to establish individual identity. Unfortunately, recognition trials of this system have not yet been reported.

Gallery and Trew[99] reported an *invariant* facial recognition technique. By using recursive geometric sub-division of the face, they have shown that it is possible to establish a very low dimensionality vector, containing the characteristic facial information, invariant to facial position or orientation. Facial asymmetry is thought to contain some recognition information[100], however, as for Gallery and Trew's technique, no results have been reported for a facial recognition study based on this approach.

2.6 Related Research

There are a number of related techniques, not explicitly used for facial recognition, which are worthy of consideration here.

Golomb *et al* have described a system called *SEXNET*, which can be used to identify the gender of a given facial image[101]. The neural network used, is very similar to that attributed to Cottrell *et al* in section 2.5.4. The network's performance compared favourably to a human panel of judges, however, manual preprocessing of the input images was required.

In the analysis of facial expression, the ways in which the face deforms has been analysed. As a result of this research, mathematical models of the face, and accurate descriptions of the possible feature movements, have been derived[102, 103]. The direct relevance of this information to facial recognition is limited, however, the fundamental facial points and their likely variations have been identified. In a similar vein, *optical flow* information has been exploited to estimate muscular movements in the face and, hence, facial expression[104].

Pentland[105] has suggested that it is possible to form a face from a number of primitive shapes. The construction of the face can be performed using thirteen primitives and represented in approximately 100 bytes of data. However, this technique was suggested as a means of constructing synthetic faces and not as a means of facial recognition.

The construction of facial caricatures has also attracted recent research interest[106]. Some researchers have examined the human ability to recognise caricatures of famous people and concluded that, in some cases, the caricature is more easily recognised than the original image[39]. This suggests that the human uses the distinctiveness regarding a particular facial characteristic, as an additional guide when performing facial recognition.

2.7 Summary

It can be concluded that the process of human facial recognition is based on primitive functions common to much of human visual processing. Some of these processes are known. However, most of the processes involved still remain unspecified. It is, therefore, not yet possible to obtain an accurate model of the human facial recognition process. Thus, in order to identify the underlying visual processing functions in action, it is necessary to observe humans performing facial recognition.

Research has revealed that differing significances are attached to the various features present in the face, when humans perform recognition. Unfortunately, only the broadest assumptions about this process are generally accepted.

- Most of the obvious facial features have some part to play in recognition.
- The eyes and mouth are of particular importance.
- The overall structure of the face is a significant factor.

Thus, a system that encorporates these three factors should be able to mimic some of the observed characteristics of human visual processing.

For this reason, the first two approaches discussed above, feature-based and measurement-based recognition, will not, on their own, be sufficient to perform facial recognition. Neural network and principal component analysis techniques, both hope to automatically capture the statistical facial differences in a more holistic manner. However, the training requirements, and the computational complexity, of both approaches make them unattractive.

The underlying three-dimensional structure of the face may well hold significant clues to facial recognition. However, it may also be true that humans are able to incorporate the 3D nature of the face when viewed in only two dimensions.

Thus, the additional hardware requirements of the 3D based approach does not yet appear to be justified.

The hybrid feature/measurement approach, as explored in Rhodes' research, appears to be one of the most promising avenues for facial recognition research. A system based on this principle, would be particularly attractive when attached to a caricature like system of comparison (*ie* additional significance is given to facial characteristics which differ substantially from the norm). However, any feature based system of facial recognition requires the accurate location of the face, and the component facial features, before comparative analysis can be performed. The following chapter discusses the ways in which this location function has been attempted.

Chapter 3

Face Detection and Facial Feature Location.

3.1 Introduction

The preceding chapter described a number of different image processing solutions to the problem of automatic facial recognition. Many of the techniques described assume that the facial size and position are already known. In order to construct a fully automatic system, it is therefore necessary to derive an automatic method of isolating the face from the rest of the image data. The spatial image parts which represent the face must be separated (or segmented) from those parts of the scene which represent the image background. To perform this task, a practical method of extracting the key image parts which form the face must be devised.

The face can be viewed in two distinct ways. Firstly, as a single self-contained object, or secondly, as a collection of smaller sub-parts, such as the eyes, nose and mouth. This chapter reviews some image processing techniques which have been used to locate the entire face as a single object. It then goes on to discuss techniques which can be used to locate the individual facial features.

A novel technique of Limited Feature Embedding will be introduced. Comprehensive results, of a full evaluation of this technique, are also given in this chapter.

3.2 Face Detection

Separation of the face from a cluttered scene, is an easy process for even a small child. However, as described in the previous chapter, the processes of data abstraction and perceptual organisation required to perform even this fundamental task, still elude researchers. To simplify the processing involved, most of the image processing solutions which have been constructed, can only perform facial segmentation in a carefully controlled environment.

Much of the research in this area has been driven by research into *model-based* encoding of facial images for videophone transmissions[107]. A number of different techniques for face location are outlined in the following section, for each system, the operating constraints and the consequent system performance are discussed.

3.2.1 Outline Tracking

When an image of a face is subjected to an edge enhancing algorithm, the output produced, indicates the strongest intensity gradients within the image[1]. Only when these edges are linked together, to form higher level shapes and objects, can the contents of the image be accurately analysed. Outline tracking attempts to make sense of the cluttered edge map produced by edge detecting a facial image.

Kelly[110] pioneered an outline tracking technique termed *planning* which searches the incoming image for a 'head and shoulders' type shape (Figure 3–1). In practice, this head-like shape appears as a series of smaller, sometimes

[1]The process of edge detection is not reviewed in this thesis as there are already a substantial number of good reviews on edge detection and related image segmentation[16, 108, 109].

unconnected, edge segments. The task of the outline tracking algorithm, is to connect these possible edge segments together, to form a plausible head shape. To reduce the computational requirement of this extensive search process, a multi-resolution approach is used.

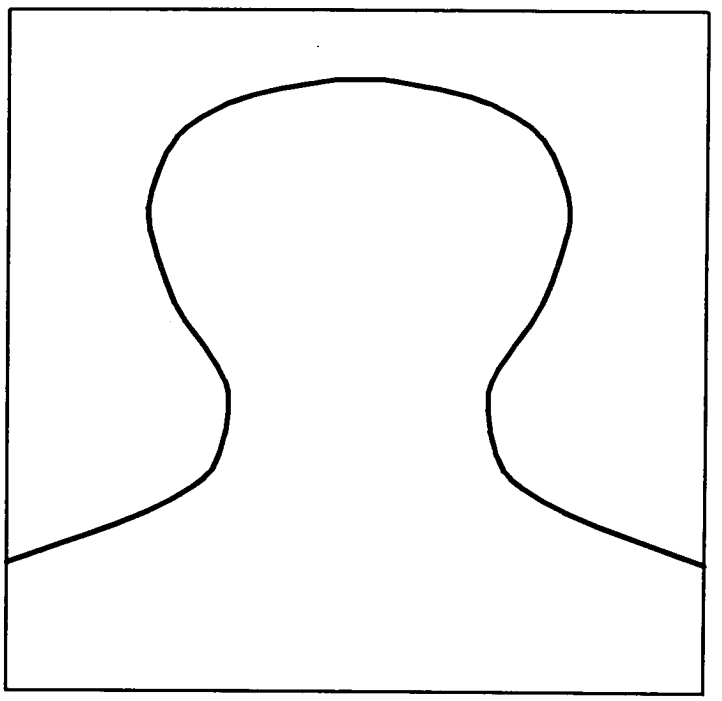

Figure 3–1: Head and shoulders template.

Initially, the image is spatially sub-sampled, a number of times, to produce a very coarse representation of the face. The primary face detection is then performed on this low resolution image. The location found for the face in the coarse image, is then used as a guide to the location of the face within the full resolution image. Computational requirements are manageable, as the dimensionality of the search process is dramatically reduced by the sub-sampling process.

In Kelly's study, the initial image size was 266 by 325 pixels. A sub-sampling factor of eight was utilised to obtain the coarse view of the face. A gradient edge detector was then applied to this lower resolution (28 by 40 pixels) image. A binary output was produced by subjecting the edge information to a threshold.

This smaller image represents a slightly smaller visual area, as the initial image size is not evenly divisible by the chosen sub-sampling factor.

A line tracking algorithm was applied to the 28 by 40 pixels binary image. Using heuristic criteria, the line tracking function linked edges, or edge segments, together in an attempt to form a 'head and shoulders' shape. If a possible target shape is located at this lowest resolution, then this information is used to located the same shape in the next coarsest facial image (*ie* the location of the face in the 28 by 40 pixel image is used as a template for finding the face in the 56 by 80 image). This process is performed on all the intermediate image resolutions. In this way, the initial estimation of the head and shoulders is refined, until an accurate location is obtained at the highest resolution.

Unfortunately, the accuracy of this method of facial location is heavily dependent on the quality of the initial edge detection stage. If this stage is noisy, and false or ambiguous edges are introduced, as often happens, then the later location can be very easily side-tracked. Despite this disadvantage, this approach is still being used for facial location[57].

Similar to Kelly's approach, research reported recently by Govindaraju *et al* [111] identified the way in which a facial outline can be extracted from an image, as a set of line segments. Their approach was to use a model of the face, constructed from four interconnected shapes. Two arcs, of differing curvatures, were used to represent both the top and the bottom of the face; and two straight lines represented the sides of the face. Again, by searching for an edge pattern which approximated this model, the facial outline was located. Govindaraju *et al* reported that this technique had been successfully used for locating facial images within newspaper photographs.

3.2.2 The Adaptive Contour Model

The use of the adaptive contour model, or *snake*[112, 113], for facial location, is somewhat similar to the outline tracking approaches described above. In common with both of these approaches, the snake attempts to locate a plausible facial outline by identifying edge segments which could form part of that outline. However, unlike the outline tracking techniques, the snake **iterates** towards an optimal solution.

The snake is initialised to the corners of the image frame. It is then drawn towards the centre, terminating only when it has settled on a strong edge pattern. Hopefully, this pattern represents the facial outline. Again pre-processing edge enhancement is used to emphasise the perimeter of the face.

The snake consists of individual curve segments connected together. It is possible for these curve segments to form either an open ended loop – where the two end points are fixed to two image points – or it can form a closed loop. Snakes can be implemented using finite element and finite difference mathematical methods, as detailed by Waite and Welsh[114].

There is no actual *a priori* knowledge given to the snake. Thus, similar to the outline tracking approach, it is possible for the snake to get trapped on other image parts, *eg* the subject's clothing, if they could form part of a pattern of edges. For this reason, image backgrounds have to be either, removed, or chosen to be uniform in colour. The facial size, within the image frame, has also to be carefully controlled.

3.2.3 Vector Quantization

Instead of considering the image as a whole, Sexton[115] pioneered a technique in which the image is spatially partitioned into a number of sub-blocks. Each of these blocks is then analysed to determine whether it is part of a human face. The

spatial distribution of these candidate face parts should reveal the likely location of the face. This process is illustrated in Figure 3–2.

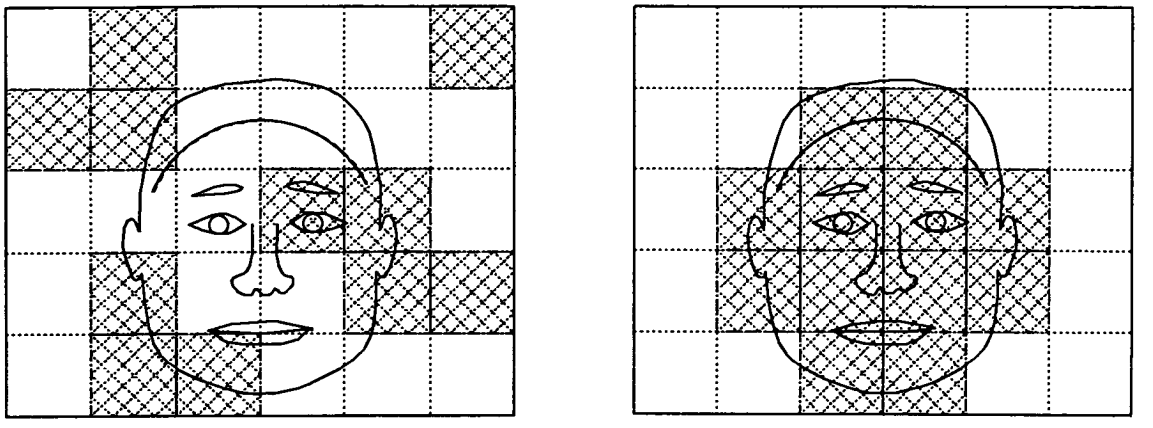

Figure 3–2: Failure and success of the Vector Quantization based face location algorithm.

Vector Quantization[1] is used as the means of deciding if a particular block is a likely facial part. This technique should be able to cope with image backgrounds as long as they do not contain patterns which could be readily confused with facial features. This technique has been successfully demonstrated on video sequences of facial images.

3.2.4 Stereoscopic Methods

The process of face detection is essentially the separation of the image data into background and foreground information. Thus, the use of three dimensional range data, in the extraction procedure, is an attractive option. Depth information will reveal which image parts are likely to be facial data, assuming that the face forms the largest single foreground object.

A system utilising this concept has been recently implemented[116]. In this implementation, a spatial quad-tree approach (a technique popular in image pro-

[1] Vector Quantization will be discussed in depth in chapter 4.

cessing [117, 118]) is used to produce the range *map* relatively efficiently. By using a quad-tree, only the image areas of interest, *ie* those image parts which contain transitions from foreground to background, are analysed in detail. Accurate 3D measurements, regarding surfaces in the image, are not obtained.

The depth map produced in this way, can then be scanned to identify the foreground image parts proximal to the camera. The transition in the data, from foreground to background, can be used to signify the facial outline. Incorporated in this algorithm, is a frame differencing stage and also some motion estimation. The accuracy of this technique is quite impressive; it is able to cope with very cluttered images, as long as a clear definition between foreground and background is preserved. Again this technique has only been tested on video sequences.

3.2.5 Feature Embedding

Conceived by Fischler and Elschlager, the *feature embedding* method of facial location exploits the concept that the face is a collection of component features[119]. This technique marks the transition between purely holistic facial location techniques and facial feature location techniques.

In their implementation, Fischler and Elschlager, defined the face as an object made-up of two eyes, a nose, a mouth, hair and left and right edges. The embedding procedure, involved the conditional placement of all these features to form a face. The location of the face is only confirmed when all the component features have been placed (or embedded) in the image, in a plausible facial pattern.

To give the placement of the features some flexibility, the features are connected by conceptual *springs*, illustrated in Figure 3–3. The features to be placed, are thus allowed to adjust slightly, in their configuration, to maximise the similarity with the target face. The springs facilitate some deformation of the initial, approximate, facial map.

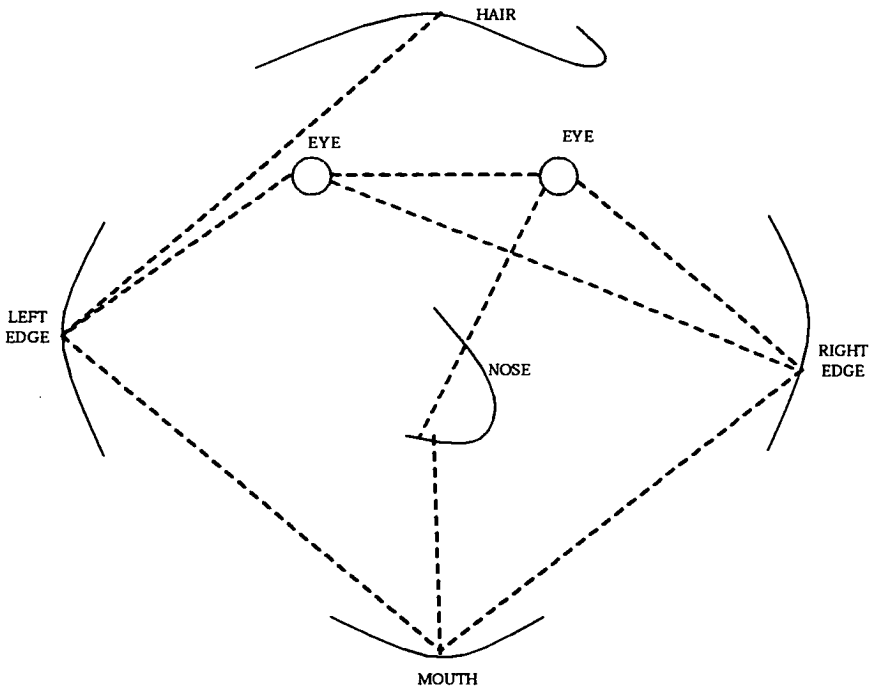

Figure 3–3: 'Springs' form inter connections between features.

To place each of the seven component features, a template matching procedure was utilised[1]. Predefined templates were stored for each of the features used. These templates take the form of pixel intensities and their inter-relationships. For example, the left edge of the face was defined as a sustained transition from light to dark in the pixel values. Most of the templates used were of a fairly crude nature.

An embedding *cost function* was used to evaluate the goodness of fit for each placement of the features. The final placement is found as a result of the optimisation of the overall, interdependent, cost function. Fischler and Elschlager implemented the optimisation as a dynamic programming problem. Their research results were very encouraging, illustrating the ability of the feature embedding function to perform successfully, even when the facial images had been substantially corrupted with noise.

---

[1]Template matching is discussed in detail in section 3.3.4.

3.3 Feature Location

The accurate location of the face is crucial to the subsequent performance of a number of facial recognition systems. In addition to this function, many systems require the exact locations of the component features of the face. However, the feature embedding technique demonstrates the *chicken-and-egg* nature of this problem. For example, if the facial outline has been successfully located then it is possible to estimate the likely positions of the internal features. Conversely, if the positions of the facial features are known, then it is possible to estimate the facial outline. However, the predictive stage involved in both of these possible approaches will, at best, be fairly inaccurate. Fortunately, there are a number of more sophisticated feature location algorithms available, which should be able to fine tune the predictive location of each feature, into an accurate placement.

3.3.1 Line Following

In a manner similar to the facial outline tracking algorithms, described above, it is possible to locate local facial features by extracting, and interpreting, edge information. Instead of searching for edges which could form a face, the techniques described here attempt to link smaller edge segments together to form the individual facial features. Again, the line segments are derived from an edge enhanced view of the face.

There are a number of different research groups who have performed work in this area[120-122]. While disagreeing over the precise methods, they all agree on the importance of identifying the features in context. The systems described in these studies all assume an *a priori* model of the face, and then search for a particular feature in a pre-determined area. The edge detection algorithms, and the actual feature criteria, vary between the different research studies.

For example, Craw *et al* [122], implemented such a technique using line segments to describe the lips, the eyebrows and the eyes. In addition to describing the line segments, they also constrained the overall size and shape of the features. For example a mouth was constrained thus -

> *A mouth has upper and lower lips, with both left and right sides. The lips may not be separated by an excessive distance and the entire shape has to have an aspect ratio within certain limits.*

If all these conditions are met, then a candidate shape would be accepted as a mouth. A more complex system of line following, based on Craw *et al*'s research, has been recently proposed by Robertson and Sharman[123].

The systems mentioned here, are all able to perform facial feature location at a reasonably accurate level. However, all these system are prone to error if the edge detection process introduces any spurious edge information.

3.3.2 The Hough Transform

The Hough transform[124] maps spatial information into a new coordinate system, or feature space. The transformation used, relates to the characteristics of the shape being sought. In the new feature space, the specified shape can be more easily located. The process involved is described below.

As each pixel is subjected to the transformation, accumulation occurs, at a particular point in the new feature space. Accumulation will only occur if the pixel under examination could form part of the specified shape. Thus, it is possible for the Hough transform to locate shapes even when they are incomplete, or noisy. The advantage of the Hough transform is that the search process is constrained to a relatively compact feature space. The Hough transform, or the generalised Hough transform[125], can be successfully used to locate complex spatial shapes even when substantial parts of the shape being sought are missing.

Nixon[52] used the Hough transform for eye location. Initially, the image was edge detected, then the Hough transform was used to locate two shapes. One shape is a circle describing the iris, and the other is a oval representing the sclera (or white) of the eye. The shapes used are illustrated in Figure 3–4.

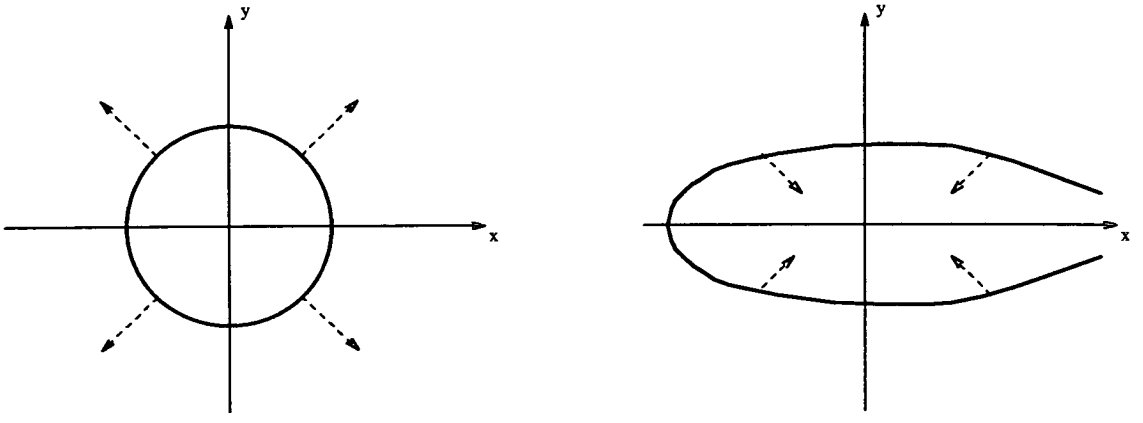

Figure 3–4: Shapes sought with the Hough transform.

On a test set of six images, correct eye location, to within ± 1 pixel, was achieved in all cases. However, the computational requirements of the transformation process are considerable.

3.3.3 Edge Grouping

Research by Nagao[126] suggested another method of grouping edge information in order to detect features. Instead of attempting to locate line segments directly, Nagao grouped the edge information, in particular spatial areas, into histograms. The analysis was performed on binary, edge detected, facial images.

For example, to locate the edges of the face, a scanning window eight pixels high, and the width of the image, is passed down over the entire image. At each vertical location, the number of edges within each column is summed. Thus, at each position of the scanning window, a histogram is produced which describes the edges present within that window. The location of two high values in the

histogram, on each side of the central axis of the face, signify the likely positions of the sides of the face. The edge grouping process is illustrated in Figure 3–5.

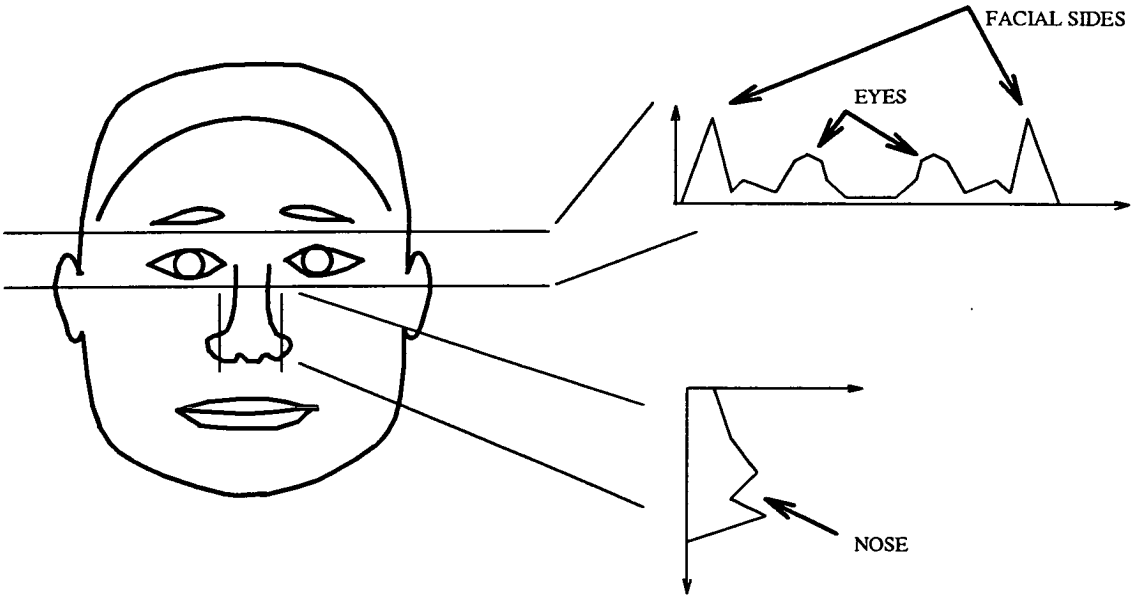

Figure 3–5: Location of eyes, nose and facial sides.

Nagao argued that each of the features (the eyes, nose, mouth and chin) has a characteristic pixel pattern, when recorded in either a vertical, or horizontal, histogram of the relevant facial part. Thus, using the likely positions of the facial sides as a guide, a number of other searches can be initiated for the other facial features.

Nagao reported a feature location success rate in excess of 90%, for images without glasses, and without beards. In common with the last two techniques described above, it also relies on the quality of the pre-processing stages to maintain this level of performance.

3.3.4 Template Matching

Template matching is a conceptually simple process, involving the comparison of a stored feature (or template) with a test image, or image part. The comparison is performed by *convolving* the template with the entire image. At each pixel

location, the similarity between the template and the image, is measured. The highest similarity, signifies the most likely placement of that template, within the image.

Hutchinson and Welsh[127, 128] compared the use of template matching and neural networks on the problem of eye location. To obtain a representative eye *template*, they averaged sixteen example eyes together. The template matching procedure involved convolving this average template with a particular search area in a number of facial images. Successful location occurred if the best match position, denoted by the pixel location with the highest similarity to the template, was within $\pm$ 2 pixels of the actual eye position, as judged by a human viewer. On a test population of 44 images a success rate of 91% was obtained. The results of their neural network experiments will be discussed in a later section.

3.3.5 Deformable Templates

Conventional template matching, as described above, can only function correctly, if the feature being sought, does not differ significantly from the standard templates used to locate it. As this is not always the case, Yuille *et al* [129] have devised a system of *deformable templates* which can be used to locate image features which are substantially different to the standard templates. Again, their research concentrated on eye location, although, the technique can be readily used for other facial features. Their algorithm is detailed below.

An initial general shape, or template, is assumed for the eye (Figure 3–6). This template is then placed near to the target eye. This process may involve the use of a different feature location technique to provide a starting point for the deformable template search. The shape of the initial template is then progressively deformed, to fit the eye in the image. An energy function controls the deformation of the template, constraining its geometric shape to certain predefined limits. These constraints are included so that the template does not lock onto

any other feature present within the image. The iterative search strategy, used by the deformable templates approach, is very similar to the adaptive contour model discussed in section 3.2.2.

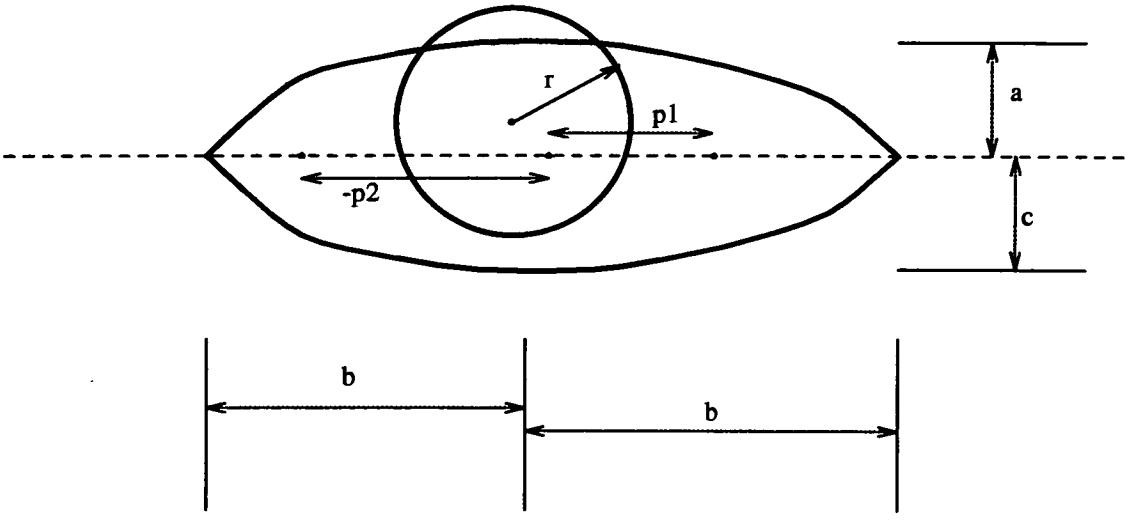

Figure 3–6: A basic geometric shape assumed for the eye.

A more general deformable template approach has been demonstrated by Bennett and Craw[130]. Their research illustrated the location of the eyes, the mouth and the facial outline, using *a priori* statistical knowledge, as a guide to the deformable template algorithm.

The performance reported by both research groups is impressive, locating facial features under differing lighting conditions, image scales and rotations. However, the computation involved is considerable.

3.3.6 Neural Networks

The *Multi Layer Perceptron (MLP)* neural network has been widely used for pattern recognition in image analysis[131-134]. The fundamental task performed by the MLP is that of data classification. However, as a full description of the MLP and its *back propagation* learning algorithm is given in[135]; and a good

introduction to neural networks, and the MLP, is given by Lippmann in [136], only a brief description of the operation of the MLP neural network is given here.

The MLP performs supervised learning, *ie* the network is *trained* to produce particular target responses for specified stimuli. During the learning process, the MLP constructs an internal representation of the patterns it is being taught to recognise. For example, research by Hutchison and others[137, 128, 138], attempted to train an MLP to recognise eyes by presenting it with a number of examples of pixel patterns representing eyes. To reinforce the internal representation, negative training data (*ie* pixel patterns representing objects which are not eyes must also be presented to the system). Using this training data, the MLP is able to construct, in its internal *weights*, a general picture of what an eye looks like. If an MLP trained in this way is then presented with a number of pixel patterns from a facial image, it should respond most positively when an *eye-like* pattern is presented. Unfortunately, the standard back propagation training algorithm, as used here, is prone to to fail if the network finds a partial solution to the classification task.

Hutchinson's research produced a successful location rate of approximately 90%, on a test set of forty-four images. This technique equalled the performance of a conventional template matching process, on the same data. A variation of this work involved the use of a Kohonen Self Organising Feature Map, coupled with the MLP. However, this approach did not produce any significant alteration in performance.

Other researchers[139, 140] have recently suggested ways in which a number of different neural network architectures could be used to locate facial features, however, they have not yet managed to produce significant results.

3.4 Brief Discussion of Approaches to Face and Feature Location

A number of different face, and feature, location techniques have been described in the preceding sections. Considering the face location schemes first, it is possible to identify a number of problems associated with each approach.

The first two approaches to face location, outline tracking and the adaptive contour model, rely heavily on an initial pre-processing edge detection stage. Inherent in this stage is the introduction of noise or misleading information. Spurious edges are caused primarily by background information, but can arise from other sources. The ability of these two techniques to be misled in their face location by this additional information, severely limits the functionality of these systems in a real-world (less constrained) environment.

The Vector Quantization technique of face detection does not suffer from this problem. However, the performance of this system has not been proven. This is also true of the stereoscopic approach described in section 3.2.4. Although the exploitation of 3D information in order to segment foreground faces from cluttered images is a promising concept, the stereoscopic approach has the practical disadvantage of requiring additional imaging equipment.

The feature embedding approach has the ability to extract a face, using its component features, in a noisy or cluttered environment. However, the computational load of the optimisational search procedure is a significant disadvantage. In addition to the computation problem, the feature embedding algorithm, as presented by Fischler and Elschalger, used a very simplistic feature location technique to perform the final feature placement. A number of more sophisticated feature location techniques are available.

Similar to facial outline location algorithms mentioned above, the line tracking and the edge grouping techniques are also dependent on accurate edge extraction. Both systems can fail dramatically when given noisy or cluttered images. For this reason, edge based face and feature location schemes have been rejected for the purposes of this study.

The deformable templates algorithm represents an elegant solution to feature location. Unfortunately, the search process involved can be easily side-tracked if not initialised in close proximity to the target feature. Again, a computationally intensive iterative search strategy is required. Deformable templates, are not suitable for feature location in a situation where only a crude prediction of the possible feature location is available.

The Hough transform approach to facial feature location has been shown to perform well where, like the eye, it is possible to specify the object being sought as a number of simple geometric shapes. However, it remains questionable whether this approach can be readily extended to locate the variety of shapes present within the other facial features. In addition to this factor, the transformation stage of the Hough algorithm represents a substantial amount of computational effort. Thus, if many different primitive shapes are being sought from within an image, the computational requirements may be excessive.

The template matching procedure and the MLP approach to feature location have similar documented performance and computational complexity. Both approaches will fail, if the feature being sought is substantially different from those examples used to train the system, although, if supplied with good training data, both approaches can perform reasonably well.

None of the face location schemes, described here, are able to provide accurate facial feature location, with a manageable level of computational complexity. The development of a novel facial location algorithm is thus essential, to achieve the goal of a low computational requirement, facial recognition device.

3.5 Limited Feature Embedding

Acknowledging that a requirement of a realistic facial recognition system is the ability to isolate faces from complex, or cluttered, images; the feature embedding approach appears to be the most promising. The advantages of the feature embedding algorithm include its ability to deal with complicated images and the fact that it does not require the input images to be preprocessed in any way.

However, there are two substantial drawbacks to the feature embedding algorithm. Firstly, a large computational load is required by the optimisational search procedure. Secondly, the templates used to locate the individual features are of a very limited nature and could therefore be easily misplaced.

By using a smaller set of more distinctive features, the optimisational search requirements could be reduced substantially. However, to maintain the performance of the original feature embedding algorithm, the accuracy, and certainty, with which each of these component features is located, must be increased. Thus, by incorporating one of the more sophisticated feature placement techniques, described in section 3.3, to a system of feature embedding based on a reduced set of features, the advantages of the original feature embedding scheme can be maintained. This algorithm would represent a more attractive and realistic solution to the problem of facial location.

The novel technique of *Limited Feature Embedding (LFE)*, proposed in this thesis, draws on the strengths of the original feature embedding algorithm, while still maintaining a level of computational complexity suitable for future hardware realisation.

3.5.1 Algorithm Overview

In order to reduce the computational requirement of the search stage of the LFE algorithm, it is essential that the number of component features, used to located the face, be kept to a minimum. The limited feature embedding algorithm, is thus based on a sub-set of the seven features used by Fischler and Elschlager (listed in section 3.2.5).

The LFE algorithm uses the conditional placements of only the eyes and the mouth to isolate the position of the face. The sides of the face have been rejected because of their dependence on the image background. The nose is not particularly distinctive and has been rejected for that reason. The variability of the position of the hair line and the lack of distinctiveness associated with it, make it a difficult feature to locate and it has also been rejected.

The LFE algorithm requires that the eyes and the mouth be located in a plausible interrelationship, an approximate triangle, before a face is deemed to be present within the image. A conditional placement of these three features, which conforms to the prespecified shape, is termed a successful *embedding.*

By using only the locations of these three features to signify the presence of a face, the computational requirements of the limited feature embedding algorithm are substantially less than those of the Fischler and Elschlager algorithm. The necessity that these three features appear in a plausible configuration, should prevent false location when no face is present within the image.

3.5.2 Implementational Details

The limited feature embedding system of facial location has been implemented in software form. The details of the algorithm, and consideration of its computational requirements, are given in the following sections.

Embedding Strategy

The limited feature embedding algorithm utilises a conditional search strategy in order to place the three target features in their correct locations. This process requires substantially less computational power than the optimisation approach adopted by Fischler and Elschlager. The algorithm uses crude predictions of the likely feature positions, to guide the local feature placement algorithm. The embedding strategy is as follows.

Step 1: Search the image for a possible right eye shape using a suitable feature location algorithm[1].

Step 2: At each likely eye position, initiate a new search to locate the matching left eye using a prediction of the probable location of that left eye, to constrain the local feature placement.

Step 3: If no corresponding left eye is found then return to **Step 1**, otherwise proceed.

Step 4: Using the possible positions of both eyes, initiate a heavily constrained search for a matching mouth.

Step 5: If no mouth is found then return to **Step 1**, otherwise, store this possible embedding of the eyes and mouth, and then continue the right eye search, *ie* return to **Step 1**.

In this way, a face is only acceptable if all three features are present, and are configured in a plausible manner. However, in the unlikely event that an incorrect embedding, or more than one embedding, is found in the image; then all of these likely faces (*ie* the different feature embeddings) are stored for further comparison.

---

[1]The *right* eye refers to the subject's right.

In practice, the complete search for the right eye is performed first. This search is constrained to a sub-part of the image within which a right eye can be located, in such a way, that a complete face can be present within the image. This search area is illustrated in Figure 3–7 – the figures given in this diagram represent pixel distances.

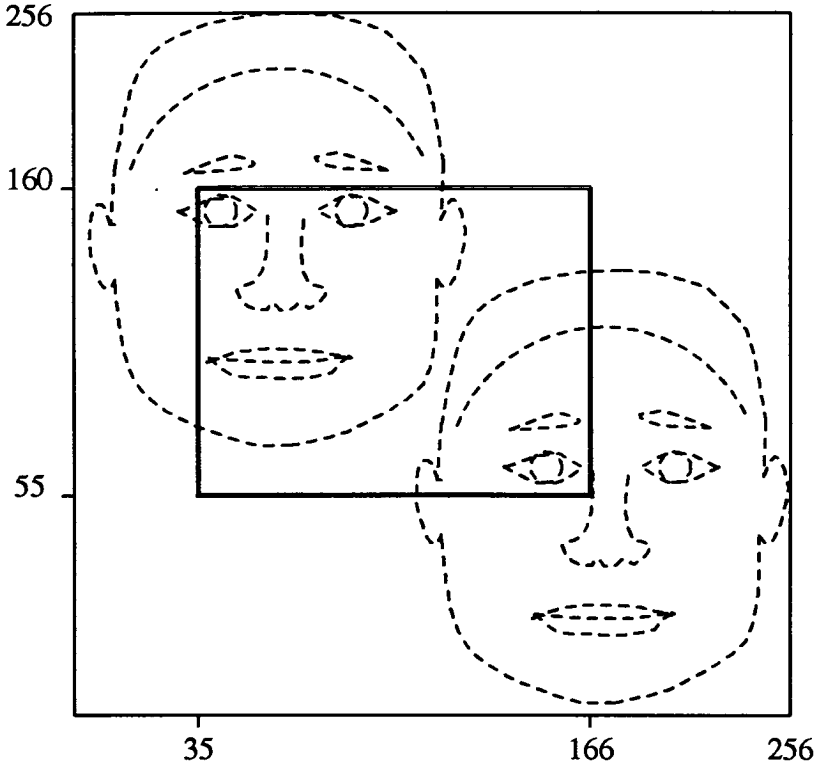

Figure 3–7: The right eye search area.

It is likely that the feature location stage will produce positive results for several pixel locations surrounding the actual feature placement. If this problem were to occur during the initial search for the right eye, then several, very similar left eye searches would have to be performed. In order to alleviate this problem, a simple pixel clustering algorithm, detailed in Appendix A, has been employed after each feature search has taken place.

As described above, a prediction of the likely position of the left eye, is used to constrain the size of the local search. The prediction is based on the location of the right eye, and also on the maximum and minimum face sizes it would

possible to fit into the image frame. Figure 3–8, illustrates the area used for the left eye search. The specified spatial configuration of the eyes, allows for only a very little rotation of the face, and places some restriction on the facial size. These restrictions are necessary to reduce the number of false faces located. Unfortunately, these measures do place constraints on the flexibility of the image capture stage of the entire system.

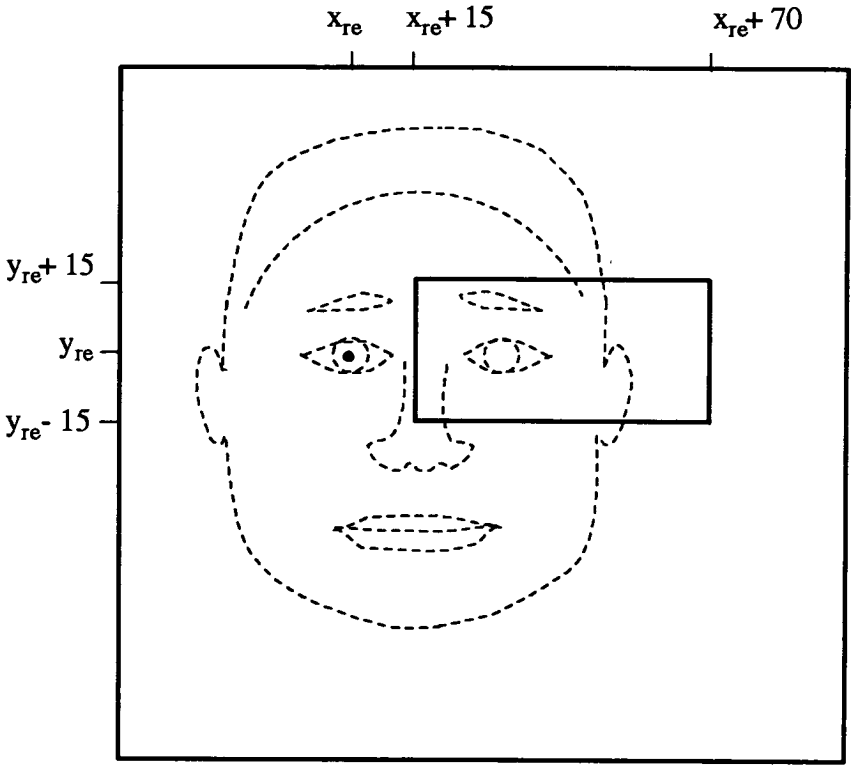

Figure 3–8: The left eye search area given a position of the right eye.

