# Optimising Multimodal Fusion For Biometric Identification Systems

A THESIS SUBMITTED TO THE UNIVERSITY OF KENT AT CANTERBURY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE SUBJECT OF ELECTRONICS ENGINEERING

BY

JACQUELINE JOHN GEORGE

November 2004

To Dad, Mum and Teta

# Contents

# Acknowledgments

I thank God, who has constantly been by my side through the best and toughest years of my life and for whom the credit of this work goes to.

The work on this thesis has been an inspiring, often exciting, sometimes challenging, but always an interesting experience. It has been made possible by many other people, who have supported me with criticism, helpful assistance and references. These people have been nothing less than instruments of the Divine in order to get me through these years and this research. It is a pleasant aspect that I have now the opportunity to express my gratitude for them.

For both my supervisors, Professor M.C. Fairhurst and Dr Farzin Deravi, I owe a debt of gratitude, which I cannot put into words. I thank them for making my dream come true and giving me the opportunity to do this research. I thank them for their patience and for bearing my bad moments. I thank them for their continuous support and precise guidance, which made this research possible.

I acknowledge the financial support I received from the Sudanese government for this research work.

I would like to express my gratitude to my friends and colleagues. I would like to thank my friend Lina for encouraging me to do this research. I would also like to thank Nick Mavity for his precious help and guidance at the start of this project and to thank Samuel Chindaro for his continuous support not only with matlab programming, but also for providing valuable advises and help through out this research. I would also like to thank Kostas Sirlantzis for his patient respond to all my questions and for his interesting conversations during coffee breaks, which I did learn a lot from, especially about the Greek history. My thanks extend to all my colleagues in the DSRG, the previous and current people, for supporting me in all sorts of ways. In particularly, I would like to thank Sanuel Haque,

Mohamed Razian, Elina Kaplani, Salem Alkaabi, Costas, Ether, Serkawt Farhan, and Thomas. Special thanks goes and to Kameran from the Photonics Lab and to Vangelis from the Embedded Systems Lab.

My special thanks goes to all the academic and technical staff in the department for all their assistance and friendliness throughout my four years in the university.

A journey is easier when you travel together. Special friends lent a hand and often a 'heart' during this long educational journey, which I owe a dept of gratitude. I would like to thank my friends Alexious louridas, Salavat Magazov and Miral Metawie. Thank you for being there, I do not know what I would have done without you. Thank you for tolerating my stupidity and madness and thank you for sharing with me the best and toughest moments of this period.

I feel a deep sense of gratitude for my uncle Mr. Brian Wilson and my aunt Mrs. Norma Wilson, my parents in local parentus, as my uncle always says. My aunt and uncle have always been there to help and guide me throughout my stay in U.K. I thank them for providing me with a family atmosphere and make me feel as if I am in Sudan. I also thank my two cousins Jen and Jess for being as my sisters and providing me with help and guidance whenever I needed it.

I have saved the best for last and the best is for those who have been there the longest, these are my family. There are no words that adequately express my appreciation and gratitude to them. I thank them for their never-ending love, encouragement, continuous prayers and long-distance support, which carried me through both the good and bad times. I thank them for understanding and accepting my absence from home for the sake of my ambitions and my career. Father, mother and grandmother I indeed thank you for everything and I dedicate this thesis to you.

Many more people have participated in various ways to ensure the success of this research. I am painfully aware of the omission of their names from this page, although I cannot express my gratitude to you all, I simply say "**Thank you**"

# Abstract

Biometric systems are automatic means for imitating the human brain's ability of identifying and verifying other humans by their behavioural and physiological characteristics. A system, which uses more than one biometric modality at the same time, is known as a multimodal system. Multimodal biometric systems consolidate the evidence presented by multiple biometric sources and typically provide better recognition performance compared to systems based on a single biometric modality.

This thesis addresses some issues related to the implementation of multimodal biometric identity verification systems. The thesis assesses the feasibility of using commercial off-the-shelf products to construct deployable multimodal biometric system. It also identifies multimodal biometric fusion as a challenging optimisation problem when one considers the presence of several configurations and settings, in particular the verification thresholds adopted by each biometric device and the decision fusion algorithm implemented for a particular configuration. The thesis proposes a novel approach for the optimisation of multimodal biometric systems based on the use of genetic algorithms for solving some of the problems associated with the different settings. The proposed optimisation method also addresses some of the problems associated with score normalization. In addition, the thesis presents an analysis of the performance of different fusion rules when characterising the system users as sheep, goats, lambs and wolves.

The results presented indicate that the proposed optimisation method can be used to solve the problems associated with threshold settings. This clearly demonstrates a valuable potential strategy that can be used to set *a priori* thresholds of the different biometric devices before using them. The proposed optimisation architecture addressed the problem of score normalisation, which makes it an effective "plug-and-play" design philosophy to system implementation. The results also indicate that the optimisation approach can be used for effectively determining the weight settings, which is used in many applications for varying the relative importance of the different performance parameters.

# List of Acronyms

| | |
|---|---|
| **DET** | Detection Error Trade-off |
| **EER** | Equal Error Rate |
| **FAR** | False Acceptance Rate |
| **FRR** | False Rejection Rate |
| **FTE** | Failure To Enrol |
| **GAR** | Genuine Accept Rate |
| **GAs** | Genetic Algorithms |
| **GEM** | Genetic Evolution Module |
| **GFAR** | Global False Accept Rate |
| **GFRR** | Global False Reject Rate |
| **GTER** | Global Total Error Rate |
| **PC** | Personal Computer |
| **PIN** | Personal Identification Number |
| **ROC** | Relative Operating Characteristic |
| **TER** | Total Error Rate |
| **TSR** | Total Success Rate |
| **USS** | Universal Stochastic Sampling |

# List of Figures

# List of Tables

# Part 1

## Research and Experimental Results

# Chapter 1

# Introduction

## 1.1 Abstract

This chapter starts by giving a brief overview of the history of biometrics, followed by a section, which defines biometrics. This is then followed by a section, which introduces a range of personal identification methods then a discussion of the general structure of a biometric system and its errors. A brief overview of the different biometric technologies is then provided, summarizing the advantages and disadvantages of each. The reasons choosing multimodal biometrics systems are also presented. Finally, the purpose of the research and the outline of the thesis are presented.

## 1.2 History of Biometrics

Biometrics in its general term is derived from the Greek words *bios* (life) and *metrikos* (to measure). It is also defined as the statistical measurement and analysis of biological observations and phenomena. In the context of system authentication the term biometrics means using the body as a "password".

Biometrics is becoming an interesting topic in computer and network security. However, the ideas of biometrics have been around for many years. One of the first known cases of humans using biometrics to identify one another was by early Chinese merchants [Moenssens71]. Joao de Barros, an explorer and writer, wrote that the Chinese merchants used a form of biometrics by stamping children's palm prints and footprints on paper with ink. In doing this, the Chinese found a way to distinguish young children from one another. This is one of the earliest known cases of biometrics in use and is still being used today.

In the 1890s, Alphonse Bertillion developed 'Bertillonage', a method of bodily measurement [Rhodes56]. He realized that there are certain elements of the body that remain fixed, such as the size of the skull or the length of the fingers. His system was used by police authorities throughout the world, until it quickly faded when it was discovered that some people shared the same measurements and based on the measurements alone, two people could get treated as one. After this, the police used finger printing, which was developed by Richard Edward Henry of Scotland Yard, instead, essentially reverting to the same methods used by the Chinese for years [Jain04a].

Although biometrics emerged from its extensive use in law enforcement [George99] [Prins98] to identify criminals (e.g., illegal aliens and forensics), it is being increasingly used today to establish person recognition in a large number of civilian applications [ATMs99].

## 1.3 What is Biometrics?

Biometrics is the science of using digital technology to identify individuals based on the individual's unique physical and/or behavioural characteristics [Jain99]. Physical characteristic include fingerprint, facial recognition, retinal and iris scanning, hand geometry. On the other hand the behavioural aspects of human beings include voice pattern and handwriting.

Any of the human physiological and/or behavioural characteristics can be used as a biometric characteristic if it satisfies the following requirements:

- *Universality*, which means that each person should posses the required feature characteristics.
- *Uniqueness,* which indicates that no two persons should have the same measured characteristics.
- *Permanence,* which means that the characteristic should be invariant over a period of time.
- *Collectability,* indicating that the characteristic is readily presentable to a sensor and is easily quantifiable.

## 1.4 Personal Identification Methods

There are two main established types of automatic personal identification methods that have been widely used: *knowledge-based* and *possessions-based.* Knowledge-based methods use "*something that I know*" for identification such as pin numbers and passwords. Possessions-based methods use "*something that I possess*" for identification such as ID cards and physical keys. The weakness of these two methods lies in the fact that knowledge can be forgotten as well as shared, stolen or guessed and possessions can be easily lost, forged or duplicated [Miller94]. In addition, they are unable to differentiate between an authorized person and an impostor using the token or the knowledge fraudulently acquired from the authorized person [Jain00]. Biometrics, on the other hand, which is "*something unique about me* " are inherently secure since they are unique features an individual has. The science of biometrics is an elegant solution to identifying an individual and avoids the problems faced by knowledge-based and possession-based security methods. In addition, they are more reliable and more capable of differentiating between an authorized person and a fraudulent impostor.

## 1.5 Biometric Authentication Systems

### 1.5.1 The General Structure of a Biometric System

A biometric system is essentially a pattern recognition system that recognises the identity of a person on the basis of a physiological or behavioural characteristic. Although the use of each biometric technology has its own specific issues, the basic operation of any biometric system is very similar. Figure 1.1(a) and 1.1(b) illustrates the enrolment stage and the recognition stage respectively, which represents the typical steps of an authentication process.

**Enrolment Stage**

This stage is performed only once, since it inserts the specific biometric characteristic into the system database. This phase either combines the knowledge-based method (e.g. PIN or name) with biometrics (e.g. fingerprint) in the case where the biometric characteristic will be stored in a central database or combines the possession-based method (e.g. smart card) with biometrics (e.g. fingerprint) in the case where the biometric characteristic will be stored on a smart card. The first step in this stage starts by the user providing either a knowledge-based method or a possession-based method depending on the application, then a data capture process is performed where the biometric sample of the user is captured using an input device. The quality of this sample is crucial for further authentications of the user, so the quality of this biometric sample must be particularly checked and the acquisition of the biometric sample must be repeated if it is not sufficient. This is the reason why this first measurement is normally guided by a supervisor who explains the use of the biometric reader.

The biometric sample in its raw format can be expected to contain a lot of noise or irrelevant information that needs to be eliminated, so the raw measurements are processed and only the important features are extracted and used. This significantly reduces the amount of data to be processed and generates a compact but expressive representation, called a "template". The process of feature extraction is not lossless and so the extracted features cannot in general be used to reconstruct the biometric sample completely.

The generated template must then be stored. Depending on the application, the template may be stored in a central database of a biometric system or recorded on a smart card issued to the individual.

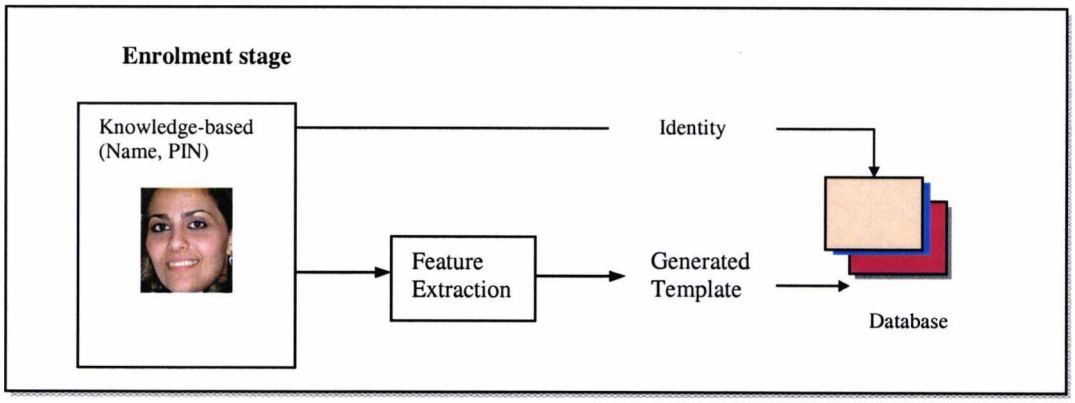



**Figure 1.1(a):** Block diagrams of the enrolment stage

**Recognition Stage**

This phase is repeated at each transaction. During this phase, the biometric device captures a current biometric sample of the user to be identified. This sample measurement is then processed and the important features are extracted to produce the same representation as the template. The resulting representation is then fed to the matcher, which compares it against the template obtained during enrolment to validate the identity of the individual. This returns a matching score $s$ that quantifies the similarity between the input and the database template representations. The final step in this stage is based on a predetermined threshold $t$, where the score $s$ is compared with the threshold $t$ to make the final decision. As a result the system will make one of the following four possible decisions:

1. A legitimate user (genuine) is accepted; this happens if the score $s$ generated from pairs of samples from the same person is higher than or equal to the threshold $t$.