Assuming successful location of the eyes, the mouth search can be heavily constrained to a small area of the image. The actual area of this search is determined by two measures of the feature inter-relationships R_1 and R_2, as given in Equations 3.1 and 3.2.

$$R_1 = \frac{x_{le} - x_{re}}{y_m - (y_{re} + y_{le})/2} \tag{3.1}$$

$$R_2 = \frac{(x_m - x_{re})/2}{x_{le} - x_{re}} \tag{3.2}$$

where -

x_{re}, y_{re} are the coordinates of the right eye.

x_{le}, y_{le} are the coordinates of the left eye.

x_m, y_m are the coordinates of the mouth.

Likely values for these two measurements, R_1 and R_2, were derived from analysis of a sample set of one hundred facial images. The ratios were obtained after the three features had been manually located. The values obtained for the sample population of one hundred faces are given in Table 3–1. The acceptable ranges of these measurements, given in Equations 3.3 and 3.4, are defined as the mean, plus and minus, three standard deviations.

Measure	Mean (μ)	Standard Deviation (σ)
R_1	0.79	0.07
R_2	0.53	0.07

Table 3–1: Statistical parameters for the sample population.

$$0.6 < R_1 < 1.0 \tag{3.3}$$

$$0.3 < R_2 < 0.7 \tag{3.4}$$

In the original Fischler and Elschlager algorithm, springs controlled the feature placements. These springs allowed the initial face map to deform, to a certain degree, to fit the features located. Conceptually, this function has been replaced in the LFE algorithm, by the constraints placed on each of the stages of the conditional search process.

Local Feature Placement

In order to perform the precise location of the three key facial features, a local feature placement algorithm is required. As concluded in section 3.4, the MLP and the template matching techniques appear to be reasonably well suited to local feature location. The documented performances of these two approaches are very similar, and hence, both approaches were further investigated to assess their suitability for this application.

Multi Layer Perceptron Learning

As described in section 3.3.6 the MLP has been successfully applied to eye location. However, this approach has the distinct disadvantage of requiring a substantial investment in time to train the network. For this reason, it was decided to investigate the use of one recent development in the field of MLP training algorithms.

As reported by Murray[141], the introduction of random noise onto the values of the synaptic weights in the MLP, during the training procedure, reduces the likelihood that the MLP will get trapped in a *local minimum* (or partial solution to the problem). If the additive noise is reduced, as the network approaches the end of its training, an optimal solution is more likely. This approach to learning is similar to the *simulated annealing* method of optimisation.

As well as introducing random noise during the learning process, Murray's algorithm uses a system of *virtual targets* in order to facilitate localised learning. This algorithm is particularly well suited to analogue VLSI implementation[71], however, a software simulation of Murray's algorithm was utilised here[1]. The algorithm, and its use in training an MLP to recognise eyes, is described below.

---

[1]The software used to perform the MLP learning was supplied by Dr A F Murray.

Unlike the conventional multi-layer perceptron, back propagation of gradient descent error signals is not performed. Instead of transmitting error signals back into the entire network, a system of local virtual targets are used. These targets are stored locally, with the synaptic weights of each neuron adapting to produce the target output. It is this aspect which makes this method of MLP learning particularly well suited to *on-chip* learning, although this has not yet been realised in hardware.

In the single hidden layer MLP, each of the input patterns, has a set of hidden layer targets associated with it. These targets are initially derived from the ideal response for each pattern. As the training takes places, these targets are gradually met. It has been shown, that with the additive noise, the network is much less likely to get trapped in local minima and the network learns more rapidly.

Once the network has been trained, it behaves in exactly the same manner as the conventional MLP. Thus, it is possible to train the network **once** to recognise the chosen feature, and then use the network weights produced to perform location.

For this study, a three layer MLP, with one hidden layer of neurons, was chosen. The input layer of 256 neurons were directly connected to the 16 by 16 pixel input matrix[1]. The first layer was connected to a hidden layer of eight neurons. This hidden layer was then connected to the single output neuron. By convention, this output is trained to give a positive response or '**1**' when presented with the target feature. In the opposite case, the desired response is a '**0**'.

The training data provided to the MLP must include examples of the type of feature it will be expected to locate. It also requires various examples of negative

[1]In practice, a template size of 32 by 32 pixels was utilised with spatial sub-sampling to 16 by 16. This reduction was required as an input layer of 1048 neurons was thought to be undesirable in computational terms.

data (*ie* image parts which do not represent the feature of interest) to reinforce the internal representation of the training feature.

To provide this training information, ten example eyes were manually extracted from different images of ten individuals[1]. These were chosen to represent as good a spread of eye types as possible. One hundred other image segments were randomly drawn from each of the ten training images, in order to represent instances of non *eye-like* shapes. The ten training eyes were repeatedly presented to the network interspaced with the non-eye patterns. This procedure allowed the network to learn in a balanced way. Details of the training parameters are given in Appendix C.

To test the system, one hundred different facial images were used. The neural network scanned the entire image in a sequential manner. At each pixel location, the present pixel pattern was used as an input to the network, the corresponding output response was then obtained. Of these responses, only the top twenty outputs were stored. It was hoped that the correct location of the eye would appear within these top twenty positions.

In order to assess the location performance, the eye placements suggested by the MLP were compared with the correct feature position (as determined by a human viewer). Table 3–2 shows the location accuracy of the neural network on the test set of images. In this table, the pixel error term refers to the tolerance at which the location failed, *ie* a location failure rate of 92% was observed for a location accuracy of ± 1 pixel.

Template Matching

There are a variety of different distance measures which can be used to objectively assess the numerical level of similarity between two different pixel patterns. This

[1]The images used for this experiment, as for all the experimental work reported in this thesis, were captured in accordance with the operating conditions described in Appendix B.

Pixel Error (±)	Failure Rate (%)
0	100.00
1	92.00
2	86.00
3	72.00
4	62.00
5	60.00

Table 3–2: Performance of MLP eye location.

measurement of similarity is an essential part of the template matching process, enabling the algorithm to establish the best placement of the chosen pixel pattern within the image. Although these measures are able to quantify the numerical difference between two image parts on a pixel to pixel basis, they cannot mimic the human ability to assess visual similarity, a point discussed by Nightingale[142].

The correlation coefficient (Equation 3.5) is thought to be the most robust and reliable measure of signal similarity[143]. Unfortunately, the computational requirements of this technique make it unattractive. One of the next best techniques is the absolute difference measure (Equation 3.6), which has been demonstrated to perform almost as well as the full correlation technique, if the template and the image have been normalised in intensity[144].

$$R(i,j) = \frac{(\sum_{l=1}^{N}\sum_{m=1}^{N} T(l,m)I(i+l,j+m))^2}{[\sum_{l=1}^{N}\sum_{m=1}^{N} T^2(l,m)][\sum_{l=1}^{N}\sum_{m=1}^{N} I^2(i+l,j+m)]} \tag{3.5}$$

$$S(i,j) = \sum_{l=1}^{N}\sum_{m=1}^{N} |T(l,m) - I(i+l,j+m)| \tag{3.6}$$

where -

$T(l,m)$ is the pixel value of the template at coordinates l,m.

$I(i,j)$ is the pixel value of the image at coordinates i,j.

N is the number of pixels in each axis - assuming a square template.

The image normalisation is performed by first calculating the mean (μ) and the standard deviation (σ) for the image patch under test. These statistics are calculated using the standard formulae, given in Equations 3.7 and 3.8.

$$\mu = \frac{\sum_{i=0}^{N}\sum_{j=0}^{N} I(i,j)}{N^2} \tag{3.7}$$

$$\sigma = \sqrt{(\frac{\sum_{i=0}^{N}\sum_{j=0}^{N} I^2(i,j)}{N^2}) - \mu^2} \tag{3.8}$$

Estimates of μ and σ are obtained by only analysing a small area of the template. In this implementation, a sampling factor of one in sixteen is used. Thus, for a template size of 16 by 16 pixels (or 256 pixels in total), only 16 of these pixels will be used in the calculation of the statistical parameters for that template.

Normalisation is performed by subtracting the mean from each sample point and then dividing by the standard deviation. This function rescales the data, such that the mean is zero and the standard deviation is one. This process should remove any scale difference in the data due to intensity variations.

When using an absolute difference, it is possible to reduce the computational load by using a threshold on the accumulated difference. If the difference is accumulated as the template is compared with each image part, then the larger this number becomes, the likelihood of this image part being the feature of interest decreases. If the search is terminated when the total difference accrued reaches a pre-determined level, then less time is wasted analysing image parts which clearly do not represent the feature of interest. Care is required in selecting this threshold so that the algorithm does not miss the correct feature location.

It is possible to refine this technique further by pre-arranging the order and the manner in which the pixel elements of the template and the image are compared[145, 146]. For example, if there was a very significant region of difference between part of the template and all other parts of the image, then it would be efficient to compare this part first. However, this approach does not appear to be suitable here, as the features being sought are not sufficiently different from the rest of the face.

To facilitate an accurate comparison between the template matching and the MLP methods of feature location, the training images selected to train the MLP were also used for the template matching trial. However, the inverse (or negative) information was not required by the template matching process. Again, a template size of 16 by 16 pixels was used. The threshold value at which the matching process stopped was determined using a heuristic approach. In this case, a value of 150 was selected.

The same test population of 100 hundred facial images was again employed. Each of the test images was scanned to locate the right eye. At each pixel location, the present image patch was compared with all of the ten right eye templates. Of these ten templates, the score of the least different template was examined. Only those locations with an absolute difference of less than the threshold were stored and of these, all but the best twenty were discarded. These potential eye locations were again compared with the correct locations, to assess the accuracy of the locations suggested. A comparative graph of these results, with the results of the MLP experiment, is shown in Figure 3–9.

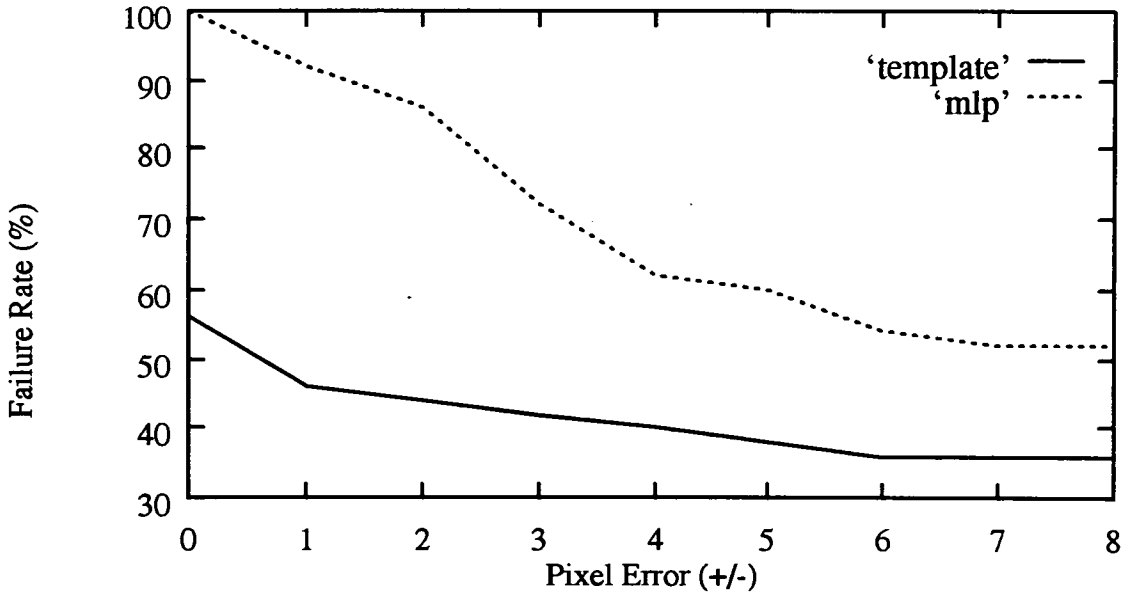

Figure 3–9: Right eye placement accuracy using MLP and template matching approaches.

On the basis of these experiments, it was decided to use template matching as the local feature location stage of the limited feature embedding algorithm. The reasons for this decision are summarised below :

- The template matching approach easily out-performs the MLP at eye location.
- The substantial investment in time required to train the MLP, is not required by the template matching approach.
- The future hardware implementation of the template matching algorithm, appears to be more feasible than the MLP.

Further reductions in computational requirements

This proposed system of face detection – performed by limited feature embedding .linked to local template matching – still represents a substantial amount of computation. This load must be reduced if the long term aim of a real-time system is to be realised. To this end, the use of spatial sub-sampling has been investigated.

Rosenfeld and Vanderburg[147] suggested a two-stage approach to template matching, termed *coarse-fine* template matching. In this algorithm, the template matching is initially performed using a low resolution version of the template. Full template matching, at the finest resolution, is only performed on those image locations which were identified as possible candidate sites, during the first stage. The low resolution view of the image was obtained by sub-sampling the template, and the image. The amount of sub-sampling is crucial to the minimisation of the computational cost[148]. A random sampling of the template has been demonstrated as more effective than a systematic sub-sampling in some cases[149].

In the LFE algorithm feature search stage, the comparison between the template and the image, is only performed on one in every four pixels in each axis(*ie*

a sub-sampling factor of two is used). In the final location stage, only invoked if there are several possible faces present within the picture, the target features are analysed at the full spatial resolution. This process is a crude example of coarse-fine template matching. A random sampling approach has been rejected because of the extra computational requirements of generating the random pattern.

Further computational reductions are obtained by only considering one in every four possible sites for the target feature, *ie* the search space is spatially sub-sampled as well as the template. However, if the template matching process returns a positive result for a given site, then the other three possible sites in that locale are also analysed. In this manner, the search efficiency is improved by wasting less effort analysing image areas some way distant from the target feature.

3.5.3 Results

The entire limited feature embedding algorithm has been implemented via software emulation. As input, the system requires a number of example features, to use in the template matching procedure, and the image data containing the face under test. The output of the system is a set of coordinates, for each of the three features, representing the best feature embedding. Again, the coordinates obtained have been compared with the exact feature placement to assess the system performance. The results of a number of different experiments are given below.

Template Sizes

Within the template matching stage of the LFE algorithm, there is a trade-off between the size of the template and the location performance. For example, if the template used is too large, then the location obtained for that template may be affected by spurious information not related to the feature of interest. However,

using a template size that is too small, the information it contains may not be distinctive enough to locate the target feature. The size of the template used is also critical to the computational requirements of the algorithm. To evaluate these factors a number of different template sizes have been generated and used within the full LFE algorithm.

A sample population of forty pictures, representing twenty individuals, has been obtained. For each of these twenty people, one additional image was captured to obtain example templates. Thus, the template matching stage of the LFE algorithm was performed using twenty different examples, for each of the three facial features sought.

The template sizes used were 48, 32, 24, 20, and 16 pixel square. For each of the different sizes, a different template matching threshold was required. An approximately linear relationship between template size and threshold was assumed. The thresholds, and program compute times, for these experiments, are given in Table 3–3. The compute times given refer to the number of seconds of *cpu* time required to perform the complete LFE algorithm on a Sun 4/25 ELC Sparcstation. The times quoted are the averages of ten different runs through the same test data.

Template Sizes Pixels	Threshold Value	Compute Times Seconds
16	40	30
20	60	23
24	90	40
32	155	63
48	345	107

Table 3–3: Thresholds and program compute times for various template sizes.

Figure 3–10 shows the location performance of each of the template sizes,

related to the pixel error associated with the placement of each feature. In this figure, a failure rate of 10% at ±2 pixels states that at least one of the three embedded features, has been misplaced by up to ± 2 pixels (from the manually obtained correct position) in four out of the forty test images.

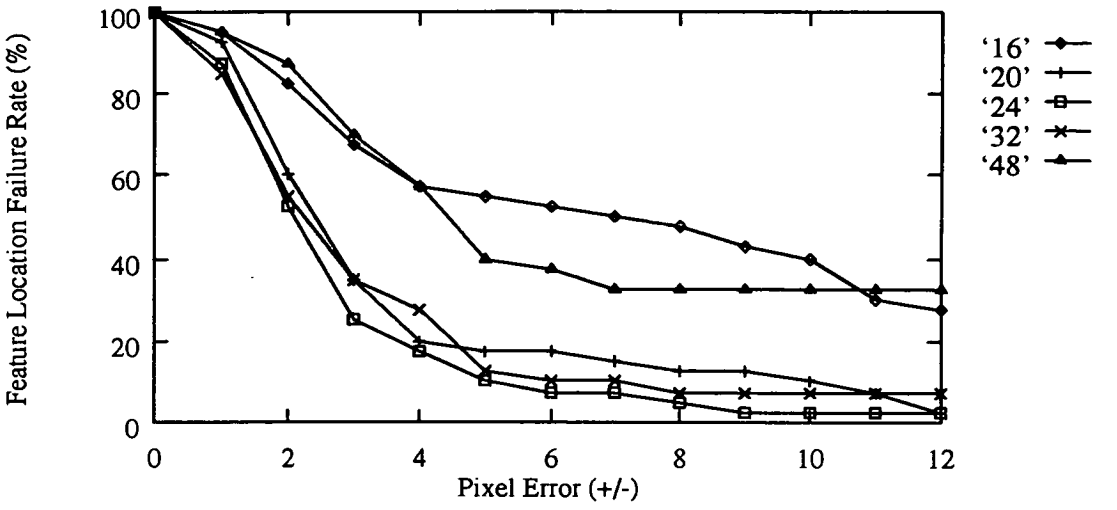

Figure 3–10: Feature placement accuracy using different template sizes.

From the graph, 48 and 16 pixel templates can be rejected as poor performers. The other three sizes of 32, 24 and 20 pixels are much more similar in their success rates. Of these three, the 24 pixel size has been selected as the best compromise between performance and compute time.

Populations

Dedicated and non-dedicated template sets

The previous experiments have all used example templates from all of the members of the test population. While a *dedicated* template set of this nature is not impossible for a viable system, it is necessary that the performance of the LFE algorithm, using a non-dedicated set of templates, be assessed.

For the first set of experiments, the entire test set of images available (*ie* twenty images of forty individuals or 800 images) was partitioned into two halves. A

template set of twenty example features were then obtained for each group of twenty individuals. Four experiments were then performed using each of the four possible combinations of test images and template examples. The four experiments are given in Table 3–4. All other factors were maintained at constant levels for this series of experiments.

Experiment Number	Test Population	Example Population
1	Group 1	Group 1
2	Group 1	Group 2
3	Group 2	Group 1
4	Group 2	Group 2

Table 3–4: Combinations used for population experiments.

Results for these experiments are given in Figure 3–11 – the numbers refer to experiments given in Table 3–4. Not unsurprisingly, the use of a dedicated population is advantageous; the observed average error rate of 13% is substantially less than the 22% of the non-dedicated experiments (both figures are at the ± 5 pixels level).

Large populations

As demonstrated by the previous experiment, a small number of feature templates may not be able to capture the possible diversity present within a particular population. If a very large population was used, then the number of templates required to produce a dedicated template set would put severe computational demands on the feature location stages of the LFE algorithm. Thus, it would be desirable to reduce the number of feature templates used. In order to simulate the possible performance of the system, in this situation, the following experimentation was performed.

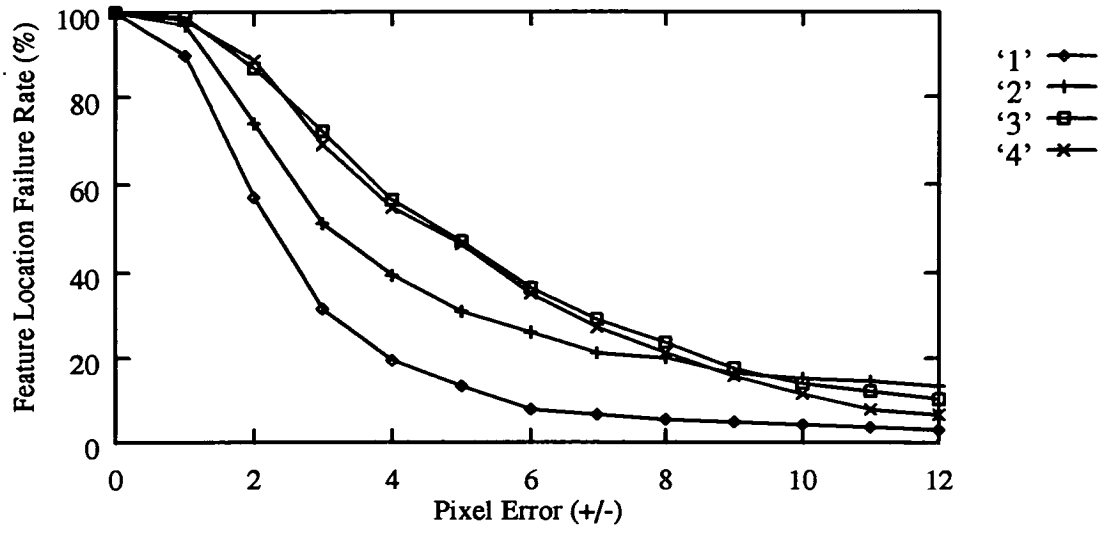

Figure 3–11: Feature placement accuracy for different populations experiments.

A test population of 300 images from 30 individuals was used. However, only five example templates were available for each feature. The five templates used were drawn from five different individuals **not** in the test set. The performance of the LFE algorithm was assessed in the same manner as for the other LFE experiments.

The results of this experiment, shown in Table 3–5, reveal a substantial decrease in system performance when compared to the experimentation reported above. An accurate comparison of the different error rates cannot be performed as the experiments use different training and test data. This experiment thus serves only as an indication of the likely performance degradation.

The inability of a template set of five examples to adequately capture the variability in feature types present within a population of only thirty people, demonstrates the importance that must be given to training the LFE algorithm. In this instance, it appears that a set of templates drawn from only five people, does not supply the LFE algorithm with sufficiently general information about possible feature variations within a larger population.

Pixel Error (±)	Failure Rate (%)
0	100.00
1	99.00
2	91.00
3	79.33
4	63.67
5	52.00

Table 3–5: Performance of reduced template set experiment.

3.6 Summary

A number of different facial location and facial feature location algorithms have been discussed. For a variety of reasons, none of these algorithms meet the specified requirements of this reseach work. A novel technique of Limited Feature Embedding, which may be suitable for future hardware implementation, has been introduced.

The performance of the LFE algorithm, in a number of different operating situations, has been analysed. The importance of the training data supplied to the algorithm has been identified. Whether the facial location accuracy achieved by this technique is sufficient to facilitate facial comparison will be considered in a later chapter.

Chapter 4

A Novel Method of Facial Parameterisation.

4.1 Introduction

As discussed in the review chapter, the decomposition of the face into its salient features is the underlying requirement of a dedicated facial recognition technique. The previous chapter described the various means whereby a face could be segmented, and the individual features located. This chapter, goes on to demonstrate the way in which these features can be reduced to form a small data set describing that face. This data set, drawn from the face, is the *feature set* which can be used to classify faces into different groups, and ultimately, recognise a particular individual. In order to determine how to perform this reduction, from facial image to data set, further consideration is given to the way in which the face is constructed from its constituent parts. On the basis of this analysis, a novel transformation of the face, from its initial pixel pattern into a low dimensionality data set, will be derived.

This chapter describes the concept of vector quantization as applied to image data compression. The use of vector quantization for facial features will then be introduced and a link to the police *photofit* technique will be established.

Various techniques for the generation of vector quantizer codebooks will be discussed, and the associated distortion measurements will also be investigated.

Finally, the construction of a complete system of facial parameterisation, incorporating facial feature and facial measurement information, will be described.

4.2 Facial Parameterisation

Fundamental to a face recognition system, and indeed to any pattern recognition process, is the accurate selection of the important characteristics within the pattern to be recognised[150]. These characteristics (or features) are selected to distinguish the chosen pattern from all others. For example, to recognise a square, the target pattern must have four equal length sides, each at 90 degrees to its adjacent sides. This kind of pattern description can be used to separate squares from all other shapes (*ie* to classify shapes into groups containing squares and non-squares). For face recognition, it is necessary to produce a set of descriptions which can be used to differentiate between facial images taken from different people.

The features used to compare faces, must be specially selected to capture the distinct ways in which faces vary; *ie* the feature set used to describe each face must include the variations in the fundamental facial characteristics which allow all of us to have slightly different faces. Thus, the underlying requirement of facial parameterisation is the selection of the correct features, which will preserve the *salient* characteristics present in the original face. There are, however, several other major factors to be considered when selecting features for an automatic system.

Certain features can vary substantially under particular circumstances. For example, an open mouth appears significantly different from a closed mouth. In this situation, an automatic system, which is not able to reconcile these two different views of the mouth, may well fail to perform correct recognition. Thus, when selecting features to be used for recognition, an attempt should be made

to incorporate features which are less prone to this type of variation (*ie* those features which exhibit a high level of constancy).

The final factor, and of equal importance in an automatic system, is the ability of the system to reliably locate, automatically, the chosen features from a large number of possible faces. A high level of accuracy is required because any *mis-registration* in the feature location stage would be likely to have an adverse effect on the later comparative analysis.

The selection of the features to be used in coding the facial information into its characteristic data set (*ie* the facial parameterisation process) must take account of the three factors described above. In practice, it is likely that there will have to be a compromise between, a set of very discriminative facial features, and the reliable acquisition of these features from a face. The following section will consider the specific facial characteristics to be included in this feature set.

4.3 Approaches to Feature Selection

Chapter 2 reviewed a number of facial recognition methods. For each of the systems described, there was an underlying set of features used to distinguish between different individuals. With reference to human facial recognition, it was concluded that a likely route for facial recognition research is a hybrid approach, linking local feature detail to parallel information about the structure of the face. A recognition system based on this approach, would use the first-order features which specify feature data, coupled to the second-order features or facial measurements.

The two level approach described above, indicates the **types** of features which can be used in the recognition process. It does not, however, suggest which particular features (within a feature type) are of primary importance. For example, establishing that measurements are important, does not solve the problem of

whether to include, either the distance between the eyes and the mouth, or the nose and the mouth, or both, or neither. Thus, some system of ranking the relative importance (or *saliency*) of the different facial characteristics still has to be arrived at.

4.4 A Novel Feature Set

It is necessary to refine further the general conclusions of chapter 2 in order to establish a novel feature set. This feature set must take account of the three factors described in section 4.2; the feature set must capture first and second order characteristics; the characteristics must exhibit a high level of constancy and it must be possible to locate the chosen features accurately, using an automated technique. In order to arrive at this feature set, each of the obvious facial features have been considered.

eyes The importance of the eyes, in the recognition process, is undeniable. The only problem is that their appearance can change dramatically when they are opened and closed, however, their exclusion from a face recognition system for this reason alone would be foolish.

mouth Again the mouth is of undeniable significance and likewise it can not be excluded because of its lack of constancy.

ears When present, the human ears can be very distinctive, however, in many cases the ears are not visible at all. For this reason, they must be excluded from the proposed feature set.

hair The position of the hair line and the hair styling and colouring are very distinctive, however, in biometric terms, the hair can be rapidly and convincingly altered, to radically alter the facial appearance. It is also very difficult to capture the hair style in an accurate way.

nose In profile view, the nose is of great importance, however, in frontal facial data, it is of less significance. In practical terms, the overall shape of the nose is not very well conveyed in the frontal view.

eyebrows The eyebrows, like the hair, can be artificially altered. They can also be partially hidden by the hair and thus difficult to locate. The eyebrows have been excluded from this feature set.

chin Only the position of the chin is likely to change with expression. The appearance could be expected to remain very constant. However, the chin is not considered to be a particularly important feature for recognition.

The analysis given above, suggests the likely facial features to be included in a first order (*ie* local detail, not structural information) recognition system. On the basis of this analysis, a partitioning of the face into eight facial parts has been devised. This facial *mapping* represents the decomposition of the face into its component feature set. The mapping used is shown in Figure 4–1 – the pixel dimensions used for each template will be detailed in the following section.

Small templates representing the eyes and the mouth have been included in the new facial data set. Larger template sizes have been used in an attempt to capture the overall appearance of the hair and the chin. The nose has been separated into two different templates, one representing the bridge of the nose, and the other, the tip (or nostrils). These two templates preserve the detail of the nose but not the overall shape. An overview of the face, centred on the tip of the nose, has been included as the final feature. It is hoped that this facial template will capture some of the general structural information about the face.

The pixel areas, chosen to represent each facial feature, have been selected in accordance with a particular facial size. Thus, before using this facial mapping, the input facial image would have to be standardised, in size, to conform to this specific partitioning. The feature sizes used have been very carefully selected in an attempt to maximise the information content of each individual feature.

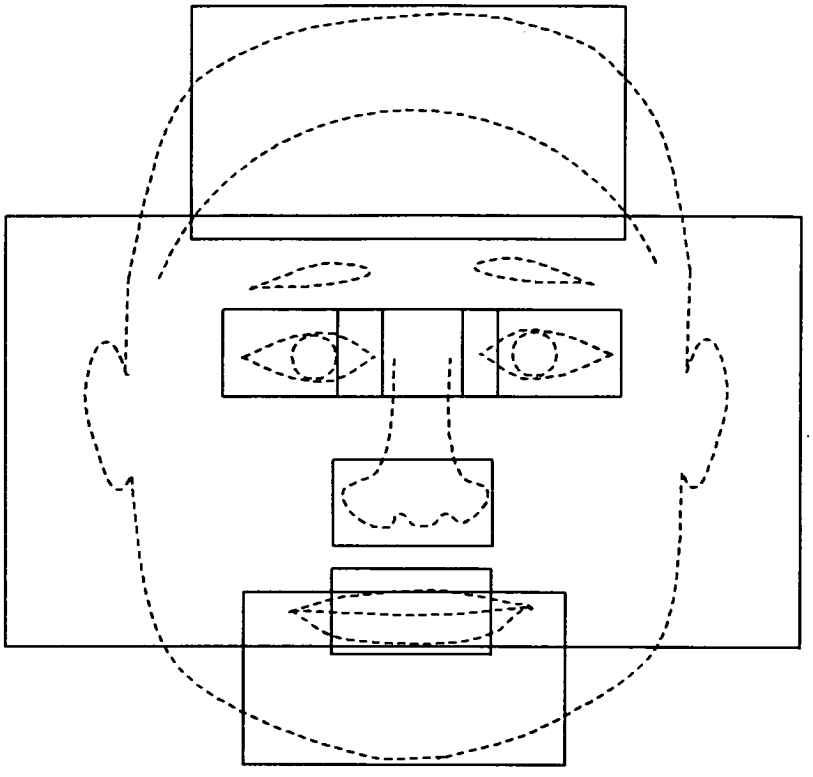

Figure 4–1: The chosen partitioning of the face.

4.4.1 Relative Importance

The partitioning of the input face into eight sub-parts, as described above, yields a data set contain eight different pixel arrays, or vectors. These pixel patterns contain the eight facial features to be used for comparative analysis. The size of each of these eight arrays is related to the physical size of the chosen feature. Thus, the array containing the face, is many times larger than that of the eye. As well as the implementational difficulties associated with this approach, these differing feature sizes do not relate to the relative importance of each of the facial features, in the recognition process. In order to standardise the template sizes used, and to relate their size to the importance of each feature, a system of differing resolutions has been devised.

By sub-sampling an image (*ie* selecting only some of the pixels within the image) it is possible to produce a smaller, coarser view of that image. For this

implementation, the eight sub-parts of the image, containing the eight features, are stored at differing resolutions. The eyes and mouth are preserved at full resolution, as are the two nose templates (the bridge and the nostrils). The chin is sub-sampled by a factor of two. The template of the hair is preserved at an interval of three pixels. Finally, the overview of the face is stored in a very crude resolution by sub-sampling one in every six pixels. The resolutions used are summarised in Table 4–1, and an example face subjected to this mapping is given in Figure 4–2. A sample set of the extracted features in shown in Figure 4–3. The features chosen, at their respective resolutions, can now all be stored in the same area of data; each template is an array of 28 by 20 pixels or 560 pixels in total.

Feature Name	Size (in pixels)	Sub-sampling Factor
Right Eye	28 × 20	1
Left Eye	28 × 20	1
Nose Bridge	28 × 20	1
Nose Tip	28 × 20	1
Mouth	28 × 20	1
Chin	56 × 40	2
Hair	84 × 60	3
Face	168 × 120	6

Table 4–1: The template sizes and resolutions.

The partitioning of the face suggested here, represents one possible way in which the face can be parameterised into a set of features. The feature sizes and corresponding pixel arrays represent an arbitrary partitioning which has been arrived at in a partially heuristic manner. However, the partitioning chosen does take account of the much of the established knowledge regarding facial recognition. The feature set produced is also well suited to a computationally constrained environment.

Figure 4–2: Original and coded views of an example face.

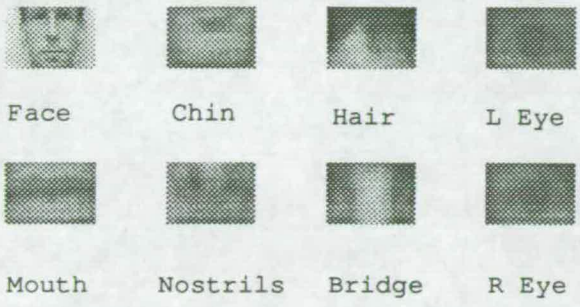

Figure 4–3: Features into which facial image is decomposed.

It can be seen, from Figure 4–2, that the decomposition of the face, into these eight features, has been performed with remarkably little loss in facial recognisability.

4.4.2 Pixel Level Feature Constancy with Time

Pixels patterns, such as the facial features selected here, have not been widely used in this kind of pattern recognition task. Of the many techniques described in chapter 2, Baron's[49] experiences with pixel correlations, give the clearest indication that facial recognition can be successfully performed using pixel level data. The full pixel level data will contain tonal information and depth cues which would be lost if binarisation of the facial image was performed. Thus, it

was decided to maintain the pixel arrays, containing each feature, in full grey scale form.

The facial features already chosen, if located accurately, may well contain sufficient information to distinguish between a number of individuals. However, it has not been shown that the pixel arrays chosen for each feature can be accurately used in comparisons between different individuals. It is likely that changes in expression and mood may well have a detrimental effect on the constancy of the spatial pixel patterns which represent each feature.

In an attempt to evaluate the reliability of this low-level greyscale pixel data, a number of experiments were performed. However, in order to perform this experimentation, the ways in which pixel arrays can be compared must first be considered.

Difference Measures

The measurement of signal differences was mentioned in section 3.5.2 in reference to template matching. Generally, such measures evaluate the overall distance between two pixel vectors, by accumulating the difference between corresponding pixels in the two vectors. In section 3.5.2 it was concluded that a normalised, absolute difference measure approximated the full correlation technique at a much lower computational cost. It was therefore chosen for the template matching procedure. As the computational requirement of this experiment is not so crucial, another difference measure has been adopted.

In order to assess the difference between two different pixel arrays, it is convenient to represent the arrays as one dimensional data vectors. The *Euclidean distance* metric measures the absolute straight line distance between such vectors, when represented as points in Euclidean space.

The Euclidean distance, D, measures the distance between the two vectors (A and B), by comparing their representative points in N dimensional space (a

and b). The formula is given in equation 4.1. In this equation, N, represents the total number of pixels in each vector, for this experiment $N = 560$.

$$D = \sqrt{\sum_{j=1}^{N}(a_j - b_j)^2} \tag{4.1}$$

The advantage of the Euclidean distance, over the standard absolute distance measure, is that the Euclidean distance measures the actual straight line distance between the two points instead of the sum of the differences in each axis. This measurement is, by definition, more accurate than the absolute difference measure.

Experimental Details

In order to evaluate the constancy of the facial features, with time, images were captured, from a sample population, on a number of different days. For this experimentation, a population of nine individuals was used, with each person providing five images. All forty-five images were manually segmented to obtain the eight facial features of interest.

To evaluate whether these chosen facial parts contained enough constant information to differentiate between people, it was necessary to show that the within-person, or *intra-person*, variation (*ie* the variation observed in the five images of each person) is less than the between-person, or *inter-person*, variation (*ie* the variation observed between the nine members of the population).

To evaluate this relationship, for one particular feature, it was necessary to compare each example of that feature type with all the other examples of that feature (*ie* each right eye was compared with all of the remaining 44 right eyes). The average between-person, and the average within-person, differences were then calculated.

The various examples of the feature under test were represented as points in 560-dimensional space. By measuring the Euclidean distance between certain points, the difference between the different feature templates was evaluated. In order to perform this measurement accurately the pixel values, within the image parts of interest, were normalised. This was done to reduce variations of intensity produced by spurious factors. The normalisation was performed in the same manner as described in section 3.5.2 (in relation to template matching).

The difference between two templates, i and j, can be denoted as D_{ij}. The intra-person Euclidean distance, D_{INTRA}, for the kth member of the population, is given below.

$$D_{INTRA_k} = \frac{\sum_{j=0}^{C_k} \sum_{i=0}^{C_k} D_{ij}}{C_k^2} \tag{4.2}$$

Where C_k is the number of images of the kth member of the population. The average intra-person distance for the whole population can be calculated using equation 4.3.

$$\overline{D_{INTRA}} = \frac{\sum_{k=0}^{N} D_{INTRA_k}}{N} \tag{4.3}$$

Where N is the number of people in the present population.

To obtain the average inter-person distance, the relative differences must be obtained for all the templates, equation 4.4.

$$\overline{D_{INTER}} = \frac{\sum_{l=0}^{N} \sum_{k=0}^{N} \sum_{j=0}^{C_k} \sum_{i=0}^{C_k} D_{ijkl}}{C_k^2 N^2} \tag{4.4}$$

Where D_{ijkl} is the distance between the ith template of the kth person and the jth template of the lth person.

Table 4–2 shows the results of this analysis, performed on all eight facial features. The quotient of D_{INTRA} and D_{INTER} has been calculated for all the eight features. If the value of this quotient, for any particular feature, approached 1.0, then that feature would be of little use in distinguishing between different faces. It must be noted that the sample size of this experiment is quite small.

The figures obtained in this experimentation, suggest that a significant amount of data is preserved on a pixel to pixel basis. Of the figures obtained, the most distinctive feature is the hair, although on such a small sample this may not be significant. It must be noted that the level of similarity obtained, for all the features used, is heavily reliant on the accuracy of the feature location.

Feature Name	Intra-Person Average ($\overline{D_{INTRA}}$)	Inter-Person Average ($\overline{D_{INTER}}$)	Quotient
Right Eye	2742	4362	0.63
Left Eye	3230	4535	0.71
Nose Bridge	1906	3005	0.63
Nose Tip	3622	4663	0.78
Mouth	4074	5452	0.75
Chin	4932	7062	0.70
Hair	3069	7414	0.41
Face	3678	5267	0.70

Table 4–2: Measured pixel variations for a sample population.

4.5 Feature Location

In the previous chapter, a constrained search strategy was established as a likely means of face location. The location of the face, was based on the successful embedding of the eyes and the mouth within the image. If the concept of a constrained feature search is extended, then this approach can be used to locate the new set of feature points required for the inter-person facial comparison.

For example, assuming that the location of the eyes and mouth are known from the first stage of processing, the nose must be above the mouth and lie between the eyes. Figure 4–4 illustrates the search areas which have been selected for each of the other features. The measurements given in the figure represent

pixel distances. The sizes of these search areas were obtained from analysis of an example set of facial images. As for the location of the eyes and mouth, the detailed fine tuning of the feature placement is performed using a computationally optimised template matching process. The template size used was 28 by 20 pixels.

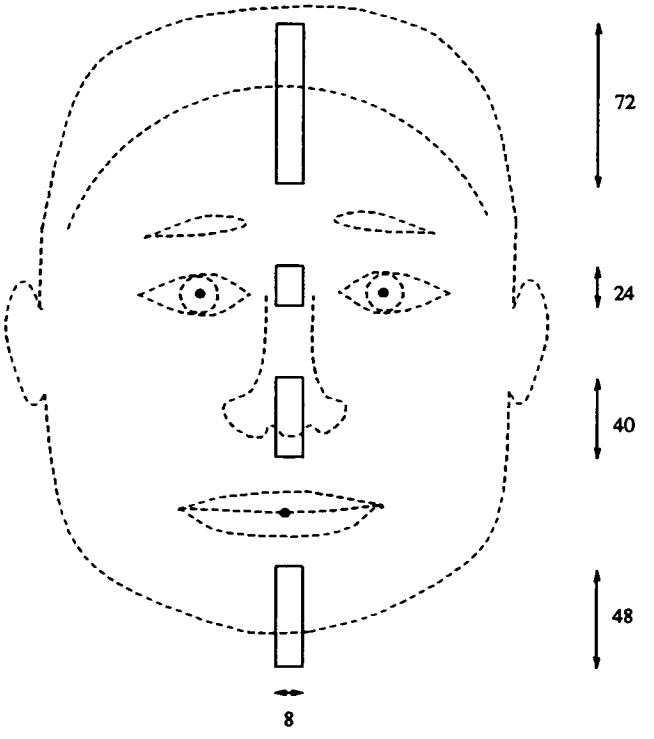

Figure 4–4: The search areas used to locate additional features.

To further ease the computational requirements of this stage, a constant set of absolute feature inter-relationships were utilised in the search algorithm. Unfortunately, for this approach to function correctly, the input faces must be standardised to the correct target size and position. To perform this standardisation, the vertical distance between the mouth and the eyes was chosen as the *benchmark* metric. It was then possible to rescale the size of the face in relation to this one measurement. A simple linear interpolation routine was utilised to scale the image, in accordance with the benchmark distance. With the chosen size for images being 256 pixels square, the distance between the eyes and the mouth should then be normalised to 69 pixels, and the centre-line of the face should be

moved to the centre-line of the image. The target size, and position, of the face is shown in Figure 4–5.

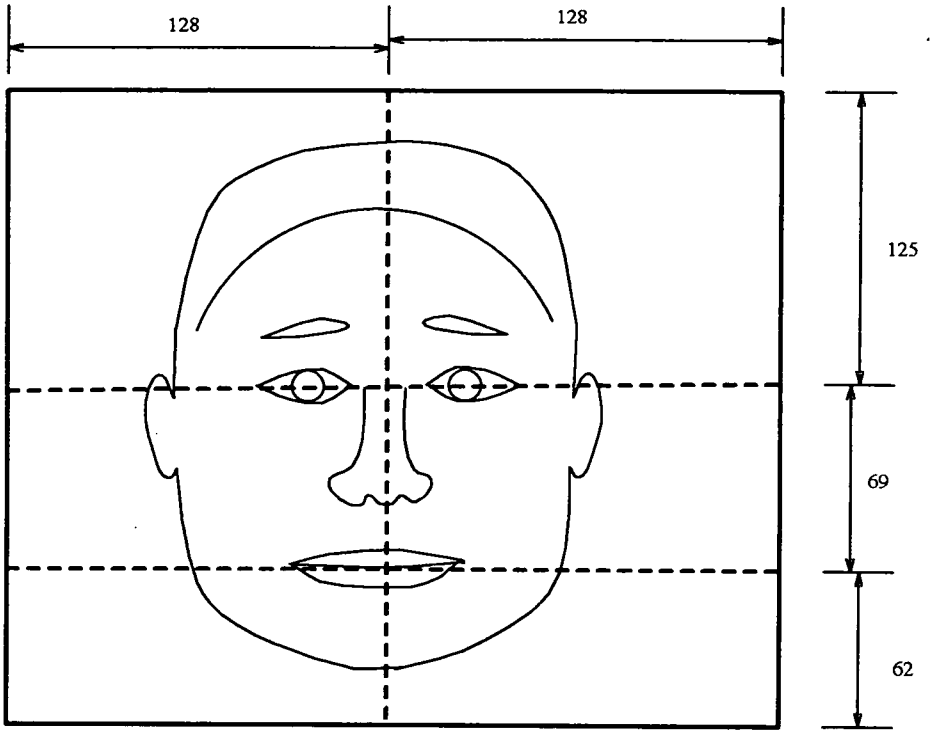

Figure 4–5: The target facial size (not drawn to scale).

4.5.1 Performance Evaluation

To test the proposed location technique, a sample set of 75 images, drawn from twenty different people, was obtained. These images had all been successfully segmented by the first stage of face location (*ie* the initial placement of the eyes and the mouth was correct, and the images had been correctly scaled in accordance with the specifications given above).

The evaluation of the location algorithm was performed on a visual basis. The feature placement was judged to be correct if the automatically determined position of the feature appeared to be within approximately $\pm$ 3 pixels of the correct position (as estimated by the human viewer). Table 4–3, gives the overall performance figures. The most common mislocations were of the hair, due to

its variability; the chin, possibly due to beards; and the left eye, due to facial rotation. Of the two total failures observed, the first, Figure 4–6, is quite badly rotated and badly centred. In the second example, Figure 4–7, the system fails because of the confusing information in the reflections on the subject's glasses.

	Correct	1 Feature Missed	2 Features Missed	3 Features Missed	Complete Failure
No. of Images	47	17	8	1	2
Percentage	62	22	11	1	3

Table 4–3: The feature location performance.

Figure 4–6: Badly rotated facial image.

Figure 4–7: Pronounced reflections on subject's spectacles.

4.6 Data Reduction

Using the system of feature location described above, it is now possible to extract the eight facial features of interest, from a facial image. The pixel patterns containing these features can then be stored as a representation of the input face. This process involves a substantial reduction in the data requirement of the facial information. The reduction is from an initial image size of 64Kbytes, down to eight sub-parts, each of 560 bytes. This represents a reduction of, approximately, 14:1. It is argued here, that this parameterisation of the face has been performed with remarkably little loss in facial recognisability.

However, the data space required to store these facial parts is still larger than the few hundred bytes stated as the desired objective of this research. Hence, the exploitation of the principle of the police *photofit* technique, has been investigated

as a possible way to further reduce the data storage requirements of this method of facial parameterisation[151].

4.6.1 Standard Facial Features Types

To one viewer, an eye may be round or oval, dark or light, or even shifty. These measures, while subjective in nature, can be used to categorise a single facial feature, in this case an eye, into a particular group. This technique of feature classification was discussed in the review chapter.

It is, however, very difficult for a number of human judges to agree about the classes, or classification, to be used for each particular feature. This problem is due to the inherent subjectivity of the process involved. If a face could be reliably dismantled into its constituent parts, and these parts could be categorised, then a very limited description could identify a person. It is this process that the police photofit technique attempts to perform.

The traditional photofit kit uses eyes, nose, mouth, hair and chin to form an overall view of a person. It has been suggested, that given a large number of different possible features, it is possible for a witness to construct an image similar to a particular individual, using a particular combination of the features available[152]. The partitioning of the face used by the photofit system, is given in Figure 4–8. The number of possible feature types is given in Table 4–4[153].

If a photofit of a person was generated accurately, then a description of the combination of features used to make up the photofit would describe that person[1]. In essence, this photofit representation is a coded view of what the face looks like. Unfortunately, the unreliability of photofits is well known; this is perhaps due to the inflexibility of the early photofit kit. Now that computerisation of

[1]It must be noted that when a photofit is constructed by the police, it is possible to enhance the facial image by the addition of shading and facial scars *etc.*[154].

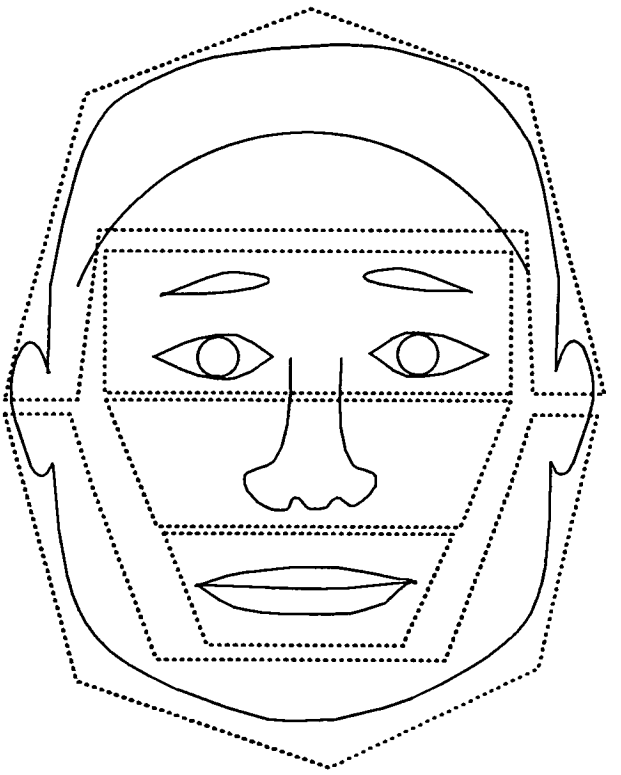

Figure 4–8: Approximate facial partitioning used by conventional Photofit.

Feature Type	Number
Eyes	124
Noses	113
Mouths	128
Chins	93
Hair/heads	219

Table 4–4: Number of different options allowed for each feature in the standard photofit kit.

the photofit process has been established[155], it may well be possible to obtain photofits which look much more realistic.

Acknowledging the practical problems of the photofit approach, it does still yield a very compact description of the target face. This representation incorporates many of the facial features, established as significant, in the facial recognition process. The underlying process of picking a best match to a particular facial feature can, to some extent, be mimicked automatically by the use of *Vector Quantization (VQ).*

4.7 Vector Quantization

Vector quantization is a standard signal processing technique which has been widely used for data compression of speech and image signals[156, 157]. The process of vector quantization yields a reasonable level of signal compression, however, there is some corresponding loss of quality. VQ has been mainly used as a coding scheme for the transmission of signals.

4.7.1 The Vector Quantization Algorithm

The essential part of the vector quantizer is the storage of a number of signal segments, at both the transmitter and the receiver. These are selected to represent possible segments of the actual input. All the entries in the store can be referred to by their index number. Using these index numbers an entire signal can be transmitted in substantially less bandwidth than that required to transmit the original signal. This reduction is possible because the number of bytes required to represent the index number is less than would have been required to transmit the entire waveform segment. The process of VQ, as applied to a 1-dimensional signal waveform, is illustrated in Figure 4–9.

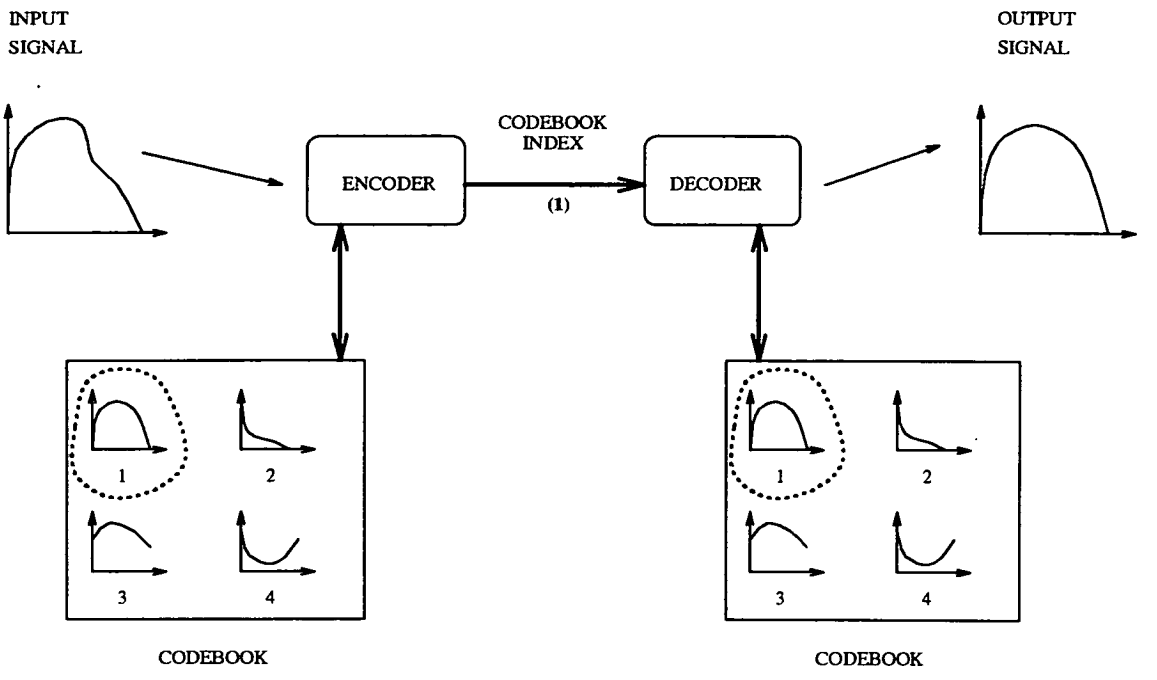

Figure 4–9: Vector Quantization in a 1D transmission system.

In operation, an input waveform of a specific number of sample points (termed a *vector*) is presented to the system. The encoder then compares this waveform segment with the stored waveforms (the codebook). The most similar of these reference waveforms, or codewords, is then selected. The index of this waveform is then transmitted to the receiver module.

Using the same codebook, the receiver then reconstructs the waveform using the index transmitted. The reconstructed waveform will only approximate the original input signal. The quality of that approximation will depend on the codebook generation. The techniques used to construct representative vector codebooks will be discussed in a later section.

As a result of the approximation, quantization errors are introduced into the reconstructed signal. Consequently, this technique is used for 'lossy' data compression. In order to minimise this *quantization noise*, it is possible to allow the codebook to adapt as the encoding process occurs[158]. This approach should ensure that the signal is reconstructed more accurately. However, this method of

error reduction will require the transmission of additional data to the receiver, effectively diminishing the data reduction achieved. The additional data is required to update the codebook at the decoder, as both codebooks must remain identical.

In order to determine which of the codewords is most similar to the waveform under test, some measure of signal similarity is required. The similarity measurements used in VQ are termed *distortion measures.* The Euclidean distance, which was described above, is the standard VQ distortion measure[156]. Other less computationally intensive measures have been suggested[159], however, the use of the Euclidean distance is justified by its superior performance.

In the preceding description, the signal waveform can be considered to be a data signal, digitised speech, or image data. In the case of image data, conceptually, the codewords could be two dimensional. For example, an image patch, of perhaps 4 by 2 pixels, could be replaced by a one byte codeword. Assuming each pixel value was represented as a one byte number, this substitution represents a data reduction of 8:1. It is this level of data compression that makes VQ an attractive method of data reduction for facial features.

4.7.2 Vector Quantization for Facial Features

For use in automatic face recognition, the concept of vector quantization has been extended from small image segments, to encorporate image patches containing entire facial features. The similarities between vector quantization and the photofit technique now become evident, as this process is described.

For example, an image segment of a particular size (as partitioned) can contain an eye. For this feature, there would be a dedicated vector codebook containing only eyes. Therefore, given an eye, the vector quantizer would then select the best match for that eye. This process is very similar to the role of the police photofit kit, in which the entire set of photofit eyes represents the eye codebook, and

the selection of the correct vector is performed manually by a human viewer. It must, however, be noted that the vector quantizer performs feature matching on the basis of a pixel level numerical similarity, this technique does not necessarily mimic human visual similarity measures.

For this approach to be extended to entire face recognition, the face has to be partitioned into a chosen set of facial features - as has been achieved - and a codebook constructed for each of the facial features in use. To minimise the quantization noise, the codebooks must represent as good a spread of likely feature types as possible.

4.7.3 Codebook Generation

Separate vector codebooks are required to encode each of the eight facial features selected. The simplest feature codebook, for a given population, would include an example of each feature, drawn from each member of that population. As the data reduction obtained using VQ is dependent on the number of vectors used, this approach to codebook design would become infeasible for large populations. Hence, a method of reducing the dimensionality of the vector codebook is required, to maintain the data reduction performance of the VQ approach.

It is desirable that the chosen codebook performs two functions.

1. The amount of quantization noise introduced must be minimised.
2. A good spread of possible feature types must be maintained.

The codebook spread is of particular importance in the use of VQ for face recognition, because of the distinctive information contained in the facial features. For example, if the codebook did not have a sufficiently broad spread of possible feature types, then substantial information could be lost when out-lying features were badly coded.

In order to maintain this desired coverage of feature types, an initial codebook representing the entire population (*ie* that has example features drawn from one image of each population member) has been used as a starting point in the derivation of a new codebook[1]. Thus, the process of codebook generation performs the dimensionality reduction of this initial codebook into a smaller set of representative vectors. Using this approach a large initial codebook, dedicated to a large population, could be reduced to a manageable size. The following section describes a number of methods of VQ codebook generation which have been investigated for vector quantization of facial features.

Statistical Clustering

Cluster analysis is a statistical approach used to partition items, or *patterns*, within a particular data space, into a certain number of groups, or *classes*[160, 161]. It performs this task by merging together, or *clustering*, the most similar example patterns into a new set of groups in the data space.

Using this concept, it is possible to reduce the dimensionality of a training set of vectors into a characteristic set of classes which maintain the initial spread of the vector space, present within the training data. The clustering algorithms yield a set of classes into which the initial training patterns have been assigned. In codebook generation, the training patterns are the initial codebook vectors, and the clusters formed with these patterns, represent the new vector set. The new vector set is obtained by merging the most similar patterns in the initial codebook together. In general, the centroid of the vectors assigned to each class, is used as the value of the new vector, representing that cluster.

Equitz[162, 163] has implemented a successful hierarchical codebook generation algorithm based on pairwise nearest-neighbour clustering. Initially, each

[1]These example features were taken from a control image of each population member. This image was **not** used in either the training or the testing of the recognition system.

member of the training set is placed in its own single member group. Then, sequentially, the two most similar groups are merged together, taking account of the error encountered at each stage. This process terminates when the training data has been reduced to the required number of vectors.

The main problem with hierarchical clustering methods is their inability to move the vectors around, after they have been placed in a particular cluster. For example, at an early stage in the clustering process, it is possible for a vector to be placed in the wrong cluster. It is then impossible for the algorithm to recover from this error, and a poor quality codebook may result. This problem can be overcome by optimisational clustering techniques, however, the computational load of full optimisational techniques is considerable.

K-Means Clustering

The K-Means algorithm[164] produces K clusters by iteratively moving the training vectors from one cluster to another, in order to form the best partitioning of the training set. The distance metric used to assess similarity is the Euclidean distance. The overall error is the summation of the distances between each vector and its cluster mean.

The algorithm assumes an initial partitioning of the training vectors into the specified number of output classes, or vectors. To refine this partitioning, each of the training vectors is compared to the centroids of all the present clusters. If, the overall error can be reduced by moving the vector into another cluster, then that function is performed, and new cluster means calculated. If the overall error cannot be reduced in this way, then no action is taken. The process terminates when the overall error reaches a minimum.

The output vector set produced by a clustering technique will represent complete coverage of the input training set, as all the examples in the training set are used to construct the centroids of the clustered classes. The K-Means algorithm

can move the vectors between clusters to reduce the overall error term, however, it is possible for the K-Means algorithm to get trapped in a local minimum without finding the optimal clustering of the given data.

The Linde Buzo Gray algorithm for VQ design

Developed by Linde, Buzo and Gray, the LBG algorithm[165][1] is one of the most widely used methods of VQ codebook generation. The algorithm is initialised with an arbitrarily selected codebook of possible vectors. It is then presented with an example data set to encode, using this initial codebook. The algorithm then refines the initial codebook, iteratively, in pursuit of a better set of vectors with which to code the example data. The algorithm is briefly described below.

The size of the vector to be used is predetermined, and the training set is partitioned into blocks of this size. In turn, each of the training examples is presented to the vector quantizer and the most similar vector selected. The error, or distortion, between that training example and the nearest vector is computed. This process is repreated for all the examples in the training set. The overall error is the summation of the error associated with each training example. If this overall error is below a predetermined 'acceptable' level of error, or it has reached a minimum, then the process terminates. If, however, neither of these two conditions have been satisfied, then the vectors in the codebook are altered.

To allow for the alteration of the codebook, the index and the associated error for each training vector are recorded during the training phase. The codebook is then improved by replacing each of the present vectors with the centroid of all the training patterns that were assigned to that particular vector. This alteration should reduce the error associated with the coding operation.

---

[1]The LBG algorithm has appeared at different times under a number of different names[166-168], however, its use for two dimensional VQ codebook design must be attributed to Linde, Buzo and Gray.

As described here, the LBG algorithm is essentially the same as K-Means clustering, a point noted by Makhoul *et al* [169]. However, there is a difference between the two algorithms. This difference lies in the way in which the codebook is altered to reduce the overall error. In the strict definition of the K-Means algorithm, the cluster groupings, and hence the output vectors, can only be altered if different training examples are assigned to them. In the LBG approach, vectors which remain unused, or little used, are reinitialised to make them more useful. This process is performed by splitting one of the most commonly used vectors into two similar copies, the unused vector is then replaced by one of these copies.

The initial starting point of LBG is recognised as being of considerable importance in achieving a good vector codebook. There a number of methods of performing this:

- Use a random set of initial values.
- Initialise each vector to the centroid of the training set.
- Select a number of the training examples, at random.

Kohonen's Self Organising Feature Map

Kohonen developed his *Self-Organising Feature Map (KSOFM)* algorithm in order to model the neural feature maps which are thought to form in the human brain[170]. The KSOFM algorithm produces a network in which neighbouring neurons have similar responses. As a clustering algorithm, KSOFM allows a reduction in the dimensionality of the input vector space, to a smaller number of reference vectors, by selecting common *features* from similar input vectors. The feature map has previously been used for VQ codebook generation[171], nearly equalling the performance of the more established LBG algorithm in one case[172].

KSOFM defines a matrix (usually two dimensional) of output units, or neurons, each of which has a set of coefficients termed a weight vector, associated

with it. The weighting vectors are allowed to adapt, in an unsupervised manner, to provide coverage of the feature space. The weight vectors should converge towards cluster centres, present within the training data, after sufficient training time. The elements of the weight vector, for each neuron correspond to the pixel values of each vector in the new VQ codebook. This is similar to the way in which the cluster centres produced by the previous two approaches represent the new vector set.

Sequentially, each training vector is presented to the Kohonen network. For each training pattern, the nearest neuron to it (in the feature space) is selected using the Euclidean distance. The weights of this *winning* neuron are then updated using an adaption formula which moves the neuron nearer to this particular training example. The amount of movement is controlled as a function of an adaption gain term and a neighbourhood gain term. The neighbourhood gain ensures that the neurons are influenced by all the training examples which are close to them in the feature space. The neighbourhood term decreases with time, as does the adaption gain. The equations and the parameters used for this experimentation are given in Appendix D.

This process repeats for a predetermined number of cycles, or until the gain terms have decreased to negligible values. Comprehensive descriptions of the LBG and KSOFM algorithms are included in [173], which is appended to this thesis.

The KSOFM method of cluster analysis has been likened to the K-Means cluster algorithm described above[174]. However, there is a difference in the way that the algorithms form the new vector values. The KSOFM assigns a neuron to a particular region, and the neuron adapts its weights to converge to a cluster centre. In this process, the weights are influenced by the underlying structure of the vectors in the feature space, and the neuron is influence by training examples which are not in its immediate locale. In the K-Means algorithm the clusters are

formed on a purely local basis and each cluster is unable to influence any other outcome.

The K-Means and LBG algorithms are iterative, and require the storage and repetitive presentation of the initial training set. The storage of training examples is not required by the KSOFM algorithm, however, as these three algorithms have been implemented here, they are all, effectively, iterative.

Three different methods of vector codebook design have been described above. All three approaches can be used to perform dimensionality reduction on a given set of vector codewords in a fundamentally similar way. Experimental results will be presented in the following chapter[1] which demonstrate the abilities of these three techniques to perform codebook generation.

4.8 Feature Measurements

The previous sections have outlined a possible system of parameterising the local facial feature data into a set of characteristic information. It is, however, necessary that some of the structural information regarding the distribution of these facial features is also preserved. A basic set of of facial measurements has already been provided by the placement of the fundamental facial features described above.

The eight facial features used to encapsulate the facial feature information, were selected with regard to a number of practical constraints and the present knowledge of feature saliencies. The entire facial feature data has been parameterised into a set of eight facial features. The measurements between these

[1]The research into VQ codebook generation has been carried out with the assistance of Mr Colin Ramsay. The LBG and KSOFM codebook generation algorithms were investigated with the aid of his software and expertise.

features should contain information pertinent to facial identity, however, the relative significances of these different measurements is not known. Thus, these facial measurements have also to be parameterised.

As this information is of a simple, scalar, measurement nature, and there are very few measurements to deal with, a rigorous method of multivariate analysis can be employed. There are a variety of different techniques available which will yield a smaller number of more distinctive measurements given an initial set[175, 176]. Some of these techniques have already been discussed in this thesis (*eg* Principal Component Analysis). The *F ratio* has been chosen as the initial technique with which to evaluate the relative distinctiveness of each of the available measurements. F ratio analysis has been used for feature evaluation in a number of different research areas, notably in automatic speaker recognition[177][178, pp 193-196].

The F ratio measures the relative distinctiveness of a particular measurement by one-way analysis of variance (ANOVA). It assesses the value of each measurement, by contrasting the variation of the measurement for several examples of the same person, with the overall variation of this measurement for the entire population. The formula for this ratio is given below, Equation 4.5[1], taken from [179].

[1]The ANOVA undertaken here, was performed using the BMDP suite of statistical programs. This software package was also used to perform the further measurement analysis reported in this section.

$$F = \frac{MS_{between}}{MS_{error}} \tag{4.5}$$

Given

$$MS_{between} = \frac{\sum_{i=1}^{g} c_i(\overline{\mu_i} - \overline{\mu})^2}{g-1}$$

$$MS_{error} = \frac{\sum_{i=1}^{g}\sum_{k=1}^{c_i}(\mathcal{M}_{ki} - \overline{\mu_i})^2}{N-g}$$

where -

$\mathcal{M}_{ki}$	=	value of the measurement in the kth image of the ith person.
$\overline{\mu_i}$	=	measurement mean for person i.
g	=	number of people.
i	=	personal index.
k	=	index of image within that person.
c_i	=	number of images of person i.
N	=	total number of cases.

The value for the F ratio, is a measure of that characteristic's variability among the test population. Low values for F represent low levels of distinctive information content. Measurements with high values of F, are highly variable among the people in the test set, and are therefore valuable for pattern recognition. The F ratio technique is similar to the inter-person and intra-person comparisons performed in section 4.4.2.

The F ratios of twelve available measurements have been obtained for a training set of 400 images (10 each from a population of 40 people). Using the values for the F ratios of the twelve measurements under test, a bar chart has been constructed, Figure 4–10. The feature numbers given in this figure, relate to the measurements shown in Figure 4–11. All the values of F obtained are highly

statistically significant with a very low probability that these values have been caused by random factors ($P < 0.0001$). This significance suggests that all the measurements available contain some discriminative power.

Of the values shown, feature M_{11}, the vertical position of the hair line has the lowest F ratio. There are two likely reasons for this.

1. A number of mislocation errors, particularly with bald people.
2. A high level of intrinsic variability of the position of the hair line, due to styling *etc.*

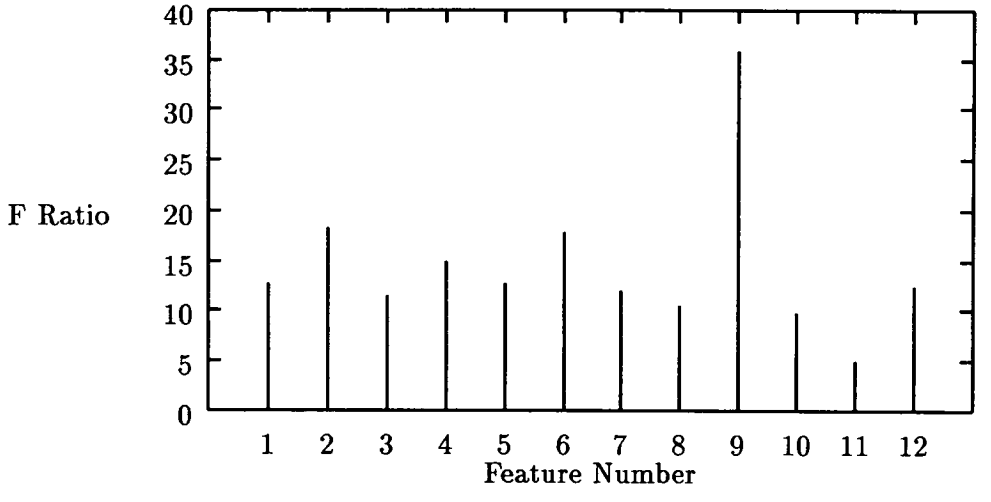

Figure 4–10: The F ratios for the twelve facial measurements.

The best available feature appears to be measurement M_9, the vertical placement of the chin. The F ratio of M_7 is slightly confusing at first instance, as this is the distance between the right eye and the mouth, and the images used, should have been standardised to this distance, see section 4.5. However, the standardising distance was the average distance between both of the eyes, and the mouth, and thus, the F ratio of 12.1, for feature M_7, is more a measure of the difference between the levels of the two eyes, than that of the position of the mouth. This hypothesis is supported by the similarity between the F ratios of measurement M_7 and measurement M_1.

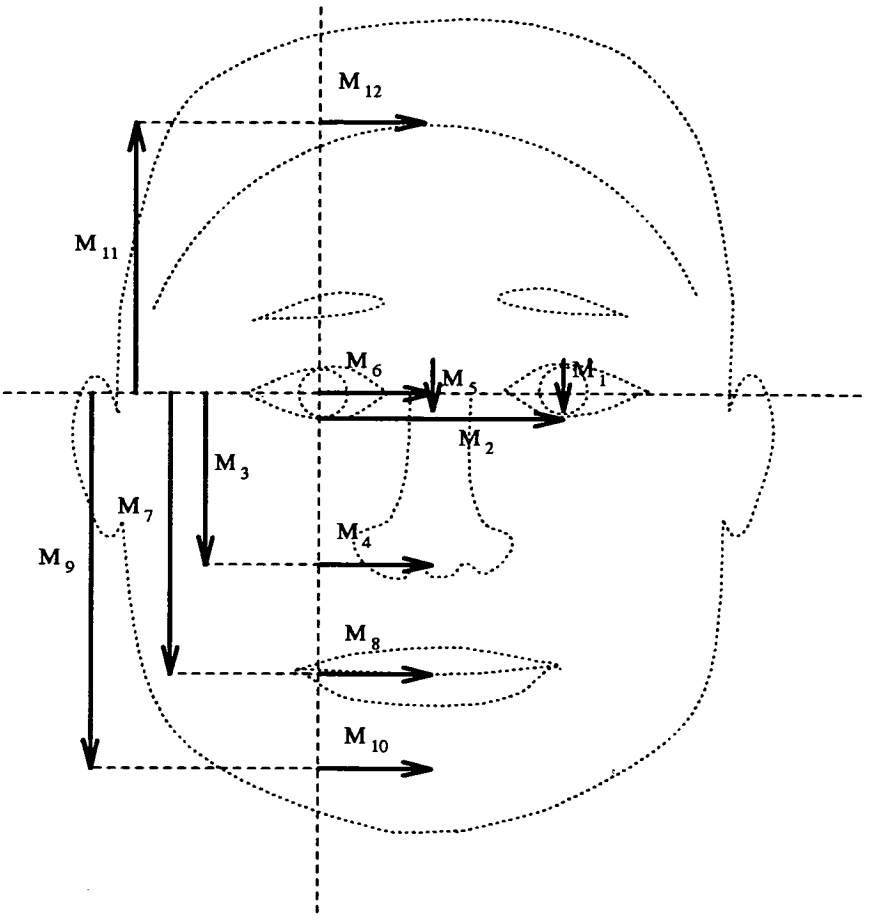

Figure 4–11: The feature measurements used.

While such analysis of feature measurements is valuable, there are two main problems associated with the use of the F ratio in feature selection.

- The F ratio is a measure of a particular feature's variability over the entire population. Thus, a high F ratio can be obtained when only a small number of people differ substantially from the population norm. It does not, therefore, follow that a feature with a high F score can be used to distinguish between a lot of different people.

- The F ratio does not identify which of the measurements are highly correlated. Thus, several measures with high F ratios may well be measuring the same facial characteristic. For example, it has already been noted that feature numbers M_1 and M_7 could be linked in some way.

Addressing the first point, the only real way in which the discriminating power a particular measurement, or feature, can be evaluated is to perform recognition experiments using that feature. However, leaving this factor aside for the moment, it is possible to assess the correlation between different measurements using scatter diagrams.

Figure 4–12 shows the inter-relationship between features M_4 and M_6, the horizontal position of the top, and the bottom, of the nose. The ordered shape of this scatter diagram betrays the high level of correlation between these two features. This correlation can be measured using the Pearson product-moment correlation 'r'[180]. For Figure 4–13, r is equal to 0.777, showing a high level of correlation. The null hypothesis that the two variables are unrelated is rejected with a high degree of certainty ($P > 1 - 0.001$). Performing the same analysis on features M_9 and M_2, a less ordered picture is obtained, Figure 4–13. The correlation value is r = -0.097, the null hypothesis of these measurements being uncorrelated features is accepted at the 5% level.

It can be seen from this analysis, that by continuing this process of linking feature F ratios with correlation figures, it would be possible to derive a small set of uncorrelated, discriminative measures. However, there is an automatic process available to perform this task, called *canonical analysis*[181]. This technique iteratively calculates F ratios, and correlation measures, to produce a limited set of features containing the best discriminative power of the original feature set. These new features can now be used to partition the decision space.

For the training set of forty people, all twelve measurements were analysed and a set of five *canonical vectors* produced. These five vectors are composed of differing proportions of the original twelve input variables. The coefficients used to produce each of the canonical vectors, are given in Table 4–5. The first vector CV_1 is derived using Equation 4.6.

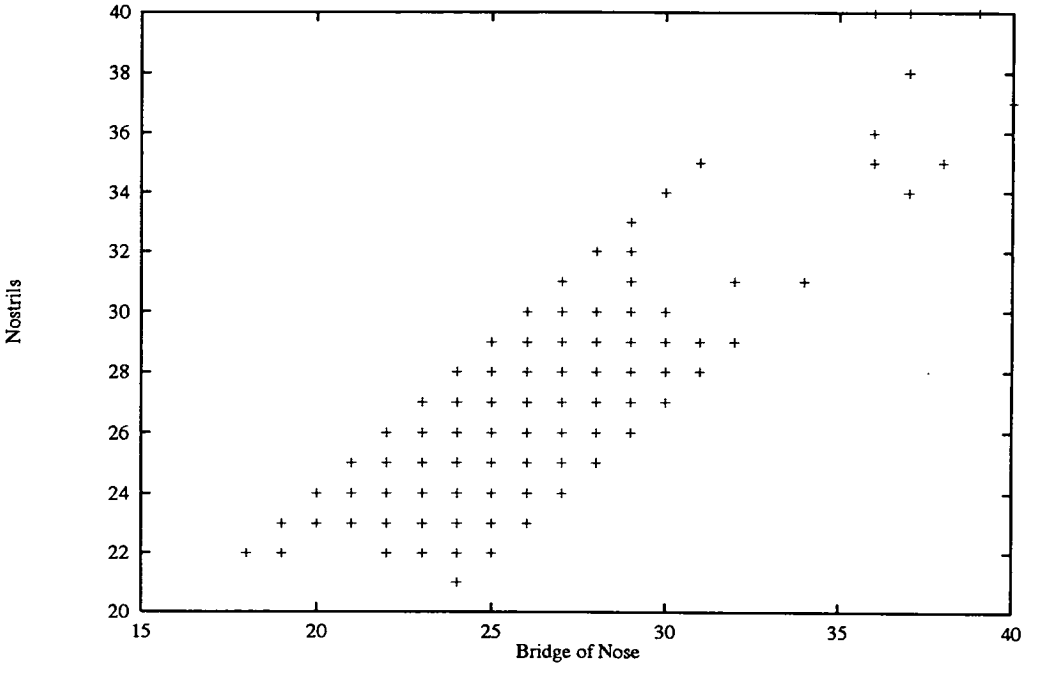

Figure 4–12: Scatter Plot showing M_4 plotted against M_6.

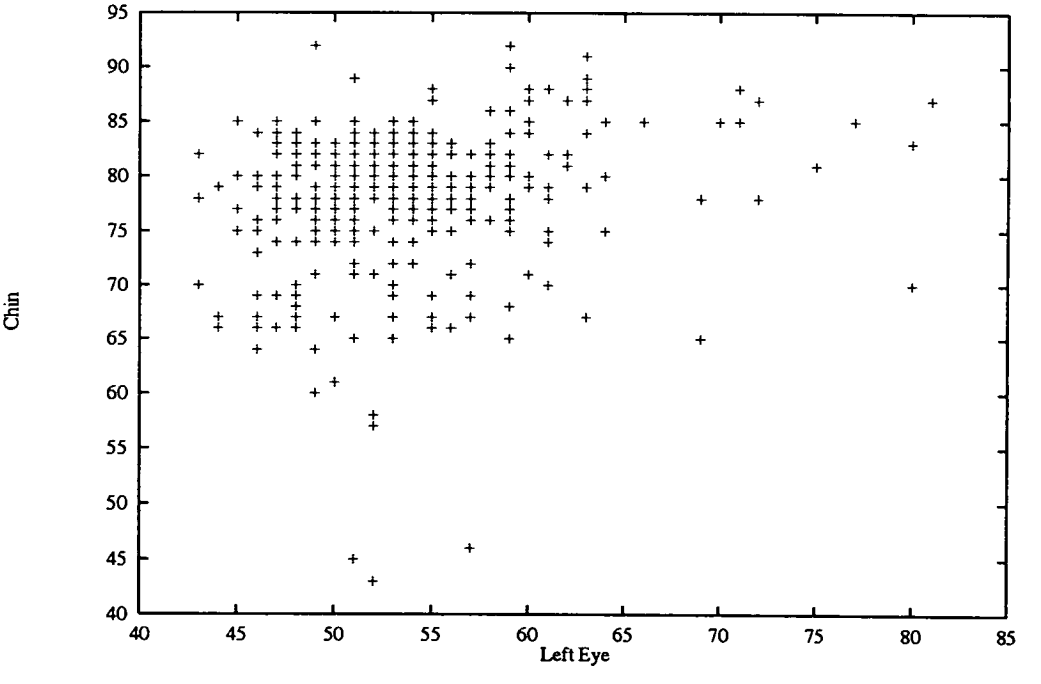

Figure 4–13: Scatter Plot showing M_9 plotted against M_2.

Measurement	CV_1	CV_2	CV_3	CV_4	CV_5
M_2	-0.09036	-0.22907	-0.20467	-0.06877	0.23663
M_3	-0.06050	0.08512	-0.04266	0.34871	-0.06181
M_4	-0.00961	-0.07050	0.16981	0.22652	-0.27883
M_5	-0.01565	0.17661	-0.40932	-0.37651	-0.55078
M_8	-0.03366	-0.02755	0.08441	-0.01195	-0.12974
M_9	-0.15511	0.00163	0.03815	-0.01557	0.03363
M_{10}	0.00301	-0.02662	-0.05206	-0.06059	-0.15421
M_{11}	0.00096	-0.03985	-0.03274	0.06688	0.11603

Table 4–5: Coefficients for the five canonical variables.

$$\begin{aligned} CV_1 = {} & M_2 \times -0.09036 + M_3 \times -0.06050 + M_4 \times -0.00961 \\ & + M_5 \times -0.01565 + M_8 \times -0.03366 + M_9 \times -0.15511 \\ & + M_{10} \times 0.00301 + M_{11} \times 0.00096 \end{aligned} \tag{4.6}$$

The canonical variables produced here, provide a compact means of assessing facial similarity between people (the details of this comparison will be given in the following chapter). However, it must be noted that the canonical analysis has only been been performed using 40 people, and the boundaries that the algorithm has drawn through the feature space on the basis of these twelve measurements, may not be readily expandable to a larger population.

4.9 Complete Facial Parameterisation

A novel method for the parameterisation of the face has been presented above. By using vector quantization, is has been shown that it is possible to radically reduce the amount of data required to represent the facial features. It has also been shown that a set of five discriminative measurements, which capture some of the structural information of the face, can be derived. The complete facial parameterisation process is summarised below.

Firstly, the features are located and the face partitioned in eight sub-parts. These eight sub-parts, are presented to a bank of eight vector quantizers, each of which is dedicated to a particular facial feature. This process is illustrated in Figure 4–14. The output of this stage is an eight dimensional vector containing the indices of the eight best matched features to the input face. In addition to this vector, the values of the five canonical variables, for that face, are also stored.

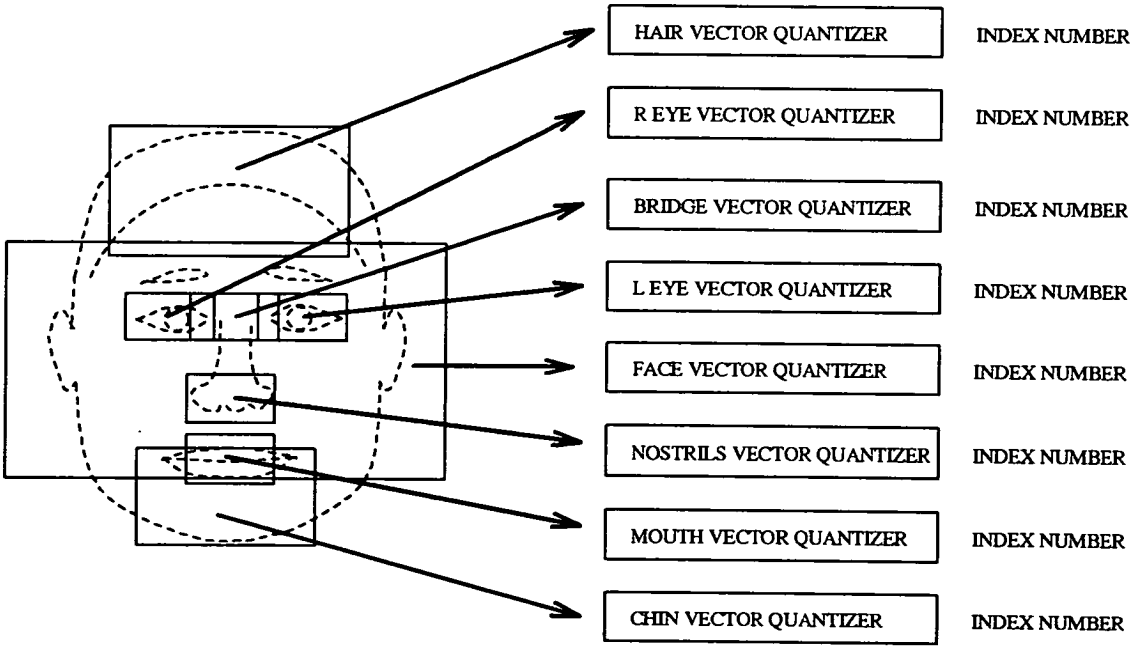

Figure 4–14: Facial parameterisation based on Vector Quantization.