2. A legitimate user (genuine) is rejected; this happens if the score *s* generated from pairs of samples from the same person is lower than the threshold *t*.

3. An impostor is accepted; this happens if the score *s* generated from pairs of samples from different persons is higher than or equal to the threshold *t*.

4. An impostor is rejected; this happens if the score *s* generated from pairs of samples from different persons is lower than the threshold *t*.

Depending on the application context, the recognition stage may either be a verification mode, an identification mode or a screening mode [Jain04c]:

In the verification mode or sometimes called the "positive identification", the user-input sample is compared against the particular claimed reference template stored in the database. It conducts a one-to-one comparison to determine whether the claim is true or not. It requires the claimed identity such as a PIN (Personal Identification Number), a user name or a smart card to be provided prior to the verification stage.

In the identification mode, the user-input identity is compared with all the templates stored in the database in order to find the closest match. It conducts a one-to-many comparison to establish a user's identity (or fails if the subject is not enrolled in the system database) without the user having to claim an identity. Thus biometric identification is a more complicated, difficult and time-consuming process than biometric verification.

The screening mode or sometimes called the "open set identification" determines whether a person belongs to a watch-list of identities. The screening watch-list consists of a moderate number of identities. The user-input identity is compared with all the templates stored in the watch-list database in order to find the closest match. In this mode the individual does not make an identity claim, and in some cases does not personally interact with the system whatsoever. Examples of the watch-list task could be comparing visitors to Parliament against a terrorist database.

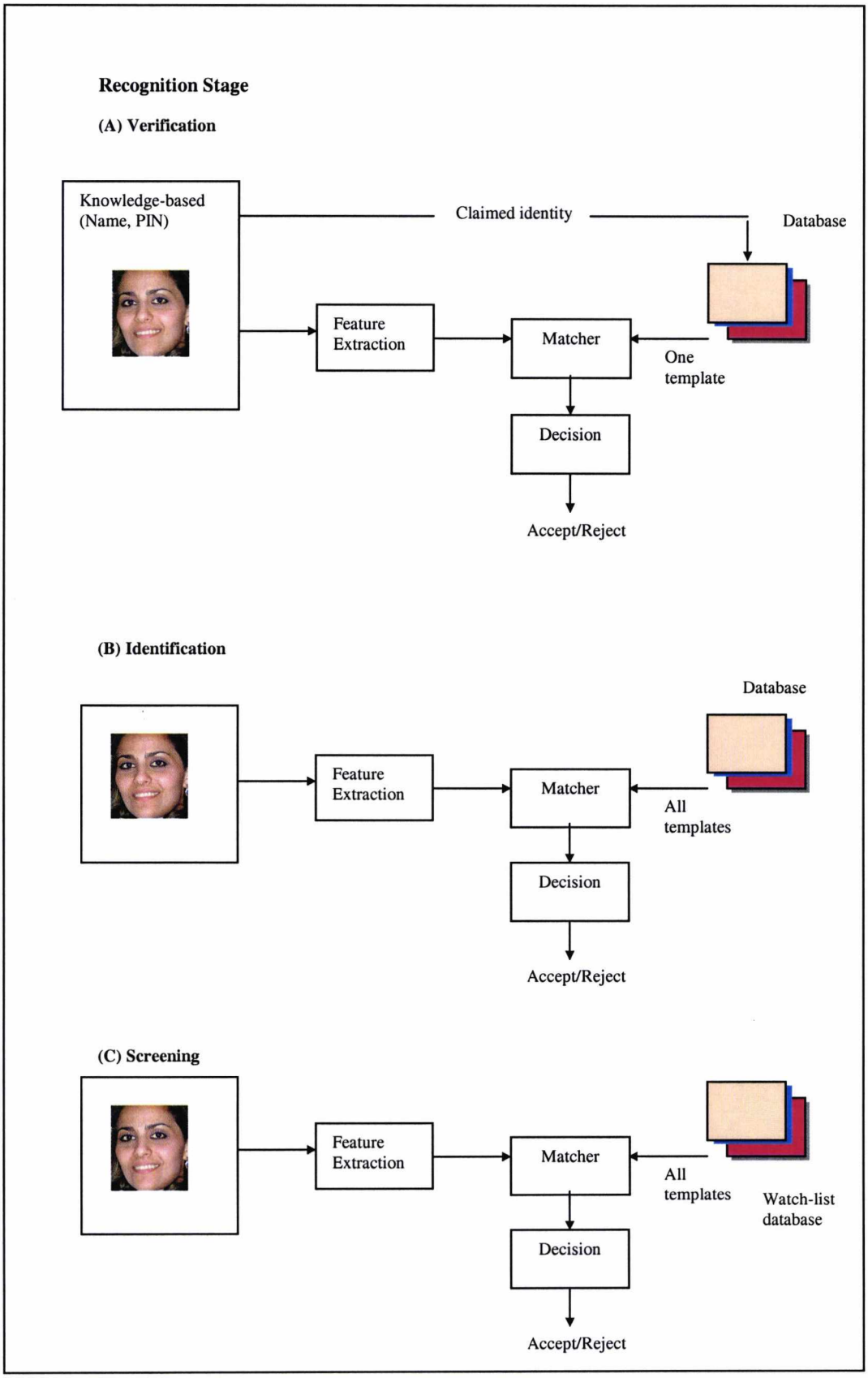



**Figure 1.1(b):** Block diagrams of the recognition stage

### 1.5.2 Biometric System Errors

Any biometric system will suffer from some specific failure and error rates occurring at the enrolment stage and the recognition stage, which will affect the system performance, and these may be characterised in different ways [Jain01]:

**Failure to Enrol (FTE)**

This failure rate indicates the percentage of times the user cannot enrol in the system. It occurs when the system rejects poor quality templates during the enrolment stage.

**False Reject Rate (FRR)**

This error is the likelihood that a legitimate user (client) is rejected during the recognition stage (because the system does not find the user's current biometric data similar enough to the master template stored in the database). This error is also known as *Type I error*. It is defined as:

$$FRR = \frac{\text{Number of false rejection}}{\text{Number of client accesses}} \qquad (1.1)$$

**False Accept Rate (FAR)**

This error is the likelihood that an impostor is accepted by the system as being a legitimate user during the recognition stage (because the system finds the impostor's biometric data similar enough to the master template of a legitimate user). It is also known as *Type II error*. It is defined as:

$$FAR = \frac{\text{Number of false acceptance}}{\text{Number of impostor accesses}} \qquad (1.2)$$

**Equal Error rate (ERR)**

This is the point at which the system performs with an equal rate of false acceptance and false rejection. This value does not have any practical use it only indicates how accurate the device is. For example, if two devices with equal error rates of 1% and 10 % then this shows that the first device is more accurate (i.e. fewer errors) than the other.

Ideally, a biometric system should produce a zero equal error rate; that is it should be able to accept all genuine users and reject all attempted forgeries. However, the performance of today's biometric technologies is far from ideal, despite impressive claims by manufacturers. This is due to inaccuracy of the deployed technology, inconsistency of the related biometric characteristics and/or skilled forgery.

Figure 1.2(a) and 1.2(b) illustrates the performance of an ideal and a typical biometrics system respectively.

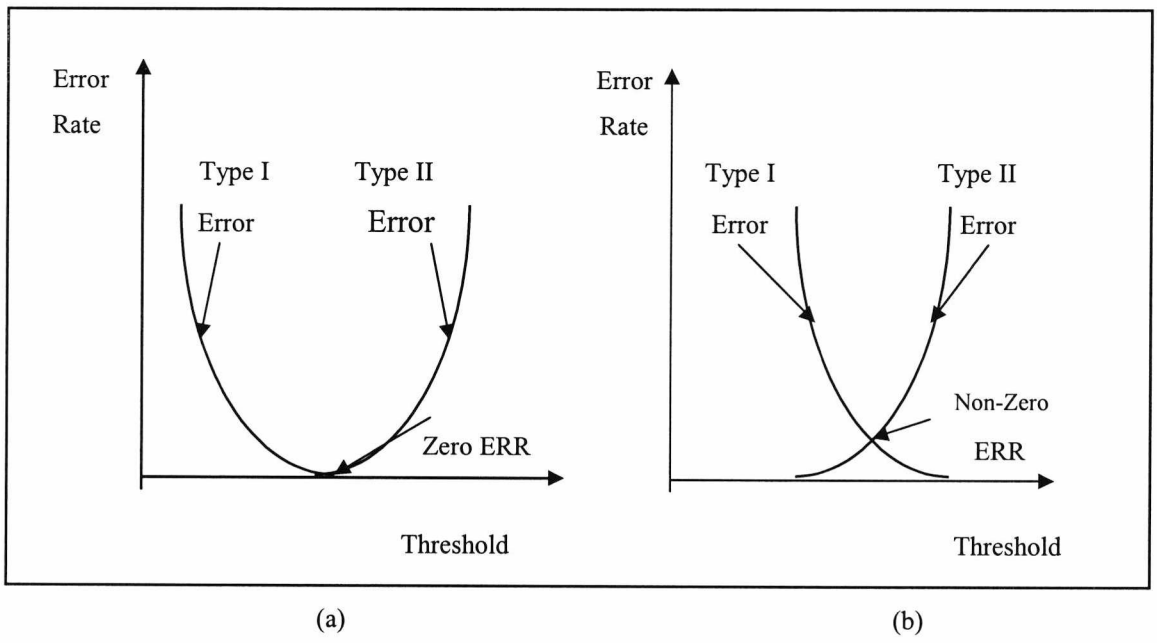



**Figure 1.2 (a):** The ideal behaviour of biometrics systems
**Figure 1.2 (b):** The typical behaviour of biometrics systems

Figure 1.2 shows that both Type I and Type II errors are functions of the system threshold $t$, if $t$ is decreased to make the system more tolerant to input variations and noise, Type II error increases. On the other hand, if $t$ is raised to make the system more secure, then Type I error increases accordingly.

## 1.6 Biometric Technology

There are various biometric technologies that are commercially available. Each has its strengths and weaknesses and the choice depends on the application. Before giving a brief introduction of the commonly used biometrics, a number of issues that need to be considered in a biometric device are addressed and these are [Prabhakar03]:

***Performance***: The overall performance of a system is evaluated in terms of its storage, processing time and the achievable recognition accuracy. The size of a template, especially when using smart cards for the storage, can be a decisive issue during the selection of a biometric system. Also the time required by the system to make a recognition decision is important, especially in real-time applications. If the processes of using a biometric system are lengthy, they could negatively affect the ability of the assets being protected to operate and fulfil its mission. For example there are challenges in using biometrics for border security. The use of biometric technologies could potentially impact the length of the inspection process. Any lengthening in the process of obtaining travel documents or entering a country could affect travellers significantly. Delays at the border affect the travellers and result in fewer people visiting the country, which might lead to loss in business for a nation. Accuracy is critical for determining whether the system meets requirements and in practice, how the system will respond. It is defined as the ability of the biometric system to discriminate between genuine and false claims of identity [Allgrove99]. For example, a very demanding authentication system may not tolerate a high degree of false acceptance. On the other hand, a credit card user will be annoyed if the system keeps on rejecting his genuine transaction.

***Acceptability***: User acceptability is a crucial consideration particularly in applications involving the general public. It indicates the extent to which people are willing to accept the use of a particular biometric identifier in their daily lives. For example, some people find biometric technologies difficult, if not impossible, to use. Still others resist biometrics because they believe them to be intrusive, inherently offensive, or just uncomfortable to use. Lack of cooperation or even resistance to using biometrics can affect a system's performance and widespread adoption. As an example, fingerprint technology may not be acceptable by some people because of its strong associations with the traditional identification of criminals [Jain99]. On the other hand, despite the low level of accuracy of signature verification [Jain99], this technique is widely used in document processing due to its high level of user acceptance.

***Circumvention***: This reflects how easily the system can be fooled using fraudulent methods. There are several methods for circumventing a system such as forcing exception processing built into the system that may not require using a biometric [Penny02]. Other method is to use verification fraud attempts to circumvent the system during the process of verification itself. Examples include forcing an individual to verify his identity to gain access, or presenting a facsimile of the actual biometric by faking it, or presenting stolen fingers that were chopped off the owner. In the latter case, most of the biometric devices available today can differentiate between a 'live' finger and an amputated one. Different types of biometrics have different degrees of difficulty of circumvention and these are summarised by Jain in [Jain99].

***Cost***: The cost of a biometric system is another factor to be considered when developing a biometric system. Not only the costs of the technology must be considered, but also the costs of the effects on people and processes. Both initial costs and recurring costs need to be estimated. Initial costs account for the engineering efforts to design, develop, test, and implement the system; hardware and software costs; network infrastructure improvements; and additional facilities required to enroll people into the biometric system. Recurring cost elements include hardware and software maintenance, hardware replacement costs, training of

personnel to enroll or verify the identities of people in the biometric system, and possibly the issuance of token cards for the storage of biometrics.