An example face with its vector quantized image is shown in Figure 4–15 – for display purposes the output of the vector quantization stage is represented by the average of the three most likely templates chosen. An additional example image is given in [182], which is appended to this thesis. It must be remembered, when viewing these images, that the VQ process does not necessarily preserve visual similarity to the human; it does, however, maintain a constant mapping from one picture to another.

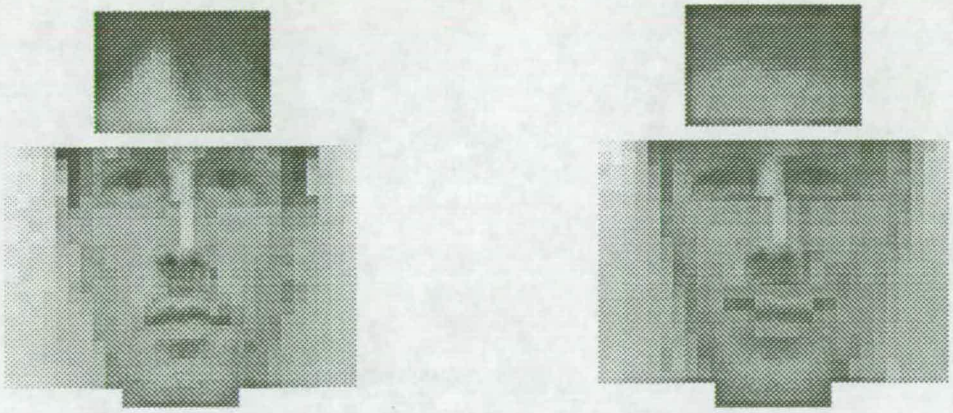

Figure 4–15: Coded and vector quantized views of an example face.

4.10 Summary

Facial parameterisation is the fundamental process required to perform automatic facial recognition. This chapter has introduced a facial parameterisation method which reduces the facial information into a very small number of bytes. The characterisation process yields a data set which incorporates both sources of distinctive facial information; the feature detail, and the facial structure. The following chapter will demonstrate how this algorithm can be used in a complete, automatic, facial recognition system.

Chapter 5

A Complete Facial Recognition System.

5.1 Introduction

This chapter will draw together the facial segmentation, and consequent feature location, with the facial feature coding – devised in the last two chapters – to form the basis of a facial recognition system. It will then outline the requirements of the final comparative stage and present the complete facial recognition algorithm.

In a real-world situation, a facial recognition device would encounter many problems associated with systematic changes of expression and appearance. This is the sort of variation which could be expected to occur on a day to day basis. To overcome these problems, a probabilistic pattern recognition technique will be introduced. This approach makes use of a large training set of images, to produce a generalised representation of the facial appearance, capturing many of the likely changes in a particular person's face.

The inclusion of a scheme of feature weightings, to signify varying levels of importance to the facial features, will be evaluated. Analysis of recognition experiments using inter-feature measurements will also be performed.

For comparison, a WISARD type device has been implemented and tested on the same data set. Full results for experimental trials on both of the competing systems will be presented in this chapter.

5.2 Problems of Facial Comparison

The final stage of a facial recognition system is the evaluation of the similarity present between different faces. In a dedicated facial recognition device, this comparative function is performed on the parameterised feature set, derived from the input face. In this study, the parameterised feature set is obtained from the facial data using the feature based vector quantization algorithm, described in the previous chapter. Unfortunately, there are a number of different circumstances which can cause problems for comparative analysis of faces.

These possible circumstances include:

- Changes in facial expression.
- Changes in hair styles or beards.
- Changes in lighting intensity and direction.
- Changes in use of spectacles.

Of these four problems, lighting can be ignored because of the experimental procedures used (described in Appendix B). Substantial changes in hair style or beard growth (*ie* removal) will have such a profound effect on the facial appearance as to make recognition practically impossible for an automatic system, such instances have been removed for this study. However, minor changes in hair style will be included in this study. An attempt is also made here to overcome the difficulties associated with the two remaining factors; facial expression and removal of spectacles.

5.2.1 Generalisation

Ignoring extreme facial expressions, the basic way in which the face changes, with a particular expression will be repeated when that same expression is again

used. If the subject is requested to adopt a fairly neutral expression, when using the device (similar to that requested for a passport photograph), the likely facial variation will be limited to only small adjustments in the face. Thus, if a number of images are captured during the training phase of the device (perhaps over a number of days) they are likely to include some of the expressional variations that are most common for that individual. Similarly, if a number of images are captured with, and without, the spectacles the subject normally wears, then the 'sum' of these images will capture someone who sometimes does, and sometimes does not, wear spectacles.

To allow for these possible variations, a new method of facial comparison has been devised. This new technique utilises a training set of several images to produce a facial signature. If the training set is sufficiently varied, as described above, then the representation of the face, contained within the personal signature, will incorporate the likelihoods of certain variants in facial appearance. In this manner, the device is able to construct a *generalised* view of the subject's face.

Considering the facial feature encoding technique in the previous chapter, it can be shown that the generalisation process can be performed using a probabilistic approach to facial comparison. The following example illustrates the process involved.

If eight images are presented to the VQ system, eight sets of feature indices will be obtained; and, if a VQ codebook of four vectors is used, then each feature index will be in the range of 1 to 4. Considering the mouth in isolation, in a random example, each possible mouth (*ie* the four codebook entries) could be expected to be matched with two of the images presented. However, if the high level of pixel-to-pixel similarity shown in section 4.4.2 is also present here, then some clustering of these vector indices would be expected, given several mouths drawn from different images of the same person. It would then be hoped that the choice of vectors would be restricted to only two, or even one, of the possible

codebook members. To evaluate whether this hypothesised clustering occurs in practice, the following experimentation was performed.

5.2.2 Feature Choice

To measure the real amount of clustering present amongst VQ coefficients, a sample set of ten individuals was used. For each person, a set of ten images were captured on several days. These images were subjected to the face segmentation and feature location stages described earlier. The output images were then visually inspected to assess the accuracy of these stages. Any mislocations detected were rectified manually. This was performed to nullify any detrimental effects which could otherwise have been introduced. For example, if the features were slightly mislocated in one of the example faces, then incorrect VQ coefficients could be chosen for that face and the comparison would be invalid.

The features drawn from these example faces were then submitted to the VQ algorithm. In each case, the vector chosen by this algorithm, should be the best match to the given stimulus feature. The codebooks utilised were created using one additional image of each of the ten people in the test population.

For each person, the ten sets of VQ coefficients, one for each of the ten images of each individual, were then combined to form a personal *vector histogram* reflecting the frequency of use of each of the ten possible codebook entries. Eight feature histograms were thus produced for each member of the sample population. An example set of feature histograms is given in Figure 5–1. These feature histograms can be thought of as probability distributions, quantifying the probability of a given individual causing a particular codebook vector to be observed.

5.2.3 Spread Evaluation

To measure the *spread* of chosen vectors it is necessary to order the histogram data. As a valid measure of ordering the vectors, in accordance with their visual

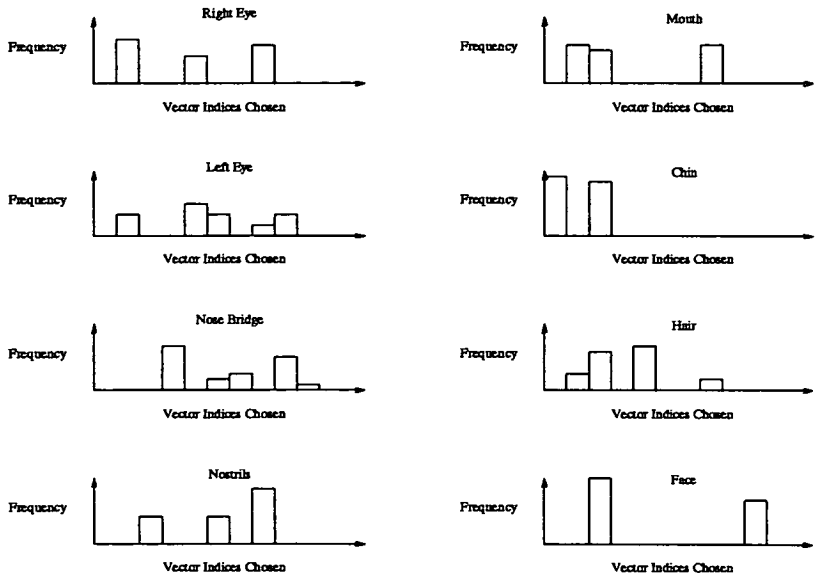

Figure 5–1: An example set of feature histograms.

similarity, has not been arrived at, it was chosen to order the histograms on the basis of magnitude. Figure 5–2 illustrates the ordering process for the same histograms as are shown in Figure 5–1.

The deviation, away from the origin, can be calculated by multiplying the frequency of each choice, by the distance away from the origin. The total of this measure can be divided by the number of samples to obtain an average deviation metric. For example, using ten samples, the ordered histograms could have values of $\{4, 3, 2, 1, 0, \ldots\}$; the average deviation would be:

$$(4 \times 1) + (3 \times 2) + (2 \times 3) + (1 \times 4) \;=\; \frac{20}{10} \;=\; 2$$

The average deviation away from the origin is a measure of constancy attributable to any one particular feature. Although the reliability of this measure is not entirely acceptable, it does yield an empirical measure of the relative feature constancy.

Table 5–1 shows the values obtained for the deviation metric for the entire population, together with the average for each feature. The values obtained for this measure must be compared to the vector spread for the whole population. To

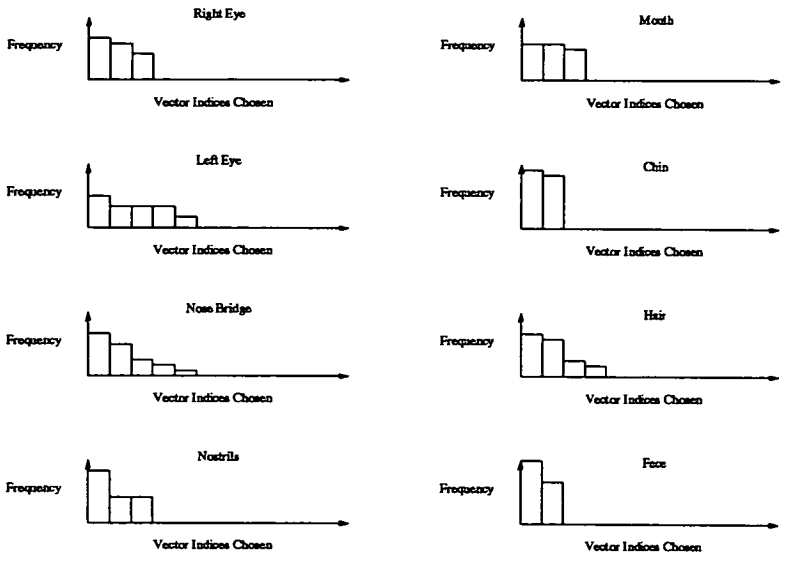

Figure 5–2: An ordered set of feature histograms.

perform this analysis, all the images of the ten subjects, were used to construct a new superset of 100 images. The same histogram analysis was performed on this data set and the results are presented in Table 5–2.

Feature	Subject Number										Average
	1	2	3	4	5	6	7	8	9	10	
Right Eye	1.0	1.3	1.0	2.0	1.7	1.6	1.0	1.7	1.4	1.0	1.4
Left Eye	1.0	1.0	1.0	1.6	1.5	1.1	1.1	1.4	1.5	1.0	1.2
Nostrils	1.6	1.7	1.6	1.2	1.7	2.2	1.2	1.5	1.7	1.3	1.6
Bridge	1.3	1.3	1.6	1.3	1.5	1.6	1.0	1.3	1.4	1.0	1.3
Mouth	1.2	1.0	1.5	1.2	1.7	1.6	1.0	1.6	1.2	1.3	1.3
Hair	1.0	1.3	1.4	1.1	1.3	1.4	1.2	1.2	1.0	1.1	1.2
Chin	1.0	1.0	1.2	1.1	1.0	1.1	1.1	2.3	1.4	1.2	1.2
Face	1.2	1.3	2.4	1.1	1.6	1.3	1.0	2.0	1.2	1.4	1.5

Table 5–1: Deviation metric for each population member.

From this brief experimentation, it can be observed that a substantial amount of VQ coefficient clustering is occurring in real images. This demonstrates the

Feature	All Population	Average
Right Eye	3.82	1.37
Left Eye	3.23	1.22
Nostrils	3.90	1.57
Bridge	3.11	1.33
Mouth	4.89	1.33
Hair	3.59	1.20
Chin	3.31	1.24
Face	4.24	1.45

Table 5–2: Deviation metric for entire population and each member average.

possible viability of a facial comparison algorithm based on pixel level constancy within the important facial features.

5.3 Inter-person Comparisons

There are two sources of information which can be exploited to perform inter-person comparisons.

1. The feature histograms drawn from several training images of the individual, which contain a generalised view of the pixel level variations in the facial features.

2. The five canonical vectors, described in the previous chapter, containing structural information regarding the face.

These two sources of information contain the first and second order facial characteristics thought to be of much importance in the recognition process. The way in which these two different sets of characteristics can be exploited for facial recognition are detailed in the following sections.

5.3.1 Probabilistic Feature Comparisons

There are a number of different ways in which the information contained in the feature histograms can be used to perform inter-person comparison. Two such approaches have been investigated[183], the algorithms used are presented in the following sections.

Winner Takes All

The storage of **one** of the feature probability distributions, or histograms, would be sufficient to identify that same feature, from that person in the future – assuming that the facial expression remained within the tolerances under which the training set of images were produced. To produce an entire feature based facial signature in this manner, the histograms for all the eight facial features must be stored.

In order to compare a new face with a stored signature, the features in the new face must be isolated and coded using the same vector quantizers as were used for training. This process yields the feature indices which represent the new face. If these feature indices are then compared with the target signature (Figure 5–3 – the arrows within this figure represent the VQ indices chosen for the facial image under test.) the overall probability of the identity of the new face being that of this particular signature can be evaluated.

Putting this strategy in a probabilistic framework, it is possible to associate a probability for each person causing a particular vector (for a given feature) to occur.

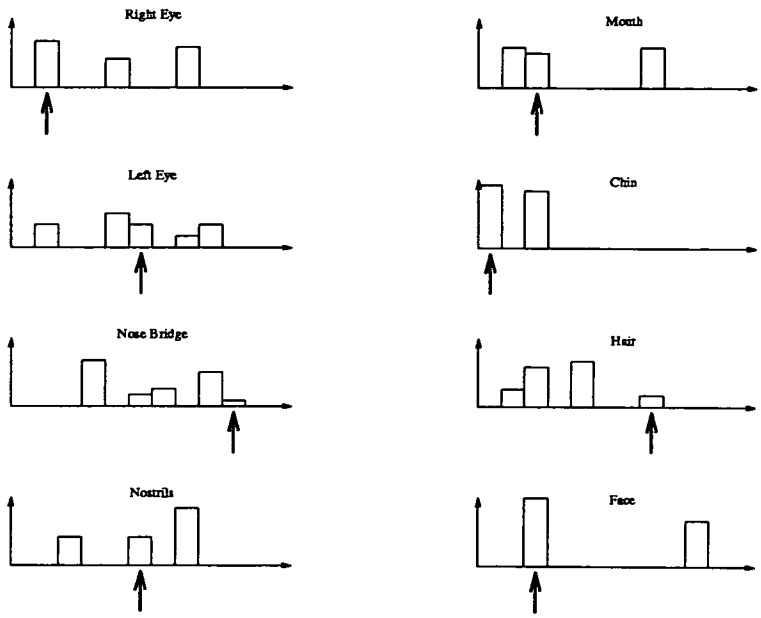

Figure 5–3: A stored personal signature compared with new stimulus face.

This probability can be denoted thus:

$$P(H_i|T_jV_k)$$

where -

H_i = the ith individual.

T_j = the jth feature (*ie* nose, mouth ...).

V_k = the kth vector for that feature.

Using Bayes theorem

$$P(H_i|T_jV_k) = \frac{P(H_i)P(T_jV_k|H_i)}{\sum_{i=0}^{N} P(H_i)P(T_jV_k|H_i)} \tag{5.1}$$

where

N = total number of individuals in population.

Assuming that $P(H_i)$, the *a priori* probability, is equal to $\frac{1}{N}$, *ie* each person is equally probable, then the overall probability, Equation 5.1, can be evaluated using values for $P(T_jV_k|H_i)$ obtained from the probability histograms produced

during training. The value obtained for $P(H_i|T_jV_k)$ is a measure of likelihood that the observation of a particular feature vector, T_jV_k, suggests any particular individual, H_i.

An accumulation of the values of $P(H_i|T_jV_k)$, for all values of j (*ie* the probabilities associated with each of the facial features), is required to produce an overall probability that all the observed feature vectors are consistent with the stimulus face being any one member of the population. There are a number of different ways in which this accumulation can be performed.

One possible manner in which to obtain an overall probability of a match, is to multiply the eight individual probabilities, of the eight features, together. At this stage, it is possible to incorporate a weighting function into the probability calculation. The weighting function can be used to control the different levels of significance given to each individual facial feature during the recognition process. Assuming a weighting vector $W = \{w_1, w_2, \ldots, w_8\}$, then the overall probability that the observed vectors were caused by subject i is given in Equation 5.2.

$$P(H_i) = \frac{1}{C} \prod_{j=1}^{8} w_j P(H_i|T_jV_k) \tag{5.2}$$

where

$$\sum_{j=1}^{8} w_j = C$$

In order to establish which member of the population is the most likely owner of the stimulus face, it is necessary to calculate the values of $P(H_i)$ for all i (*ie* for all the population members). The highest value of $P(H_i)$ would indicate the most likely match for the face under test.

Unfortunately, substantial variations in the face can cause problems for this method of probability accrual. For example, when coding a new face, a vector could be chosen which was not chosen during the training phase for that particular person. This produces a zero probability for that feature, which in turn,

results in a zero probability for the entire face, regardless of the probabilities associated with the other features. This effect can be caused by large changes in a particular feature *eg* closure of the mouth or eye, *etc.* Thus, it may be that a feature is, in reality, very similar to those recorded during training, although, its pixel level representation has been substantially changed. It would be undesirable for the system to fail to recognise someone because of such a change, therefore, it is necessary that the system incorporates a mechanism whereby a very unlikely occurrence (*eg* a closed eye) can be related to the overwhelmingly positive information drawn from the other features.

One way to overcome this zero probability problem is to sum the the feature probabilities, Equation 5.3.

$$P(H_i) = \frac{1}{C} \sum_{j=1}^{8} w_j P(H_i|T_j V_k) \tag{5.3}$$

This approach is less affected by rare variations in certain facial features.

The zero probability problem is primarily a result of sparse training data. The probability of someone closing their eye may well be small, but it is certainly not zero. If the codebook vectors were ordered then it would be legitimate to smooth the probability histograms, producing small finite probabilities for some of the more unlikely facial feature variants. However, as this is not the case here, an alternative way of performing, essentially the same task, has been devised.

Total Difference

When comparing the features drawn from a new image, with the VQ codebook, a difference measure is computed between the new feature and each codebook entry. If, instead of selecting the best one and comparing it with the probability histograms (as described in the *winner takes all* approach), all the difference information is used, then it is possible to produce a more accurate measure of similarity between a new face and a stored signature. The following example illustrates this principle.

Considering only one feature in isolation (with a possible set of five vectors) and, given a training set of ten images, a personal signature can be derived. This personal signature will contain the probability histogram information. When presented with a new example of that person, the system can produce a set of differences for the chosen feature relating to all the possible vectors. These differences are signified D_k, where D_k represents the numerical difference between the new feature and the kth member of the vector codebook. These two pieces of information are presented in Table 5–3.

Vectors	V_1	V_2	V_3	V_4	V_5
Probabilities T_jV_k	5	2	0	3	0
Pixel Differences D_k	56	68	52	72	87

Table 5–3: Example feature difference information.

Here, the most likely vector choice is V_3 and the training frequency value of **0** would be used in the winner takes all approach. However, in difference terms, the new feature is also very close to vector V_1 and this similarity is ignored by only selecting the best output. If the reciprocal of the difference (scaled in some manner) is used as a measure of similarity, then a total match score between the new feature and that feature's appearance during training can be derived using all the probability information. This calculation is shown for feature T_1 in equation 5.4.

$$\begin{aligned} Match\ Score \;=\; & P(T_1V_1) \times \frac{K}{D_1} + P(T_1V_2) \times \frac{K}{D_2} + P(T_1V_3) \times \frac{K}{D_3} + \\ & P(T_1V_4) \times \frac{K}{D_4} + P(T_1V_5) \times \frac{K}{D_5} \end{aligned} \tag{5.4}$$

Where K is a constant scaling factor. In practice, K was assigned the least value of D_k.

In this way, important information revealed in the further examination of the vector scores is not lost. It is then possible to sum these feature scores over all

eight facial features, again incorporating a weighting function, to obtain a total facial similarity score.

Initially, the weighting function used by these two approaches was assumed to be equal for all eight features. However, it is possible to use this weighting vector to accentuate the importance of the more salient features in the face. Results for recognition trails using these two alternative approaches to feature comparisons, and a number of different weighting strategies, are given in section 5.7.

5.3.2 Measurement Comparisons

As described in section 4.8, the canonical analysis technique yields a discriminative set of uncorrelated measurement vectors. It is suggested that each of these individual vectors contain information which can be used to distinguish between different faces. As the vectors are uncorrelated, an unweighted Euclidean distance can be used to link all five measures together to form one overall similarity metric.

In practice, the mean canonical vectors for each person, derived from the facial measurements, are stored during training. To recognise a new face, the same measurements are then extracted from the new test face and converted into their canonical representations. A match *score* is produced by calculating the Euclidean distance between the new values and the stored values, for each population member. The lowest distance signifies the best match. This score is analogous to the facial feature probability, given above.

The facial feature inter-relationship measurements, as contained in the canonical vectors, should be stored along with the feature probability information, to form a full *personal signature*. This facial signature now incorporates first and second order facial information. The inter-person comparisons should then be performed by combining the measurement score to the feature score to produce

one, overall, recognition score. Unfortunately, there still remains the matter of how much importance to assign to either the measurement information or the feature probabilities during comparison.

5.3.3 Data Reduction

At this stage, it is worth evaluating the data storage requirement of the proposed personal signature, as this is a key factor in the design of this facial recognition algorithm. The salient facial data has been compressed in a number of ways since it was initially captured in a 64 Kbyte image.

To store the measurement information, one byte is required for each of the five canonical vectors. The storage of each probability distribution is dependent on two factors; firstly, the number of training images and, secondly, the number of possible vectors in the codebook. However, assuming twenty vectors, and ten training images, each histogram *bin* could be required to store any number between 0 and 10 (the minimum and maximum possible frequencies), hence requiring 4 bits of data storage. Thus, each feature histogram would require 4 bits for each possible vector choice (*eg* with 20 vectors, $4 \times 20 = 80$ bits, or 10 bytes, of data storage would be required). The total data storage requirement of the personal signature is then 8×10 (for the eight features) + 5 (for the canonical vectors) or 85 bytes of data.

The data reduction ratio is a substantial figure at $\sim 770 : 1$. This level of reduction will allow for rapid comparison and cost effective storage of the facial information. Even with this amount of reduction, the personal signatures stored still incorporates both first order (pixel level) feature data and second order structural information (in the form of the inter-feature measurements). Added to this factor, is the ability to weight the different features when performing comparisons.

5.4 FAMFIT: A Complete Facial Recognition Algorithm

The overall face recognition system introduced in this thesis is the realisation of the block diagram given in Figure 5–4. This incorporates the segmentation stages described in chapter 3, linked to the encoding algorithm introduced in chapter 4, followed by the facial signature production and comparison outlined above. A completely novel facial recognition algorithm has now been formed from these discrete functional elements. This **F**eature **A**nd **M**easurement based **F**acial **I**dentification **T**echnique has been christened **FAMFIT**.

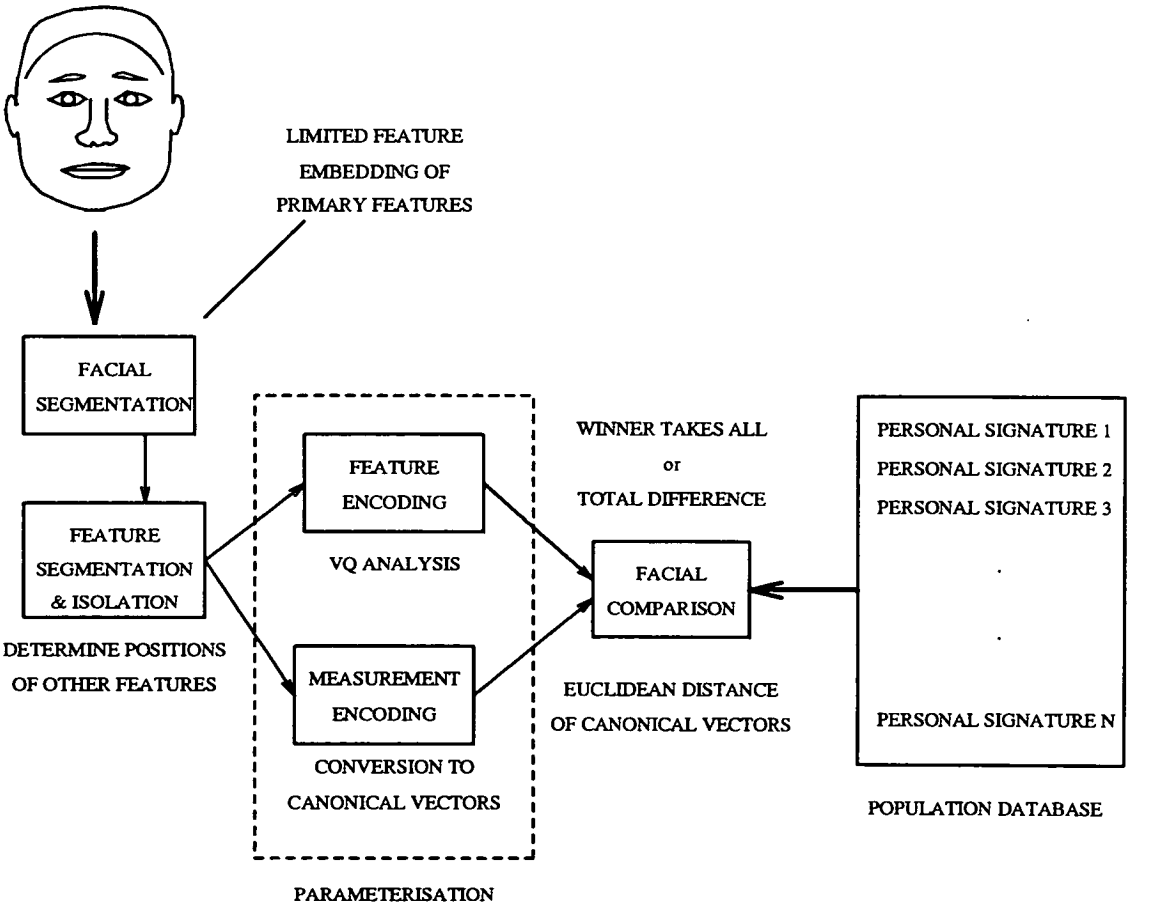

Figure 5–4: The complete algorithm for FAMFIT.

The algorithm produced takes account of the system constraints, laid out in chapter 1, in the following ways:

- The algorithm should have the ability to deal with images captured in a relatively unconstrained environment.
- The probabilistic comparison stage exploits the good generalisation provided by the training process.
- The design of the algorithm incorporates much of the present knowledge regarding human facial recognition ability.
- The data reduction achieved by the facial parameterisation is sufficient to facilitate rapid comparisons and cost-effective storage.
- The computational load required by the new algorithm is not beyond the computational capabilities of present technology.

It now only remains for the proposed algorithm of automatic facial recognition to be tested on a sample population. However, before doing this, another important method of facial recognition must be considered.

5.5 WISARD

The WISARD system, developed by Wilkie, Stonham and Aleksander, is possibly the only widely known facial recognition system. The principle of operation is very closely linked to the intrinsics of neural networks. However, the simplicity underlying the WISARD system has meant that it has been available as a real-time hardware device for many years[184, 185] and its operation is well understood[186].

5.5.1 Details of Operation

WISARD is a general purpose pattern recognition device based on a distributed RAM architecture. In operation, the input image is broken up into groups of several pixels. The bit pattern contained in a group of n pixels can be used to generate an n-bit memory address (termed an n-tuple). In the training phase, each RAM location is initialised to '**0**'. For each n-bit address generated from the image, the value of '**1**' is assigned to that memory location. Hence, the pattern of the input picture defines a set of ones and zeros in the memory locations of the RAM.

In the test phase, the entire image is processed in the same way as for the training images. The number of bit matches within the RAM, constitutes the level of similarity between the test image and the training image (or images). For a number of objects, $\mathcal{X}$, the same number, $\mathcal{X}$, RAMs (termed discriminators) are required. These discriminators can then be dedicated one each, for each of the objects.

Each discriminator is trained in isolation, receiving only examples of the object it is expected to recognise. For example, in an alphabetical character recognition system, the 26 discriminators would be presented with various examples of **their** letter. An internal representation of the relevant character would be constructed in each of the RAMs.

To perform recognition, a new character could be presented to each of the discriminators. The output of each RAM would be the number of bit-level matches observed between the new character and the internal representation of that discriminator. The highest output value would then be chosen as the best match to the test character. This function is illustrated in Figure 5–5 for a very simple example. In this case, the **A** character would be chosen as the most likely match, because its discriminator is producing the most positive output.

One of the strengths of WISARD, not illustrated in this example, is the abil-

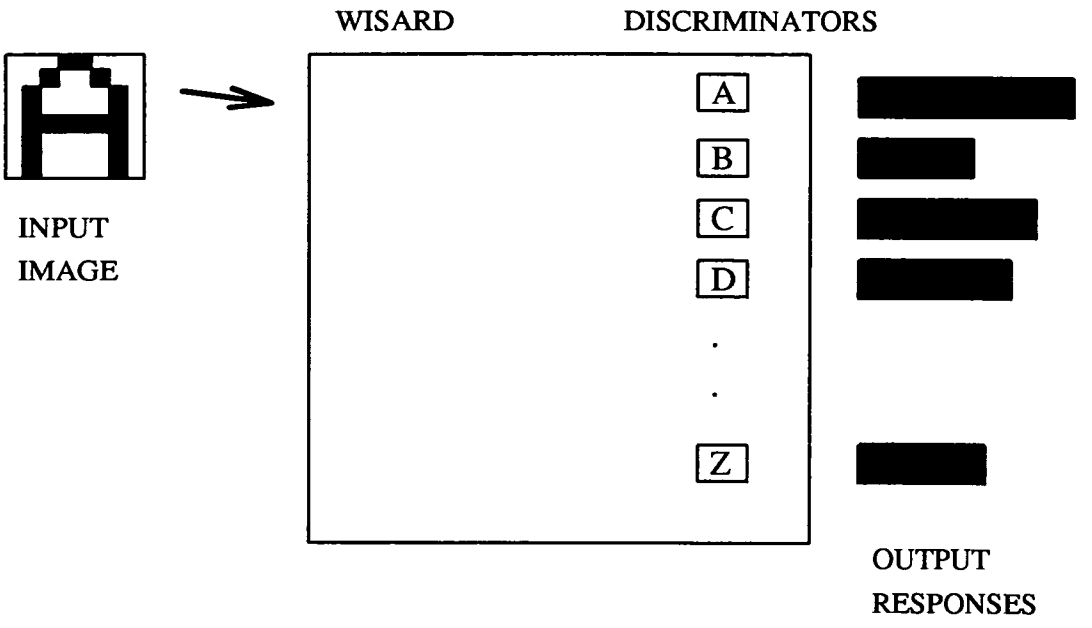

Figure 5–5: The WISARD system in operation.

ity to generalise the input pattern. For example, it would be possible to train a discriminator with a number of different patterns representing the same character (*eg* different fonts) and the internal representation for each character, would represent the **generalised** pixel pattern for that character.

The patterns that can be recognised using this device are obviously not limited to characters. However, in the implementation of WISARD described above, a binary input is necessary (*ie* in image terms, only black and white can be used). Also, when breaking up the input image into n-tuples, a random encoding is used to reduce the correlation of the input data facilitating efficient data storage.

5.5.2 WISARD for Faces

In research reported by Stonham[187], WISARD's performance at face recognition was evaluated. In Stonham's experimentation, each discriminator was trained to recognise the face of one member of the test population. A large number of images, between two and four hundred, of each person were used in the

training phase of the device. This substantial training set allowed the WISARD to construct a relatively general internal representation of each person.

The training process was performed in real-time, with the subject standing in front of the camera for several seconds as the discriminator constructed its internal representation. The training was terminated when the discriminator output gave a consistently high response.

In Stonham's experiment, he used a population of sixteen, and training of about 20 seconds – or until the response of the discriminator was in excess of 95% for the correct stimulus. During this training, the subject changed their facial position and expression, hence the internal representations constructed by the device were largely invariant to these changes[1]. Testing was performed by presenting new images of the sixteen members of the population to all the discriminators, and recording the output responses obtained.

In this experimentation, the outputs of Stonham's WISARD indicated the correct subject with 100% accuracy. Similar performance results for facial recognition using WISARD have been presented more recently[188, 189].

5.5.3 Differences between WISARD and FAMFIT

WISARD is a very different method of face recognition from the FAMFIT algorithm which has been introduced in this thesis. WISARD is fundamentally a *dumb* recognition system capable of recognising very different objects given the correct training. A WISARD device could be trained to recognise turnips as easily as faces. However, in all cases it will assign as much importance to the background information as to the foreground object. Its performance at face recognition, reported by Stonham, is all the more remarkable considering the crude binary input images used.

---

[1] Rotation of the face away from the camera was not permitted.

The FAMFIT system extracts a number of key facial characteristics and stores these as the facial signature. The advantage of this approach is that a substantial amount of data reduction can be performed. However, the FAMFIT system does require facial segmentation stages which are not needed by the WISARD approach.

The manner in which data is stored by WISARD is also very different from the FAMFIT approach. The entire contents of the discriminator must be stored for future recognition. The data space required for each person's face (as stored in the discriminator) is substantial; a 64Kbyte image would require 32Kbytes of data storage. Whereas, the FAMFIT system requires approximately 85 bytes of data for each facial signature. The calculation of data storage for the WISARD system, assumes binarisation of the image prior to its placement in the discriminator and an n-tuple sample size of four.

Despite the substantial differences between WISARD and FAMFIT, WISARD must be considered as a suitable benchmark as it is the most successful face recognition system available to date.

5.6 Performance Analysis

To perform comparisons between WISARD and the FAMFIT system a number of different experiments have been undertaken. However, in order to validate the experimental results, further consideration must be given to the way in which the relative performances of the two systems can be analysed. Before doing this however, precise definitions of the two parallel functions of recognition and verification must be given. In human terms:

Recognition is the function performed each time a human sees a familiar face, and can retrieve that person's name (or some similar identifier to that face) from their entire database of all the people they know.

Verification is the function performed by the person at passport control when they compare a passport photograph with the person standing in front of them.

A number of measures have been devised to evaluate the performance of automatic systems, when mimicking the recognition and verification tasks.

5.6.1 Recognition

By convention, in an automatic recognition system, the output of the system is the name of (or reference to) a particular member of the sample population. This is the most likely, or most similar, match to the input stimulus available in the population database. In performance testing, the correct match to each input image is known, and therefore, it is possible determine whether the output produced by the system is the correct match for that input. The percentage of successful matches is a first order, crude, measure of a recognition system's performance.

Knowing that in most systems, the similarity between the stimulus and all the members of the population is computed to arrive at the most likely output, a further measure of system performance can be found be exploiting this information. If the similarity measures produced for each population member, are compared with each other, then the system can output the entire set of population names, in descending order of likelihood. As the desired output is known, then it is possible to detect its position in the output ranking list. The average position of the correct output for an example test set is termed the *average rank*. The problem with this performance measure, is its vulnerability to extreme, out-lying, results. For example, a system which has a 'first place' recognition rate of 90% may well have a poor average rank figure, if a substantial number of the 10% that failed to be recognised are very badly missed.

5.6.2 Verification

For verification purposes, the test image is only compared with the personal signature of one individual. In effect, the subject claims to be a particular person. The verification of this claim is performed by subjecting the similarity measure, produce by the recognition comparison, to a threshold. If the similarity is below this threshold, the claim is rejected, and if it is above the threshold, then the subject is accepted. The threshold placed on the similarity metric is termed the *acceptance threshold.*

The performances of identity verification systems are conventionally measured in accordance with the two possible ways in which such systems can fail[2]. Firstly, it can fail by incorrectly rejecting a genuine claimant. This error is termed the TYPE I or 'false reject' error. Secondly, the system can fail by falsely verifying the identity of an imposter. This second type of error is termed TYPE II or 'false accept' error. In general, the two error rates are quoted as ratios (or percentages) of the total number of tests.

There is, however, a problem associated with system evaluation based on the levels of the False Acceptance or False Rejections Ratios (FAR & FRR). By definition, the two ratios are inter-related. For example, as the acceptance threshold is lowered the FAR will increase and the FRR will decline. In order to assess this inter-relationship, the graphs of the TYPE I and II errors are often superimposed. In the ideal case, Figure 5–6, there is no crossover between the two error rates and the acceptance threshold can be placed somewhere in this middle region. However, in most practical systems, Figure 5–7, the error rates have to be traded-off against each other depending on the exact system requirements. The percentage error rate at which the two error graphs cross-over, the *Equal Error Rate (ERR)* is often quoted as a comparative measure.

The means by which it is possible assess performance, for both verification and recognition experiments, have now been presented. However, it must be not-

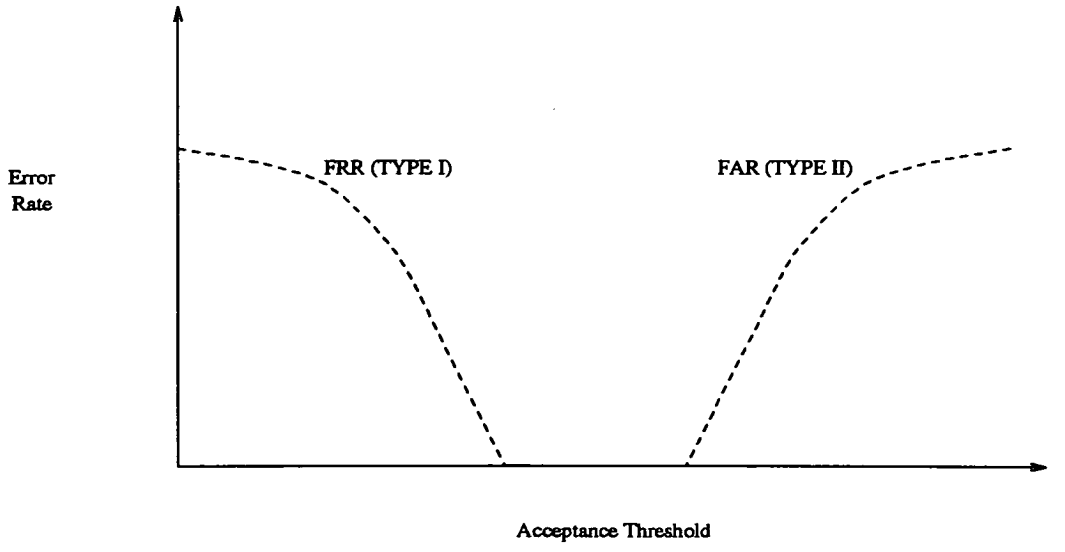

Figure 5–6: Ideal FAR and FRR curves.

ed, that without accurate knowledge about the data acquisition and experimental procedures, the performance figures obtained are largely without meaning. For this reason, the same experimental conditions have been adhered to throughout the experimental analysis presented within this thesis. The experimental conditions used for this study are outlined in Appendix B.

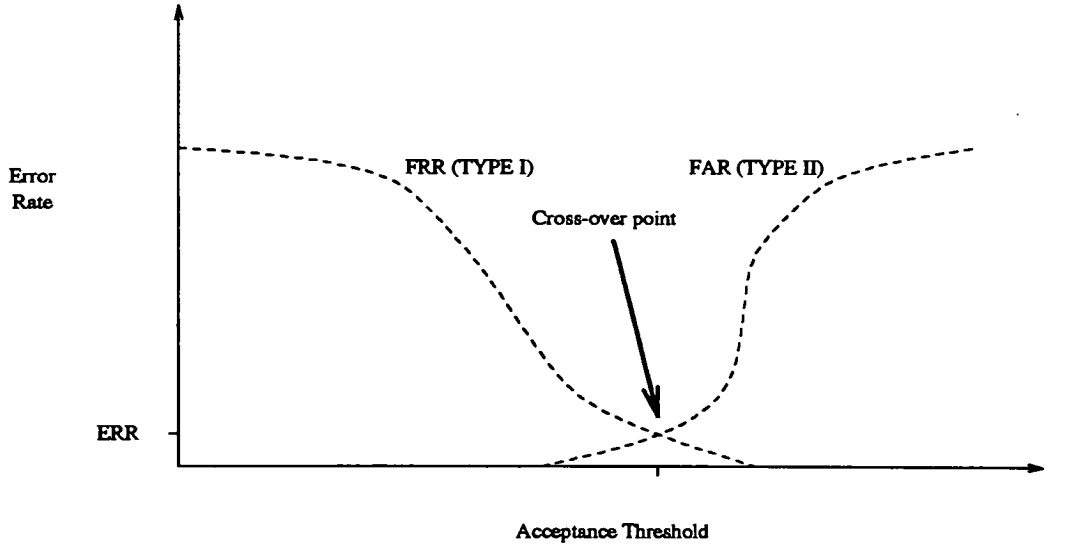

Figure 5–7: Practical FAR and FRR curves.

5.7 Experimental Results

To evaluate the performance of the proposed FAMFIT system, a number of different experiments have been devised. The experimental results, presented in this section, reveal the strengths and weaknesses of the system under different operating conditions. The relative performance of a number of the different system variants, described in the preceding sections, has also been examined.

In order to evaluate the performance of the comparison stage of the algorithm, in isolation, the initial facial segmentation function has been performed manually. In this way, the true error rate of the recognition stage can be measured. However, as a result of the software implementational details, it was only possible to specify the positions of the three primary features correctly. Therefore some of the system errors could be caused by minor mis-locations in the other feature points, although this was not deemed to be a major factor, as significant feature mis-locations of this kind were found to be very infrequent. Results are reported in section 5.7.7 which quantify the system's recognition performance when primary feature location is performed automatically using the limited feature embedding algorithm. This experimentation represents the complete FAMFIT system in fully automatic operation.

The initial experiments were performed using the FAMFIT system as it has been described in the preceding sections. Initial vector quantization codebooks were constructed from one additional image of each population member. The recognition score was measured using the second approach described in section 5.3.1, the *total difference* method. A uniform weighting scheme has been used for the initial experiments. At this stage, the system used does **not** incorporate the feature inter-relationship information discussed in section 4.8.

For practical reasons, the population used in this trial is entirely male. There is no reason why the FAMFIT system reported in this thesis, cannot be expanded to include females, however, it would be anticipated that additional vector quantizer codebooks would be required to adequately encode the feature types present in the female population. WISARD does not require any modification to perform recognition on a mixed gender population.

5.7.1 Training Sessions

Despite the probabilistic approach to personal comparisons, and the experimental constraints used, changes in expression and other extraneous factors, may well have a measurable effect on the proposed system. For example, it may be that on any given day, a particular subject's face can appear differently from on other days, as a result of changes in mood or any other factors. Detailed analysis of possible *systematic* changes in the face on a *day-to-day* basis is required to evaluate whether the system will be able to perform robustly. Biometric recognition studies which have ignored such systematic changes, by using data acquired at one *sitting*, cannot claim to have modelled real-world system operation.

To perform this evaluation, a test population of 10 people was used. For these ten people, three images were captured on five consecutive days – *ie* a total of 150 images were used in this trial. Two alternative strategies were implemented to produce the facial signature data. These two strategies were devised in order to determine whether training and testing on different days, had a significant effect on the system's performance. The two strategies were as follows:

1. Training using all six images from the first two days and testing on the images from the other three days. Six images are used for training.
2. Training with one image from each of the five days and testing on all the others. Only five images are used for training.

The relative performances of the two approaches are given in Table 5–4, this table

gives the average rank figure and the cumulative success rates for recognition in the first three positions. The first place recognition percentage corresponds to the number of successes, s, divided by the total number of tests, n.

	Experiment 1 2 day training	Experiment 2 all days training
Average Rank	1.36	1.15
1^{st} Place Recognition	83.33%	91.00%
2^{nd} Place Recognition	88.88%	96.00%
3^{rd} Place Recognition	93.33%	98.00%

Table 5–4: Comparative test results for different training times.

Training on all days, as opposed to just two, appears to perform better. This suggests that the facial signatures obtained from all the test days, contained a more generalised view of each person's face than the other approach. It can also be argued that these results suggest that there is some variation within images taken from different sittings, which the system is not able to cope with. However, some statistical evaluation of these performance figures is necessary to assess their true validity.

It is necessary to assess whether the observed difference in performance is due to a real difference in system performance, or merely an artifact resulting from random effects. To perform this evaluation, it is convenient to first assume that there is no intrinsic difference between the two experimental results, and then endevour to disprove this test hypothesis[190]. This assumption is termed the null hypothesis, H_0.

In this case, the null hypothesis is : $H_0 : p_1 = p_2$, where p_1 and p_2 are the respective success rates of experiments 1 and 2. With $p_1 = \frac{s_1}{n_1}$, where s_1 is the

absolute number of successful recognitions and n_1 is the total number of tests relating to the first place recognition rate for Experiment 1[1].

A standard test statistic is[191] :

$$z = \frac{p_1 - p_2}{\sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{5.5}$$

where p is the pooled success rate $p = (s_1 + s_2)/(n_1 + n_2)$ and $q = 1 - p$.

Assuming sufficiently large values for n_1 and n_1 then the normal distribution can be used to approximate the difference between p_1 and p_2. The value for z can be used to read off the probability of the difference being genuine from the normal curve[2].

For this experiment:

$$p = \frac{75 + 91}{90 + 100} = 0.87$$

$$z = \frac{0.83 - 0.91}{\sqrt{0.87 \times 0.13(\frac{1}{90} + \frac{1}{100})}} = -1.637$$

The value of $z = -1.637$ is significant at the 10% level (*ie* there is only a 1 in 10 chance that the observed performance difference between experiments 1 and 2 is *not* a result of genuine performance difference) but this result is not significant at the 5% level.

This experimentation suggests that while the system performs better with training and test images taken from the same sessions, it does achieve a reasonable amount of generalisation, demonstrated in its ability to recognise faces using training and test images from different sessions.

---

[1]The first place recognition rate is being used as the comparison metric, however, it would also be possible to use the mean and standard deviation of the rank figure.

[2]This measure of significance assumes that n_1 and n_2 relate to independent random samples. This assumption is not adhered to here as the same test data is used in both experiments, thus the value for z can only give an **indication** of the likely significance of the performance figures obtained.

5.7.2 Spectacle Wearing

The wearing and removal of spectacles represents a problem for a facial recognition device. There are now a large number of people who sometimes do and sometimes do not wear their spectacles. Therefore, the problems associated with spectacles must be addressed, if the proposed facial recognition system is to be viable.

Using the FAMFIT facial partitioning, the wearing of spectacles affects the visual appearance of the eyes, the bridge of the nose and the overall face. Thus, the removal of the spectacles could result in alterations within four out of the eight facial features used in the facial parameterisation process. To evaluate the performance of the system, as a function of spectacle wearing, a population of eight people was used. Five images of these people wearing glasses, and ten without, were acquired from five different sessions. Three experiments were then conducted using five training and ten test images. In all cases the five training images used were drawn one from each of the five sittings.

Ex. 1 All five training images with glasses, all 10 test images without.

Ex. 2 All five training images without glasses, 5 test images with and without.

Ex. 3 Two training images with and 3 without, 7 test images without and 3 with.

The performance rates for these three experiments are shown in Table 5–5. The recognition rates for all three experiments are very good, however, this is partly due to the small population size. There is no obvious difference between experiments 1 and 2 (in terms of their first place recognition rates), however, experiment 3 has a higher recognition rate than both 1 and 2. Using the statistical test given in the previous section, a figure of $z = 1.664$ was obtained, this is significant at the 5% level. It can be concluded that the system is able to

construct a general facial signature, when given training examples of faces with, and without, spectacles.

	Experiment 1 only glasses	Experiment 2 without glasses	Experiment 3 mixture
Average Rank	1.20	1.10	1.03
1^{st} Place Recognition	91.25%	91.25%	97.50%
2^{nd} Place Recognition	95.00%	98.75%	98.75%
3^{rd} Place Recognition	96.25%	100.0%	100.0%

Table 5–5: Training with and without spectacles.

5.7.3 Comparison Metrics

In section 5.3.1 two different methods of calculating the most likely match to a given stimulus were presented. Firstly, there was the *winner takes all* approach, based on a weighted sum of the eight individual feature probabilities. Secondly, there was the *total difference* approach, used in the experiments outlined in the preceding sections, which calculated an overall similarity match, given the relative distance measures of all the possible codebook entries. The recognition performances of both of these approaches has been analysed on the entire test population.

The full population of forty people was used for this experimentation. The facial signatures were constructed using the first ten images and the testing performed using the second ten images. The test set thus contained 400 images. The initial vector codebooks described in section 4.7.3 (constructed from example features taken from all the population members) were used for this experiment. The recognition results are detailed in Table 5–6. Comparing the first place recognition scores, a z figure of 1.06 is obtained, this is not significant. Thus, neither of the two competing approaches can be described as superior.

	Winner Takes All	Total Difference
Average Rank	1.43	1.37
1^{st} Place Recognition	86.00%	88.50%
2^{nd} Place Recognition	93.00%	92.75%
3^{rd} Place Recognition	96.00%	95.50%

Table 5–6: Comparative test results for different accumulation methods.

5.7.4 Vector Codebooks

The following experiments analyse the performance of the proposed system when given different sets of vector codebooks.

Codebook Generation

Section 4.7.3 discussed three methods of clustering vectors into new groupings. The three approaches considered were K-Means clustering, Kohonen's Self Organising Feature Map and the Linde-Buzo-Gray technique. Using the example features from all forty people, these three methods have been used to construct vector codebooks containing only 20 examples of each feature. All three of these techniques should be able to perform this function while still maintaining a good coverage of the initial feature space.

To evaluate the performance of the three clustering operations, the three, twenty member codebooks generated were used in a complete recognition trial using all 40 people. The performance figures of these three experiments are given Table 5–7. There is a substantial variation between these three techniques. The K-Means approach is significantly poorer than KSOFM (at the 5% level) and the KSOFM is significantly poorer than the LBG method (at the 1% level).

In order to explain these results, it is helpful to examine the population wide feature histograms for these three approaches. Figure 5–8 shows how frequently

	LBG	K-Means	KSOFM
Average Rank	1.46	2.82	1.99
1^{st} Place Recognition	83.50%	64.00%	70.25%
2^{nd} Place Recognition	90.00%	76.75%	81.50%
3^{rd} Place Recognition	94.75%	83.00%	88.50%

Table 5–7: Recognition performance for the three clustering methods of codebook design.

the vectors have been utilised during training of the 40 subjects for the K-Means technique. The graphic illustrates the feature histograms, as discussed in section 5.2.2, in a 3 dimensional manner, with the lightness of each square representing the frequency of use of that particular vector choice. In this representation, *black* signifies a well used vector, with lighter tones signifying less utilisation. Figures 5–9 and 5–10, illustrate the same information for the KSOFM and LBG techniques, respectively.

For good facial parameterisation, it would be desirable that a good utilisation of all the available vectors occurred. This is not the case for the K-Means clustering vector set, where many of the possible vectors are almost completely unused. To some extent, this is also true for the KSOFM vector set, in particular, the choice of hair vectors appears to have a very poor distribution. The reasons for this variation in vector choice are complex. However, the most likely situation is that the two clustering techniques have maintained very good coverage of the entire data space, by clustering the most similar, and in some cases the most commonly used, initial vectors into only one or two vectors in the new codebook.

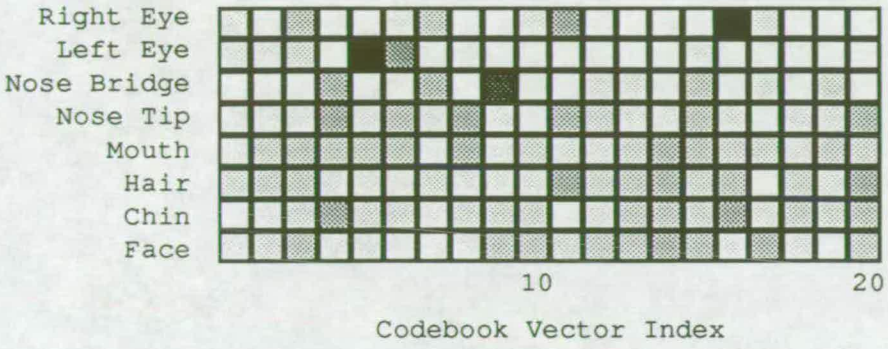

Figure 5–8: Feature choice histogram for K-Means clustering.

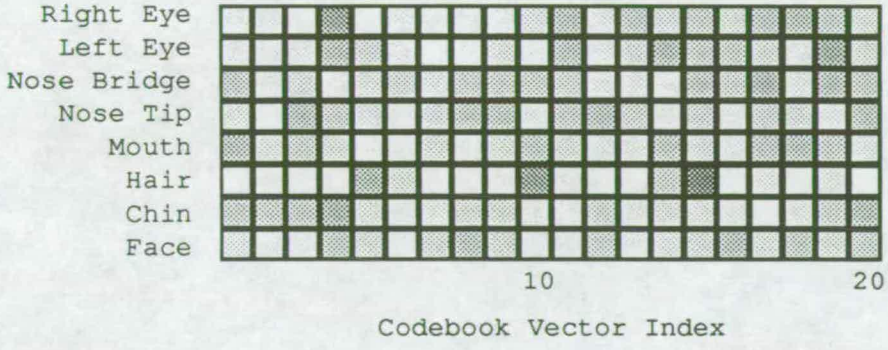

Figure 5–9: Feature choice histogram for KSOFM codebook design.

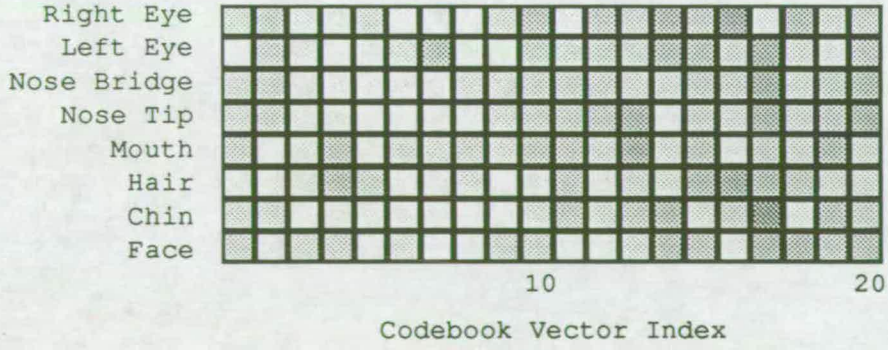

Figure 5–10: Feature choice histogram for LBG codebook design.

Thus, assuming a *normal* distribution of features, not enough new vectors will have been assigned to the majority of features which lie close together. Therefore, when used to encode faces, too many features are assigned to the most general vectors in the codebooks. In this way, distinctive information regarding these features is lost, hence the relatively low recognition rates. The LBG method of vector quantization codebook design does appear to have succeeded in achieving a good coverage of the initial feature space, while still maintaining sufficiently accurate encoding of the most common feature types.

It must be noted that the KSOFM algorithm used here may be capable of significantly better performance if the training of the network was performed using different training parameters. Indeed, Allinson[192] reported good recognition results when using Kohonen's SOFM for a similar purpose. Further experimentation would be necessary to determine whether the KSOFM performance can be improved upon.

Compared to the results reported in the previous section, the degradation in performance produced by halving the number of available codebook vectors is not as drastic as could have been expected. The first place recognition rate using the LBG codebooks is only 3-5% lower than using the full forty vectors. This suggests that codebooks of a substantially lower dimensionality than the population size can be successfully used for recognition, if the codebooks are designed in a suitable manner.

Dedicated and non-dedicated codebooks

A further set of experiments were performed to assess the difference between the use of a dedicated vector set (*ie* containing vector examples drawn from the test population) and a non-dedicated set. This analysis was performed by splitting the test set into two separate halves, and performing recognition on the four permutations of populations and codebooks. Again ten test images were used for

training and testing. The results of these four experiments are shown in Table 5–8. There is a significant degradation in performance when using a non-dedicated vector set.

Training Population		Test Population	
		First Half	Second Half
First Half	Average Rank	1.21	1.63
	1^{st} Place Recognition	92.50%	80.00%
	2^{nd} Place Recognition	96.50%	89.00%
	3^{rd} Place Recognition	96.50%	89.50%
Second Half	Average Rank	1.48	1.15
	1^{st} Place Recognition	75.00%	91.00%
	2^{nd} Place Recognition	87.00%	95.50%
	3^{rd} Place Recognition	94.00%	98.50%

Table 5–8: Recognition rates using different populations.

5.7.5 Feature Weighting Strategies

For the earlier experiments a uniform weighting of the different features was assumed. There are, however, numerous ways in which the available feature information can be weighted when it is accumulated. In order to evaluate whether an improvement in system performance is possible, two different weighting functions have been investigated.

The two strategies are as follows:

1. weight the inner facial features highly (*ie* eyes, nose and mouth).
2. weight the outer facial features highly (*ie* hair, chin and face).

The actual values used for the weighting function, w_j, are given in Table 5–9. Using these two strategies the recognition rates reported in Table 5–10 have

been achieved – the results of the same trial using a uniform weighting scheme, reported above, are reiterated here. The observed improvement between uniform weights and the outer features biased weighting vector is not very significant.

Feature	Uniform	Favour Inner	Favour Outer
Right Eye	1	5	1
Left Eye	1	5	1
Nostrils	1	5	1
Bridge	1	5	1
Mouth	1	5	1
Hair	1	1	5
Chin	1	1	5
Face	1	1	5

Table 5–9: The numerical values used for the weighting function, w_j.

	Uniform	Favour Inner	Favour Outer
Average Rank	1.37	2.07	1.20
1^{st} Place Recognition	88.50%	73.75%	91.00%
2^{nd} Place Recognition	92.75%	83.75%	96.00%
3^{rd} Place Recognition	95.50%	87.75%	98.50%

Table 5–10: Recognition rates for the three weighting schemes.

Table 5–10 suggests that the outer features are of more use in facial recognition than would previously have been expected. To fully investigate this result, a series of eight other experiments, which measure the recognition rate of each of the eight features *in isolation*, was performed. The results of this experimentation are given in Table 5–11.

The three templates containing information for the face, the hair and the chin do appear to contain more discriminative information than the other templates.

Feature	Recognition Rate
Right Eye	30.00%
Left Eye	25.00%
Nostrils	28.75%
Bridge	23.75%
Mouth	30.00%
Hair	40.00%
Chin	57.00%
Face	38.50%

Table 5–11: Individual features' recognition scores.

This is not in accordance with many of the facial recognition studies, reviewed in chapter 2, which emphasised the importance of the inner facial features[1]. This experimentation represents one of the few objective studies of the relative recognisability of the individual facial features.

5.7.6 Feature Measurements

The inter-feature measurement information has been encapsulated in the five canonical vectors. The comparison of these vectors for different facial images, represents a simple, measurement only, facial recognition technique. To evaluate the information content of the canonical vectors chosen, a simple experiment was performed.

[1]It must be noted that the overall facial template, *the face*, contains crude information pertaining to these inner features.

Using the same test and training data as for the experiments detailed above, the following recognition results were obtained, Table 5–12. The first place recognition rate of a third is very much poorer then the feature based recognition described above. The recognition rate achieved is also poorer than some of the other measurement only facial recognition systems, reviewed in chapter 2. The main reasons for this are as follows:

- Only a small set of feature measurements is being used here.
- Some minor inaccuracy in feature location is likely.
- The image resolution limits the accuracy of the measurements obtained.

Average Rank	6.25
1^{st} Place Recognition	33.00%
2^{nd} Place Recognition	44.25%
3^{rd} Place Recognition	52.00%

Table 5–12: Recognition based on measurement information only.

Acknowledging these problems, it is still likely that the inclusion of the measurement information can enhance the performance of the VQ feature based system. The encorporation of this information is the realisation of the feature and measurement system, FAMFIT, proposed in this thesis.

The relative measurement information, between a test face and a particular personal signature takes the form of a Euclidean distance of the five canonical vectors. The feature information is in the form of the accumulated *match score* (as obtained by the total difference approach). In order to relate these two figures, the Euclidean distance was rescaled and negated. This value, representing the measurement information, was then added to the *match score*, representing the feature information, to obtain one overall recognition score. The relative proportions of these two components were determined using a heuristic approach.

Rating the canonical vector based difference, at approximately one fifth of the importance of the feature based data, then the system was seen to improve, Table 5–13. This perceived improvement is not, however, significant.

	Features Only	FAMFIT
Average Rank	1.20	1.16
1^{st} Place Recognition	91.00%	92.25%
2^{nd} Place Recognition	96.00%	97.25%
3^{rd} Place Recognition	98.50%	98.50%

Table 5–13: Recognition rates using features and measurements.

5.7.7 Automatic Feature Location

The first place recognition rate of 92.25% reported for the FAMFIT system has been achieved using the correct locations, *ie* manual placement, of the three primary facial features (eyes and mouth). However, for the FAMFIT system to be of practical use, this function must be performed automatically. If the limited feature embedding system of feature location, as introduced in chapter 3, is adopted as the first stage of processing the fully automatic FAMFIT system is realised.

Using this approach, the recognition rate drops to 84.5%, Table 5–14. This degradation in performance is not surprising, as the accuracy of the primary feature location stage is crucial to the entire system. This degradation in performance is significant at the 1% level.

5.7.8 WISARD

To perform a fair comparison between the WISARD and the FAMFIT systems, the same training and test data must be used. To this end, each of the forty

	Automatic Location	Manual Location
Average Rank	1.85	1.16
1^{st} Place Recognition	84.50%	92.25%
2^{nd} Place Recognition	90.75%	97.25%
3^{rd} Place Recognition	93.75%	98.50%

Table 5–14: Recognition performance using manual and automatic feature location.

WISARD discriminators were trained to recognise one member of the forty people in the population (using the same ten images of each person used to derive the facial signatures of the FAMFIT system). Thus, the contents of the WISARD's discriminators can be considered to be its facial signatures.

In testing, each of the four hundred test images were presented to all of the forty discriminators. The observed outputs were then used to obtain average rank and recognition rates – as for the FAMFIT system. Table 5–15 contains the comparative results of the two systems using identical training and test data. The difference between these two systems is very significant (at the 0.01% level). The fully automatic FAMFIT system, evaluated in the previous section, is also significantly better than the WISARD system, although only at the 5% level.

	FAMFIT	WISARD	Automatic Location FAMFIT
Average Rank	1.16	2.07	1.85
1^{st} Place Recognition	92.25%	80.00%	84.50%
2^{nd} Place Recognition	97.25%	88.00%	90.75%
3^{rd} Place Recognition	98.50%	92.25%	93.75%

Table 5–15: Recognition rates using FAMFIT and WISARD.

The FAMFIT system appears to be performing very much better than the

WISARD system. There are, however, a number of considerations which make this comparison a rather biased trial.

- WISARD requires substantially more than the 10 training images used here, in order to produce a proper internal representation.
- Initial manual location of the three primary features was performed prior to presentation to feature based system, this process is not required by WISARD.

Addressing the first point; while it is true that WISARD is being under trained in this trial, it may be true that **both** systems would improve given more training data. It cannot be categorically stated that WISARD's performance would improve disproportionally given more training data.

Assuming that a more accurate way of locating the primary feature points could be implemented automatically, then the FAMFIT system's performance reported here using manual location would be possible.

A third consideration is pertinent when comparing these two systems, that is the data requirement of the facial signature. In this comparison the new FAMFIT approach scores heavily, producing very good data reduction completely unparalleled by the WISARD system. The redundancy present within the WISARD representation is argued to facilitate recognition with different expressions and facial positioning, this function has been replicated in the FAMFIT system by using probabilistic comparisons.

5.7.9 Verification

A personal verification system can be implemented by subjecting the output recognition scores of both WISARD and FAMFIT to an acceptance threshold. As described earlier, there are two likely ways in which the verification function can fail, measured as the FAR and FRR. To compare both systems, a number of

different thresholds have been used, and the resultant FRR and FAR recorded at each point.

Figure 5–11, illustrates the FAR and FRR for both facial recognition techniques. The absolute threshold values have been removed so that the two systems can be compared. The FAMFIT technique has an equal error rate (the point where FRR and FAR errors cross-over) of ~7%. For the WISARD system the equal error rate is higher at ~10%. Again the arguments regarding the biased nature of this trial, as described in the previous section, can be applied here.

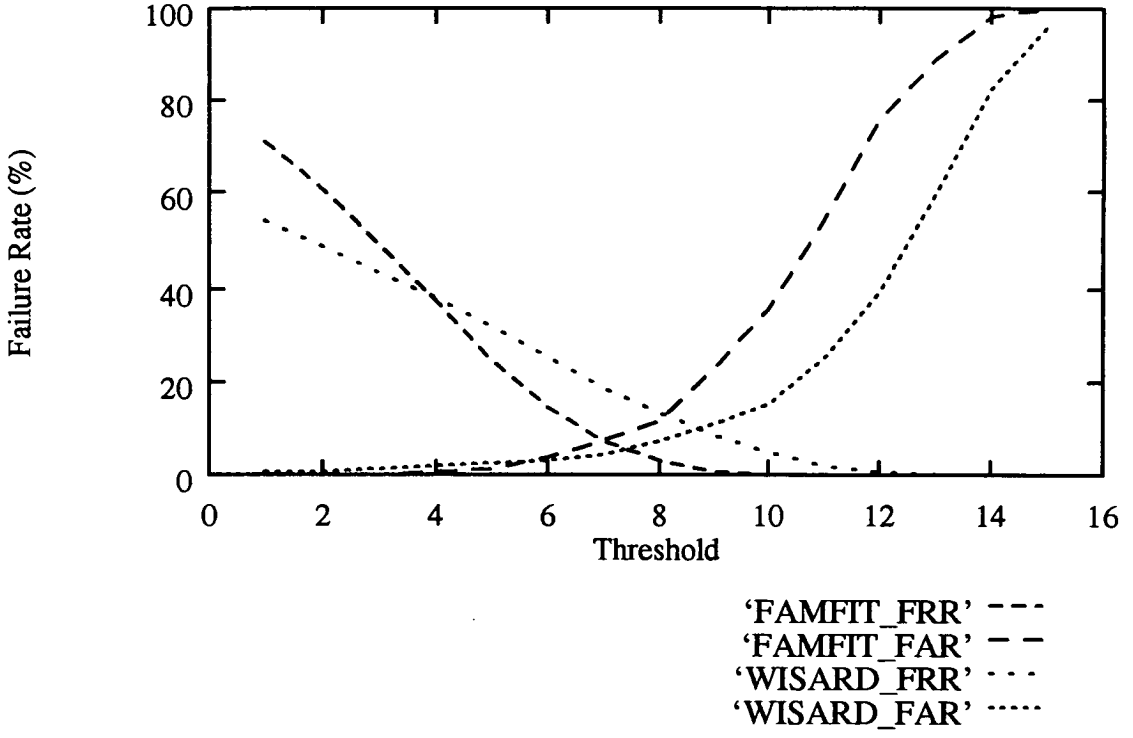

Figure 5–11: Verification curves for both systems.

5.8 Summary

The FAMFIT system of automatic facial recognition has been constructed from a number of different functional modules. Fundamental to the system is the provision of a very low data requirement, personal signature. This personal signature must be able to preserve a generalised view of the training stimulus. The major results of the experimental research into the operation of the FAMFIT system are presented in Table 5–16.

Experiment		Recognition Rate
Training Days	two days	83.33%
	all days	91.00%
Codebooks	LBG	83.50%
	K-Means	64.00%
	KSOFM	70.25%
Weights	favour outer	91.00%
	uniform	88.50%
Measurements only		33.00%
Features only		91.00%
Automatic Location		84.50%
WISARD		80.00%
FAMFIT		92.25%

Table 5–16: Summary of experimental results.

The FAMFIT system demonstrates the viability of a largely feature based method of automatic facial recognition. The vector quantization technique has been successfully exploited to yield substantial saving in data storage requirements. The generalisation, performed within the facial parameterisation process, has been successfully demonstrated. Therefore, it must be suggested that the

training data supplied to the device should reflect as many different variations as is possible.

The importance of vector codebook generation has been established, and the improvements that can be made in this area have been demonstrated. The crucial role of accurate feature placement, in a feature based system of facial recognition, has been illustrated. The fully automatic FAMFIT system, incorporating limited feature embedding, is still able to significantly out-perform the WISARD system, on the same trail data.

The FAMFIT system has also been used to provide objective evidence of the relative saliency of the different facial features. In this process, some of the established assumptions regarding feature saliency have been challenged. However, the important result is that these relative saliencies have been exploited in the construction of the FAMFIT system.

The performance of the measurement based element of the FAMFIT system is disappointing. It was hoped that the inclusion of measurement information would significantly increase the performance of the FAMFIT recognition technique, unfortunately, this is not the case. Recommendations regarding possible improvements to this entire FAMFIT algorithm are given the following chapter.

Chapter 6

Conclusions.

This concluding chapter will summarise, and attempt to gauge the significance of the results of the experimental work reported in this thesis. In this summary a number of possible refinements to the algorithm are suggested as future work. As another possible avenue for future research, the hardware implementation of the feature based facial recognition technique (introduced in this thesis) will be discussed. Practical considerations regarding the operation of a facial recognition device will also be considered.

6.1 Feature Location

The accurate placement of the facial features is crucial to a feature based recognition algorithm of this nature. For this system, the placement function was performed in two stages; firstly, the three primary features were located using the LFE algorithm, and secondly, the four other feature points were located assuming a particular target facial size.

The advantages of the LFE algorithm over its competitors are considerable. The LFE approach exploits *a priori* knowledge of the facial configuration coupled to a fine-tuning local placement algorithm, in order to locate the face as a collection of its component features. This approach is better suited to cluttered images

than the various edge based techniques discussed. The predictive placement of the facial features has been exploited to reduce the possible computational requirements.

The local placement of all of the features is performed using computationally optimised template matching. The algorithm chosen exploits many different aspects of the matching task to reduce the computational load. The template matching approach is demonstrably better than the competing neural network technique.

The performance of the FAMFIT system was analysed using both automatic and manual placement of the primary features. The poor results of the automatic approach suggest that the LFE algorithm is somewhat less than optimal. Some extraneous factors, including facial rotation and spectacles, have been identified as possible causes of these errors. Any improvements in the LFE algorithm could be expected to lead to a direct improvement in the recognition rate of the full FAMFIT system, *ie* the recognition rate could be expected to rise from the 84.5% of the fully automatic FAMFIT system, to approach the 92.25% of the manual location FAMFIT.

The performance of the LFE algorithm has been demonstrated here with facial images which have uniform backgrounds. However, the algorithm should be well suited to performing feature location when there is cluttered background information present. It is unfortunate that time was not available to test this aspect of the algorithm.

6.2 Facial Parameterisation and Comparison

The overriding requirement of the system is that it should be able to produce a low data rate generalised personal signature. By exploiting measurement and feature information, an approach suggested by human facial recognition research,

the FAMFIT system has demonstrated its ability to perform successful automatic facial recognition and verification.

The use of Vector Quantization to produce very low data rate descriptions of facial features has been shown to be successful. The use of VQ in other similar image pattern recognition tasks should be further investigated in the future.

The probability based system of training and comparison has been vindicated as a suitable method of performing inter-person comparisons. The feature histograms produced are able to capture expressional and temporal changes in the face, and exploit them, to construct a generalised facial signature. Indeed, these personal signatures can also store faces with, and without, spectacles[1]. The canonical variables approach to the measurement information, ensures that the maximum information content is preserved within the stored measurements. The feasibility of extending FAMFIT to a large population has been demonstrated by the consideration of different VQ codebook generation techniques.

The recognition performances of both the automatic and the manual location, versions of the FAMFIT system are significantly better than the WISARD device. In addition to this performance difference, the data space required to store each FAMFIT personal signature is a tiny fraction of that required to store the contents of each WISARD discriminator. The verification performance difference is less pronounced.

To improve on the best performance figure of FAMFIT, *ie* 92.25% on 400 test images, there are a number of different refinements which could be attempted.

- A considerable increase in the number of training images supplied to the FAMFIT system would increase the data storage requirements of the personal signature information, however, it may be that the more generalised signature produced would have considerable advantages for recognition.

---

[1] Assuming that the facial features can still be located.

- An increase in the amount of measurement information stored in the personal signature could well increase the recognition performance. Unfortunately, this would require more computational work in feature location.
- An increase in the number of features stored might improve the FAMFIT recognition rate. For example, the eyebrows and the ears, could be included as additional features, however, the ability to locate both of these features is very dependent on the subject's hair style.
- Further experimentation could be undertaken to determine whether the system performance could be improved using different feature resolutions and different feature weighting strategies.
- The inter-person comparison stage could be performed in a number of other ways than the two suggested in this thesis. This area could be another possible application area for neural network classification techniques.
- The inclusion of 3D information regarding the placement of the key features could well contribute to the recognition abilities of the device. The information content of this 3D data is now under serious investigation elsewhere. The additional computational and imaging equipment requirements would make this option unattractive unless the performance gain was substantial.
- An increase in image resolution from the present 256×256 pixel images could be used to increase the accuracy of the measurement information and the definition of the feature pixel data. These improvements could well lead to a measurable increase in recognition performance.

The results presented in chapter 5, which demonstrated the facial recognition rates based on each of the individual features, provide an objective measure of the differing significances which can be attached to each of the facial features. The continuation of this research is not strictly relevant to the further development of the proposed algorithm, however, it could deliver some profound conclusions regarding relative feature saliencies.

6.3 Real-time Implementation

A major objective of the research reported in this thesis, is to devise a successful automatic facial recognition algorithm, suitable for implementation in a real-time hardware device. Towards this goal, the algorithm which has been devised avoids the use of excessive processing capabilities, wherever possible.

Similar image processing research has led to the design of dedicated hardware devices capable of performing low-level image processing tasks (reviewed in [193]). However, none of these devices appear to be particularly suited to the implementation of the FAMFIT algorithm, thus, the use of custom hardware has been considered.

The integration of custom hardware onto the same chip as an image sensor has been pioneered by Anderson *et al* [194] as a possible method of implementing a similar image processing task. However, this actual device is not suitable here, as the image has to be stored while the different processing stages are performed, a facility not available on that particular single chip implementation.

The proposed algorithm incorporates a number of different processing stages requiring differing amounts of computational resource. Hence, it may be suitable to implement certain stages in custom hardware, with the less computational intensive tasks being performed in software, on a general purpose microprocessor (or microcontroller). A possible design of a device based on this principle would require a microcontroller and some specific custom hardware coupled to an image sensor and image store. A possible schematic for this device is shown in Figure 6–1. As a first stage in determining which parts of the FAMFIT algorithm would require hardware implementation, the entire algorithm has been broken down into its fundamental processing stages.

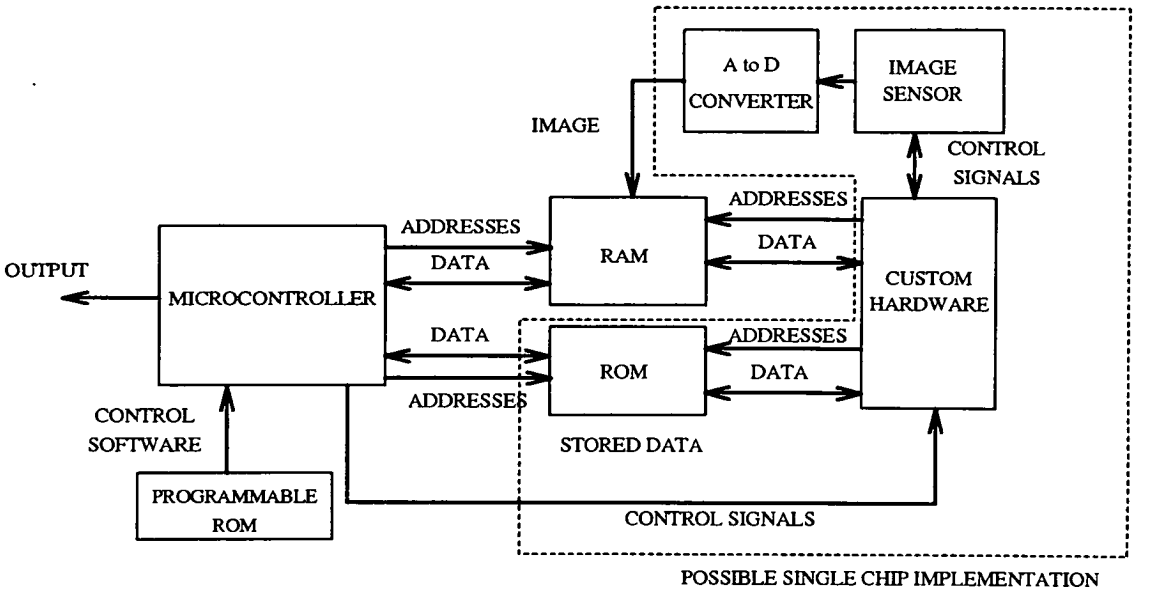

Figure 6–1: Possible implementation of an image processing algorithm.

6.3.1 Algorithmic Steps

The following algorithmic steps are performed to process and recognise a single face (the computational requirements of the various training functions have not been considered in depth, as these functions are only performed once).

1. Locate the primary features using Limited Feature Embedding. This involves the loading up of a number of example templates and comparing them with an image region centred on each pixel location in the search area. After each search the candidate sites are then clustered together.

2. Rescale the image in an interpolative manner, so that the face fills the image frame.

3. Locate the secondary features (hair, chin and nose), again template matching is performed using a number of previously obtained example templates.

4. Extract and store the pixel patterns for all of the eight features used. This involves sub-sampling the image to the correct resolution for each feature.

5. Calculate the inter-feature measurements.

6. Perform vector quantization on each of the eight features. This function necessitates loading in the relevant VQ codebook for each feature and then identifying the best match.

7. Calculate the canonical representations for the inter-feature measurements.

8. Evaluate the difference between the VQ coefficients chosen and the stored feature histograms for all the members of the present population.

9. Calculate the total Euclidean distance between each population member and the test face, using the canonical vectors.

In order to estimate the computational load required by these different algorithmic steps, typical compute times for each of the stages have been obtained.

6.3.2 Compute Times

Table 6–1 gives the compute times for the different algorithm steps (the step numbers refer to the steps described in the previous section). The timings correspond to *cpu* seconds for an optimised compilation of the algorithm software, running on a Sun 4/25 ELC Sparcstation. In this experimentation, a population size of forty people was used, with each feature codebook having twenty members.

The feature based inter-person comparison has been performed using the *total difference* approach. Of the two approaches discussed this one is likely to be the most computationally intensive and thus the worse case result.

6.3.3 Computational Details

For a real-time implementation of the proposed device, the total compute time of the device, of 36.5 seconds, is excessive. In order to determine how this function can be performed in real-time, the main computation steps have been analysed in detail.

Step	Description	Timing (secs)
1	Right eye search	17.2
	Cluster points	0.1
	Left eye search ($\times$ 7)	9.8
	Clustering	< 0.1
	Mouth search ($\times$ 4)	3.6
		30.8
2	Rescale Image	0.7
3	Locate Hair	0.7
	Locate chin	0.5
	Locate nose tip	0.5
	Locate nose bridge	0.4
		2.1
4	Extract features and	
5	Calc measurements	0.1
6	VQ all 8 features	0.4
7	Calc canonical vectors	< 0.1
8	Compare features and	
9	Compare measurements	2.3
All		36.5

Table 6–1: Compute times for component algorithmic stages.

Feature Location

For each individual feature, a template matching function is required at each pixel location in the search area. Table 6–2 gives the number of pixel locations investigated to find each feature pattern. The left eye search may have to be performed a number of times, if several isolated candidate sites are found for the right eye; typically seven separate searches are initiated for the left eye. The size of the mouth search is dependent on the locations of the eyes, a typical figure is given in Table 6–2. This search may also have to be performed a number of times (typically four searches are required). The number of times the left eye and the mouth searches have to be performed is data dependent. These additional searches were included in the algorithm compute times, given in Table 6–1.

Feature	Search locations
Right Eye	3438
Left Eye	263
Mouth	~150
Hair	144
Chin	96
Nose Bridge	48
Nose Tip	80

Table 6–2: Number of search locations for each feature.