### 1.6.1 Commercial Biometric Technologies and their Applications

In this section a brief description of each of the available biometric technologies is provided, stating the advantages and disadvantages of each of them and presenting some of their typical applications. A detailed description of fingerprint, voice and face technologies is provided as they were used in this project.

Biometric applications fall into three main groups [Prabhakar03]: *commercial applications*, such as ATMs, Internet access, e-commerce, cellular phones, computer network logins, physical access control, electronic data security, medical records management and distance learning. *Government applications*, such as national ID cards, driving licence, passport control, border control and social security and *forensic applications,* such as criminal investigation, corpse identification, terrorist identification and missing children.

The commercial applications require positive recognition and may use the biometric system either in verification or identification mode. The government and forensic applications consist mainly of identification. There are two types of identification systems; one type, which is mainly used for government applications, is designed to ensure that a person's biometric information is not present in a database. The expected result of this search is a non-match. Comparing a person's biometric information against a database of all who are registered in a public benefits program, for example, can ensure that this person is not faking documentation to register under multiple identities. The other type, which is mainly used for forensic applications, is designed to check whether a person's biometric information is present in a database or not. For example comparing visitors to Parliament against a terrorist database results in either a visitor being on the database or not. In this section some examples of different applications of each biometric technology is provided.

### 1.6.1.1 Fingerprint

Fingerprint recognition is one of the oldest biometric techniques. Fingerprints as a biometric characteristic are unique (fingerprints of identical twins are different and so are the prints on each finger of the same person), highly permanent (the formation of a fingerprint which is the pattern of ridges and valleys on the surface of a fingertip is determined during the first seven months of fetal development) and easily collectable if present (i.e. not damaged, burned etc.) [Jain97][Ross03][Uludag04][Maltoni03]. However, they have some limitations: the performance of the currently available fingerprint recognition systems is affected by some environmental conditions (e.g. dirt on the sensor) and occupational factors (e.g. manual workers may have a large number of cuts and bruises on their fingerprints that keep changing). They are also generally regarded as highly unacceptable in some applications and social contexts because of their strong associations with the traditional identification of criminals.

**Fingerprint Sensing**

Based on the mode of acquisition, a fingerprint image may be classified as off-line or live-scan. An off-line image is typically obtained by smearing ink on the fingertip and creating an ink impression of the fingertip on paper. The inked impression is then digitised by scanning the paper using an optical scanner or a high-quality video camera. A live-scan image on the other hand, is acquired by sensing the tip of the finger directly, using a sensor that is capable of digitising the fingerprint on contact.

There are a number of live-scan sensing devices that can be used to detect the ridges and valleys present in the fingertip (ridges are the lines that the fingerprint pattern is made off while valleys are the spaces between the ridges). The most common live-scan sensing devices are based on optical, capacitive (or silicon) and ultrasound sensors [Maltoni03]:

The optical method is the most common method at present. At the centre of the optical scanner, a CCD- Camera (charged coupled device) is used [Newham95]. A CCD-Camera is simply an array of light sensitive diodes called photosites which generate an electrical signal in response to light photons. Each photosite records a pixel, a

tiny dot representing the light that hit that spot. Collectively, the light and dark pixels form an image of the scanned finger. An analog-to-digital converter in the scanner system processes the analog electrical signal to generate a digital representation of this image. In general the finger will be placed on a glass plate and the CCD camera takes the picture. The CCD system has an array of LEDs (light-emitting diodes) to illuminate the ridges and valleys of the finger.

The optical fingerprint sensors can withstand to some degree temperature fluctuations, they are also relatively cheap and they can provide resolutions up to 500 dpi. However, they suffer from some drawbacks such as the size of the platen, which must be of a sufficient size to achieve a quality image. The latent prints (leftover prints from previous users) must also be cleaned otherwise they can cause image degradation, as severe latent prints can cause two sets of prints to be superimposed. Also, the coating and CCD arrays can wear with age, reducing accuracy.

The capacitive method is one of the increasingly popular methods. Like the optical scanner the capacitive scanner generates an image of the ridges and valleys that make up a fingerprint. They are based on the capacitance of the finger. The capacitive sensor is made up of one or more semiconductor chips containing an array of tiny cells. Each cell includes two conductor plates, covered with an insulating layer. The two conductor plates form a basic capacitor, an electrical component that can store up charge. The surface of the finger acts as a third capacitor plate, separated by the insulating layers in the cell structure and, in the case of the fingerprint valleys, a pocket of air. Varying the distance between the capacitor plates (by moving the finger closer or farther away from the conducting plates) changes the total capacitance (ability to store charge) of the capacitor. This capacitance is converted through an analog-to-digital converter into an 8-bit grayscale digital image

Capacitive sensors generally produce better image quality, with less surface area, than optical sensors. Capacitive fingerprint sensors are integrated into many devices such as mobile phones and laptop computers due to its small size. However, due to

the reduction in the sensor size more care should be taken to ensure that enrolment and verification are done carefully. A poor enrolment may not capture the centre of the fingerprint, and subsequent verifications are subject to the same type of placement. Also, the fingerprint bitmap obtained from the capacitive sensor is affected by the finger moisture as the moisture significantly influences the capacitance. This means that people with unusually wet or dry fingers have problems with capacitive fingerprint sensors since wet fingers produces black images whilst dry fingers make the image pale.

Ultrasound technology, though considered perhaps the most accurate of the fingerprint technologies, is not yet widely used. The ultrasonic fingerprint sensors use ultrasound to monitor the finger surface. The user places his/her finger on a piece of glass and the ultrasonic sensor moves and reads the whole fingerprint. It measures the distance based on the impedance of the finger, the platen, and air. Ultrasound is capable of penetrating dirt and residue on the platen and the finger, countering a main drawback to optical technology.

**Fingerprint Processing**

Fingerprints are not compared and usually are not stored as bitmaps. Fingerprint matching techniques can be placed into two categories: minutiae-based and matching pattern-based [Prabhakar03a] [Hong88]. The Minutiae-based technique requires the location of the minutiae to be calculated with respect to the core (see Figure 1.3) during the process of feature extraction [Prabhakar03b]. Minutiae are individual unique characteristics within the fingerprint pattern that can be defined as the discontinuities that interrupt the smooth flow of ridges [FingerScan]. Many types of minutiae exist such as ridge ending, ridge bifurcation, bridges or islands as shown in Figure 1.3. A ridge ending is defined as the ridge point where a ridge ends abruptly. A ridge bifurcation is defined as the ridge point where a ridge diverges into branch ridges. Bridges is where small ridges join two longer adjacent ridges and island is a long ridge occupying a middle space between two divergent ridges.

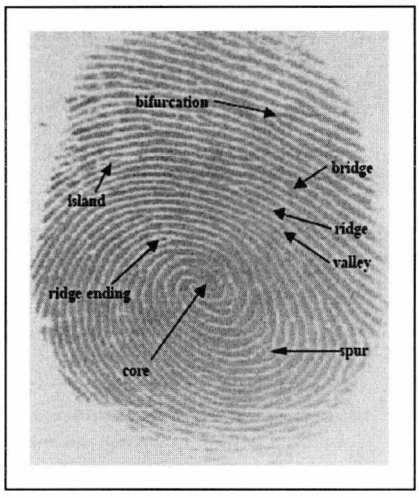



**Figure 1.3:** A Fingerprint Pattern

Although a minutiae-based representation is characterized by representing distinctive information about the fingerprint, a reliable automatic minutiae extraction can be problematic in low quality fingerprints [Jain01]. This method also does not take into account the global pattern of ridges and valleys and it is affected by wear and tear.

On the other hand, the pattern matching technique extrapolates data from a particular series of ridges. This data is used as the basis during the comparison stage. It requires that a segment of the same area be found and compared. Pattern matching performs better in the case of anomalies caused by scars, sweat, or dirt as compared to minutiae matching.

**Fingerprint Applications**

Fingerprint technology has been used in the areas of financial transaction and network security, examples of its applications include:

The Bank of America in 1999 used fingerprints to give customers access to their online banking services [Press99]. Before using the system, the customer enrols his/ her fingerprint on a chip attached to a multi-application smart card. During authentication the customer places his/her finger on a scanning device attached to a personal computer and the software then matches the fingerprint from the scanner

against the image stored in the smart card in order to make a decision on whether to accept or reject the customer.

A number of vendors have developed finger scanners resembling a computer mouse. Scanners built into computer keyboards have also been produced [Davies94].

#### 1.6.1.2 Speaker recognition

Speech contains information about the identity of the speaker. A speech signal includes also the language that is spoken, the presence and type of speech pathologies, the physical and emotional state of the speaker. Often, humans are able to extract the identity information when the speech comes from a speaker they are acquainted with.

The principle of speaker recognition is to analyse the voice of the user in order to store a voiceprint that is later used for recognition. The recording of the human voice for speaker recognition requires a human to say something. In other words the human has to show some of his/her speaking behaviour. Therefore, voice recognition fits within the category of behavioural biometrics. [Furui97][Bimbot97][Campbell97].

Speaker recognition has several drawbacks; it is not permanent since it changes over time due to age, medical conditions (having a cold), emotional state (e.g. stress), etc. The performance of voice-based recognition systems is also affected by several factors such as the background noise, the quality of the microphone used and the variation in tone due to disposition. However, the main advantage of the voice technology is that it does not require any special and expensive hardware. A microphone is used which is a standard accessory of any multimedia computer. The speaker recognition is also not intrusive for users and is easy to use.

**Text-dependent vs. Text-independent Speaker Recognition**

Speaker recognition systems are classified as text-dependent (fixed-text) and text-independent (free-text). In text dependent systems, during enrolment the user is asked to pronounce a phrase and the voice is then processed and stored in a

template. During authentication the user is asked by the system to pronounce the same phrase. In text-independent systems, during enrolment the system records the pronunciation of multiple phrases (e.g. numbers). In the authentication phase the system randomly chooses a phrase and asks the user to pronounce it.

The two main advantages of text-independent systems over text-dependent systems are: first the user does not have to remember a fixed phrase and second the system cannot easily be "spoofed" with the replaying of recordings of the user's speech.

**Speaker Recognition Techniques**

There are a few methods that are used for speaker verification. Text-dependent methods are usually based on template-matching techniques. In this approach, the input utterance is represented by a sequence of feature vectors, generally short-term spectral feature vectors. The time axes of the input utterance and each reference template or reference model of the registered speakers are aligned using a dynamic time warping (DTW) algorithm and the degree of similarity between them, accumulated from the beginning to the end of the utterance, is calculated. An alternative is to model the statistical variation in the spectral features. This is known as Hidden Markov Modeling (HMM), which has shown to outperform the DTW-based methods. The Hidden Markov Model (HMM) can efficiently model statistical variation in spectral features. Therefore, HMM-based methods were introduced as extensions of the DTW-based methods and have achieved significantly better recognition accuracies [Naik89].

In text-independent speaker verification, methods that look at long-term speech statistics or consider individual spectral vectors as independent of each other have been proposed. Examples of these are: the average-spectrum-based method and the vector quantization (VQ) methods. The average-spectrum-based method uses a weighted ceptral distance measure where the phoneme effects in speech spectra are removed by averaging the spectra. In the vector quantization method, VQ codebooks consisting of a small number of representative feature vectors are used as an efficient means of characterizing speaker-specific features. A speaker-specific codebook is generated by clustering the training feature vectors of each speaker. In

the recognition stage, an input utterance is vector-quantized using the codebook of each reference speaker and the VQ distortion accumulated over the entire input utterance is used to make the recognition decision.

**Speaker Recognition Applications**

Speaker recognition has been used in different applications. It has being integrated into security systems [Cole95] for online banking, bill payment and electronic commerce [SAFLINK99]. Speaker recognition has also made an impact in the penal system. This technology has been used for inmates on parole, juvenile inmates, and those under house arrest [Das1].

#### 1.6.1.3 Face

Facial images are probably the most common biometric characteristic used by humans to make a personal identification [Jain99]. They are the least intrusive and most socially acceptable from the user perspective [Millller94]. However, they have some limitations: the facial recognition systems are usually very sensitive to variation in illumination and to faces with different positions or expressions. They also perform poorly when the database size increases and require a large amount of storage for the database.

There are two main types of commercial facial recognition systems; the most common uses video, while the other uses thermal imaging.

Video face recognition technology analyses the unique shape, pattern and positioning of facial features [IBG99]. A video camera is used to capture an image from a distance of up to a few feet away from the user. A number of points on the face such as the position of the eyes and the mouth are usually mapped out.

On the other hand, the facial thermogram uses an infrared camera to scan a person's face and then digitise the thermal patterns [Ross94]. The patterns are created by the branching of blood vessels in the face. As the blood is hotter than the tissue surrounding it, it radiates heat that can be picked up at a distance.

**Facial Recognition Techniques**

There are four common approaches used to identify and verify users [Hsu02], this include eigenfaces, automatic face processing, neural network and local feature analysis.