Essentially the same routine is used to perform all of these different feature searches. As the most computationally intensive stage of the algorithm, a hardware implementation of this function would seem to be essential. The following pseudo-code illustrates the processing stages involved at each pixel location.

```
A. Extract an X by Y pixel block from the incoming image

B. Calculate a sample mean and standard deviation for that block

  For i = 1 to X step 4        (sub-sample block)
     For j = 1 to Y step 4
        Accumulate x(i,j)
        Accumulate x(i,j)^2

  Mean = sum x(i,j) / (X/4) * (Y/4)
  SD = sqrt((sum x(i,j)^2 / (X/4) * (Y/4)) - Mean^2)

C. Normalise the image patch to these statistics

  For i = 1 to X step 2
     For j = 1 to Y step 2
         x(i,j) = x(i,j) / SD - Mean
                          (data now in range 0-1)

D. Compare this patch with the pre-normalised templates

  For k = 1 to Z                (Z templates)
     For i = 1 to X step 2
        For j = 1 to Y step 2
           DIFF = |x(i,j) - template(k,i,j)|
   if DIFF > THRES then end

  if DIFF < LEAST_DIFF then LEAST_DIFF = DIFF

E. Store LEAST_DIFF for this location
```

The comparison stage described above, D, is implemented here using full floating point precision. Further investigation would be required to establish whether it could be performed sufficiently accurately using fixed point or integer arithmetic. The use of less precise arithmetic would yield a considerable saving in computational load.

If an acceptable level of accuracy could be obtained using a simplified arithmetical approach, it would be possible to implement the three main stages, B, C and D, using simple hardware. Four adders would be required to perform address generation, *ie* to extract the relevant pixel value at the correct time. Another adder would be required for the actual calculations. If only powers of two were used for the template sizes, then the divisions in stage B could be performed as shifts. A number of register memories would also be required to store the partial results at each stage. The square root function could be performed using a suitable look-up table.

Of all the computational stages, stage D – which is likely to be performed a number of times (depending on the number of templates being used) – will predominate in computational time. However, within all of the three main stages, the access time for the RAM (*ie* the time taken to extract each pixel value from the image store) is likely to take longer than the relevant computation performed on that element. Thus, the use of a fairly fast RAM would be advisable. On this basis, the hardware implementation of the feature location template matching stage is all most certainly feasible.

A further refinement of this algorithm would be to *window* the template over the search area, *ie* remove one column from the left and read the next column from the right when scanning across the image. In this way, substantial computational saving could be gained by not having to normalise and compare the whole template, at each location. Again, further experimentation with the software model of the algorithm would required to determine whether this is a viable approach.

Vector Quantization

Since the vector quantization algorithm being employed here is of a non-adaptive nature, the computation involved is not very complex. The pseudo-code for this function is given below.

```
For k = 1 to 8          (the eight different features)
   For l = 1 to 20          (20 vector codebook)
     For i = 1 to 20
       For j = 1 to 28
         DIFF = (x(i,j) - codeword(k,l,i,j))^2
         Store DIFF(l)
               (store diff for Total Difference comparison)

     if DIFF < LEAST_DIFF then LEAST_DIFF = DIFF

   store LEAST_DIFF for feature k
```

The kernel of this operation is very similar to the functions performed in the feature location stage. Thus, it would probably be possible to implement this function using the same hardware primitives.

Inter-person Comparison

The inter-personal comparison stage of the proposed algorithm is the other computationally intensive part of the complete system. However, the time taken to perform these comparisons is dependent on the number of people in the population. The code given below illustrates the operational steps required to perform recognition (it is assumed that the vector differences and canonical variables have been calculated already).

```
For i = 1 to N      (N population members)
  Read in histograms and canonical vectors (HIST and CV)

  For j = 1 to 8    (8 features)

    For k = 1 to T     (T VQ codebook entries)
       if HIST(j,k) < MIN then MIN = HIST(j,k)

    For k = 1 to T
      Accumulate HIST(j,k) * MIN / DIFF(k)  (TOTAL DIFF)
                    (DIFF(k) stored during VQ)

  For j = 1 to 5       (5 canonical vectors)
    Accumulate (CV(j) - NEW_CV(j))^2      (MEAS DIFF)

  MEAS_DIFF = sqrt(MEAS_DIFF)

  SCORE = TOTAL DIFF + A1 * ( A2 - MEAS_DIFF)
                           (A1 and A2 are scaling constants)

  store SCORE(i) for ranking list
```

When performing verification, *ie* comparing a new face with only **one** stored signature, then it is likely that this function could be performed in software. However, for recognition, the decision to perform this stage in hardware or not, would be dependent on the proposed population size for the device.

The real-time implementation of the proposed algorithm would appear to be feasible if a small number of refinements were to be made to the algorithm. However, these modifications should be performed within the software version of the algorithm, and the modified algorithm should again be fully tested before the design of hardware is attempted.

The proposed design could probably be implemented using the configuration suggested in Figure 6–1. The custom hardware could possibly take the form of a field programmable or dedicated gate array chip, designed to perform the feature location and vector quantization stages. The other functions could then be programmed in software to run on the microcontroller. This software would be stored in the programmable ROM. The population personal signatures, and

the feature templates, would have to be stored in the other read-only memory. The output would provide an identifier to the stimulus face.

6.4 Practical Operating Conditions

In addition to the precise details of the algorithm used, there are a number of other factors which should be considered when producing an automatic facial recognition device.

The future commercial development of biometric systems is conditional on the public acceptance of the concept of biometric data storage and retrieval. As a result of its unobtrusive nature, automatic facial recognition can be viewed as a strong contender in the biometric market. However, the performance of the FAMFIT system, as reported in this thesis, cannot yet rival the published performance figures obtained for some of the other available biometric systems.

It is an obvious extension of biometric research to encorporate several biometric systems together to form one very secure system[195, 196]. In such systems, the identity verification is conditional on a number of different personal characteristics being consistent with the claimant's identity. The strength of FAMFIT system, in this context, is that it can provide verification at an approximate error rate of 7%, while requiring less than 100 bytes of storage.

For automatic face recognition systems to gain public acceptability on their own, it is necessary that they be able to operate as unobtrusively as possible. It is, therefore, a requirement of such systems that they are able to perform facial recognition without placing any substantial constraints on the facial positioning *etc.* The LFE system of facial location has attempted to provide this flexibility, however, the computational search requirements do impinge on the performance achieved, especially in respect of facial rotation.

The data used for the FAMFIT experiments did require the subject to adjust their position until they thought that their face was inside the image frame. One possible way to perform this function in an operational device, would be to require the subject to look at a particular object, or perhaps, view their own image in a two-way mirror.

The time taken to perform the identity verification is also a considerable factor in system design. It would seem to be unacceptable to expect a customer to wait more than a few seconds to establish their identity; the delay required for automatic credit card authorisation appears to be publically accepted at present. This constraint must be applied to any real-time implementation of an automatic facial recognition device.

Any face recognition system would have to be able to cope with gradual changes in the face likely to occur through aging. Thus, the system must be able to adapt its internal representation of the subject's face, if they appear to change. It would be possible to incorporate an adaptive element into the FAMFIT system. The system could be allowed to alter slightly the feature histogram values, if the new instance of the face was slightly different from the stored signature. In this way, the system would gradually change the internal representation of the face, eventually *forgetting* the earlier training images. However, if the system incorrectly recognised someone and then altered the wrong personal signature accordingly, the whole system could require retraining.

6.5 Discussion and Concluding Remarks

A plausible method of automatic facial recognition has been designed and implemented in software. The performance of the system, reported in this thesis, is easily as good as any of the other reported facial recognition studies using similar amounts of data. The verification performance figures obtained, are not competitive with the best of the other methods of biometric identification.

The FAMFIT system is an interesting solution to the facial recognition problem. Automatic facial recognition based on facial feature classification and facial structure analysis has not previously been attempted. By using data compression on both feature and measurement information, the system is able to construct a very small personal signature, containing many of the intrinsic facial characteristics. The weakness of the system, both in performance and computational requirements, lies in the feature location stage. If the initial feature location could be performed accurately, in an automated way, then more of the full potential of the system could be realised.

During this research the problems of feature location and facial parameterisation have been largely decoupled, by performing the initial feature location manually. This has allowed the two halves of the system to be developed without impediment. It is perhaps true to say that too much time was spent on developing the parameterisation stage, with not enough effort devoted to the feature location problem. However, as yet, few other researchers have produced reliable methods of facial feature location.

The results presented in this thesis do not only suggest that facial feature classification can be used as the basis of an automatic facial recognition device but also that such a device could be implemented in real-time hardware. The improvements to the algorithm, suggested in this chapter, could well go some way to bridging the gap between facial recognition and the other biometric approaches. I hope that I have demonstrated that this research field is deserving of continued endeavour.

References

[1] D M Bowers. 'Access Control and Personal Identification Systems'. In *Proceedings of the Carnahan Conference on Security Technology: Electronic Crime Countermeasures*, pages 13–16, Lexington, Kentucky, May 1988.

[2] J R Parks. 'Biometrics: The People Sensors'. *Sensor Review*, 9(2):79–84, April 1989.

[3] J A Barry III. 'Back to the Future With Biometrics'. *Security Management*, 34(4):83–85, 1990.

[4] R L Maxwell and L J Wright. 'A Performance Evaluation of Personnel Identity Verifiers'. Technical report, Sandia National Laboratories, Albuquerque, New Mexico, 1987.

[5] J P Holmes, R L Maxwell, and L J Wright. 'A Performance Evaluation of Biometric Identification Devices'. Technical report, Sandia National Laboratories, Albuquerque, New Mexico, 1990.

[6] W H Bruce. *Fingerprint Comparison By Template Matching.* PhD thesis, University of Edinburgh, 1992. Unpublished.

[7] B L Miller. *Personal Identification News – 1990 Biometric Industry Directory.* Warfel & Miller, Washington, 1990.

[8] R Plamondon and G Lorette. 'Automatic Signature Verification and Writer Identification — The State of the Art'. *Pattern Recognition*, 22(2):107–131, 1989.

[9] B Hussien, R McLaren, and A Bleha. 'An Application of Fuzzy Algorithms in a Computer Access Security System.'. *Pattern Recognition Letters*, 9(1):39–43, January 1989.

[10] D Renshaw, P B Denyer, G Wang, and M Lu. 'ASIC Image Sensors'. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pages 3038–3041, April 1990.

[11] P B Denyer, D Renshaw, G Wang, M Lu, and S Anderson. 'On Chip CMOS Sensors for VLSI Imaging Systems'. In *Proceedings of the VLSI '91 Conference*, pages 4b.1.1 – 4b.1.10, Edinburgh, August 1991.

[12] K S Fu and A Rosenfeld. 'Pattern Recognition and Image Processing'. *IEEE Transactions on Computers*, 25:1336–1346, December 1976.

[13] V Bruce and P Green. *Visual Perception: Physiology, Psychology and Ecology.* Lawrence Erlbaum Associates, London, 1985.

[14] D Marr. *Vision.* Freeman, 1982.

[15] D Marr and E Hildreth. 'Theory of Edge Detection'. *Proceedings of the Royal Society of London*, B207:187–217, 1980.

[16] R C Gonzalez and P Wintz. *Digital Image Processing*, pages 334–368. Addison-Wesley, second edition, 1987.

[17] H D Ellis. 'Introduction to Aspects of Face Processing: Ten Questions in Need of Answer.'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 3–13. Martinus Nijhoff, 1986.

[18] J Morton and J H Johnson. 'CONSPEC and CONLERN: A Two-Process Theory of Infant Facial Recognition'. *Physiological Review*, 98:164–181, 1991.

[19] R D Walk. 'Perception'. In M Rutter, editor, *Developmental Psychiatry*, pages 177–179. First American Psychiatric Press, Washington, 1987.

[20] M A Jeeves. 'Plenary Session. An Overview. Complementary Approaches to Common Problems in Face Recognition.'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 445–452. Martinus Nijhoff, 1986.

[21] E De Haan and F Newcombe. 'What Makes Faces Familiar'. *New Scientist*, 129(1755):49–52, 9th Feb 1991.

[22] H D Ellis. 'Recognising Faces'. *British Journal of Psychology*, 66(4):409–426, 1975.

[23] H D Ellis and A W Young. 'Are Faces Special?'. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 1–26. North-Holland, 1989.

[24] D C Hay and A W Young. 'The Human Face'. In A W Ellis, editor, *Normality and Pathology in Cognitive Functions*, pages 173–202. Academic Press, London, 1982.

[25] V Bruce and A Young. 'Understanding Face Recognition'. *British Journal of Psychology*, 77:305–327, 1986.

[26] A W Young, K H McWeeny, D C Hay, and A W Ellis. 'Matching Familiar and Unfamiliar Faces on Identity and Expression'. *Psychological Research*, 48:63–68, 1986.

[27] V Bruce, M Burton, and I Craw. 'Modelling Face Recognition'. *Philosophical Transactions of the Royal Society of London*, B335:121–128, 1992.

[28] V Bruce. 'The Structure of Faces'. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 101–106. North-Holland, 1989.

[29] J Sergent. 'Structural Processing of Faces'. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 57–91. North-Holland, 1989.

[30] G Davies, H Ellis, and J Shepherd. 'Face Recognition Accuracy as a Function of Mode of Representation.'. *Journal of Applied Psychology*, 63(2):180–187, 1978.

[31] K R Laughery, J F Alexander, and A B Lane. 'Recognition of Human Faces : Effects of Target Exposure Time, Target Position, Pose Position and Type of Photograph'. *Journal of Applied Psychology*, 55(5):477–483, 1971.

[32] K R Laughery, C Duval, and M S Wogalter. 'Dynamics of Facial Recall'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 373–387. Martinus Nijhoff, 1986.

[33] A J Parker and P Williamson. 'Patterns of Cerebral Dominance in Wholistic and Featural Stages of Facial Processing'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 223–227. Martinus Nijhoff, 1986.

[34] J Shephard, G Davis, and H Ellis. 'Studies in Cue Saliency'. In G Davies, H Ellis, and J Shepherd, editors, *Perceiving and Remembering Faces*. Academic Press, 1981.

[35] H D Ellis, J W Shepherd, and G M Davies. 'Identification of Familiar and Unfamiliar Faces From Internal and External Features: Some Implications for Theories of Face Recognition.'. *Perception*, 8:431–439, 1979.

[36] R L Atkinson, R C Atkinson, and E R Hilgard. *Introduction to Psychology*, page 150. Harcourt Brace Jovanovich, London, 1982.

[37] N D Haig. 'How Faces Differ — A New Comparative Technique.'. *Perception*, 14:601–615, 1985.

[38] N D Haig. 'Investigating Face Recognition with an Image Processing Computer'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 410–425. Martinus Nijhoff, 1986.

[39] V Bruce. *Recognising Faces*, chapter 3, pages 37–58. Lawrence Erlbaum Associates, 1988.

[40] M Endo. 'Perception of Upside-Down Faces : An Analysis From the Viewpoint of Cue Saliency'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 53–58. Martinus Nijhoff, 1986.

[41] I H Fraser and D M Parker. 'Reaction Time Measures of Feature Saliency in a Perceptual Integration Task'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 45–52. Martinus Nijhoff, 1986.

[42] L D Harmon. 'The Recognition of Faces'. *Scientific American*, 229:70–82, 1973.

[43] G Rhodes. 'Looking at Faces: First-Order and Second-Order Features as Determinants of Facial Appearance'. *Perception*, 17:43–63, 1988.

[44] A J Goldstein, L D Harmon, and A B Lesk. 'Identification of Human Faces'. *Proceedings of the IEEE*, 59(5):748–760, May 1971.

[45] G Della Riccia and A Iserles. 'Automatic Identification of Pictures of Human Faces'. In *Proceedings of the Carnahan Conference on Security Technology: Electronic Crime Countermeasures*, pages 145–148, Lexington, Kentucky, 1977.

[46] K R Laughery, B T Rhodes Jr, and G W Batten Jr. 'Computer-Guided Recognition and Retrieval of Facial Images'. In G Davies, H Ellis, and J Shepherd, editors, *Perceiving and Remembering Faces.* Academic Press, 1981.

[47] V Bruce and M Burton. 'Computer Recognition of Faces'. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 487–506. North-Holland, 1989.

[48] R J Baron. 'Strengths and Weaknesses of Computer Recognition Systems'. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 507–511. North-Holland, 1989.

[49] R J Baron. 'Mechanisms of Human Facial Recognition'. *International Journal of Man Machine Studies*, 15:137–178, 1981.

[50] R L Russel, R L Routh, J R Holten III, and M Kabrisky. 'A Face Recognition System Based on Cortical Thought Theory'. In *Proceedings of the IEEE 38th National Aerospace and Electronics Conference*, volume 4, pages 1377–1385, 1986.

[51] T Sakai, M Nagao, and T Kanade. 'Computer Analysis and Classification of Photographs of Human Faces'. In *Proceedings of the 1st USA-JAPAN Computer Conference*, pages 55–62, Montvale USA, 1972.

[52] M Nixon. 'Eye Spacing Measurement for Facial Recognition'. *Proceedings of the SPIE: The International Society for Optical Engineering*, 575:279–285, 1985.

[53] K H Wong, P W M Tsang, and H W Law. 'A Human Face Recognition System'. In *Proceeding Electronic Imaging West Conference*, volume 2, pages 603–605, Anaheim, California, March 1988.

[54] D Tock, I Craw, and R Lishman. 'A Knowledge Based System for Measuring Faces'. In *Proceedings of the British Machine Vision Conference*, Oxford, September 1990.

[55] Y Kaya and K Kobayashi. 'A Basic Study of Human Face Recognition'. In *Proceedings of the International Conference Frontiers of Pattern Recognition*, pages 265–289, Honolulu, January 1971.

[56] T Kanade. *Computer Recognition of Human Faces.* Birkhauser, Basel, 1977.

[57] K H Wong, H H M Law, and P W M Tsang. 'A System for Recognising Faces'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1638–1642, Glasgow, April 1989.

[58] T Kohonen, E Reuhkala, K Makisara, and L Vainio. 'Associative Recall of images.'. *Biological Cybernetics*, 22:159–168, 1976.

[59] A J O'Toole, R B Millward, and J A Anderson. 'A Physical System Approach to Recognition Memory for Spatially Transformed Faces'. *Neural Networks*, 1(3):179–199, 1988.

[60] A J O'Toole and H Abdi. 'Connectionist Approaches to Visually-Based Facial Feature Extraction'. In G Tiberghien, editor, *Advances in Cognitive Science Vol 2*, pages 123–140. Ellis Horwood Limited, 1989.

[61] L Sirovich and M Kirby. 'Low-Dimensional Procedure for the Characterisation of Human Faces'. *Journal of the Optical Society of America A*, 4(3):519–524, 1987.

[62] R J Clarke. *Transform Coding of Images*, pages 72–134. Academic Press, London, 1985.

[63] A K Jain. 'Image Data Compression: A Review'. *Proceedings of the IEEE*, 69(3):349–388, March 1981.

[64] M Kirby and L Sirovich. 'Application of the Karhunen-Loeve Procedure for the Characterisation of Human Faces'. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, January 1990.

[65] I Craw and P Cameron. 'Parameterising Images for Recognition and Reconstruction'. In P Mowforth, editor, *Proceedings of the British Machine Vision Conference*, pages 367–370, Glasgow, September 1991. Springer-Verlag.

[66] I Craw. 'Recognising Face Features and Faces'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[67] M Turk and A Pentland. 'Face Processing: Models For Recognition'. *Proceedings of the SPIE: The International Society for Optical Engineering*, 1192:22–32, 1989.

[68] M Turk and A Pentland. 'Eigenfaces for Recognition'. *Journal of Cognition Neuroscience*, 3(1):71–86, 1991.

[69] I Craw, A Aitchison, and P Cameron. 'Principal Component Analysis of Face Images'. Unpublished paper from the Department of Mathematical Sciences, University of Aberdeen, 1992.

[70] M A Shackleton and W J Welsh. 'Classification of Facial Features for Recognition'. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 1991.

[71] A F Murray. 'Analog VLSI and Multi Layer Perceptions - Accuracy, Noise and On-chip Learning'. In *Proceedings of the International Conference on Neural Networks*, Munich, Germany, 1991.

[72] G W Cottrell and M Fleming. 'Face Recognition Using Unsupervised Feature Extraction.'. In *Proceedings of the International Neural Networks Conference*, volume 1, pages 322–325, Paris, July 1990. Kluwer Academic Press.

[73] G W Cottrell and P Munro. 'Principal Components Analysis of Images via Back Propagation'. *Proceedings of the SPIE: The International Society for Optical Engineering*, 1001(2):1070–1077, 1988.

[74] G W Cottrell and J Metcalfe. 'EMPATH: Face, Emotion, and Gender Recognition Using Holons'. In R P Lippmann, J E Moody, and D S Touretzky, editors, *Advances In Neural Information Processing Systems 3*, pages 564–571. Morgan Kaufmann, 1990.

[75] 'PiCard - Automatic Face Recognition'. SD-Scicon UK Ltd, Pembroke Ho., Camberley, Surrey, 1990.

[76] R M Rickman and T J Stonham. 'Coding Facial Images for Database Retrieval Using a Self Organising Neural Network'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[77] S Wang, A C Schreiber, and S Rousset. 'Connectionist Modelling of a Cognition Model of Face Identification : Simulation of Context Effects'. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2, pages 549–556, Washington, 1989.

[78] P J B Hancock. 'GANNET: Design of a Neural Net for Face Recognition by Genetic Algorithm'. In *Proceedings of the IEE Workshop on Genetic Algorithms, Neural Networks and Simulated Annealing*, Glasgow, May 1990.

[79] W A Phillips and L S Smith. 'Conventional and Connectionist Approaches to Face Processing by Computer'. In A W Young and H D Ellis, editors, *Handbook of Research on Face Processing*, pages 513–518. North-Holland, 1989.

[80] D E Pearson and J A Robinson. 'Visual Communication at Very Low Data Rates'. *Proceedings of the IEEE*, 73(4):795–812, April 1985.

[81] M W Whybray and E Hanna. 'A DSP Based Videophone for Hearing-Impaired Using Valledge Processed Pictures'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1866–1869, Glasgow, May 1989.

[82] L D Harmon and W F Hunt. 'Automatic Recognition of Human Face Profiles'. *Computer Graphics and Image Processing*, 6:135–156, 1978.

[83] C J Wu and J S Huang. 'Human Face Profile Recognition by Computer'. *Pattern Recognition*, 23(3/4):255–259, 1990.

[84] A Coombes, R Richards, A Linney, E Hanna, and V Bruce. 'Shape Based Description of the Facial Surface'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[85] L D Harmon, M K Khan, R Lasch, and P F Ramig. 'Machine Identification of Human Faces'. *Pattern Recognition*, 13(2):97–110, 1981.

[86] S R Arridge, J P Moss, A D Linney, and D R James. 'Three Dimensional Digitisation of the Face and Skull'. *Journal of Maxillo-facial Surgery*, 13:136–143, 1985.

[87] A T Deacon, A G Anthony, S N Bhatia, and J P Muller. 'Evaluation of a CCD Based Facial Measurement System.'. *Medical Informatics*, 16(2):213–228, 1991.

[88] Y Suenaga and Y Wantanabe. 'A Method for the Synchronised Acquisition of Cylindrical Range and Color Data.'. *Transactions of the IEICE*, E74(10):3407–3416, October 1991.

[89] J T Lapreste, J Y Cartoux, and M Richetin. 'Face Recognition From Range Data by Structural Analysis.'. In *Proceedings NATO Research Workshop on Syntactic and Structural Pattern Recognition*, pages 303–314, Sitges, October 1986. Springer Verlag.

[90] T Abe, H Aso, and M Kimura. 'Automatic Identification of Human Faces by 3-D splines of surfaces - Using Verticles of B-Spline Surface'. *Systems and Computers in Japan*, 22(7):96–105, 1991.

[91] O Nakamura, S Mathur, and T Minami. 'Identification of Human Faces Based on Isodensity Maps'. *Pattern Recognition*, 24(3):263–272, 1991.

[92] A C Aitchison and I Craw. 'Synthetic Images of Faces — An Approach to Model-Based Face Recognition.'. In P Mowforth, editor, *Proceedings of the British Machine Vision Conference*, pages 226–232, Glasgow, September 1991. Springer-Verlag.

[93] A Shashua. 'Illumination and View Position in 3D Visual Recognition.'. In J E Moody, S J Hanson, and R P Lippmann, editors, *Advances In Neural Information Processing Systems 4*, pages 404–411. Morgan Kaufmann, 1992.

[94] K Aizawa, Y Yamada, H Harashima, and T Saito. 'Model-Based Synthesis Image Coding System — Modeling A Person's Face and Synthesis of Facial Expressions'. In *Proceedings of the IEEE Global Telecommunications Conference*, volume 1, pages 45–48, 1987.

[95] K Aizawa, H Harashima, and T Saito. 'Model-Based Analysis synthesis Image Coding (MBASIC) System for a Person's Face'. *Signal Processing: Image Communication*, 1:139–152, 1989.

[96] S R Cannon, G W Jones, R Campbell, and N W Morgan. 'A Computer Vision System for Identification of Individuals'. In *Proceedings of the International Conference Industrial Electronics Control and Instrumentation*, volume 1, pages 347–351, Milwaukee, 1986.

[97] M D Kelly. *Visual Identification of People by Computer.* PhD thesis, Stanford University, 1970.

[98] X Jia and M S Nixon. 'On Developing an Extended Feature Set for Automatic Face Recognition'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[99] R Gallery and T I P Trew. 'An Architecture for Face Classification'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[100] R L Sherman. 'Obtaining Information Characterising a Person or Animal'. *UK Patent Application - GB 2231699 A* , 1990.

[101] B A Golomb, D T Lawrence, and T J Sejnowski. 'SEXNET: A Neural Network Identifies Sex From Human Faces.'. In R P Lippmann, J E Moody, and D S Touretzky, editors, *Advances In Neural Information Processing Systems 3*, pages 572–577. Morgan Kaufmann, 1990.

[102] I Pilowsky, M Thornton, and B B Stokes. 'Towards the Quantification of Facial Expressions with the Use of a Mathematical Model of the Face.'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 340–348. Martinus Nijhoff, 1986.

[103] F I Parke. 'Parameterised Models for Facial Animation'. *IEEE Computer Graphics and Applications Magazine*, pages 61–68, November 1982.

[104] K Mase. 'Recognition of Facial Expression From Optical Flow.'. *Transactions Institute of Electronic, Information and Communications Engineers*, E74(10):3474–3483, October 1991.

[105] A P Pentland. 'Perceptual Organisation and the Representation of Natural Form'. *Artificial Intelligence*, 28:293–331, 1986.

[106] P J Benson and D I Perrett. 'Synthesising Continuous-Tone Caricatures.'. *Image and Vision Computing*, 9(2):123–129, April 1991.

[107] W J Welsh. *Model-Based Coding of Images.* PhD thesis, Essex University, 1991.

[108] R D Boyle and R C Thomas. *Computer Vision A First Course*, pages 32–80. Blackwell Scientific, 1988.

[109] J Canny. 'A Computational Approach to Edge Detection'. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[110] M D Kelly. 'Edge Detection in Pictures by Computer Using Planning'. In B Meltzer and D Michie, editors, *Machine Intelligence Vol 6*, pages 397–409. Edinburgh University Press, Edinburgh, 1970.

[111] V Govindaraju, D B Sher, R K Srihari, and S N Srihari. 'Locating Human Faces in Newspaper Photographs'. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 549–554, San Diego, June 1989.

[112] M Kass, A Witkin, and D Terzopoulos. 'Snakes: Active Contour Models'. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[113] J B Waite and W J Welsh. 'An Application of Active Contour Models to Head Boundary Location.'. In *Proceedings of the British Machine Vision Conference*, pages 407–412, Oxford, September 1990.

[114] J B Waite and W J Welsh. 'Head Boundary Location Using Snakes'. *British Telecom Technology Journal*, 8(3):127–136, July 1990.

[115] G Sexton. 'Automatic Face Detection for Videoconferencing'. In *Digest of the IEE Colloquium on Low Bit Rate Image Coding*, London, May 1990.

[116] H Harasaki, M Yano, and T Nishitani. 'Background Separation/Filtering for Videophone Applications'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1981–1984, Albuquerque, New Mexico, April 1990.

[117] R J Schalkoff. *Digital Image Processing and Computer Vision*, pages 339–344. J Wiley & Sons, 1989.

[118] P J Burt. 'Smart Sensing Within a Pyramid Vision Machine'. *Proceedings of the IEEE*, 76(8):1006–1015, January 1988.

[119] M A Fischler and R A Elschalger. 'The Representation and Matching of Pictorial Structures'. *IEEE Transactions on Computers*, 22:67–92, 1973.

[120] K Tsui and P Nickolls. 'Automatic Feature Extraction of Human Facial Images'. In *Proc ICSC '88*, pages 505–512, 1988.

[121] M J Conlin. 'A Rule-Based High-Level Vision System'. *Proceedings of the SPIE: The International Society for Optical Engineering*, 726:314–320, 1986.

[122] I Craw, H Ellis, and J R Lishman. 'Automatic Extraction of Face-Features'. *Pattern Recognition Letters*, 5:183–187, 1987.

[123] G J S Robertson and K C Sharman. 'Object Location Using Proportions of the Direction of Intensity Gradient (PRODIGY)'. Unpublished paper from the Department of Electrical and Electronic Engineering, Glasgow University, 1992.

[124] J Illingworth and J Kittler. 'A Survey of the Hough Transform'. *Computer Vision, Graphics and Image Processing*, 44:87–116, 1988.

[125] D H Ballard. 'Generalising the Hough Transform to Detect Arbitrary Shapes'. *Pattern Recognition*, 13(2):111–122, 1981.

[126] M Nagao. 'Picture Recognition and Data Structure'. In F Nake and A Rosenfeld, editors, *Graphic Languages*, pages 48–68. North-Holland, 1972.

[127] R A Hutchinson and W J Welsh. 'Comparison of Neural Network an Conventional Techniques for Feature Location in Facial Images'. In *Proceedings of the IEE International Conference on Artificial Neural Networks*, London, October 1989.

[128] W J Welsh and R A Hutchinson. 'Image Coding Using an Analysis-Synthesis Technique'. In *Digest of the IEE Colloquium on Low Bit rate Image Coding*, London, May 1990.

[129] A L Yuille, D S Cohen, and P W Hallinan. 'Feature Extraction From Faces Using Deformable Templates'. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 104–109, San Diego, June 1989.

[130] A Bennett and I Craw. 'Finding Image Features Using Deformable Templates and Detailed Prior Statistical Knowledge'. In P Mowforth, editor, *Proceedings of the British Machine Vision Conference*, pages 233–239, Glasgow, September 1991. Springer-Verlag.

[131] K Fukushima. 'A Neural Network for Visual Pattern Recognition'. *Computer*, pages 65–75, March 1988.

[132] B Widrow, R G Winter, and R A Baxter. 'Layered Neural Nets for Pattern Recognition'. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7):1109–1118, July 1988.

[133] D G Elliman and R N Banks. 'Shift Invariant Neural Net for Machine Vision'. *Proceedings of the IEE*, 137(3):183–187, June 1990.

[134] M W Roth. 'Survey of Neural Network Technology for Automatic Target Recognition'. *IEEE Transactions on Neural Networks*, 1(1):28–43, March 1990.

[135] D Rummelhart, G Hinton, and R Williams. 'Learning Internal Representations on Error Propagation'. In *Parallel Distributed Processing.* The MIT Press, Cambridge Massachusetts, 1986.

[136] R P Lippmann. 'An Introduction to Computing with Neural Nets'. *IEEE ASSP Magazine*, pages 4–22, April 1987.

[137] C Nightingale and R A Hutchinson. 'Artificial Neural Nets and Their Application to Image Processing'. *British Telecom Technology Journal*, 8(3):81–93, July 1990.

[138] R A Hutchinson. 'Development of an MLP Feature Location Technique Using Preprocessed Images.'. In *Proceedings of the International Neural Networks Conference*, volume 1, pages 67–70, Paris, July 1990. Kluwer Academic Press.

[139] J M Bishop. 'A Hybrid Network for Feature Extraction'. In *Proceedings of the International Neural Networks Conference*, volume 1, page 50, Paris, July 1990. Abstract only.

[140] J M Vincent, J B Waite, and D J Myers. 'Precise Location of Facial Features by a Hierarchical Assembly of Neural Nets.'. In *Proceedings of the IEE 2nd International Conference on Artificial Neural Networks*, pages 69–73, Bournemouth, November 1991.

[141] A F Murray. 'Analogue Noise-Enhanced Learning in Neural Network Circuits'. *Electronic Letters*, 27(1):1546–1548, August 1991.

[142] C Nightingale. 'Image Processing in Visual Communications'. In D E Pearson, editor, *Image Processing*, pages 283–305. McGraw-Hill, London, 1991.

[143] J P Secilla, N Garcia, and J L Carrascosa. 'Template Location in Noisy Pictures'. *Signal Processing*, 14:347–361, 1988.

[144] D I Barnea and H F Silverman. 'A Class of Algorithms for Fast Digital Image Registration'. *IEEE Transactions on Computers*, 21(2):179–186, February 1972.

[145] R N Nagel and A Rosenfeld. 'Ordered Search Techniques in Template Matching'. *Proceedings of the IEEE*, 60:242–244, February 1972.

[146] A Margalit and A Rosenfeld. 'Reducing the Expected Computational Cost of Template Matching Using Run Length Representation.'. *Pattern Recognition Letters*, 11(4):255–265, April 1990.

[147] A Rosenfeld and G J Vanderburg. 'Coarse-Fine Template Matching'. *IEEE Transactions on Systems, Man and Cybernetics*, 2:104–107, February 1977.

[148] G J Vanderburg and A Rosenfeld. 'Two-Stage Template Matching'. *IEEE Transactions on Computers*, 26(4):384–393, April 1977.

[149] E R Davies. 'Tradeoffs Between Speed and Accuracy in Two-Stage Template Matching'. *Signal Processing*, 15:351–363, 1988.

[150] B F J Manly. *Multivariate Statistical Methods*, pages 1–16. Chapman & Hall, 1986.

[151] K Sutherland, D Renshaw, and P B Denyer. 'A Novel Automatic Face Recognition Algorithm Employing Vector Quantization'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[152] H Ellis, J Shepherd, and G Davies. 'An Investigation of the Use of the Photo-Fit Technique for Recalling Faces'. *British Journal of Psychology*, 66(1):29–37, 1975.

[153] Home Office / University of Aberdeen. *The Aberdeen Index to Photo-Fit.* Home Office, 1982.

[154] DCI D MacNeil. *Personal Communication*, 1991. Lothian and Borders Police, Fettes Avenue, Edinburgh.

[155] 'E-FIT Facial Recognition'. Aspley Ltd, 12-14 Smug Oak Business Centre, Lye Lane, St. Albanes, Herts., 1991.

[156] R M Gray. 'Vector Quantization'. *IEEE ASSP Magazine*, pages 4–29, April 1984.

[157] N M Nasrabadi and R A King. 'Image Coding Using Vector Quantization: A Review'. *IEEE Transactions on Communications*, 36(8):957–971, January 1988.

[158] M Goldberg and H-F Sun. 'Image Sequence Coding Using Vector Quantization'. *IEEE Transactions on Communications*, 34(7):703–710, July 1986.

[159] V J Mathews and M Khorchidian. 'Multiplication-Free Vector Quantization Using L_1 Distortion Measure and its Variants'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1747–1750, Glasgow, April 1989.

[160] B Everitt. *Cluster Analysis.* Heinemann Educational Books, London, 1980.

[161] D J Hand. 'Cluster Analysis'. In *Discrimination and Classification*, pages 155–185. J Wiley & Sons, 1981.

[162] W Equitz. 'Fast Algorithms for Vector Quantization Picture Coding'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 725–728, Dallas, Texas, April 1987.

[163] W H Equitz. 'A New Vector Quantization Clustering Algorithm'. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(10):1568–1575, October 1989.

[164] J A Hartigan. *Clustering Algorithms*, pages 84–112. J Wiley & Sons, 1975.

[165] Y Linde, A Buzo, and R M Gray. 'An Algorithm for Vector Quantization Design'. *IEEE Transactions on Communications*, 28(1):84–95, January 1980.

[166] S P Lloyd. 'Least-Squares Quantization in PCM'. *IEEE Transactions on Information Theory*, 25:129–137, 1982.

[167] E W Forgy. 'Cluster Analysis of Multivariate Data: Efficiency vs Interpretability of Classifications'. *Biometrics*, 21:768, 1965. Abstract only.

[168] R O Duda and P E Hart. *Pattern Classification and Scene Analysis*, page 227. J Wiley & Sons, 1973.

[169] J Makhoul, S Roucos, and H Gish. 'Vector Quantization in Speech Coding'. *Proceedings of the IEEE*, 73(11):1551–1588, November 1985.

[170] T Kohonen. *Self-Organisation and Associative Memory*, chapter 5, pages 119–157. Spinger-Verlag, Berlin, 1989.

[171] A K Krishnamurthy, S C Ahalt, D E Melton, and P Chen. 'Neural Networks for Vector Quantization of Speech and Images.'. *IEEE Journal on Selected Areas in Communications*, 8(8):1449–1457, October 1990.

[172] J D McAuliffe, L E Atlas, and C Rivera. 'A Comparison of the LBG Algorithm and Kohonen Neural Network Paradigm for Image Vector Quantization'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2293–2296, Albuquerque, New Mexico, April 1990.

[173] C S Ramsay, K Sutherland, D Renshaw, and P B Denyer. 'A Comparison of Vector Quantization Codebook Generation Algorithms Applied to Automatic Face Recognition'. Accepted for presentation at *British Machine Vision Conference*, Leeds, 1992.

[174] DARPA. *Neural Network Study*, chapter 8, pages 87–95. AFCEA International Press, 1988.

[175] A N Mucciardi and E E Gose. 'A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition Properties'. *IEEE Transactions on Computers*, 20(9):1023–1031, September 1971.

[176] M Michael and Wen-Chun Lin. 'Experimental Study of Information Measure and Inter-Intra Class Distance Ratios on Feature Selection and Orderings.'. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(2):172–181, March 1973.

[177] W S Mohn. 'Two Statistical Feature Evaluation Techniques Applied to Speaker Identification'. *IEEE Transactions on Computers*, 20(9):979–987, September 1971.

[178] A M Sutherland. *Automatic Speaker Verification Based on Waveform Perturbation Analysis.* PhD thesis, University of Edinburgh, 1989.

[179] W J Dixon. *BMDP Statistical Software Manual*, 1988.

[180] R G D Steel and J H Torrie. *Principles and Procedures of Statistics*, chapter 10, pages 183–193. McGraw-Hill, 1960.

[181] M James. *Classification Algorithms*, pages 94–126. Collins, 1985.

[182] K Sutherland, D Renshaw, and P B Denyer. 'Automatic Face Recognition'. In *Proceeding of the IEE Intelligent Systems Engineering Conference*, Edinburgh, 1992.

[183] K Sutherland, D Renshaw, and P B Denyer. 'Probabilistic Pattern Analysis for Facial Recognition'. Accepted for presentation at *International Conference in Automation, Robotics and Computer Vision*, Singapore, 1992.

[184] I Aleksander and T J Stonham. 'Guide to Pattern Recognition Using Random Access Memories'. *IEE Journal of Computers and Digital Techniques*, 2(1):29–40, 1979.

[185] I Aleksander, W V Thomas, and P A Bowden. 'WISARD: A Radical Step Forward in Image Recognition'. *Sensor Review*, pages 120–124, July 1984.

[186] I Aleksander. 'Emergent intelligent Properties of Progressively Structured Pattern Recognition Nets'. *Pattern Recognition Letters*, 1:375–384, 1983.

[187] T J Stonham. 'Practical Face Recognition and Verification with Wisard'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 426–441. Martinus Nijhoff, 1986.

[188] R B Starkey and I Aleksander. 'Facial Recognition for Police Purposes Using Computer Graphics and Neural Networks'. In *Digest of the IEE Colloquium on Electronic Images and Image Processing in Security and Forensic Science*, London, May 1990.

[189] R B Starkey and I Aleksander. 'Facial Recognition for Police Purposes Using Digital Neural Networks'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[190] R E Walpole and R H Myers. *Probability and Statistics for Engineers and Scientists*, chapter 7, pages 238–280. Macmillan, second edition, 1978.

[191] W Mendenhall and R J Beaver. *Introduction to Probability and Statistics*, page 300. PWS-Kent, eighth edition, 1991.

[192] N M Allinson, A W Ellis, B M Fluke, and A J Luckman. 'A Connectionist Model of Familiar Face Recognition'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[193] K W J Findlay. *Algorithms for Low Cost VLSI Stereo Vision Systems, with Specific Application to Intruder Alarms.* PhD thesis, University of Edinburgh, 1992. Unpublished.

[194] S Anderson, W H Bruce, P B Denyer, D Renshaw, and G Wang. 'A Single Chip Sensor and Image Processor for Fingerprint Verification'. In *Proceedings of the IEEE Custom Integrated Circuits Conference*, pages 12.1.1 – 12.1.4, San Diego, CA, May 1991.

[195] J F Fleming. 'Identity Verification'. *UK Patent Application - GB 2229305 A* , 1990.

[196] J Carter and M Nixon. 'An Integrated Biometric Database'. In *Digest of the IEE Colloquium on Electronic Images and Image Processing in Security and Forensic Science*, London, May 1990.

Appendix A

Pixel Clustering Algorithm.

When using a template matching algorithm it is likely that a number of positive responses will be recorded around the actual spatial location of the object being sought, Figure A–1. In order to find a unique position for the chosen object it is necessary that all the positive responses, in a particular area, be examined and the best one selected. For this research, a simple pixel clustering algorithm has been devised to perform this task.