Eigenfaces is an MIT technique that utilized two dimensional grayscale images representing distinctive characteristics of a facial image [Turk91]. Any facial image can be represented by combining many (100+) eigenfaces, and it is the coefficients representing that combination, which make up the template that is used to determine if the presented face is the claimed face. The advantage of this method is its speed and efficiency. However, it has problems identifying faces in different levels and pose positions.

Automatic face processing uses distances and ratios between common facial features [Ponti99]. This is the simplest technique and the least robust, and does not tend to be used as much as the others. Its advantages are simplicity and it is less affected by poor lighting conditions.

Neural Network processing uses the neural network to determine whether the presented face features are similar enough to the enrolled face features [Miros99]. It has the theoretical ability to be very intelligent and adaptive to changes.

Local-feature analysis records the relative locations of as many as 80 prominent facial landmarks, such as eyes, eyebrows, mouth, tip of the nose, bridge of the nose, and cheekbones [Visionics99]. In operation, the system compares facial features from a test subject, along with slight variations to account for changes in expression, with a database of these relative distances. Local-feature-analysis systems can also accommodate head orientations that vary on either side of a direct frontal image.

**Face Applications**

The use of video-based face recognition for consumer applications has grown considerably in the last few years. Examples of its commercial applications are:

In America, an ATM system automatically takes a picture every time a customer cashes a cheque [ATM99]. The customer first has to enrol in the system, but no bank account or driver's licence is needed. In order to cash a cheque, customers enter their Social Security numbers. This information combined with the biometric picture, creates a real-time, permanent record of the transaction.

Facial recognition is also used in some casinos as a way of identifying suspicious players. A surveillance camera captures an image of the individual's face and compares it to a digitised photo database of "known cheaters"[Beiser99a].

#### 1.6.1.4 Iris

The human iris is unique to each individual and even one's left and right irises are not the same [Negin00]. Studies showed that the iris remains stable over decades of life, making it a very distinct biometric. Iris recognition technology involves the use of a camera to capture a digital image of the eye and process it to locate the iris and compute the iris code which is then compared with the data collected during enrolment. The initial available results on accuracy and speed of iris-based identification are promising and point to the feasibility of a large-scale recognition using iris information [Mai03][Zhu02]. However, the main issue is that iris scanning requires a certain amount of user participation since a user must stay still at certain spot during the process of data capture. Examples of iris recognition application include:

The installation of an ATM (Automatic Teller Machine) that includes iris scans as an alternative to passwords or PINs at the Bank United of Texas in may 1999 [Iris99].

Iris scan cards have been employed by the Schiphol Privium scheme at the Amsterdam airport to speed up the passport and visa control procedures. Passengers who are enrolled in the scheme insert their card at the gate and look into the camera, the camera acquires the image of the traveller's eye and processes it to locate the iris and compute the Iris code which is then compared with the data residing in the card to complete the user verification [CNN02].

### 1.6.1.5 Signature

Signatures have formed the basis of legal identification in documents and have been a means of proof of identity in financial transactions. The main attractiveness of signature recognition systems is that it is widely accepted by the public [Allgrove99]. However, they have several weaknesses: although each person has a unique style of handwriting, no two signatures of a person are exactly identical, they are a behavioural biometric that change over a period of time and are influenced by physical and emotional conditions of the signatories. Further, professional forgers may be able to produce signatures that fool the system.

Signature biometrics is often referred to as dynamic signature verification. With this technique, the manner in which someone signs is as important as the static shape of his/her finished signature. For example the angle at which the pen is held, the time taken to sign, the velocity and acceleration of the signature, the pressure exerted and the number of times the pen is lifted from the paper all can be measure and analyzed as unique behavioural characteristics.

In a signature recognition system, a signature data is captured via a special pen or tablet or both. The pen-based method incorporates sensors inside the writing instruments while the tablet method relies on sensors imbedded in a writing surface to detect the unique signature characteristics. When a person signs his or her name on the digitized graphics tablet, the system analyzes the signature dynamics such as speed, relative speed, stroke order, stroke count, and pressure. The signature dynamics information is then encrypted and compressed into a template.

Despite its user friendliness and lack of invasiveness, signature recognition has not yet dominated the market, like other biometric technologies (especially fingerprint recognition). Some documented applications include the Chase Manhattan Bank (the first known bank to adopt signature recognition technology) and the Internal Revenue service for verification purposes in tax returns that have been filed online [Das].

#### 1.6.1.6 Retinal Scanning

The retina scanning is regarded as highly unique since no two persons or even the same pair of eyes has the same web of capillaries running through the retina [Hill78]. Retina scanning is very accurate and has been used in very demanding authentication applications. However, its weakness is that it relies on a relatively complicated operation since the technology requires co-operative, well trained and patient users to stand close to the device and focus on a target while a low-intensity beam of light is shot into the eyeball to record the pattern of veins in the eye. This makes it unacceptable by the user. Besides that, due to its high cost and difficult sample collections, retina scanning is still highly referred to government use and the highest security situations. It has been used in prisons in both Pennsylvania and Florida in U.S for making positive identification of prisoners prior to release or transfer [Beiser99b].

#### 1.6.1.7 Hand Geometry

Hand geometry systems use an optical camera and light-emitting diodes with mirrors and reflectors to capture two orthogonal two-dimensional images of the back and sides of the hand.

Hand geometry systems are highly acceptable and have been widely deployed in various applications, such as access control and employee attendance applications [Sidlauskas88]. The main advantage of hand geometry is that it is not affected by dirt, cuts and dryness of the hand. However, it have a few drawbacks, one of which is the high possibility of some people having the same hand geometry as in the case of identical twins or between the same family members. Other disadvantages include the bulky size of the hand geometry devices, which makes them unsuitable for certain applications (such as laptop computers) and their expensive cost.

An example of hand geometry application is its use in 1996 Summer Olympic Games in Atlanta to identify approximately 150,000 athletes, staff and other participants [George96].

The brief overview given in the previous sections shows that there is not a single technology that out-performs all the others in all operational environments and that an application must be analysed in detail in order to select the most appropriate biometric to adopt. In this sense, each biometric technique is admissible and there is no optimal biometric characteristic. For example, it is well known that both the fingerprint-based technique and the iris-based techniques are more accurate than the voice based technique [Jain99]. However, in a tele-banking application, the voice-based technique may be preferred since it can be integrated seamlessly into the existing telephone system.

## 1.6 Limitations of any Unimodal Biometric System

The successful installation of biometric systems in various civilians applications does not imply that biometrics is a fully solved problem. Biometric systems that operate using any single biometric characteristic have the following limitations [Jain04a]:

***Non-universality***: As mentioned previously a human physiological or behavioral characteristic can be used as a biometric characteristic if each person possesses the required feature characteristic. In reality this is not possible since there is always a subset of users that are unable to enroll in any given system for different reasons. For example, people who are mute cannot use the voice system and people lacking fingers or hands from congenital disease cannot use fingerprints or hand geometry systems. It was recently reported by the National Institute of Standards and Technology that some 2% of the population is unable to provide a fingerprint sample suitable for enrolment into a typical biometrics system [NIST00] [Jain04b].

***Difficult replacement***: In some situations a biometric cannot be easily replaced. If a biometric is destroyed as result of a disease, surgery or injury, or stolen, it may not be replaced [Schneier99]. With a credit card, the bank can issue the user a new card with a new number. But a user has only a limited number of biometrics and they are not easy to replace.

***Easy to spoof***: Each biometric is subjected to attacks where an impostor will attempt to imitate the biometric characteristic of a legitimate enrolled user in order to circumvent the system.

***Lack of permanence***: The passage of time might give a rise to a situation where the biometric data acquired from an individual during authentication may be different from that used to generate the template during enrolment and hence affects the matching process. This variation may be a result of several reasons such as a fundamental change in the way in which the feature was presented to the device during authentication or damage to the feature over the period (e.g. scars on the finger).

***Noisy data:*** The data captured from a sensor might be noisy or distorted. Noisy data is a result of several factors one of which is the faulty or improperly maintained sensors (e.g. the accumulation of dirt or previous fingerprints on a fingerprint sensor). Another factor is the existence of unfavourable ambient conditions such as a poor illumination of a user's face in a face recognition system. An example of a noisy data is a fingerprint with a scar or a voice altered by a cold. This noisy data affects the performance of the system and can result in a user being incorrectly rejected.

***Non-acceptability***: Not all biometrics are highly acceptable by the public. For example, the fingerprint technology is not always highly acceptable because of its strong associations with the traditional identification of criminals, while in some countries women are not allowed to reveal their faces. In this case the face technology is not a good means of identification.

***Intra-class variations*** are generated when different biometric samples of the same feature are generated from the same person [Pankanti01]. This happens when the biometric data acquired from an individual during authentication is different from the data that was used to generate the template during enrolment thereby affecting the matching process. This variation is either caused by the incorrect interaction of the user with the sensor or as a result of modifying the sensor characteristics during the verification phase. Some intra-class variations are natural, for example two

signatures of a person are not always exactly identical. Intra-class variation results in a user being incorrectly rejected.

## 1.7 Multimodal Biometric System

Some of the limitations of a unimodal biometric system can be overcome by using multiple biometric modalities [Hong99]. This could be by using multiple sensors for the same biometric (e.g. optical and solid-state fingerprint sensors), multiple representations and matching algorithms for the same biometric (e.g. multiple face matchers like PCA and LDA), or multiple biometric traits (e.g. face and fingerprint). Using multiple sensors solves the problem of noisy data, but all the other problems associated with unimodal biometric systems remain. The multiple representation and matching algorithms for the same biometric improves the recognition performance of the system. However, all these methods suffer from many of the problems faced by unimodal systems. A multimodal biometric system based on different traits is expected to be more robust to noise, address the problem of non-availability or unreliability of any particular trait in a given situation, the non-acceptability of a particular trait for an individual user or user group, improve the matching accuracy and provide reasonable protection against spoof attacks. Hence, the development of biometric systems based on multiple biometric traits is adopted as a practical solution for many recognition applications.

## 1.8 Purpose of Research

The main aim of this thesis is to investigate the fusion of multimodal biometric verification system, which in turn lead to the evaluation of their performance on multimodal biometric systems. The evaluation of these systems raises a number of problems and challenges. One of which is the insufficient availability of multimodal databases representing the features of a large population. Another challenging problem is the fusion of the multiple modalities and the question of how should the outputs of the verification experts based on individual modalities be combined to achieve lower error rates and whether to combine the soft outputs or fuse the hard

decisions. Finally, the two most challenging problems in fusing multiple modalities are score normalization and setting the thresholds of both single experts and the fusion rule to achieve lower error rates. These are the principle issues, which will be addressed in this thesis.

## 1.9 Thesis Organization

This thesis consists of nine chapters, which are described, in more details below:

Chapter 1 is an introductory chapter, which sets the scene for the thesis.

Chapter 2 provides a review of research studies in the field of multimodal biometric systems regarding the fusion approaches and multimodal databases. The chapter gives an overview of the different architectures and different levels of data fusion suggesting using the parallel architecture and fusing the data at the decision level.

Chapter 3: describes the data collection exercise that was undertaken. The chapter describes the biometric devices that were chosen for the evaluation and the test protocol used to capture the biometric samples.

Chapter 4: describes the database formulated from the exercise that was carried out and provides a preliminary analysis of it.

Chapter 5: provides a brief review of the commonly used fusion rules at the decision level in multimodal person recognition systems. It also describes the most commonly used score normalization methods and proposes a novel method of score normalization. The chapter explains the experimental set-up used for calculating the error types and provides a comparison between the hard and soft decision rules when characterizing the system users as lamb, sheep, goats and wolves.

Chapter 6: gives a general overview of the genetic algorithm and their different parameters. The chapter discuss the reasons of using genetic algorithms in the field of biometric recognition as an optimisation technique instead of other techniques.

Chapter 7: proposes a novel approach based on the use of genetic algorithms to solve problems associated with score normalization and weights/ threshold settings.

Chapter 8: summarizes the work presented in this thesis and presents the main conclusions that have been drawn from the work. The chapter also suggests some future research.

## 1.10 Summary

In this chapter a brief overview of the field of biometrics was provided, summarizing the advantages and disadvantages of each biometric and pointing to the potential advantages offered by multimodal biometric systems. The chapter also presented the purpose of the research and the organisation of the thesis with a brief description of each chapter.

The next chapter provides a review of multimodal biometrics systems and states their challenges.

# Chapter 2

# Multimodal Biometric Systems Concepts

## 2.1 Introduction

In this chapter the multimodal fusion approaches and databases that have been explored in recent research studies are briefly reviewed. The issues and challenges of these fusion techniques are presented and the publicly available medium and large-scale multimodal databases are described stating their limitations.

## 2.2 Information Fusion in Biometrics

As mentioned in the previous chapter, recognition based on any modality alone may not be very robust whilst fusing information from a number of different biometric modalities may well provide higher and more consistent performance levels.

Information fusion is a term that refers to any area exploiting the combination of different sources of information, either to generate one united representational format, or to reach a decision [Barvin81]. This includes areas such as: team decision

theory, integration of multiple sensors, multi-modal data fusion, combination of multiple classifiers and distributed decision making.

There are several advantages in using information fusion to reach a decision, such as:

- By using complementary information (e.g. fingerprint, audio and video) the error rates can be reduced.
- Robustness and reliability. The system is operational even if one or several sources of information are missing or malfunctioning.
- Using several cheap sensors rather than one expensive one can reduce the cost of implementation.

Since the aim of this research was to combine information from different classifiers of multiple biometric devices, different strategies for combining multiple classifiers was investigated. There is a large number of combination methods reported in the literature [Canuto00] [Fairhurst97] [Kittler98] [Rahman99] [Xu92]. In the following sections the architecture/topology of the classifiers and the different levels of information fusion are reviewed.

## 2.3 Multiple Classifiers System Architectures/Topologies

The architecture of a system describes the way the components are organized within the system. There are different architectures/topologies for combining classifiers; in this section the two basic ones [Dasarathy94] are discussed:

**Serial topology**

As shown in Figure 2.1, the *serial* classifier architecture consists of a set of $m$ classifiers whose decisions are combined in series or tandem [Roli02]. This architecture is well suited to deal with situations where the different classifiers have a ternary {accept, reject, undecided} decision scheme [Kamel03]. A scheme in

which the classifiers cannot decide on the input pattern they are presented with. If the current classifier is undecided, information is passed to the next expert in the sequence. The inputs represent the feature sets that help the classifiers in making a decision. For this serial scenario to be effective, the classifiers have to have a varying ability of generalization. This architecture is suitable for combining decisions from classifiers with varying ranges of effectiveness and modelling sequential decision refining from one sensor to the next.

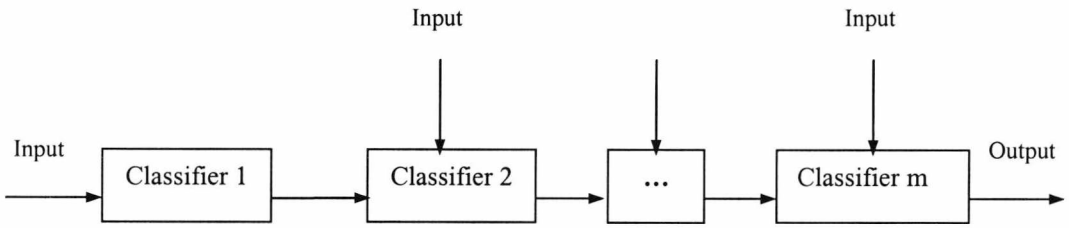



**Figure 2.1**: The serial multi-classifier architecture

**Parallel topology**

As shown in Figure 2.2, the *parallel* classifier architecture consists of a set of *m* classifiers that are consulted in parallel. The decisions of the various classifiers are combined in parallel by the combining/fusion module. This architecture is suitable for combining decisions or scores from classifiers that are capable of operating simultaneously and independently of one another.

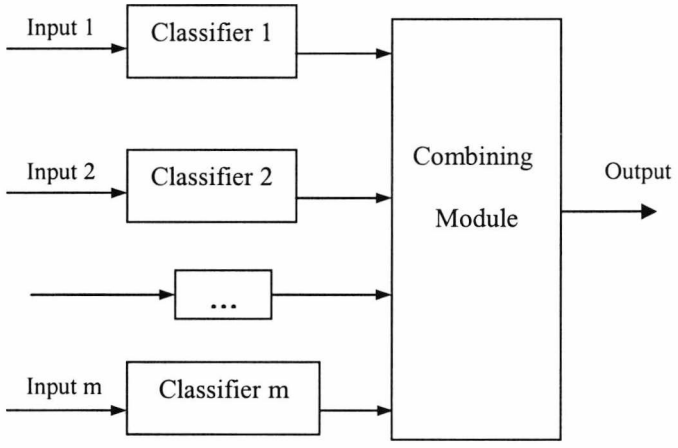



**Figure 2.2**: The parallel multi-classifier architecture.

There are also hybrid combinations of these two basic schemes that can be used, such as parallel-serial or serial-parallel architectures, which are also referred to as layered architecture. These combinations are more complex than the previous two and fall outside the scope of this work.

The choice between the two basic architectures described was not only based upon the descriptions presented above, but also on the following reasons:

1. So far research on multiple classifier systems has principally focused on the parallel architecture and it has been extensively applied to the field of pattern recognition [Chibelushi99] [Fairhurst97].

2. General methodologies and clear foundations are mostly available for parallel architecture.

3. The parallel architecture is less complex than the serial one.

4. As a serious drawback, any serial network is vulnerable to link failure.

Taking into account the descriptions of the basic architectures and the reasons mentioned above, a parallel architecture was adopted for this work.

## 2.4 Fusion Levels in Biometrics

A biometric system has four important modules. The sensor module that acquires biometric data from a user; the feature extraction module processes the acquired biometric data and extracts a feature set to represent it; the matching module compares the extracted feature set with the stored templates using a classifier or matching algorithm in order to generate matching scores; in the decision module the matching scores are used either to identify an enrolled user or verify a user's identity.

Multimodal biometric systems that verify a user's identity are categorised into four system architectures according to the strategies used for information fusion [Ross01]:

- Fusion at the Sensor Level
- Fusion at the Feature Extraction Level
- Fusion at the Matching Score Level
- Fusion at the Decision Level

That is, the systems are classified depending on how early in the authentication process the information from the different biometric sensors is combined. Biometric authentication is a chain process, as illustrated in Figure 2.3.

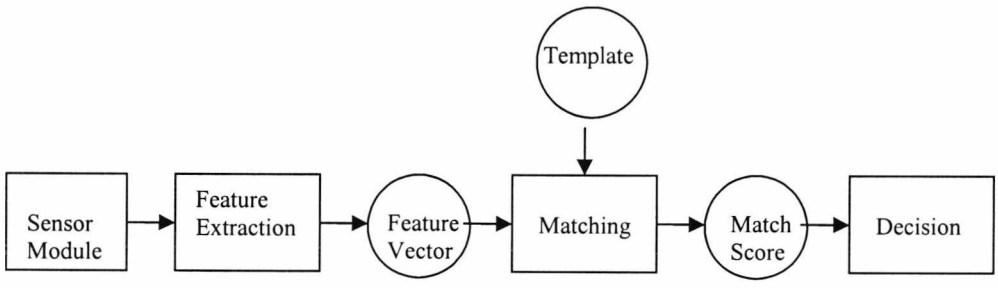



**Figure 2.3**: The authentication process chain

Fusion at the feature extraction level stands for immediate data integration at the beginning of the processing chain, while fusion at the decision level represents late integration at the end of the process.

The following subsections describe each of these levels in detail and a review on related research activities.

### 2.4.1 Fusion at the Sensor Level

In this architecture, Figure 2.4, the raw data streams coming out of different sensors are combined. To accomplish this combination there are two main methods, but these depend on the application it is used for. For example, the weighted summation rule can be used to combine the data from two microphones (to reduce the noise), while mosaic construction can be used to generate one image out of images provided by several cameras each looking in different parts of the same object [Hong98]. In sensor level fusion, the data obtained from the different sensors must be compatible, and this may not always be possible (e.g., it may not be possible to fuse face images obtained from cameras with different resolution).

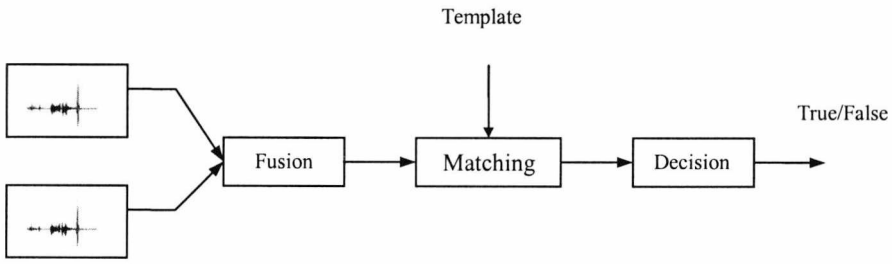



**Figure 2.4**: Fusion at the Sensor level

An extensive literature search did not reveal any significant recent research on this fusion strategy. This suggests that fusion at the sensor level is much less preferable than the other strategies.

### 2.4.2 Fusion at the Feature Extraction Level

In this architecture Figure 2.5, the information extracted from the different biometric sensors is encoded into a joint feature vector, which is then compared to an enrolment template (which itself is a joint feature vector stored in a database) and then a decision is made.

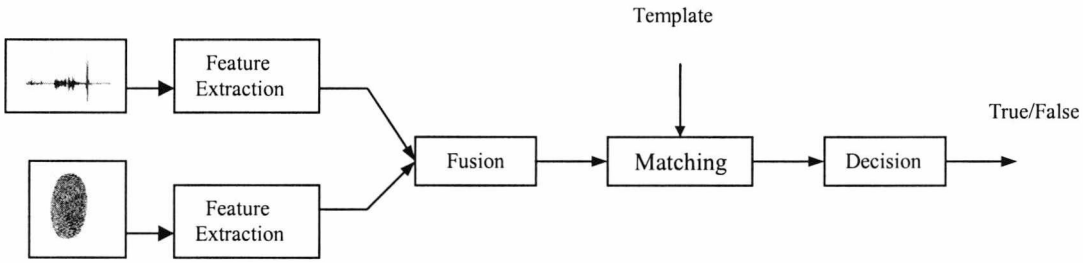



**Figure 2.5**: Fusion at the feature extraction level

Fusion at the feature level is often difficult because the feature sets used by different biometric modalities may either be inaccessible or incompatible. There are two methods used for combining the extracted features, but these certainly depend on the features themselves. If the features are commensurate, i.e. having a common measure, the combination can be achieved by the weighted summation rule. If the features are not commensurate then a simple fusion scheme consisting of concatenating the feature vector is employed [Brooke94].

Fusion at the feature level has been presented by some researchers such as Luettin, Marcel and Sanderson and Kumar.

Luettin in [Luettin97] combined speech and (visual) lip information using feature vector concatenation. In order to match the frame rates of both feature sets, speech information was extracted at 30 fps instead of the usual 100 fps. In text-dependent configuration, the fusion process resulted in a minor performance improvement,

however, in text-independent configuration, the performance slightly decreased; suggesting that feature vector concatenation in this case is unreliable.

Marcel et al in [Marcel02] proposed to use the skin colour as an additional feature to the face image. The verification method is based on Multi-Layer Perceptrons. For each client, an MLP is trained to classify an input to be either the given client or not. The input of the MLP is a feature vector formed by the concatenation of the face feature vector with the skin colour vector. The output of the MLP is either a client or an impostor. Experiments were carried out on the XM2VTS database and the results show good improvement when using the skin colour information.

Sanderson et al in [Sanderson02] evaluated the performance of feature vector concatenation fusion and several non-adaptive opinion fusion methods such as the weighted summation fusion, Bayesian and SVM post-classifiers, for combining face and speech information under the presence of audio noise. Experiments were conducted on the VidTIMIT database. The results showed that the performance of the feature concatenation fusion approach was relatively more robust than the three post-classifier approaches. However, for most SNRs the performance was worse than the face expert, suggesting that while in this case feature concatenation fusion is relatively robust to the effects of noise, it is not optimal.

Kumar et al in [Kumar03] described a hand based verification system that combines the geometric features of the hand with palmprints at the feature and match score levels. Experiments were conducted on 100 users. Interestingly, in their experiments, fusion at the match score level resulted in a better performance than fusion at the feature level.

In addition to the fact that was revealed in the literature suggesting the unreliability of feature concatenation, there were two other reasons that limited the use of both the data and the feature fusion in this work and these were:-

1. In this work it was desired to explore the use of multiple modalities which means it was not possible to fuse the raw data.

2. Since it was desired to explore the use of commercial off the self-devices the information available regarding the algorithms used was inadequate, which limited the use of the features.

3. The aim of this work was to separate the design of the specialized classifiers, which is very application dependent from the fusion problem.

### 2.4.3 Fusion at the Matching Score Level

In a multimodal biometric system built on this architecture, Figure 2.6, the feature vectors are created independently for each sensor and then compared to the enrolment templates, which are stored separately for each biometric trait. Based on the proximity of feature vector and template, each subsystem now computes its own matching score. These individual scores are finally combined into a total score, which is handed over to the decision module.

Fusion at the matching score level is generally preferred due to the ease in accessing and combining the scores. Different strategies are used to combine the scores. They range from a simple sum rule to sophisticated statistical methods.