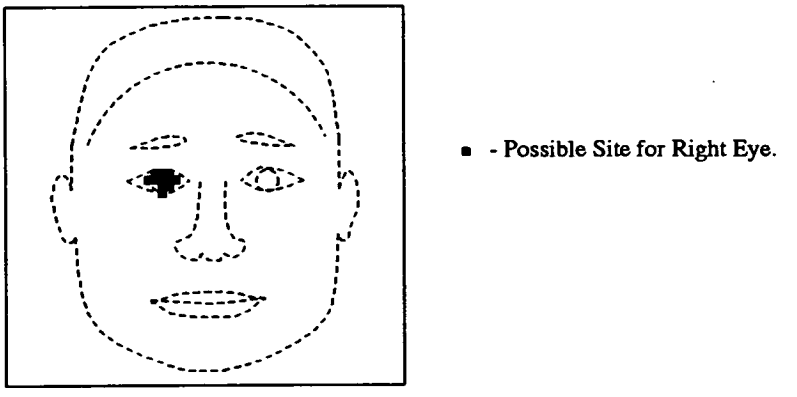

Figure A–1: Example output from the template matching stage.

A.1 Algorithmic Details

It is assumed that the template matching algorithm records positive matches for the chosen object as points in a 'dummy' image, as shown in Figure A–1. It is also assumed that the corresponding match score, as measured by the chosen difference metric, is stored for each of these sites.

The task of the clustering algorithm is then to locate these points, and then identify where several candidate sites can be clustered down to a single object placement. The following algorithm stages are performed to cluster the possible match sites.

Step 1. Scan the entire dummy image, or search area (if a constrained search is being used), for a possible site.

Step 2. Examine the match scores for a small area around this site to establish the best location in that area.

Step 3. Remove all but the location of the best site from the dummy image and store the location of this site.

Step 4. Return to **Step 1** until the entire image has been completed.

In **Step 2** of the algorithm, a search area of 10 pixels in each direction was used. This was included as a safeguard against the extreme case where the template matching algorithm had produced positive responses for a very large number of the available locations.

Appendix B

Experimental Conditions.

The experimental conditions described in this appendix were adhered to throughout the experimental work included in this thesis. Images were captured using the ASIS 1010 video camera and framegrabber in the setup shown in Figure B–1. The images were captured at a 256 $\times$ 256 pixel resolution, and digitised to 8 bits (or 256 grey levels). The images were captured on several different days and at different times of day.

B.1 Population

The population used was drawn from the staff and students at the University of Edinburgh, Department of Electrical Engineering. The members of the population were all male and in the range of 20-45 years of age. Several members of the population had beards, or spectacles, or both.

B.2 Image Posing

The subjects were seated approximately 1.5m from the camera, this distance was not allowed to be varied. The subjects were requested to adjust the height of their chair until their entire face was contained in the visual display. This process was performed without intervention by the operator. The subjects were requested to assume a fairly neutral expression when the images were taken. A

very small number of *rogue* images, containing extreme facial expressions, were removed from the trial data. A uniform white background was used for the image capture.

B.3 Lighting

As illustrated in Figure B–1, a flash unit was used to illuminate the scene. This flash was automatically triggered by the framegrabber. The bounced flash reduced the harsh lighting conditions likely to occur with direct lighting. The photographic condition known as *red-eye* was also removed by using a bounced flash.

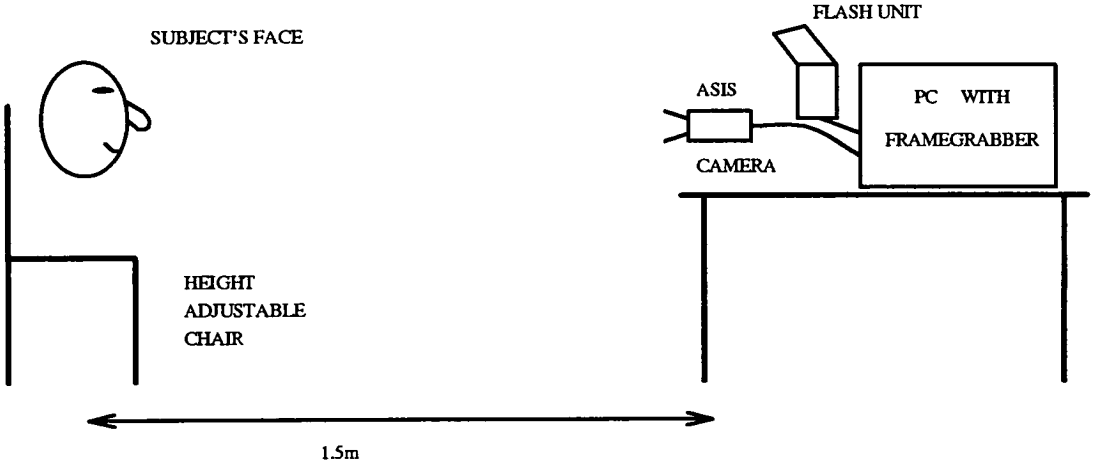

Figure B–1: Experimental apparatus.

The flash was covered with a visual light absorbing filter which allows near-infrared light to pass through it. The ASIS camera responds well to both near-infrared light and visible light. The filter was used so that the illumination of the scene could be performed using a very bright flash, without disconcerting the subject. In this case, the flash used was so bright that it completely swamped any other light sources in the room.

Appendix C

MLP Virtual Targets.

The following specific information was used in the training and testing of the *virtual targets* neural network approach to eye location.

C.1 Network Topology

The network consisted of a three layer multi-layer perceptron (as illustrated in Figure C–1) with the following number of neurons in each layer:

- 256 inputs (one each for the 16 × 16 pixels in the input pattern).
- 8 hidden units
- 1 output (trained to produce a '1' when stimulated with an eye).

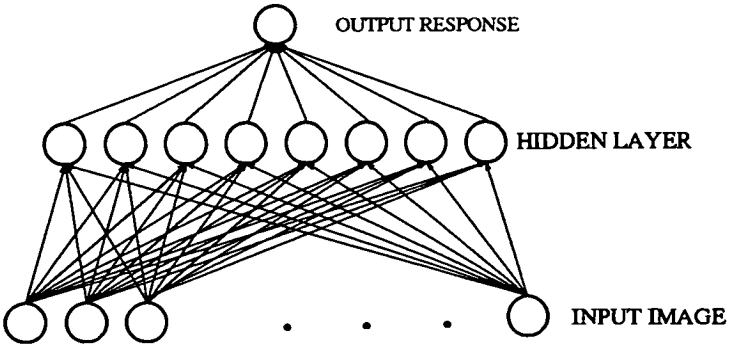

Figure C–1: A three layer multi-layer perceptron.

C.2 Training Parameters

A number of different training parameters have to be specified in order to allow the network to learn in the correct manner. The following values were selected in a largely heuristic way. A detailed discussion regarding the magnitude of the noise level and the manner in which it is injected into the learning equation, is given in the references cited in the text.

Target Learning Speed = 0.1 - This factor controls the amount by which the virtual targets are adapted after each cycle.

Weights Learning Speed = 0.1 - This factor controls the amount by which the synaptic weights are adapted after each cycle.

Acceptance Criterion = 0.1 - Maximum error permitted at the output neuron (*ie* values at the output neuron of less than 0.1 are accepted as zeros and values greater than 0.9 are accepted as ones).

Initial Noise Level = 0.1 - The initial level of random noise injected onto the neuron outputs. This values decays as the maximum error falls.

The learning process was allowed to terminate when all the test patterns gave responses within the acceptance criterion. In order to overcome any transient success, caused by noise, the acceptance criterion of 10% had to be sustained for five epochs before the learning process was deemed to have been successful.

C.3 Training Data

The network was presented with examples of both positive and negative training data. This process allows the hidden layer of the network to construct a generalised internal representation of the target pixel pattern. In this case, the positive information consisted of pixel patterns containing right eyes from ten different population members. The negative information consisted of 100 pixel patterns drawn from the same images so that they did not containing eye information. These patterns were selected at random from the image at least 10 pixels away from **each** of the eyes, so as to avoid confusion during training.

In total, the network was required to learn the target output for 2000 test patterns. To avoid letting the network fall into a rapid local minimum the learning process was performed incrementally. In total, a training period of approximately 300 epochs (or cycles), was required for the network to achieve a solution – this value is the average training time from several training sessions initiated from different random start points.

Appendix D

Kohonen's Self Organising Feature Map.

The following equations and parameters were used to perform vector quantization codebook generation using a KSOFM.

The network consisted of twenty neurons each connected to its neighbours. In this case, a two-dimensional matrix of five by four neurons was chosen. As each vector, $x(t)$, is presented to the network, the nearest neuron to it is identified. This neuron's weights, w_{ij}, are then updated in the following manner, equation D.1.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t,\mathcal{D})\alpha(t)(x_i(t) - w_{ij}(t)) \tag{D.1}$$

where $\eta(t,\mathcal{D})$ is the neighbourhood gain function, and $\alpha(t)$ is the adaption gain function:

$$\alpha(t) = \mathcal{A}_1 e^{\frac{-t}{\mathcal{T}_1}}$$

$$\eta(t,\mathcal{D}) = e^{\frac{-\mathcal{D}}{\mathcal{N}}}, \quad \mathcal{N} = \mathcal{A}_2 + \mathcal{A}_3 e^{\frac{-t}{\mathcal{T}_2}}$$

The following values for these various constants were used in this experimentation, Table D–1. The network was trained for twenty epochs (*ie* twenty presentations of each input pattern).

Constants	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{T}_1$	$\mathcal{T}_2$
Values	0.5	5.0	0.001	20.0	5.0

Table D–1: Constants used in KSOFM.

Appendix E

Publications

A number of publications have appeared as a result of the research work reported in this thesis.

1. K Sutherland, D Renshaw and P D Denyer 'A Novel Facial Recognition Algorithm Employing Vector Quantization' Presented at *IEE Colloquium on Recognition and Storage of Faces*, London, January 1992.

2. K Sutherland, D Renshaw and P D Denyer 'Automatic Face Recognition' Presented at *IEE Intelligent Systems Engineering Conference*, Heriot-Watt University, Edinburgh, August 1992.

3. K Sutherland, D Renshaw and P D Denyer 'Probabilistic Pattern Analysis for Facial Recognition' Accepted for Presentation at *International Conference in Automation, Robotics and Computer Vision*, Singapore, September 1992.

4. C S Ramsay, K Sutherland, D Renshaw and P D Denyer 'A Comparison of Vector Quantization Codebook Generation Techniques Applied to Automatic Facial Recognition' Accepted for presentation at *British Machine Vision Conference*, Leeds, September 1992.

further software study is performed on an even larger database. It is likely that the present performance of the algorithm can be enhanced further by the inclusion of feature measurement information and a system of feature weightings.

The recognition performance of ~89% and the verification error cross-over point of ~7% cannot yet challenge the performance of some other biometric systems, however, the authors believe that the construction of a viable biometric facial recognition device is a realisable goal.

References

[1] I Aleksander, W V Thomas, and P A Bowden. WISARD: A radical step forward in image recognition. *Sensor Review*, pages 120–124, July 1984.

[2] R J Baron. Mechanisms of human facial recognition. *International Journal of Man Machine Studies*, 15:137–178, 1981.

[3] I Craw and P Cameron. Parameterising images for recognition and reconstruction. In *Proceedings British Machine Vision Conference BVMC '91*, pages 367–370, Glasgow, September 1991.

[4] I Craw, H Ellis, and J R Lishman. Automatic extracion of face-features. *Pattern Recognition Letters*, 5:183–187, 1987.

[5] H D Ellis, J W Shepherd, and G M Davies. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8:431–439, 1979.

[6] M A Fischler and R A Elschalger. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22:67–92, 1973.

[7] I H Fraser and D M Parker. Reaction time measures of feature saliency in a perceptual integration task. In H D Ellis, editor, *Aspects of Face Processing*, pages 45–52. Martinus Nijhoff, 1986.

[8] R M Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, April 1984.

[9] N D Haig. How faces differ — a new comparative technique. *Perception*, 14:601–615, 1985.

[10] L D Harmon. The recognition of faces. *Scientific American*, 229:70–82, 1973.

[11] R A Hutchinson and W J Welsh. Comparison of neural network an conventional techniques for feature location in facial images. In *Proceedings IEE International Conference on Artifical Neural Networks*, London, October 1989.

[12] J T Lapreste, J Y Cartoux, and M Richetin. Face recognition from range data by structural analysis. In *Syntactic and Structural Pattern Recognition*, pages 303–314, Sitges, October 1986. Springer Verlag.

[13] B B Megdal. *VLSI Computational Structures Applied to Fingerprint Image Analysis*. PhD thesis, Californian Institute of Technology, 1983.

[14] N M Nasrabadi and R A King. Image coding using vector quantization: A review. *IEEE Transactions on Communications*, 36(8):957–971, January 1988.

[15] M Nixon. Eye spacing measurement for facial recognition. *Proceedings of the Int Soc for Optical Engineering SPIE*, 575:279–285, 1985.

[16] J R Parks. Biometrics: the people sensors. *Sensor Review*, 9(2):79–84, April 1989.

[17] W A Phillips and L S Smith. Conventional and connectionist approaches to face processing by computer. In A W Young and H D Ellis, editors, *Handbook of Research on Face Processing*, pages 513–518. North-Holland, 1989.

[18] Picard - automatic face recognition. SD-Scicon UK Ltd, Pembroke Ho., Camberley, Surrey.

[19] R Plamondon and G Lorette. Automatic signature verification and writer identification — the state of the art. *Pattern Recognition*, 22(2):107–131, 1989.

[20] G Rhodes. Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, 17:43–63, 1988.

[21] A Rosenfeld and G J Vanderburg. Coarse-fine template matching. *IEEE Transactions Systems, Man and Cybernetics*, SMC-2:104–107, February 1977.

[22] R L Russel, R L Routh, J R Holten III, and M Kabrisky. A face recognition system based on cortical thought theory. In *IEEE 38th National Aerospace and Electronics Conference (NAECON 1986)*., volume 4, pages 1377–1385, 1986.

[23] T Sakai, M Nagao, and T Kanade. Computer analysis and classification of photographs of human faces. In *1st USA-JAPAN Computer Conference*, pages 55–62, Montvale USA, 1972.

[24] W Shangrui, A-C Schreiber, and S Rousset. Connectionist modelling of a cognitive model of face identification: Simulation of context effects. *International Joint Conference on Neural Networks IJCNN '90*, 2:549–556, 1989.

[25] T J Stonham. Practical face recognition and verification with wisard. In H D Ellis, editor, *Aspects of Face Processing*, pages 426–441. Martinus Nijhoff, 1986.

[26] A M Sutherland and M A Jack. Speaker verification. In M A Jack and J Laver, editors, *Aspects of Speech Technology (Edits 4)*. University of Edinburgh Press, Edinburgh, 1987.

[27] K Sutherland, C S Ramsay, D Renshaw, and P B Denyer. A comparison of vector quantization codebook generation algorithms applied to automatic face recognition. Submitted to *British Machine Vision Conference*, Leeds, 1992.

[28] K Sutherland, D Renshaw, and P B Denyer. A novel automatic face recognition algorithm employing vector quantization. In *Digest of IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[29] D Tock, I Craw, and R Lishman. A knowledge based system for measuring faces. In *British Machine Vision Conference BMVC '90*, Oxford, September 1990.

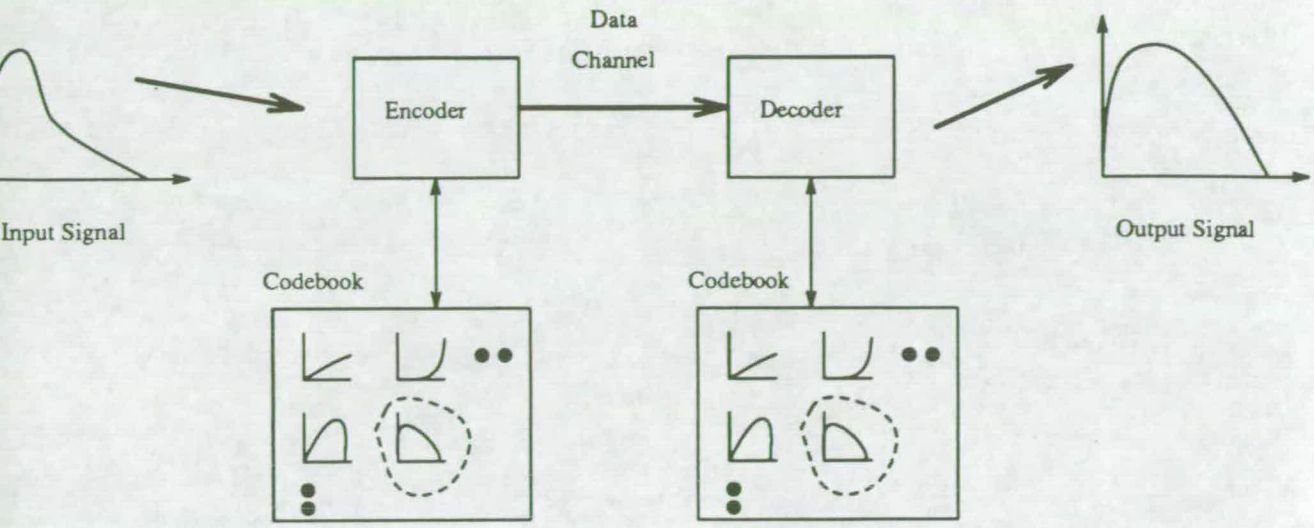

Figure 1: Vector Quantization.

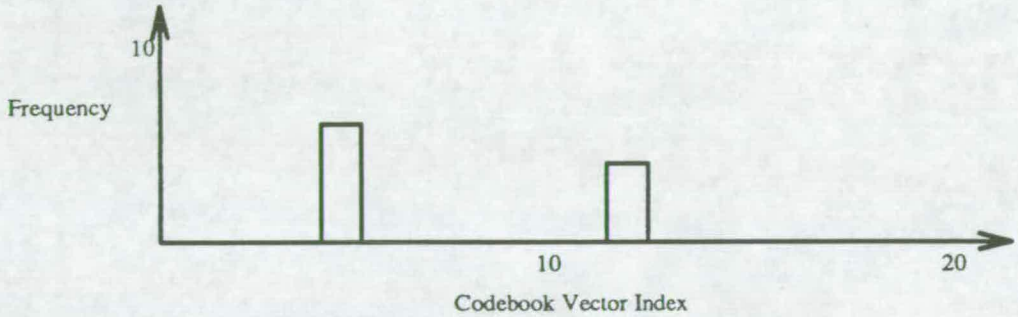

Figure 2: Vector histogram for example training set.

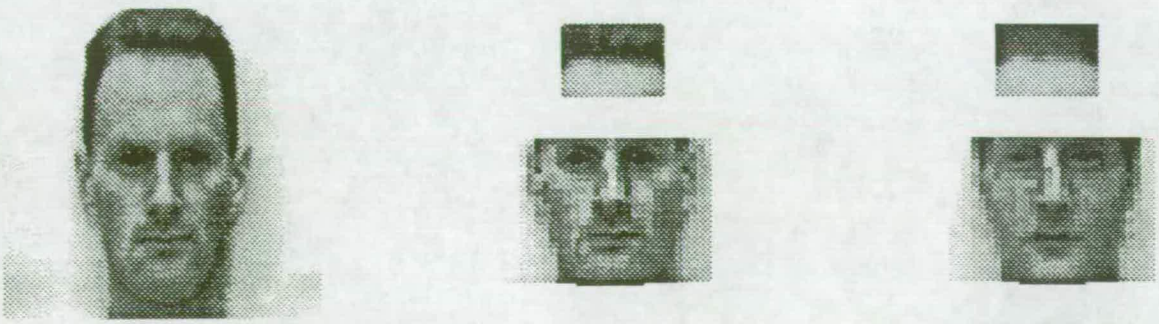

Figure 3: The three stages of facial coding: original input, spatially sub-sampled, vector quantized.

PROBABILISTIC PATTERN ANALYSIS FOR FACIAL RECOGNITION.

K Sutherland, D Renshaw and P B Denyer

Integrated Systems Group,
Department of Electrical Engineering,
University of Edinburgh,
Edinburgh, EH9 3JL.
UK

ABSTRACT

The comparison of facial images is an attractive method of automatic personal identification. This paper introduces a probabilistic approach to facial comparison which is able to cope with some changes in facial appearance.

Crucial to this approach is the construction of a generalised internal representation of each face. Two different methods of comparing new faces with the stored population database are described and evaluated.

1 INTRODUCTION

A fail-safe method of automatic personal identification is now required to control access to the ever increasing amount of personal, and sensitive, information which is being stored electronically. The use of *Biometrics*, *ie* the physical characteristics which uniquely identify us all, has been recognised as a very good method of determining, or verifying, personal identity. There is now substantial commercial interest in the development of automatic biometric systems to facilitate fail-safe access control [1, 2].

There are a variety of different personal characteristics which can be exploited in order to identify different human individuals. Algorithms have been developed to perform recognition using fingerprints [3], dynamic signature matching [4] and voice patterns [5]. However, biometrics have not yet established themselves as the most popular systems of access control; one possible reason for this, is that the public is uneasy with the intrusive nature of some of these different approaches.

Automatic facial recognition can, theoretically, be performed in a completely unintrusive manner; requiring only a passive role from the subject under test. However, the problem with using facial information, is that the face is fundamentally changeable. Thus, the challenge for a facial recognition algorithm, is to parameterise the face in such a way as to nullify the effects of changes in expression and general facial appearance. A facial recognition system must be able to capture the intrinsics of the face in such a way as to be invariant to the likely variations in facial appearance for a given individual.

A large number of different techniques have been suggested as possible methods of parameterising the face; these include facial measurements [6], facial contours [7], principal component analysis [8] and various neural networks approaches [9, 10]. None of these techniques have explicitly addressed the problem of facial variability with expression, or any other factors, and hence these techniques require that the images used are captured in a strictly controlled environment. Only WISARD [11] has been demonstrated as a possible system of facial recognition with the ability to differentiate between different subjects in a reliable manner. The problem with the WISARD approach is that the amount of data required to store the facial pattern is substantial.

In [12] the authors presented a novel technique of facial parameterisation which preserved facial recognisability, but was able to reduce the facial information into less than two hundred bytes of data. This paper describes the probabilistic pattern recognition techniques which are required to facilitate recognition, in the situation where only minimal constraints need to be placed on the posing of the face in the images used.

The following section outlines the basic parameterisation algorithm, then the pattern recognition process is described in detail. Comprehensive results are presented for two different techniques of probabilistic comparison.

2 SYSTEM OVERVIEW

The complete algorithm for facial recognition is illustrated in Figure 1. The different stages of the algorithm process the facial information into a form in which it can be readily compared with other facial images. The facial segmentation stage was performed using a variant of Fischler and Elschalger feature embedding algorithm [13]. In the cases where this technique failed, the images were manually segmented. This correction was performed to eliminate possible recognition failures caused by errors in the location stage; hence the results presented in this paper, relate only to the performance of the parameterisation and the comparison stages of the algorithm.

The intrinsic function of the facial recognition process is that of parameterisation, *ie* the extraction of the fundamental characteristics of the face to be recognised. In [12], the authors presented an algorithm which distilled the face into a compact facial signature while still maintaining much of the *recognisability* of the initial input face.

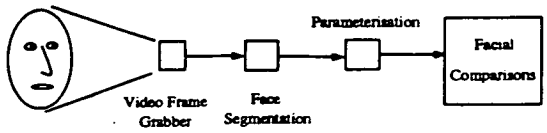

Figure 1: Facial Recognition.

In brief, the system identified seven fundamental features: each eye, the nostrils, the bridge of the nose, the mouth, the chin and the hair. In addition to storing these individual facial parts, the facial signature also incorporated a coarsely subsampled view of the entire face. The partitioning of the face used in this algorithm is given in [14]. The justification for the use of these facial parts, as opposed to the rest of the facial data, has its basis in psychological research into the human ability of recognising faces.

Having extracted these eight facial parts for storage and. discarded the rest of the image data, there still remains the task of data reduction. The signal processing function of Vector Quantization (VQ) has been chosen to provide the requisite reduction. The use of VQ for image coding is widespread [15]. However, recognition analysis using VQ is much more novel.

The process of vector quantization involves the substitution of a waveform segment with an approximately similar waveform drawn from a codebook of possible segments (or vectors). The data reduction results from the transmission of only the index of the vector and not all the data points which constitute that vector. The signal is reconstructed by re-substituting the index number with the vector or waveform segment. The generation of the vector codebooks required for each facial feature is a specialised process which has been discussed in [14].

In this algorithm, the VQ technique is used to construct a codebook for each of the facial features used (*eg* nose, mouth *etc*). Thus, a bank of eight vector quantizers is used to code the initial representation of the face into the personal signature of that individual. The VQ of the complete face yields a discrete representation for use in the subsequent recognition stage. The derivation of this *personal signature* is illustrated in Figure 2.

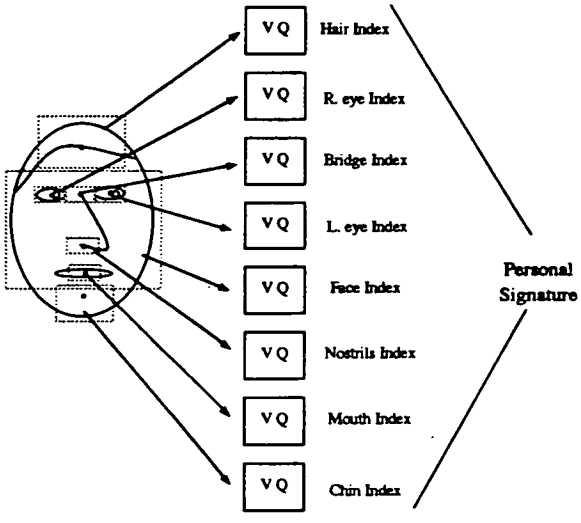

Figure 2: Personal Signature.

The personal signature obtained in this way is an eight dimensional vector, containing VQ indices pertaining to the eight facial features. However, this personal signature only relates to one facial image, the manner in which several of these facial signatures can be combined to form a full personal signature is described below.

TRAINING

In order to achieve the goal of this research - which is to perform facial recognition in a manner largely invariant to facial expression - it is necessary that the personal signatures of each member of the population are able to capture the likely variations in the person's facial appearance. To facilitate this accrual of information, a system of *feature histograms* has been devised.

The histograms for each feature contain bins relating to the different feature examples present within the VQ codebooks. During the training phase of the device, a number of different images of the same person are presented to the system (these images should contain some variations in expression and facial appearance). For each image, the eight dimensional vector is derived using the algorithm mentioned above. This information is then placed into the histograms for each feature. As this process continues, for a number of images, some accumulations are likely to occur in the feature histograms. These accumulations are related to the most likely feature types chosen for each population member. However, the **spread** of the histograms relates to the variation in facial appearance present within the training set of images. In this way, a generalised view of each individual is constructed in a unique set of feature histograms. An example set of histograms is illustrated in Figure 3.

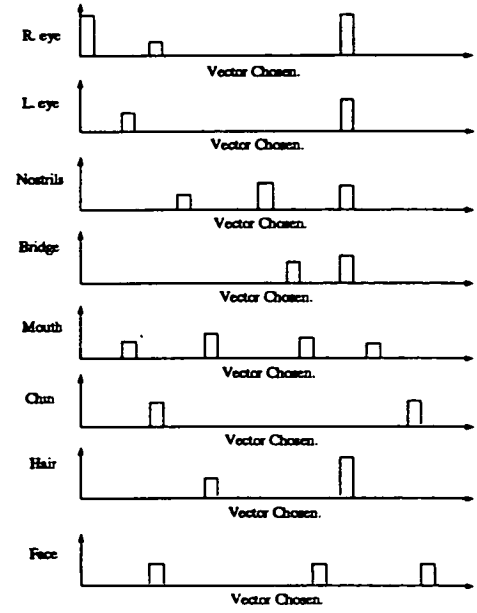

Figure 3: Feature Histograms.

These feature histograms can be considered as probability distributions; relating the observation of a particular feature type, to the occurrence of a given individual. A population database can be constructed by collecting together the feature

histograms for each person. The **recognition** of a particular person's face can be achieved by comparing a new image with the population database of signatures. This process is described in the following section.

3 PERSONAL COMPARISONS

There are a number of different approaches which can be utilised to compare these facial signatures. Two such approaches are described in detail below.

WINNER TAKES ALL

To compare a new face with a stored signature, the features in the face are subjected to the VQ process to produce the feature indices (in the same way as for the training set). If these indices are then overlaid with a target signature, the overall probability of the identity of the new face being that of this particular signature can be evaluated.

Putting this strategy in a probabilistic framework, we can associate a probability with each person causing a particular vector to occur, this can be denoted thus:

$$P(H_i|A_jV_k)$$

where -

H_i = the ith individual member of the population.
A_j = the jth feature (*ie* nose, mouth ...).
V_k = the kth vector for that feature in the codebook.

Using Bayes theorem

$$P(H_i|A_jV_k) = \frac{P(H_i)P(A_jV_k|H_i)}{\sum_{i=0}^{N} P(H_i)P(A_jV_k|H_i)} \quad (1)$$

where

N = total number of individuals in population.

If we then assume that $P(H_i)$, the *a priori* probability, is equal to $\frac{1}{N}$, *ie* each person is equally probable, then the overall probability, Equation 1, can be evaluated using values for $P(A_jV_k|H_i)$ obtained from the probability histograms produced during training. We now have a measure of likelihood that the observation of a particular feature vector suggests any particular individual. It is now necessary that we produce an overall probability that **all** the observed feature vectors came from any of the members of the population.

The most obvious manner in which to obtain an overall probability of a match, is to sum the eight individual probability of the eight features. We can incorporate in this a weighting function to denote the important given to each individual facial feature. Assuming a weighting vector $W = \{w_1, w_2, \ldots, w_8\}$, then the overall probability that the observed vectors were caused by subject i is given in Equation 2. For this study a uniform weighting function has been adopted.

Unfortunately, substantial variations in the face can cause problems for this method of probability accrual. For example, when encoding a new face, a different vector can be chosen to any chosen during the training phase of that particular person. This results in a zero probability for that feature. This effect can be caused by large changes in the feature, *eg* closure of the mouth or eye, *etc.* Thus, it may be that a feature is in reality very similar, although, its pixel level representation has been substantially changed. In this case, we do not wish to fail to recognise someone because of this change, therefore, it may be necessary to devise a way in which to relate a very unlikely occurrence (*eg* a closed eye) to the overwhelmingly positive information drawn from the other features.

$$P(H_i) = \frac{1}{M}\sum_{j=1}^{8} w_j P(H_i|A_jV_k) \quad (2)$$

where

$$\sum_{j=1}^{8} w_j = M$$

TOTAL DIFFERENCE

When comparing the features drawn from a new image, with the VQ codebook, a difference measure is computed between the new feature and each codebook entry. If, instead of selecting the best one and comparing it with the probability histograms (as described above in the *winner takes all* approach) we use all the difference information, it is possible to produce a *total difference* measure of similarity between a new face and a stored signature. The following example illustrates this principle.

Considering only one feature in isolation, with a possible set of five vectors; and, given a training set of ten images, we can derive a personal signature on the basis of the probability histograms. When presented with a new example, of that person, the system can produce a set of differences for the chosen feature relating to all possible vectors. These two pieces of information are presented in Table 1.

Vectors	V_1	V_2	V_3	V_4	V_5
Probabilities	5	2	0	3	0
Pixel Differences	56	68	52	72	87

Table 1: Example test data.

Here the most likely vector choice is V_3 and the training frequency value of 0 would be used in the complete probabilistic face recognition. However, in difference terms, the new feature is also very close to vector V_1 and this similarity is ignored by only selecting the best output. If we use the reciprocal of the difference (scaled in some manner) as a measure of similarity, then a total match score, *MS*, between the new feature and that feature's appearance during training can be derived. This calculation is shown for feature A_1 below.

$$\begin{aligned} \text{MS} = {} & P(A_1V_1)\times\frac{k}{D_1} + P(A_1V_2)\times\frac{k}{D_2} + P(A_1V_3)\times\frac{k}{D_3} \\ & + P(A_1V_4)\times\frac{k}{D_4} + P(A_1V_5)\times\frac{k}{D_5} \end{aligned}$$

Where k is a scaling factor and D_n is the difference between the new feature and the n^{th} member of the codebook.

Again, a summation of the probabilities relating to the eight different features is required to obtain an overall personal recognition score.

This second approach is able to take account of the similarities between the different vectors in the VQ codebook in order to facilitate a fuller comparison between the subject under test, and all the other members of the population. Results for recognition trials using these two different approaches are given below.

4 RESULTS

A test population of forty males was used, with ten images of each person used for training (*ie* the generation of the facial signature) and ten images kept for testing. The trial thus consisted of 400 test presentations. For each presentation, the output ranking of likelihood was recorded. From this data, and from knowing the correct response, it is possible to obtain a first place recognition rate (*ie* the proportion of tests in which the correct signature was selected as the most similar to the test stimulus). By analysing the output ordering of responses an average rank figure can also be obtained.

The subjects were requested not to assume extreme facial expressions when their photographs were being taken, however, accurate posing of the face was not performed. Some of the subjects in the trial wore spectacles. The test results are given in Table 2.

	Winner Takes All	Total Difference
Average Rank	1.43	1.37
1^{st} Place	86.00%	88.50%
2^{nd} Place	93.00%	92.75%
3^{rd} Place	96.00%	95.50%

Table 2: Recognition Rates for Both Techniques.

Using a simple statistical test [16] the difference between the two different approaches was not found to be significant, *ie* the variation in recognition rates of between 86% and 88% can be attributed to random factors.

5 CONCLUSIONS

A successful method of facial recognition has been described in this paper. The first place recognition rate, of between 86% and 88%, is very encouraging, given the relatively unconstrained nature in which the images were captured. There is no significant difference between the two different comparison algorithms introduced in this paper.

The facial recognition performance reported in this paper cannot yet rival the performance of the other biometric techniques available, however, the unobtrusive nature in which facial information can be captured remains a considerable attraction to face recognition systems. This paper has demonstrated the way in which probabilistic information can be used to aid the comparison process. It is hoped, that with further developments, this approach to facial recognition will be able to compete with the other biometric techniques available.

ACKNOWLEDGEMENTS

Mr K. Sutherland is supported by a Science and Engineering Research Council studentship.

REFERENCES

[1] J A Barry III. 'Back to the Future With Biometrics'. *Security Management*, 34(4):83-85, 1990.

[2] J R Parks. 'Biometrics: the people sensors'. *Sensor Review*, 9(2):79-84, April 1989.

[3] B B Megdal. *VLSI Computational Structures Applied to Fingerprint Image Analysis*. PhD thesis, Californian Institute of Technology, 1983.

[4] R Plamondon and G Lorette. 'Automatic Signature Verification and Writer Identification — The State of the Art'. *Pattern Recognition*, 22(2):107-131, 1989.

[5] A M Sutherland. *Automatic Speaker Verification Based on Waveform Perturbation Analysis*. PhD thesis, University of Edinburgh, 1989.

[6] K H Wong, H H M Law, and P W M Tsang. 'A system for recognising faces'. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1638-1642, Glasgow, April 1989.

[7] O Nakamura, S Mathur, and T Minami. 'Identification of Human Faces Based on Isodensity Maps'. *Pattern Recognition*, 24(3):263-272, 1991.

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

[9] G W Cottrell and J Metcalfe. 'EMPATH: Face, Emotion, and Gender Recognition Using Holons'. In R P Lippmann, J E Moody, and D S Touretzky, editors, *Advances In Neural Information Processing Systems 3*, pages 564-571. Morgan Kaufmann, 1990.

[10] R M Rickman and T J Stonham. 'Coding Facial Images for Database Retrieval using a self organising neural network'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[11] T J Stonham. 'Practical Face Recognition and Verification With Wisard'. In H D Ellis, M A Jeeves, F Newcombe, and A Young, editors, *Aspects of Face Processing*, pages 426-441. Martinus Nijhoff, 1986.

[12] K Sutherland, D Renshaw, and P B Denyer. 'A Novel Automatic Face Recognition Algorithm Employing Vector Quantization'. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[13] M A Fischler and R A Elschalger. 'The Representation and Matching of Pictorial Structures'. *IEEE Transactions on Computers*, 22:67-92, 1973.

[14] C S Ramsay, K Sutherland, D Renshaw, and P B Denyer. 'A Comparison of Vector Quantization Codebook Generation Algorithms Applied to Automatic Face Recognition'. Accepted for presentation at *British Machine Vision Conference*, Leeds. 1992.

[15] N M Nasrabadi and R A King. 'Image Coding using Vector Quantization: A Review'. *IEEE Transactions on Communications*, 36(8):957-971, January 1988.

[16] W Mendenhall and R J Beaver. *Introduction to Probability and Statictics*, page 300. PWS-Kent, eighth edition, 1991.

A Comparison of Vector Quantization Codebook Generation Algorithms Applied to Automatic Face Recognition.

C. S. Ramsay† K. Sutherland† D. Renshaw and P.B. Denyer
Integrated Systems Group, Electrical Engineering Department,
University of Edinburgh, Edinburgh, EH9 3JL.

Abstract

Automatic facial recognition is an attractive solution to the problem of computerised personal identification. In order to facilitate a cost effective solution, high levels of data reduction are required when storing the facial information. Vector Quantization has previously been used as a data reduction technique for the encoding of facial images.

This paper identifies the fundamental importance of the vector quantizer codebooks in the performance of the system. Two different algorithms – the Linde-Buzo-Gray algorithm and Kohonen's Self Organising Feature Map – have been used to obtain two sets of facial feature codebooks. For comparison, the system performance has also been analysed using a codebook dedicated to the test population. It has been shown that by using a *good* codebook generation algorithm it is possible to substantially reduce the dimensionality of the vector codebooks, with remarkably little degradation in system performance.

1 Introduction

Automatic facial recognition is now receiving an increasing amount of research interest [1], largely due to its possible applicability to the automatic personal identification task. As such, automatic face recognition is attempting to stake its claim among the other biometric systems available (notably fingerprint, hand geometry, dynamic signature matching and voice pattern recognition).

In order to establish biometric systems as likely tools for personal identification we must instill public confidence in the concept of biometric storage and retrieval. In this area automatic face recognition scores heavily over the other likely biometrics, as the least intrusive and the most *natural* personal identification process. It is also important that any prototype system has a demonstrably high accuracy. To obtain such high accuracy a likely advancement would be to incorporate the use of a number of different biometrics into one identification system, making the final response conditional on all the available information.

However, in practice, the overriding factor in achieving a *marketable* system will be the data reduction obtained, since this controls the speed at which multiple comparisons can be performed and, ultimately, the cost, as data storage

†These authors have SERC studentships.

requirement is a prime factor in this area. The available level of data reduction is now of particular importance given the increasing use of smart-card technology and the ever increasing likelihood that we will all eventually carry personal biometric information with us in this way. If a facial image is to be one of many pieces of biometric information stored on our smart-card, then accurate, high data reduction, parametrisation of the face is required.

The authors have previously introduced a Vector Quantization (VQ) based facial recognition technique and presented results for a recognition experiment using 30 individuals [2]. In this paper we would like, firstly, to present further results on a 33% larger data set and, secondly, to investigate the use of VQ codebook generation techniques to reduce further the data space required to store the intrinsics of the face (or the *facial signature*). The codebook generation techniques used here to reduce the overall number of vectors are Kohonen's neural Self-Organising Feature Map and the Linde-Buzo-Gray algorithm. However, firstly, a brief overview of the entire facial recognition algorithm will be presented.

2 System Overview

The intrinsic function of a pattern recognition process is that of parameterisation, *i.e.* the extraction of the fundamental characteristics of the object to be recognised. In [2] the authors presented an algorithm which distilled the face into a compact facial signature, while still maintaining much of the *recognisability* of the initial input face.

In brief, the system identifies seven fundamental features; each eye, the nostrils, the bridge of the nose, the mouth, the chin and the hair. In addition to storing these individual facial parts, the facial signature also incorporates a coarsely sub-sampled view of the entire face. The partitioning of the face used here is shown in Figure 1. These facial features are automatically located using a template matching algorithm based on Fischler and Elschalger's *feature embedding* approach[3]. Manual correction of the location failures was performed.

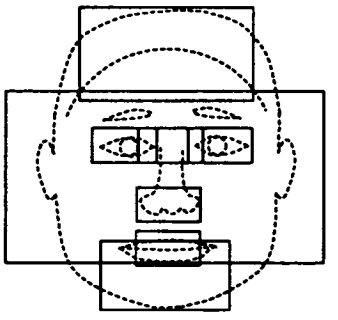

Figure 1: Facial Features Exploited to Differentiate between Individuals.

Having extracted these eight facial parts for storage and discarded the rest of the image data, there still remains the task of data reduction. The signal

processing function of VQ has been chosen to provide the requisite reduction. The use of VQ for image coding is widespread [4]. However, recognition analysis using VQ is much more novel.

Applied to faces, VQ performs a function analogous to the police *photofit* system. In turn, each of the eight selected features is compared with a codebook of standard examples (or *vectors*) of only that feature; using the normalised euclidean distance as a metric, the most similar of these standard examples is chosen as the most likely match. Then to store that feature, we need only to keep the *index* of the standard feature and not all the data points which constitute that vector.