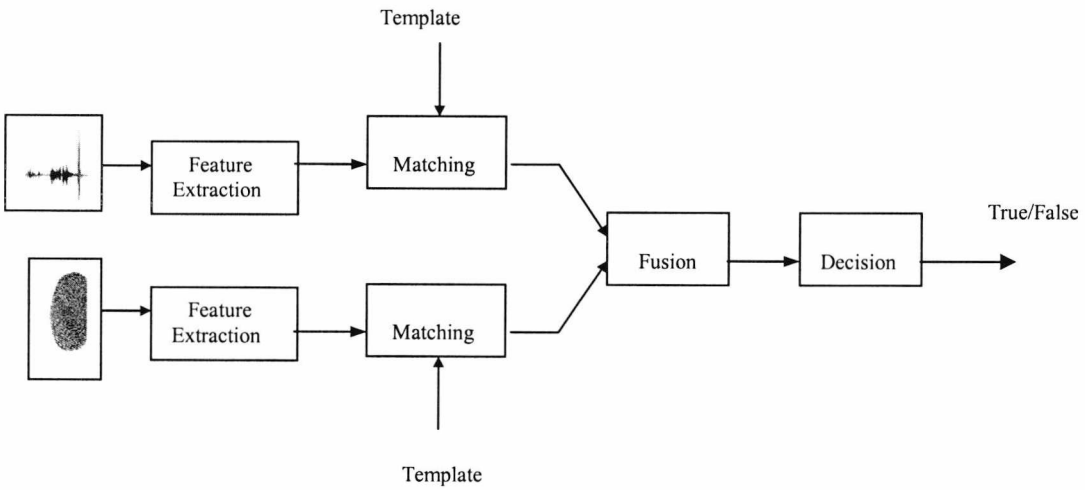



**Figure 2.6**: Fusion at the matching score level

Fusion at the matching score level has been presented by the majority of researchers to enhance recognition accuracy and improve the robustness of the system.

Kittler et al in [Kittler98] proposed a multimodal person verification system using three experts: frontal face, face profile and voice. The frontal face expert is based on template matching, the face profile expert used a chamfer matching algorithm and the voice expert was based on the use of text-dependent person dependent HMM models for isolated digits. The outputs of all the three experts were normalized scores (scores between zero and one). The authors conducted a comparative study of the performance of several combination schemes namely; the product rule, sum rule, min rule, max rule and majority voting. By assuming the joint probability distributions to be conditionally independent, the outcome of the comparative study showed that the sum rule outperformed the other combination schemes with an EER of 0.7 %.

Jain et al in [Jain99] developed a multimodal biometric system that uses three classifiers: face, fingerprint and speech. The scores from the modalities were combined using the product rule. Experiments were conducted on a database of 50 users, which was acquired in a laboratory environment, the results obtained demonstrated that the overall system performance improves by integrating multiple biometric indicators.

Ross et al combined in [Ross01] the matching scores of three modalities (Face, Fingerprint and Hand geometry) to enhance the performance of a biometric system. Three different techniques (Sum rule, decision tree, linear discriminant analysis) were used to combine the matching scores. Experiments indicated that the sum rule with normalized scores resulted in the best performance.

Roli et al. reported in [Roli02] an experimental comparison between fixed and trained fusion rules on a multimodal person-identity verification task, involving two basic modalities: speaker voice and frontal face image. The experiment used five fixed fusion rules (sum, majority vote and three rules based on order statistics

operators (OS)), and two trained rules (Behavioural Knowledge Space and the weighted average method). The experiments were carried out on the XM2VTS database. The results showed that the trained rules in particular the weighted average method provided significant improvements over the fixed rules when they were trained on the test set, this means that the advantages of the trained rules depends on the quality and the size of the training set. The results also showed that among the trained rules, the weighted average method outperformed slightly the BKS rule, while among the fixed rules, the vote rule exhibited good performance. In contrast, the effectiveness of OS rule appeared to be poor.

Snelick et al [Snelick03] developed a general testing framework that allows system designers to evaluate multimodal biometric systems by varying different factors such as the biometric modalities, normalization schemes, fusion methods and sample databases. The authors illustrated their testing methodology by evaluating the performance of a multimodal biometric system that used face and fingerprint classifiers. In this paper several normalization techniques like min-max, z-score, median and MAD, and tanh estimators were used to transform the scores into a common range. The normalized scores were then combined using fusion methods like simple sum of scores, maximum score, minimum score, sum of probabilities and product of probabilities. Their experiments showed that the min-max normalization followed by the sum rule fusion method provided better recognition performance than the other schemes. The results also show that multimodal biometric systems out perform single-mode biometric systems.

Wang et al in [Wang03] designed an identity verification system based on the fusion of face and Iris data. Two different fusion strategies were used. The first strategy computed the weighted and the unweighted sum and compared the result to a threshold. The second strategy treated the matching distances of face and iris classifiers as a two-dimensional feature vector and used both the fisher's discriminant analysis and the neural network with radial basis function (RBFNN) to classify the vector as being genuine or impostor. Results showed that the fusion based on the RBFNN produced the highest verification accuracy and that the

weighted sum rule is the best approach when compared with the sum rule and Fisher rule.

The literature revealed an increase in the performance of the system by fusing the output scores of the different classifiers, however it highlighted the problem of selecting a normalization method that maps these scores into a common interval [0, 1] before fusing them. The literature also revealed that weighting varies the importance of matching scores of each modality, thus increasing the system performance.

### 2.4.4 Fusion at the Decision Level

In this fusion strategy, a separate authentication decision is made for each biometric modality. These decisions are then combined into a final vote, as shown in Figure 2.7:

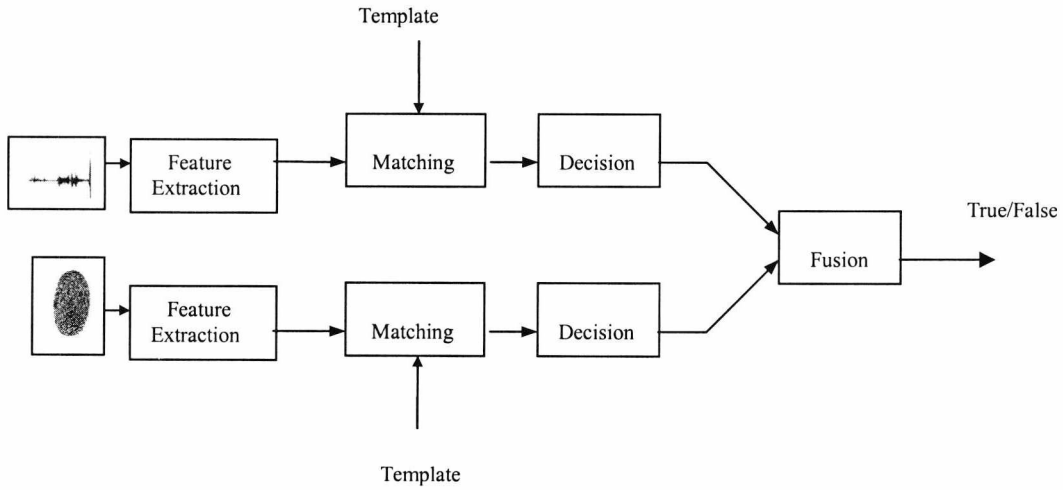



**Figure 2.7**: Fusion at the decision level

Fusion at the decision level is too rigid since limited amount of information is present at this level. The most common strategies for combining the distinct decisions into a final authentication decision are the voting techniques (AND, OR, Majority Voting).

This fusion strategy has been presented by a small number of researchers such as Diekmann and Hong.

Dieckmann et al in [Dieckmann97] proposed a decision level fusion scheme, which integrates face and voice data that are analysed by three different classifiers: face, lip motion and voice. For both cases of recognition (identification and verification) a field test was done where 66 individuals participated in it. The decision was made when two of three (majority voting) of the classifiers results lead to the same person and exceeds a given threshold. The experiments showed that the recognition rate using the majority-voting rule is higher than the recognition rate in any single modality alone.

Hong et al. presented in [Hong99] ways of combining information from two modalities: face and fingerprint images at various levels. Two levels of fusion were considered; score-level fusion, where the Bayesian method was used and a decision-level fusion, where both the OR and AND rules were used. Experimental results showed that the performance of a biometric system was improved by integrating multiple biometrics than by using either the finger or the face alone.

The literature revealed that combining the decisions using the voting techniques was the preferred method among researchers and that combining multiple modalities increase the performance of the system.

Several different names have been given to the mentioned fusion levels by researchers. Sanderson et al [Sanderson04] have classified information fusion in biometric systems into two broad categories: pre-mapping fusion and post-mapping fusion. Silsbee in [Silsbee96] referred to pre-mapping fusion and post-mapping fusion as pre-categorical integration and post-categorical integration, respectively, while Wark in [Wark00] referred to the terms as input level or early fusion and classifier level or late fusion, respectively. However, in this work it was decided to classify the information fusion into two main levels: feature fusion level and decision fusion level where the decision fusion is sub-divided into hard decision fusion and soft decision fusion. Hard decision fusion is a decision made by the system that returns either a 0 (reject) or a 1 (accept), while soft decision fusion is a

decision made by the system that returns a score that normally lies in the [0,1] interval. Figure 2.8 shows the two levels of fusion and the method to accomplish the fusion at each level. As previously mentioned the fusion at the feature level was not considered for the reasons mentioned earlier, the decision level on the other hand was the one used in this work which will be explained in details later in this thesis.

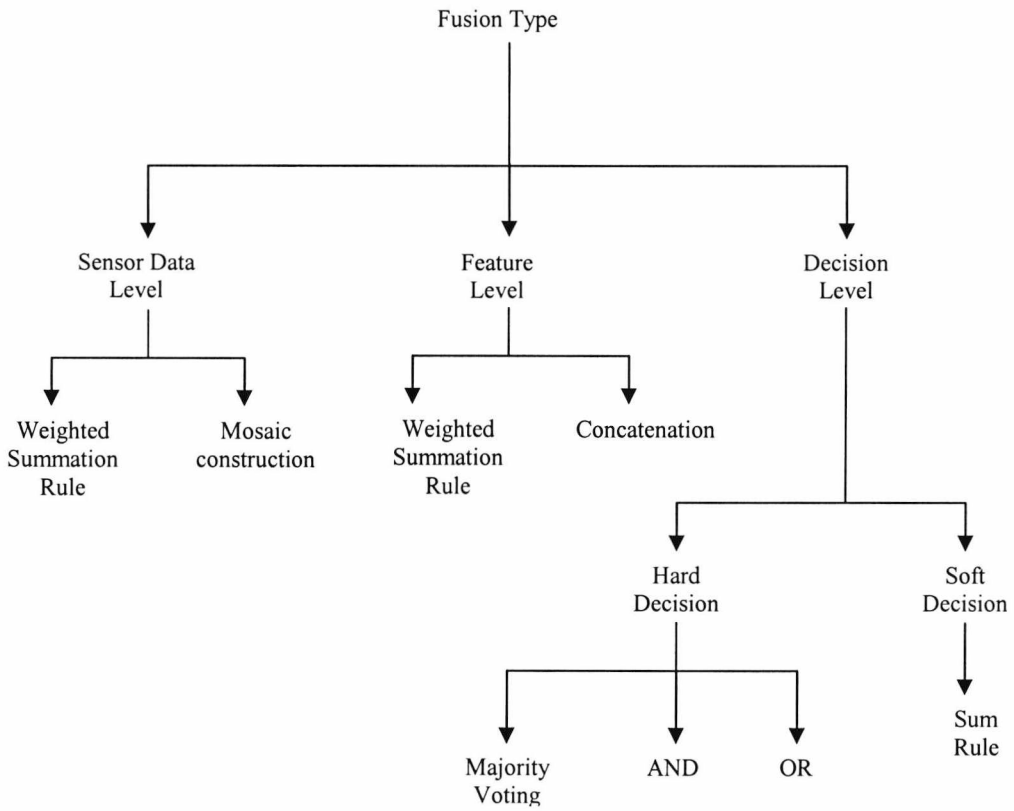



**Figure 2.8**: The hierarchy of fusion types

## 2.5 Multimodal Biometric Databases

One of the important factors in evaluating the performance of automatic recognition systems based on the biometric characteristics of individuals, in both identification and verification mode, is the availability of a large multimodal biometric database acquired under real conditions for testing the algorithms. The main problem involved in the development of a multimodal biometric database is the availability

of a large number of individuals concerned in offering its biometric features. Often, the acquisition of the biometric features is accomplished in different moments and in different conditions, which suppose high degree of collaborativeness for the participants. For that reason, the number of existing public databases for the performance evaluation of recognition systems based on multiple biometric modalities is quiet limited. At the present there are quiet a few medium and large-scale multi-modal databases and these are:

The BT-DAVID (British Telecommunication-Digital Audio-Visual Integrated Database) audio-visual database contains full-motion video, showing a full-face and a profile view of talking subjects, together with the associated synchronous sound [Chibelushi96]. It includes audio-visual material from more than 100 subjects including 31 clients recorded on 5 sessions spaced over several months. The utterances include the English digit set, English alphabet E-set, vowel-consonant-vowel syllables, and phrases for the control of a video-conferencing session. The scenes include variable scene background complexity and illumination. Portions of the database include lip highlighting.

The M2VTS (Multimodal Verification for Teleservices and Security Applications) database is another multimodal database that contains audio-visual material from 37 different subjects [M2VTS]. It provides 5 shots for each person, the shots consist of the registration of audio and video of the person counting from 0 to 9 and rotating the head in the sequence [0, -90, 0, 90] degrees.

The VidTIMIT database comprises video and audio recordings of 43 volunteers (19 female and 24 male), reciting short sentences. It was recorded in 3 sessions, with a mean delay of 7 days between Session 1 and 2 and 6 days between Session 2 and 3. For the audio 10 sentences were chosen from the test section of the NTIMIT corpus [Jankowski90] for each person. The first six sentences were assigned to Session 1. The next two sentences were assigned to Session2 with the remaining two to Session 3. The first two sentences are the same for all the volunteers, with the remaining eight generally different for each person. For the face, each person performed an extended head rotation sequence in each session, which allows for extraction of

profile and 3D information. The sequence consists of the volunteer moving his/her head to the left, right, back to the centre, up, down and finally return to the centre.

Another important resource available nowadays is the extended M2VTS (XM2VTS). This database contains audio-visual material from 295 subjects [Messer99][Cheung04] taken over a period of 4 months. On each visit (session) two recordings (shots) were made. The first shot consisted of speech whilst the second consisted of rotating head movements. This database includes high quality colour images, 32 KHz 16-bit sound files, video sequences and a 3D Model.

The BANCA database was captured in four European languages: English, French, Spanish and Italian in two modalities (face and voice) [BANCA]. For recording, both high and low quality microphones and cameras were used [Bailly-Bailliere03]. The subjects were recorded in three different scenarios, controlled, degraded and adverse over a period of three months. In total 208 people were captured, half men and half women.

The MCYT database is a large bimodal database that contains fingerprints and signatures of 330 different subjects [MCYT]. It includes a significant number of samples of each modality, under different levels of control to cope with the inherent variability of each feature at the acquisition process. The fingerprint database contains 79200 fingerprint samples acquired from 330 individuals, for each individual ten-print fingerprint, 12 samples of each fingerprint are acquired using two different sensors (optical and capacitive). The signature database on the other hand, contains 16500 signature samples, where 25 client signatures and 25 highly skilled forgeries (with natural dynamics) were obtained for each individual. Both on-line information (pen trajectory, pen pressure and pen attitude) and off-line information (image of the written signature) are considered in the database. A full description of the algorithms used for enrolment and verification can be found in [Ortega-Garcia02] [Ortega-Garcia03].

Finally, BIOMET is a large database that contains five different modalities: audio, face images (3 cameras), hand image, fingerprint and on-line signature. For the face images, a camera prototype designed to suppress the influence of the ambient light,

a 3D acquisition system prototype, and a standard digital camera were used. Three different sessions, with three and five months spacing between them, were realized. For the video sequences and face images, the persons were asked not to take off their glasses. The number of persons participating in the collection of the database was 131 for the first session, 106 for the second, and 92 for the last one. The proportion of female and male subjects was balanced in all sessions. 10% of people enrolled were students (with a mean age of 20), others' age varies from 35 up to 60 years.

Most of the research that has been done on multimodal biometric system was carried out mainly on the M2VTS database and the XM2VTS database. The multimodal biometrics experiments mentioned in [Duc97] [Duc97a] [Jourlin97] [Kittler98] [Ben-Yacoub98] [Pigeon99] [Kittler02] [Messer99] [Bengio01] [Bengio02] have all been carried out over these two databases.

The BT-DAVID and the M2VTS are medium size databases. The VidTIMIT database apart from its medium size it is not publicly available, it is only licensed for employees of IDIAP. The XM2VTS is a large database, however, it was not possible to use it for two reasons, first because the controlled recording environment was not realistic enough compared to the real world situations, such as making a transaction through an ATM in a variety of surroundings and second because the database consist of audio and video material and it was desired to evaluate the performance of a system with more than two biometric modalities. The BANCA is also a large database, however, it was not used because it was not available at the start of this research and also because it consists of two modalities (face and voice) and as mentioned it was desired to evaluate the performance of more than two biometric modalities. The MYCT is a large database but it will be publicly available in January 2005. The BIOMET on the other hand, is a large database with multiple modalities. The reason for not using it is because it was not available at the start of this research.

## 2.6 Summary

A review of research studies in the field of multimodal biometric systems has been presented regarding the fusion approaches and the multimodal biometric databases. The literature demonstrated that the overall performance of the system can improve by integrating multiple biometric systems than by using a single biometric alone. A brief overview of the fusion approaches used in combining multimodal biometric system revealed the obvious preference of combining multiple biometric at the decision level (both the score level and the decision level) than combining them at the feature level. The literature also highlighted the problem of selecting a normalization method that maps the output scores of different classifiers into a common interval [0,1] before fusing them.

An overview of the publicly available medium and large-scale multimodal databases was provided and the reasons for not using them in this work were presented.

The next chapter describes the data collection exercise that was undertaken at the University of Kent as part of this research.

# Chapter 3

# Data Collection Exercise

## 3.1 Introduction

As stated previously the limited number of available multimodal biometric databases for evaluation introduced the idea of developing a multimodal biometric database. In this chapter a data collection exercise to acquire data from a range of biometric devices was undertaken. The chapter also describes the biometric devices that were chosen for the evaluation. A description of the test protocol used to capture the biometric samples and to evaluate a set of biometric measurements is also presented.

## 3.2 Biometric Devices Selection

Biometrics encompasses a wide range of techniques based on a variety of physical or behavioural personal characteristics [Fitzegerald89] [Miller94]. Examples of physical characteristics are face, fingerprints and hand geometry. Typical behavioural characteristics are voice and hand-written signature dynamics.

Biometrics systems based on physical characteristics are generally more intrusive than behavioural-based systems. However, the latter are more error-prone than the former owing to the time variations exhibited by behavioural characteristics. Beside

temporal variation considerations, other factors affecting the selection of either physical or behavioural characteristics are cost, size, user-friendliness and reliability of the data capture equipment [BWG02]. None of the two categories of biometric characteristics wins on all fronts, as a result, the choice of either approach is often application-driven. It was decided to combine the use of physiological and behavioural biometric for grater quality, thus we chose to use the three modalities: fingerprints, face and voice [Wayman00]. Fingerprint modality was chosen for its long track record of reliability, face and voice modalities were chosen for the fact that these are part of the natural human messaging modalities [Brunelli95], and their hardware are cheap (microphone and camera), to the extent that some of them are offered as standard accessories of personal computers and workstation.

After selecting the modalities to be used, the next step was to look into the suitable systems available in the market. The aim was to find relatively cheap commercial off-the-shelf, user-friendly and reliable software. This is because using stated system successfully would demonstrate the usefulness of multimodal approach. After a long investigation it was decided to use the following software as shown in Table 3.1

**Table 3.1** The software used for the project

| Modality | Software |
|---|---|
| Face | FaceIt |
| Voice | VeriVoice |
| Fingerprint | SecuGen |

FaceIt was chosen for its ranking as the world's most advanced face recognition engine, Verivoice was chosen because it was freely provided by the vendors to be used in this research and Secugen was chosen for its user-friendliness and simple way of usage. The description and operation of each of the software used is given below.

### 3.2.1 FaceIt

FaceIt is developed by Visionics Corporation [Visionics]. At the time of this research it was claimed to be a highly commercial face recognition engine. Its software development kit can be used to perform both face identification (1-to-many searches) and face verification (1-to-1 matching).

The system works by analysing particular features of the face, such as the distance between the eyes and the nose, and the shape and location of the cheekbones [Visionics]. Skin colour and gender are not factors in the process, and the technology is designed to compensate for glasses, hats and beards.

**Figure 3.1:** FaceIt SDK detects human faces
by finding the area enclosed by the circle

The enrolment process is quick and fairly simple, it lasts about 12 seconds. Users pose in front of the camera until the window on the computer screen shows an acceptable image of the entire face. Then the process of capturing the images for enrolment begins. During this phase, users are instructed to vary the angle of the face slightly. The camera rapidly takes nine images and displays each of them in the enrolment window. The size of each template is quiet large, which is roughly between 3 to 4 kilobyte. The system then creates a facial template and enrolment is complete.


**Figure 3.2:** The face enrolment process

Face identification and verification is carried out based on the degree of similarities between the tested template of the user and his reference templates stored in the database. The process of verification lasts about 10 seconds. The decision threshold can be varied depending on the required degree of accuracy

The main advantage of FaceIt software over other face recognition engines is that it allows for tests of "liveness" of its captured images. This is particularly important in order to avoid false acceptance by attempts to use still photographs presented in front of the video camera.

### 3.2.2 VeriVoice

The VeriVoice software development kit is used for user verification, which is based on 1-to-1 matching [Verivoice]. It operates based on voice recognition and is designed particularly for access control applications, such as access to financial databases, computer networks, research facilities and other controlled environments.

VeriVoice software is restricted to text dependent samples. The enrolment process lasts about 3 minutes. The software prompts the user to repeat 12 different samples read from the given predefined texts. One reference template of size 16 Kbytes is then derived from the information provided by the twelve captured samples for each user.

In the verification process, the user is prompted to repeat a string of random digits. The software then prompts a score stating whether the person has passed or failed the process. This verification process takes less than one second.

The advantage of this software is that prompting for a randomly generated sequence of numbers during the verification process enhances security by eliminating hacking with digital recorders.

### 3.2.3 SecuGen

SecuGen software development kit is a biometric tool based on fingerprint recognition using minutiae matching [Secugen]. It is capable of performing both user identification and verification tasks.

The software can only be used with its own mouse, which has a fingerprint platen on its left side where the user places his/her thumb as illustrated in the figure below

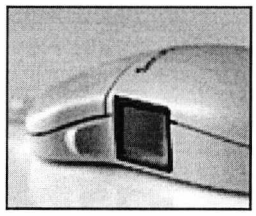


**Figure 3.3:** SecuGen EyeD Mouse

In the enrolment process, the user places his/her thumb on the mouse. The device sensor scans the user's finger and captures the live, 71 Kbytes fingerprint image. A series of algorithms developed by SecuGen extracts minutiae points from the image and converts the data into a unique mathematical template. This unique template, which is 400 bytes long, is then encrypted and stored to represent the user.

For verification, an enrolled user states a claimed identity (i.e. enters a user ID) and places his/her finger on the device sensor. A new fingerprint image of the user is captured. Minutiae data is extracted from the fingerprint and converted into a template. This template is then compared to the user's pre-enrolled template for a match. If the templates match, the user is verified positively. This process takes roughly about one second.

This software has two advantages. Firstly, fingerprint images are never stored. When a fingerprint is captured, only a portion of the minutiae are sampled and then processed by an extraction algorithm and converted into a secure template. After the template is formed, the fingerprint image is deleted. All fingerprints are used in the form of templates enrolment and matching. Secondly, Fingerprint images cannot be reconstructed from minutiae or templates. The minutiae sampled from a fingerprint do not have enough information to recreate an image of the fingerprint. Additionally, minutiae cannot be extracted from a template because the mathematical conversion from minutiae to template is irreversible. As a final measure of security, templates are secured using advanced encryption to prevent data from being "hacked".

## 3.3 Test Protocol

After selecting the biometric devices to be used, the next task was to define a test protocol for conducting technical testing in order to capture biometric samples from a set of users and to evaluate the set of biometric measurements. It was decided to use the emerging guidelines in "Best Practices in Testing and Reporting Performance of Biometric Devices" [BWG00] for the data collection exercise.

The first step was to decide whether the biometric authentication would be verification or an identification process. Due to the biometric software devices, the verification process was chosen.

To form the basis for developing an appropriate test protocol that specifies the appropriate environmental controls, volunteer selection and test size, the choice of an evaluation type had to be determined. There are three basic types of evaluation of biometric systems [Phillips00] [BWG00] [Court03]:

- **Technology evaluation:** The goal of this evaluation is to compare competing algorithms from a single technology. Testing of all algorithms is carried out on a standardised database collected by a "universal" sensor.

Nonetheless, performance against this database depends upon both the environment and the population in which it is collected.

- **Scenario evaluation:** The goal of this evaluation is to determine the overall system performance in a prototype or simulated application. Testing is carried out on a complete system in an environment that models a real-world target application of interest. Each tested system will have its own acquisition sensor and so will receive slightly different data. Care is required that the data collection across all tested systems is in the same environment with the same population.

- **Operational evaluation:** The goal of this evaluation is to determine the performance of a complete biometric system in a specific application environment with a specific target population.

Since the principal goal of this research was to evaluate and test the biometric modalities in an environment that models a "real-world" application rather than testing algorithms or determining the performance of a biometric system on a specific application with a specific population, the scenario evaluation was chosen for the exercise.

## 3.4 Modelled Scenario

The scenario modelled for the exercise is that of verification in which a single attempt is matched against a single stored template. The use of each biometric technology has its strengths and weakness depending upon the application in which it is used. Although each use of biometric is clearly different, some striking similarities emerge when considering applications as a whole. All applications can be partitioned according to at least seven categories. [Wayman98] [Wayman99] [EWA01]. The seven categories that suited the exercise are identified and these are:

**Cooperative users versus Non-cooperative**

This refers to the behaviour of the deceptive user (impostor). In "verification" applications, the user claims an enrolled identity; this means that the deceptive user is cooperating with the system in an attempt to be recognized as someone s/he is not. This is called "cooperative" application. In "identification" application, the user makes no claim to identity, thus requiring the search of the entire enrolled database. This is called "non-cooperative" application.