It is desirable to train a facial recognition device on a number of examples of each person it is expected to recognise, in order to allow it to construct internal representations of each person. To facilitate this, a system of *feature histograms* has been devised; these reflect the ways in which the subject's face has varied during training. To obtain the histograms we use the VQ technique described above to encode each feature, from each training image, then that *match* is recorded in the histogram. Thus, given 20 possible vectors and ten training images one particular *histogram* could look like this, Figure 2.

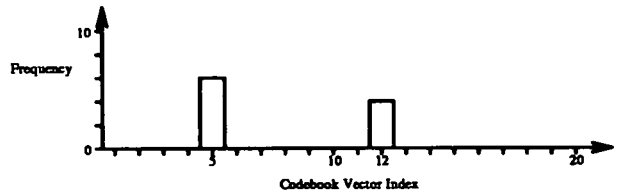

Figure 2: One Feature Histogram for One Person's Training Images.
This feature has been mapped six times to codeword 5 and four times to codeword 12 during training.

In order to differentiate one person from another their respective histograms would have to show significant differences. If we look at the histograms of all eight features for two different people, Figure 3, it can be seen that there is only a low level of variability within training for each person, yet there is a substantial difference between these characteristics between these two people. Low *within person* and high *between person* variabilities are essential for a good recognition system.

By storing the histograms obtained in the manner described above it is argued that we have the requisite facial data to perform recognition. Thus, the feature histograms can be thought of as our facial signatures. The problem with this approach is that, to date, the VQ codebook entries have been extracted from a control image of each member of the test population. In this way the data storage requirement, as determined by the number of vectors, is dependent on the population size. Thus, if we were to enlarge the population there would be a consequent rise in data storage requirements. To combat this, the following section outlines the ways in which the codebook dimensionality can be reduced.

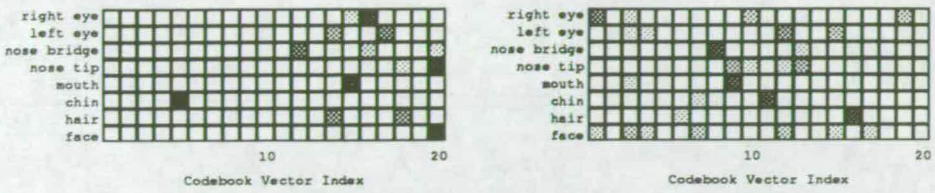

Figure 3: Feature Histograms for Two Population Members.
In this graphic all eight feature histograms are shown, with grey tone signifying the frequency of use of each codebook vector. *Black* represents a well used vector, with lighter tones indicating less use.

3 Codebook Generation

The process of VQ relies on having a good set of vectors in its codebook. For image compression, these vectors are chosen to minimise the overall pixel level error introduced. However, for face recognition, other considerations are also important. For example, distinctiveness may be more significant than pixel error when encoding facial features. Fortunately, there are many different algorithms available which perform codebook generation [5].

As an initial starting point, each feature codebook has been constructed using one sample vector drawn from a control image of each member of the test population, thus reflecting only the variation present within this test population. To obtain a reduction in data storage, the two most common VQ codebook generation techniques have been used to reduce the codebook dimensionality. In this study the test population contained 40 members and thus the initial codebooks (for each of the eight features) contained 40 vectors. A reduction to half this number has been chosen as a sufficiently demanding test of the codebook generation algorithms available. The task of the codebook generation technique is to perform the requisite dimensionality reduction while still containing the system's error rate within acceptable limits.

3.1 Kohonen's Self-Organising Feature Maps

Neural network clustering techniques have been employed in a variety of application areas, such as pattern recognition, optimisation and, notably, VQ codebook design for image compression [6, 7]. Kohonen developed his Self-Organising Feature Maps (KSOFM) in order to model the neural feature maps which are thought to form in the human brain [8]. KSOFMs result in a network where neighbouring output units have similar responses : *topological ordering* has occurred. This ordering can be used to reduce search requirements in VQ applications, though this is not an aspect of the KSOFM which has been exploited here. As a clustering algorithm, KSOFM allows a reduction in the dimensionality of the input vector space to a smaller number of reference vectors.

KSOFM defines a matrix (usually two dimensional) of output units, or neurons, each of which has a weight vector, w_j, associated with it. Input

vectors from a training set are presented sequentially and the weight vectors are adjusted as described below. The weight vectors converge towards cluster centres after sufficient training time. The reason that topological ordering occurs is that each input vector adjusts not only one weight vector, but a *neighbourhood* of weight vectors.

The KSOFM algorithm is defined thus :

Step 1. Initialise all weight vectors to random values.

Step 2. Apply new input vector, $x(t)$.

Step 3. Calculate the Euclidean distance from $x(t)$ to all output nodes j

$$d_j = \sum_{i=0}^{N-1} (x_i(t) - w_{ij}(t))^2$$

where N is the dimensionality of the input vector.

Step 4. Select the output node with the smallest distance d_i and label it as the winning unit, j^*.

Step 5. Update weight vectors of all nodes in the matrix according to the equation

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t, \mathcal{D})\alpha(t)(x_i(t) - w_{ij}(t))$$

where $\eta(t, \mathcal{D})$ is the neighbourhood gain function, and $\alpha(t)$ is the adaption gain function.
$\eta(t, \mathcal{D})$ is a function which defines a neighbourhood on the matrix round the winning neuron, decreasing exponentially with distance $\mathcal{D}$ from the winning neuron, and which shrinks over time. $\alpha(t)$ decreases exponentially with time, and $0 \leq \alpha(0) \leq 1$.

Step 6. Repeat **Step 2** to **Step 5** until the entire training set has been presented e times. e is the number of *epochs*, which is set before training starts.

In this application the input vectors, x, come from the original feature codebooks containing 40 vectors, and the matrix was a 5×4 array of neurons producing a codebook of 20 vectors.

3.2 Linde-Buzo-Gray Algorithm

The Linde-Buzo-Gray (LBG) algorithm [9] is the most commonly used codebook design algorithm due to the fact that it was the earliest proposed method and consistently outperforms other methods in a variety of applications [10, 11]. It is an iterative technique which repeatedly moves codewords to cluster centroids in an effort to find a codebook which will display the lowest error when encoding the training data (*i.e.* a modified version of the K-Means clustering algorithm). The basic algorithm is as follows :

Step 1. Initialise the codebook.

Step 2. For each vector, x, in the training set, calculate the Euclidean distance from it to each vector v_j in the codebook.

$$d_j = \sum_{i=0}^{N-1} (x_i - v_{ij})^2 \tag{3}$$

where N is the dimensionality of the vectors.
The minimum distance selects the closest vector, v_j^*, in the codebook. Assign x to the cluster around v_j^*.

Step 3. Replace each codeword with the centroid of the vectors in the training set that have been assigned to it. If any codewords are unused, they are discarded and replaced with new codewords which are more likely to be used in the next iteration. In this work, this has been done by taking the most commonly used codewords and splitting them to create two close copies of the original.

Step 4. If the total error in clustering the training data is still decreasing by a significant amount, return to **Step 2.** Otherwise, stop.

This algorithm should optimise the codebook so that the sum of the distances from each vector of the training set to its nearest codeword is a minimum. It is possible, however, that in certain conditions the algorithm will reach a local minimum rather than a global minimum. This can be seen to be true due to the fact that the final codebook will change if the initial codebook is different.

There are a number of recognised methods for generating an initial codebook, the most common of which is to populate the codebook with vectors chosen randomly from the training set. Another method is to use a codebook generated from other clustering techniques, such as the KSOFM technique described above, as the initial codebook. The method utilised here was to populate the codebook with 20 replicas of the centroid of the entire training set. In this way the codebook is gradually filled with useful codewords, as unused codewords are replaced in **Step 3**.

There are some similarities between the ways in which the two techniques outlined above perform dimensionality reduction. However, the codebooks produced can be significantly different. To illustrate this Figures 4 and 5 show the two 20 vector *face*[†] codebooks generated from the same original input of 40 vectors. It can be clearly seen that both codebooks contain composite faces formed by the merging of several of the input faces together, but that they are quite different from each other.

[†]There are parallel codebooks for each of the other seven features used in the recognition algorithm.

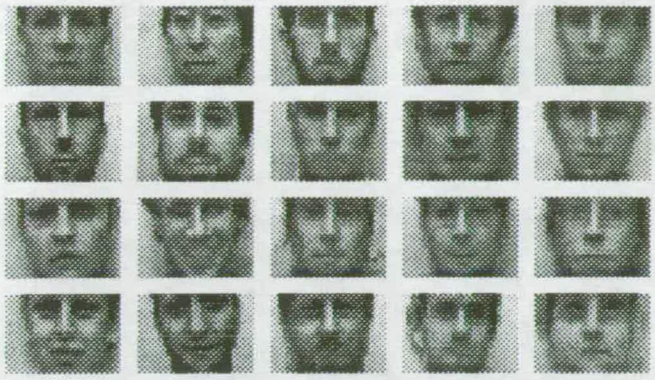

Figure 4: Codebook Generated using LBG.

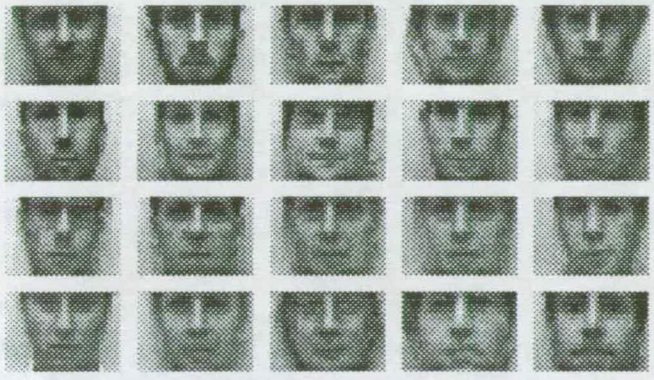

Figure 5: Codebook Generated using KSOFM.

4 Experimental Results

At present the algorithm for facial recognition is implemented as a suite of software programs. The system cannot yet function in real-time and thus the test images used are stored on disk for repetitive analysis. The images used were captured with a video camera under largely controlled conditions. However, the subjects were allowed to vary their expressions during the several days during which images were being captured. In this way, the results obtained for the system will be closer to a real-world implementation than some other, highly controlled, studies.

As mentioned earlier, a test population of 40 males was used, with ten images of each person used for training (*i.e.* the generation of the facial signature) and ten images kept for testing. The trial thus consisted of 400 test presentations. For each presentation, the output ranking of likelihood was recorded. From this data, and from knowing the correct response, it is possible to obtain a first place recognition rate (*i.e.* the proportion of tests in which the correct signature was selected as the most similar to the test stimulus). By analysing

the output ordering of responses an average rank figure can also be obtained.

4.1 Dedicated codebook

To provide a benchmark, and to demonstrate the best possible performance, a set of dedicated codebooks were used. These codebooks were constructed using the same vectors as were used to train the other codebook generation algorithms. The feature histograms were thus 40 vectors wide for each of the eight features. For this experiment the average rank and the cumulative success rates of the first three output positions is given in Table 1.

Average Rank	1.37
1^{st} Place Recognition	88.5%
2^{nd} Place Recognition	92.75%
3^{rd} Place Recognition	95.5%

Table 1: Recognition rates for a 40 member population with 400 test presentations.

4.2 Dimensionality Reduction

Essentially, here, we are performing the same experiment as above, but using codebooks of half the size. Such a significant reduction would be expected to cause a substantial reduction in recognition performance unless, as described in section 3, the algorithms used to derive the new codebook entries reflect the initial characteristics of all 40 vectors. The experiment is thus a parallel comparison between the two different sets of codebook vectors when applied to the recognition function. Table 2 gives the relative performances between the two approaches.

	KSOFM	LBG
Average Rank	1.99	1.46
1^{st} Place Recognition	70.2%	83.3%
2^{nd} Place Recognition	81.5%	90.0%
3^{rd} Place Recognition	88.5%	94.7%

Table 2: Two methods of codebook design.

4.3 Discussion

If we consider the first experiment using a dedicated codebook, the performance results are very encouraging. The first place recognition rate of 88.5% does not yet rival the other biometric systems. However, the results reported here represent one of the very few significant population studies of a practical automatic face recognition system reported to date and, as such, give an important indication of the future viability of automatic face recognition.

Considering the dramatic reduction in codebook dimensionality, the second set of recognition results are remarkably good. The LBG approach performs

significantly better than the KSOFM method. Its overall recognition rate is approaching that of the 40 vector system reported in section 4.1. To explain the variation in the performance of these two codebook generation algorithms more detailed consideration of their mechanics is required.

In general the codebook generation algorithms perform reduction by averaging the most similar vectors together, while still maintaining good coverage of the initial vector space. If the most similar vectors are also the most frequently used vectors, then the clustering process may reduce the good spread of vector choice required to maintain good differentiation when encoding the population. To investigate whether this is in fact the case feature histograms have been drawn up for the entire population.

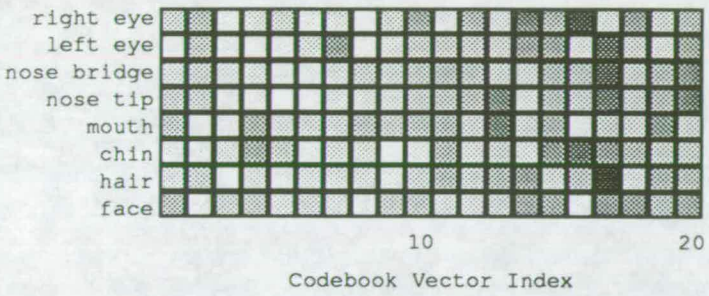

Figure 6: Population Feature Histograms for LBG.

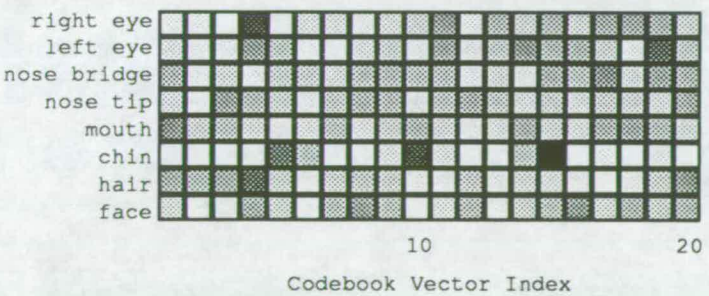

Figure 7: Population Feature Histograms for KSOFM.

Figures 6 and 7 illustrate the feature histograms for the LBG and KSOFM algorithms respectively. The spread of vector choice is much more even for the LBG vector set. We believe this is the underlying factor which explains the relatively poor performance of the KSOFM approach. In effect, KSOFM has maintained *too good* coverage of the entire feature space at the expense of differentiation between the most common feature types.

5 Conclusions and Future Work

The results of the initial research into the use of VQ for automatic facial recognition are encouraging. This approach has been shown to work for facial recog-

nition and undoubtedly has uses in other areas of image pattern recognition.

The dimensionality of the codebooks used by VQ based recognition has been identified as a potential limiting factor, however, this research has shown the important improvements which can be obtained in this area by using different codebook generation algorithms. However, we accept that further experimentation, with different populations, is required to validate the results reported here. It is further recommended that much larger populations are required to adequately test potential facial recognition systems.

References

[1] V Bruce and M Burton. Computer recogniton of faces. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 487–506. North-Holland, 1989.

[2] K Sutherland, D Renshaw, and P B Denyer. A novel automatic face recognition algorithm employing vector quantization. In *Digest of the IEE Colloquium on Facial Recognition and Storage*, London, January 1992.

[3] M A Fischler and R A Elschalger. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22:67–92, 1973.

[4] N M Nasrabadi and R A King. Image coding using vector quantization : A review. *IEEE Trans. on Comms.*, COM-36(8):957–971, August 1988.

[5] R M Gray. Vector quantization. *IEEE ASSP Mag.*, 1(2):4–29, April 1984.

[6] N M Nasrabadi and Y Feng. Vector quantization of images based upon the Kohonen self-organising feature maps. In *Proc. Int. Joint Conf. on Neural Networks (IJCNN-88)*, pages 101–108, July 1988.

[7] T-C Lee and A M Peterson. Adaptive vector quantization using a self-development neural network. *IEEE J. on Selec. Areas in Commun.*, 8(8):1458–1471, October 1990.

[8] T Kohonen. Clustering, taxonomy, and topological maps of patterns. In *Proc. 6th Int. Conf. on Pattern Recognition*, pages 114–128. IEEE, October 1982.

[9] J Linde, A Buzo, and R M Gray. An algorithm for vector quantizer design. *IEEE Trans. on Comms.*, COM-28(1):84–95, January 1980.

[10] J D McAuliffe, L E Atlas, and C Rivera. A comparison of the LBG algorithm and Kohonen neural network paradigm for image vector quantization. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-90)*, pages 2293–2296. IEEE, April 1990.

[11] W Equitz. Fast algorithms for vector quantization picture coding. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-87)*, pages 725–728. IEEE, April 1987.

A Novel Automatic Face Recognition Algorithm Employing Vector Quantization

K. Sutherland, D. Renshaw and P.B. Denyer [1]

1 Introduction.

Face Recognition is the most widely used and natural means of personal identification. However the automation of this fundamental human visual processing function is a highly challenging task requiring a good knowledge of human vision linked to multi-dimensional signal processing.

The ultimate goal of the research reported here is a functional automatic face recognition system performing at high accuracy in a real-time implementation. Crucial to this aim is the distillation of the intrinsics of the object face into the minimum amount of data, allowing for cost effective storage and rapid inter-person comparisons.

The benefit of automatic facial recognition over the other available biometrics must lie in its passive, non-intrusive ability to verify personal identification. However in addition to verification, the system described here can also perform identification.

This paper introduces a novel means of facial coding allowing an entire face to be represented in less than two hundred bytes of information. This data reduction is achieved while still preserving many of the intrinsic facial recognition features. The system could thus reduce an input face into a sufficiently small amount of information that it could be stored on a smart-card.

The algorithm used to perform the data reduction of the face is described and the results, in verification and recognition trials, are presented for a software implementation of the algorithm. The pre-processing image stages are not discussed in this publication.

2 Facial Features.

The most efficient manner in which to record information is to select the important aspects and ignore the irrelevant data. For face recognition (and storage) this means that we must isolate the facial parts (or features) of importance and separate them from the rest of the image data. This leaves us with the difficult problem of arbitrarily judging the importance of different facial parts in the recognition procedure. Human visual processing must hold the key to this process.

Fortunately the role of feature saliency in human face recognition has been researched [1]. However there is still no single framework for understanding the human performance of feature-based face recognition. Thus, for the system proposed here, it has been decided to extract a number of features from the face for which there is a substantial reason for their inclusion in the face recognition procedure.

In addition to the saliency of the feature, its variability with time and expression must also be considered. Within the selection of features a system of variable resolutions has been devised in order to facilitate a measure of relative importance.

Of the inner facial features; the eyes and mouth are of considerable importance [2] and are thus selected at full resolution. The entire face, the hair and the chin are spatially sub-sampled so that only a crude representation of these outer features is preserved. Only parts of the nose are preserved at full resolution because of its relative unimportance in frontal face recognition. The nose has been reduced to two smaller features, one representing the bridge of the nose and the other the tip of the nose and the nostrils. The spatial relationships between the features are also preserved because they are also thought to contain significant information relevant to human facial recognition [3].

[1]The authors are with the Integrated Systems Group, Electrical Engineering Department, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh, EH9 3JL.

In this manner an input image containing an entire face is broken up into eight features of interest: the eyes, the bridge of the nose, nostrils, mouth, chin, hair and the entire face. This represents the first stage of the facial coding process already yielding a substantial data reduction.

A system of limited feature embedding, similar to that used by Fischler and Elschlager [4] has been implemented to form part of the feature location required during the isolation of the features of interest. The problems associated with this process are not considered here.

3 Vector Quantization of Facial Features.

A lossy data compression technique termed Vector Quantization (VQ) has been widely used for speech and image transmission [5, 6]. However its use in image pattern recognition systems has been very limited.

The process of vector quantization involves the substitution of a waveform segment (from an image or a signal) with an approximately similar one drawn from a codebook of possible waveforms, termed vectors. The data reduction results from the transmission of only the index of the vector and not all the data points which constitute that vector. The signal is reconstructed by re-substituting the index number with the vector (or waveform segment). The data compression is obtained because less data is required to represent the index than the full vector.

Inherent in the process of VQ is a quantization error introduced by the approximate nature of the vectors. To minimise this error and perform the reconstruction function as accurately as possible the codebook of vectors must represent a wide variation of possible waveforms. There are a number of possible training algorithms available to ensure this.

For coding of facial features the important factor must be to replicate the diversity of feature types present within the population, while still maintaining a manageable number of vectors. In effect this is the function performed by the police photofit technique. The construction of a target face from the facial parts available is the process of vector quantization in action. With the witness selecting the most likely features and the difference between this photofit and the actual person being analogous to quantization noise.

In this application the vector quantization of the facial features is performed after these features have been extracted from the entire image. One Vector Quantizer is dedicated to each of the eight facial features used. For example, the subject's right eye is submitted to a Vector Quantizer trained only on examples of other right eyes, the quantizer then selects the best match from those training features. The VQ process thus yields a set of indices (coefficients) for all eight features representing the most likely vectors used to code the subject face.

To capture variations in the face a number of training images are utilised for each member of the population. From this training process a data set of probabilities is produced for each of the vectors used to code each of the facial features. For example, if a number of training images of the same person are presented to the system then an accumulation would be expected to occur in the facial features, in the codebook, most similar to those of the actual person. These probabilities are stored with information regarding the spatial inter-relationships of the features in a *signature*. This signature uniquely describes each member of the population. Figure 1 illustrates the three stages of facial coding.

4 Comparisons.

By submitting each member of the population to the training procedure, sample signatures for the entire population can be obtained. This set of individual signatures can then be used to form a population database. Identification and verification of a new test face are then performed by sequentially comparing the signature of the new face with all those in the database.

The comparative analysis is performed by using the VQ coefficients obtained for the facial features of the test image to obtain a probability measure for those features belonging a particular individual.

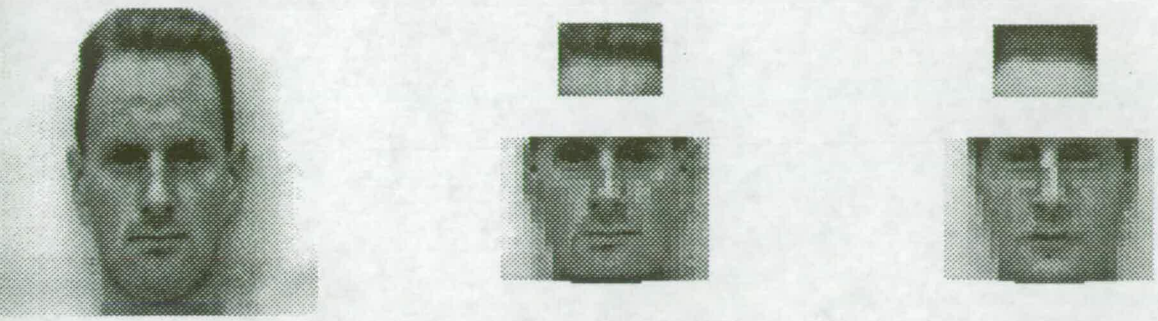

Figure 1: The three stages of facial coding: Original Input, Spatially sub-sampled, Vector Quantized.

A multiplicative accumulation is used to obtained the probability that all eight features present are a plausible representation of the individual under test. Thus probabilities for the test face representing each member of the population are obtained. The highest probability *score* is used to locate the most likely match for the test face.

At this stage the use of the data regarding facial feature inter-relationships has not been integrated with the VQ coefficient analysis. Thus the results presented in the following section relate to the comparison using only the VQ derived data. The integration of this feature positional information remains as future research.

5 Experimental Results.

For this trial a sample population of thirty individuals was obtained. For each member of the population 21 images were obtained on a number of different days. The images were frontal face information only and the size and orientation were kept approximately constant throughout the experiment.

In the trial, ten images of each person were used to construct the database of signatures. With the other ten used to test the system. The additional one image of each person was manually segmented and used to form the vector quantizer codebook for each feature.

In the recognition experiment each of the three hundred test images was coded and the best match located from the database. The algorithm produced the three people within the population most similar to the test face. The results for this trial are presented in Table 1.

Position	Cummulative Success Rate
1^{st}	89.19%
2^{nd}	92.33%
3^{rd}	95.47%

Table 1: Identification Performance of VQ Technique.

For verification a threshold was placed on the similarity between each test face and the signatures in the database. Each of the three hundred images is compared with all of the population's signatures. In this way a large number of imposter tests are performed. By convention the results of such experiments are presented as type I (False Reject Ratio FRR) and type II (False Acceptance Ratio FAR) errors[7].

The values obtained for type I and II errors are presented in Figure 2. The threshold used is plotted on a log scale as it relates heavily to the probabilities obtained in the training phase of the system. A cross over value of ~7% is obtained.

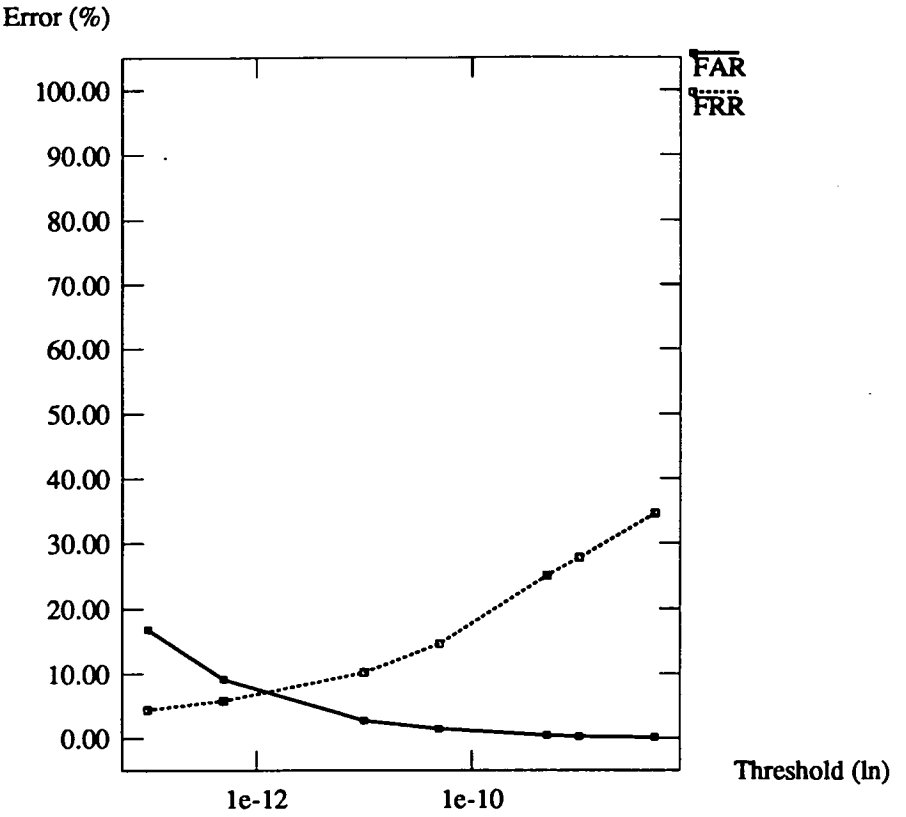

Figure 2: Verification Performance.

6 Conclusions.

The system described is a novel solution to the problem of reliable encoding of facial data for subsequent recognition. The use of Vector Quantization linked to statistical comparison has shown its ability to cope with facial variations on a day to day basis.

Compared to the other facial recognition approaches [8] which have been submitted to comparible trials, the technique introduced here has a broadly similar performance with the substantial advantage of a much smaller data storage requirement.

The recognition performance of ~89% and the verification error cross over point of ~7% cannot yet challenge the performance of some other biometric systems (e.g. fingerprint, voice or retinal scan recognition). However, the authors believe that the construction of a viable biometric facial recognition device is a realisable goal.

References

[1] I H Fraser and D M Parker. Reaction time measures of feature saliency in a perceptual integration task. In H D Ellis, editor, *Aspects of Face Processing*, pages 45–52. Martinus Nijhoff, 1986.

[2] H D Ellis, J W Shepherd, and G M Davies. Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8:431–439, 1979.

[3] G Rhodes. Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, 17:43–63, 1988.

[4] M A Fischler and R A Elschalger. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22:67–92, 1973.

[5] R M Gray. Vector quantization. *IEEE ASSP Magazine*, April 1984:4–29.

[6] N M Nasrabadi and R A King. Image coding using vector quantization: A review. *IEEE Trans on Communications*, 36(8):957–971, January 1988.

[7] J R Parks. Biometrics: the people sensors. *Sensor Review*, 9(2):79–84, April 1989.

[8] V Bruce and M Burton. Computer recogniton of faces. In A W Young and H D Ellis, editors, *Handbook of Research of Face Processing*, pages 487–506. North-Holland, 1989.

AUTOMATIC FACE RECOGNITION

K Sutherland, D Renshaw and P B Denyer.

University of Edinburgh, Scotland.

1 INTRODUCTION

The study of Biometrics, *ie* the measurements or characteristics which uniquely identify us all, has received a substantial amount of recent research work. This area represents an emerging field of multi-disciplinary research linking science and engineering with psychology and linguistics.

The increased interested in biometrics is due to the requirement for, fail-safe, personal identification in computerized access control. A number of different biometrics are being investigated at the present time, including fingerprint [13], speaker recognition [26], and signature dynamics [19].

In human terms, face recognition is the most widely used and natural means of personal identification. However, the automation of this fundamental human visual processing function is a highly challenging task requiring a good knowledge of human vision linked to multi-dimensional signal processing.

The ultimate goal of the research reported here is a functional automatic face recognition system performing at high accuracy, in a real-time implementation. Essential to this task is the distillation of the intrinsics of the object face into the minimum amount of data, allowing for cost effective storage and rapid inter-person comparisons. The benefit of automatic facial recognition over the other available biometrics must lie in its passive, non-intrusive ability to verify personal identity.

A review of competing facial recognition techniques is presented in this paper. We then go on to introduce a new method of feature-based facial coding allowing an entire face to be represented in less than two hundred bytes of information. Crucial to this coding process is the use of an *a priori model* of the face, which helps to guide the location and storage of the most important facial parts. The data reduction is thus achieved while still preserving many of the intrinsic facial recognition features.

The algorithm used to perform the data reduction of the face is described in this paper. Results, for verification and recognition trials, are presented for a software implementation of the algorithm.

2 AUTOMATIC FACE RECOGNITION

The most efficient manner in which to record information is to select the important aspects and ignore the irrelevant data. For face recognition (and storage) we must isolate the facial parts (or features) of importance and separate them from the rest of the image data. This leaves us with the difficult problem of arbitrarily judging the importance of different facial parts in the recognition procedure. The study of human visual processing must yield some indicators to this process.

The review below describes the broad approaches adopted by different research groups in the pursuit of a general method of facial characterisation and subsequent inter-face comparison. Some of the techniques described below attempt to use our knowledge of human visual processing to aid facial recognition, while other techniques represent the application of a more general image pattern recognition technique to face recognition.

Probably the earliest approach to be investigated, and that used by the police in the first stage of obtaining a suspect's picture, is that of qualitatively describing the facial features and their distribution on the face. This was researched by Harmon in 1973 [10] with the descriptions being provided by a panel of judges. The recognition of the subjects was based on a measure of the similarity of the faces in terms of the features given. The performance was reasonable, but such a qualitative system of facial analysis is very difficult to mimic in an automatic system.

For many years researchers have suggested that the measurements between the facial features (*ie* the nose, the right eye *etc*) are critical in the process of face recognition. However, different groups have produced slightly different sets of measurements [23, 29] which they argue preserve the intrinsics of the face. The underlying problem for a measurement based approach is that the information content of the mea-

urements is dependent on the accuracy of the feature location algorithm. Only in recent years has it been possible to perform this function automatically and even now only with questionable certainty and considerable computational load.

The methods used for facial feature location fall into two general categories: Firstly those based on the edge detected images (*ie* automatically obtained line drawings) of the face, and secondly on the actual grey scale digitized view of the face.

line fitting of the edge map of the face to the predefined facial features has been shown to work [4], but accuracy is still far too low for a realisable system. A related technique is the use of the Hough transform; this has been shown to work fairly well for eye location [15] although the associated computational cost is substantial. A multi-layer perceptron neural network has been used for successful eye location in a constrained facial area [11].

As yet it does not appear that any of the facial measurement techniques can perform sufficiently well to justify their use in a commercially available system.

Baron [2] described an optical computer which, instead of using measurements, performed comparisons on pixel level data. To this end, the system correlated five sub-parts of the input face with the same five face parts stored from an earlier exposure. This proved very successful at recognition on a small sample size, although extensive image normalisation was required. The five feature areas were manually selected.

Probably the most well know system used for facial recognition is WISARD [1]. As well as being able to distinguish between facial expressions, it has also been shown to function as a face recogniser on a population of 16 [25]. WISARD is a general purpose pattern recogniser and therefore when used for facial recognition it does not represent an efficient solution.

The ever expanding area of neural networks has also spawned several face recognition algorithms [17, 24]. A very recent development in this area is the PiCard identity verification system [18] which has been successfully tested on a sample population of 100 subjects.

There are a number of other techniques under investigation for the characterisation and recognition of faces [3, 22, 12], but as yet they have not proven themselves as viable face recognition systems.

To exploit fully the relative constancy of facial images (*ie* that all faces are fundamentally similar) we must adopt a model-based approach. A feature-based system of facial recognition which utilises information pertinent to the known facial features is one such model-based approach.

3 FACIAL FEATURES USED

In a feature-based facial recognition technique the information content of the facial parts must be estimated in order to determine their value in facial recognition. Fortunately, the role of feature saliency in human face recognition has been widely researched [7, 9]. However, there is still no single framework for understanding the human performance of feature-based face recognition. Thus, for the system proposed here, it has been decided to extract a number of features from the face for which there is a 'substantial' reason for their inclusion in the face recognition procedure.

In addition to the saliency of a feature, its variability with time and expression must also be considered. Within the selection of features a system of variable resolutions has been devised in order to facilitate a measure of relative importance.

Of the inner facial features; the eyes and mouth are of considerable importance [5] and are thus selected at full resolution. The entire face, the hair and the chin are spatially sub-sampled so that only a crude representation of these outer features is preserved. Only parts of the nose are preserved at full resolution because of its relative unimportance in frontal face recognition. The nose has been reduced to two smaller features, one representing the bridge of the nose and the other the tip of the nose and the nostrils. The spatial relationships between the features are also preserved because they are also thought to contain significant information relevant to human facial recognition [20].

In this manner an input image containing an entire face is broken up into eight features of interest: the eyes, the bridge of the nose, nostrils, mouth, chin, hair and the entire face. This represents the first stage of the facial coding process already yielding a substantial data reduction, while still maintaining much of the characteristic information contained in the face [28].

4 FEATURE LOCATION

If we are to successfully encode the facial features it is essential that we locate, and isolate them from the rest of the image. To this end, a system of model-based feature embedding, similar to that used by Fischler and Elschlager [6] has been implemented.

	Correct	1 Feature Missed	2 Features Missed	Complete Failure
Number	47	17	8	2
%	62	22	11	3

TABLE-1: Feature Location Performance.

Firstly, the eyes and mouth are located using three inter-dependent, constrained searches. If a plausible spatial pattern of two eyes and a mouth is present within the image, then the face is judged to have been successfully located. Further searches are then initiated to locate the remaining features. The local feature positioning, within the search area, is performed using a computationally optimized normalised template matching function [21]. The performance of the feature location system based on this approach is given in Table 1. The results presented are for a sample set of 74 images drawn from twenty different individuals. Of the two complete failures recorded, one had glasses with large reflections which misled the eye location stage, and the other face was very badly rotated.

The images used were captured in a fairly constant environment using controlled lighting. However, accurate positioning of the subject and control of facial expression were not performed. The subjects were merely asked to sit looking at the camera. In this way, it was hoped to introduce some of the problems which would be present in a real-world instance of a face recognition device.

5 VECTOR QUANTIZATION OF FACIAL FEATURES

A lossy data compression technique termed Vector Quantization (VQ) has been widely used for speech and image transmission [8, 14]. As yet, its use in image pattern recognition systems has been very limited.

The process of vector quantization involves the substitution of a waveform segment (from an image or a signal) with an approximately similar one, drawn from a codebook of possible waveforms, termed vectors. The data reduction results from the transmission of only the index of the vector and not all the data points which constitute that vector. The signal is reconstructed by re-substituting the index number with the vector (or waveform segment). The data compression is obtained because less data is required to represent the index than the full vector, one dimensional VQ is summarised in Figure 1.

Inherent in the process of VQ is an error, termed *quantization noise*, introduced by the approximate nature of the vectors. To minimise this error, and perform the reconstruction function as accurately as possible, the codebook of vectors must represent a wide variation of possible waveforms. There are a number of possible training algorithms available to ensure this [27].

For the encoding of facial features the important factor must be to replicate the diversity of feature types present within the population, while still maintaining a manageable number of vectors. In effect this is the function performed by the police 'photofit' technique. The construction of a target face from the facial parts available is the process of vector quantization in action; with the witness selecting the most likely features and the visual difference between this photofit and the actual person being analogous to quantization noise.

In this application, the vector quantization of the facial features is performed after these features have been extracted from the entire image. One vector quantizer is dedicated to each of the eight facial features used. For example, the subject's right eye is submitted to a vector quantizer trained only on examples of other right eyes, the quantizer then selects the best match from those training features. The VQ process yields a set of indices (coefficients) for all eight features representing the most likely vectors used to code the subject face.

To capture possible variations in the face with time a number of training images are utilised for each member of the population. From this training process a data set of probabilities is produced for each of the vectors used to code each of the facial features. For example, if a number of training images of the same person are presented to the system then an accumulation would be expected to occur in the facial features, in the codebook, most similar to those of the actual person. Figure 2 illustrates the manner in which the allocation of vectors to training faces occurs. In this figure a histogram for one of the eight features can be seen for a training set of ten images of one person. An accumulation has occurred in two of the 'bins' in the histogram, this is an indication of the variability of the pixel data with constitutes that subject's right eye in the training images used.

The probabilities obtained for the training images are stored with information regarding the spatial interrelationships of the features in a *signature*. This signature uniquely describes each member of the population and can be stored in less than two hundred bytes of data. Figure 3 illustrates the three stages of facial coding.

6 COMPARISONS

Signatures for the entire population can be obtained by submitting each member of the population to the training procedure. This set of individual signatures can then be used to form a population database. Identification and verification of a new test face are then performed by sequentially comparing the signature of the new face with all those in the database.

The comparative analysis is performed by using the VQ coefficients, recorded for the facial features of the test image, to obtain a probability measure for those features belonging a particular individual. A weighted accumulation is used to obtain the probability that all eight features present are a plausible representation of the individual under test. Thus, probabilities for the test face representing each member of the population are obtained. The highest probability *score* is used to locate the most likely match for the test face.

At this stage the use of the data regarding facial feature inter-relationships has not been integrated with the VQ coefficient analysis. Thus, the results presented in the following section relate to the comparison using only the VQ derived data. The integration of this feature positional information remains for future research.

7 EXPERIMENTAL RESULTS

For this trial a sample population of thirty individuals was utilised. For each member of the population 21 images were obtained on a number of different days. The images used had all been successfully passed through the feature location stage, *ie* the effect of feature location errors have been removed from this experiment.

In the trial, ten images of each person were used to construct the database of signatures, the other ten were used to test the system. The additional one image of each person was manually segmented and used to form the vector quantizer codebook for each feature.

In the recognition experiment, each one of the three hundred test images was coded and the best match located from the database. The algorithm produced the three people within the population most similar to the test face. The results for this trial are presented in Table 2. The average rank figure for this set of images was found to be 2.15, from a possible set of 30 people (this measure was badly affected by a small number of outlying results).

Position	Cummulative Success Rate
1^{st}	89.19%
2^{nd}	92.33%
3^{rd}	95.47%

TABLE-2: Identification Performance of VQ Technique.

For verification a threshold was placed on the similarity between each test face and the signatures in the database. Each one of the three hundred images was compared with all of the population's signatures. In this way a large number of imposter tests are performed. By convention the results of such experiments are presented as type I and type II errors[16]. Type I errors relate to the number of false rejects, *ie* the number of people who have failed to be recognised as who they are, and it is termed the False Rejection Ratio (FRR). Type II errors are the number of false acceptances, *ie* imposters who are falsely recognised as someone else, similarly named the False Acceptance Ratio (FAR). When this analysis was performed for this system, a cross-over value of approximately 7% was obtained, representing the best compromise between the FAR and the FRR.

8 CONCLUSIONS

The system described above is a novel solution to the problem of reliably encoding facial data for subsequent recognition. The use of vector quantization, linked to a statistical comparison stage, has shown its ability to cope with facial variations on a day to day basis.

The algorithm presented here exploits *a priori* knowledge about the face, in a model-based way, to produce a very small data set representing each person. Without adopting a similar model-based approach it is unlikely that any other technique will be able to match this level of data reduction.

The analysis of the complete facial recognition algorithm, which has been presented in this paper, represents one of the few detailed studies performed in this area. The substantial image data, drawn from a semi real-world situation, is a true test of this algorithm's performance as a possible, future face recognition system. Few similar studies have been attempted and thus a comparison between this algorithm and the other competing techniques discussed in this paper cannot be performed.

It is recommended that before fabrication of a hardware device based on this algorithm is initiated, a