**Overt versus Covert**

If the user is aware that a biometric identifier is being measured, the use is overt. If unaware the use is covert.

**Habituated versus Non-habituated**

This applies to the intended users of the application. Users presenting a biometric trait on a daily basis are considered habituated. Users who have not presented the trait recently are considered "non-habituated".

**Attended versus Non-attended**

This refers to whether the use of the biometric devices during enrolment will be supervised and guided by a supervisor or not.

**Standard versus Non-standard environment**

If the application will take place indoors at standard temperature and other environmental conditions, particularly where lighting conditions can be controlled, it is considered a "standard environment "application. Outdoor systems are considered "non-standard environment" applications.

**Public versus Private**

This refers to the users of the system if they are members of the general population (public) or employees (private).

**Open versus Closed Usage**

If the system will be required to exchange data with other biometric systems run by other management, then it is open. Otherwise, it is closed.

The exercise to be undertaken is classified as a cooperative, overt, supervised, non-habituated, standard environment, public, closed application. Cooperative because those wishing to defeat the system will attempt to be identified as someone already in the system. It is overt because all volunteers will be aware that they are required to give a biometric measure during enrolment and verification transactions. It is supervised and in a standard environment because collection of the biometrics will take place in a normal office environment and under the supervision of a supervisor. It is non-habituated because the separation between enrolment and verification transaction is one to two months, so the level of habituation will be quite low. It is public because the trial is open to 200 volunteer from different gender and different age. It is closed because we will not exchange the biometric information gathered with any other systems.

## 3.5 Device Set-up

Before recruiting people for the exercise a set up of the devices was necessary. As already mentioned the enrolment and testing procedure was to be conducted indoors in a standard office environment (3m by 4m room) designated as the "biometrics laboratory". Figure 3.4 shows the distribution of the biometric devices within the biometrics laboratory. As can be seen from the figure, two PC machines are used, One supporting the fingerprint device and the other supporting both the voice and the face devices.

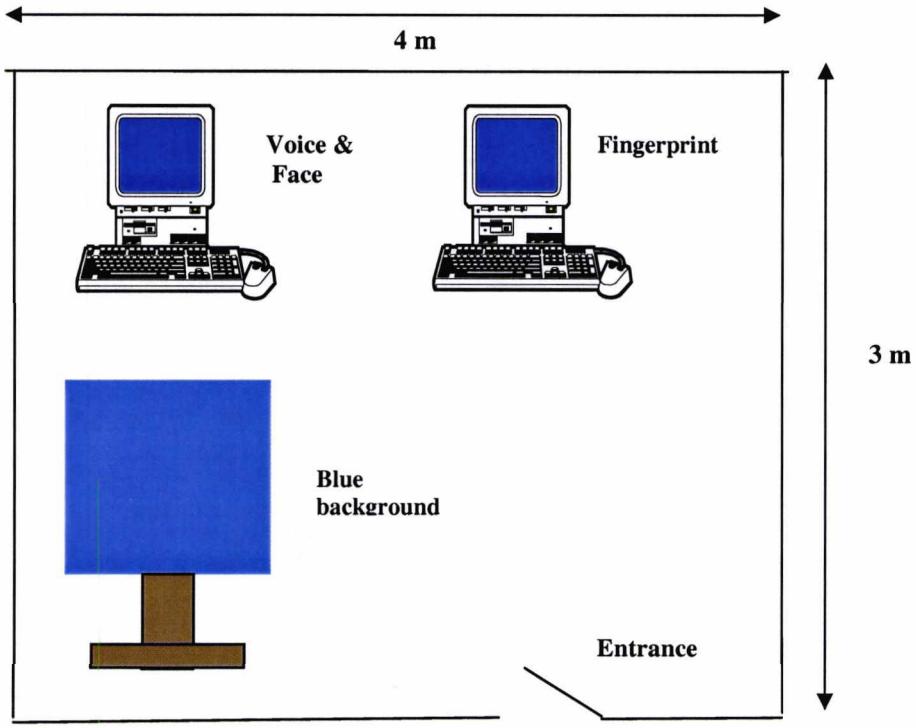



**Figure 3.4:** Layout of the biometric devices within the biometric laboratory

Both PCs were Pentium III with 256 MB RAM and a 20 GB of hard disk space. For face enrolment and verification a standard webcam camera was placed on the PC to capture human faces. A plain blue background was placed behind the sitting subject. The decision threshold of the FaceIt software ranges from 0.0 to 10.0. The manufacturer default value 8.7 was set for the trial.

For voice enrolment and verification a desktop microphone was used to capture voice samples from users. Speakers were also used in order to read out instructions to users. The decision threshold of the VeriVoice software ranges from –1476 to 323. For enrolment and verification the manufacturer's default value 0 (zero) was set.

The Fingerprint decision threshold ranges from 0 to 9. For enrolment and verification it was set to its manufacturer default value 5.

All vendor recommendations regarding positioning, illumination and background noise were taken into consideration. Some pre-trial tests using the testing team were

carried out to determine environmental and other factors that may cause problems and to find solutions to these problems.

## 3.6 Volunteer Crew

After setting up the devices and testing them, a call for volunteers was issued. To encourage the volunteers to participate, a modest payment was offered to each. All those responding were invited to participate [Mansfield02], though some withdrew when they could not attend an appointment for enrolment. A further call was issued to achieve slightly over 200 participants and finally we managed to recruit 221 volunteers for the exercise. The age and gender profile is shown in Figure 3.5

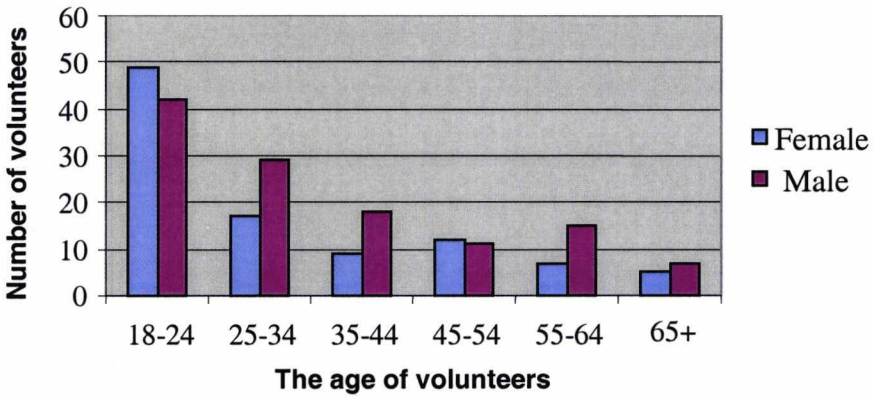



**Figure 3.5:** Age and gender of volunteer crew

This volunteer crew were recruited mainly from students and staff at the university of Kent as well as some volunteers from the city of Canterbury. It is a mix of people working in different environments some accepting the technology and others not and it is a mix of students, workers, housewives and retired people.

Before enrolment, participants were informed of the purpose of the exercise, what was required of them, and what information will be collected and stored as well as a brief description of the biometric modalities that will be collected. The main aim of informing the participants before the enrolment was to ensure participants the

intention of the project and help the participant to understand the project. Each participant was also required to give written consent (see appendix A) to participation and to supply brief personal information to assist in the data analysis.

## 3.7 Enrolment

On arriving at the "biometric laboratory", the supervisor provides the volunteer with a brief explanation of the capture process, which normally takes a few minutes. The aim of doing so is to make the volunteer aware of the enrolment process and also to allow the volunteer to recover to his/her normal state if s/he had cold fingers (when cold outside) or was out of breath (hurried to make their appointment).

Each volunteer was allocated a PIN for the trial. To avoid the possibility of the volunteer mistyping their PIN and producing another valid PIN, the ISBN error-detection scheme was used [Mansfield01a] [Mansfield02]. The 4-digit PINs abcd have the property that 4a+3b+2c+d is exactly divisible by eleven. This detects all single digit errors and transportations. From the available PINs, the set used was as widely spaced as possible, in the range 1000-9999, giving robustness against more complex typing errors. This set was stored in the biometric devices allowing the system to provide feedback if a wrong PIN was entered.

During the enrolment phase, each volunteer would use his/her PIN when attempting to enrol on each biometric system under test and up to three enrolment attempts per device was permitted. If the subject fails to enrol on any of the devices after three attempts then it is regarded as "Failure to Enrol ". Once the subject has successfully enrolled, s/he was asked to verify against his/her stored template to check if the subject can be reliably verified. Three attempts at verification were made whether the subject fails or passes any of them.

The order of enrolment on the devices being tested was randomised. However, in order to avoid keeping the participant moving around the test site, the order of enrolment was chosen in a way that the biometric modalities of each PC machine

would follow the order. For instance, if the first biometric modality is voice, automatically the second modality is face and finally fingerprint.

At the end of the session each volunteer was informed about the second session and its procedure.

## 3.8 Test Data Collection

This session was done at least 30 days after the first session. During this session subjects were required to enter the same PIN used during enrolment and follow the verification process. They were required to make three attempts to verify against their previously created template. Every attempt, even the ones that failed, were reported and saved, along with the user details.

The order of verification on the devices being tested was randomised and not correlated with the order of use on the previous session.

Overall this data collection exercise took almost 6 months; it started in November 2001 and finished in May 2002

## 3.9 Summary

In this chapter we described a data collection exercise that was undertaken at the University of Kent as part of this research. In this exercise both the physiological and behavioural biometrics were used, the fingerprint was chosen for its reliability, the voice and face for their low-cost hardware and high acceptability. The data collection protocol was linked to a scenario-testing regime and was developed within the emerging guidelines for best practice in biometric testing.

The next chapter provides some preliminary analysis on the multimodal database that was collected.

# Chapter 4

# The Multimodal Database and Some Preliminary Analysis

## 4.1 Introduction

This chapter starts by describing the database resulting from the data collection exercise that was carried out. A preliminary analysis of the database is provided, which provides an initial overview of the performance and viability of the three different modalities when considered as individual options.

## 4.2 The Multimodal Database

The database contains in addition, to the biometric information, basic personal and demographic information provided by the participant. The database consists of 221 individuals as previously shown in Figure 3.5 of whom 45 % are females and 55 % are males. Their ages range from 18 years to 65 years and above. The database contains all the information concerning both enrolment and verification; the enrolment date and the attempts to achieve it, as well as the verification date, the

successful and failure attempts were all noted. A sample of the database entities is shown in appendix (B).

## 4.3 Preliminary Analysis of the Database

In this section a preliminary analysis of the performance of the modalities adopted is presented. The analysis focuses mainly on performance comparisons between verification based on single individual biometric modalities and approaches, which exploit the opportunity to collect authentication data from more than one biometric source. In this analysis the modalities were combined using the AND rule, which will be exploited more later in this thesis. It is important to emphasise that the analysis presented here does not, in itself, provide a complete picture, since the focus was entirely on Type I verification errors. Nevertheless, the results presented are important in that they provide clear quantitative estimates of the performance potentially achievable in a real practical scenario with typical users and using commercially available devices. It should be noted that the analysis provided is an initial observation in respect of the data gathered. It should also be noted that the results through out this thesis are presented with a 95 % confidence level calculated as described in [Mansfield02] [Wayman99a] [Shen97]. The formulas used are presented in Appendix (C).

### 4.3.1 Failure to Enrol

Some users may experience problems at the enrolment stage itself, perhaps because of unfamiliarity with the acquisition infrastructure, or difficulty in generating sample data (e.g. an image of sufficient quality for accurate processing, and so on) [BWG00]. The failure to enrol rate was estimated as the proportion of volunteers who could not be enrolled under the pre-determined enrolment policy described in the previous chapter, which permits the volunteer up to three attempts to enrol successfully. Failure to enrol after the three failed attempts was regarded as final failure to enrol and no further attempts were made.

To indicate the effectiveness of the proposed multimodal systems in overcoming the deficiencies in any single modality for any particular potential enrolee, three experiments were carried out.

**Experiment 1:** Failure to enrol in single modalities

Table 4.1 shows the percentage of subjects who failed to enrol successfully with respect to each of the three biometric devices [Bengio02].

**Table 4.1:** Failure to enrol rates for each modality

| **Modality** | **Failure –to- enrol rate (%)** | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Error rates** | 2.7 ± 2.1 | 10.0 ± 4.0 | 1.4 ± 1.5 |

The table shows a considerable variability in the extent to which a satisfactory enrolment can be achieved. The Fingerprint and Face biometrics generated a relatively small failure to enrol rate, but the Voice modality proved significantly less reliable in achieving satisfactory enrolment.

**Experiment 2:** Failure to enrol in dual modalities

If failure to enrol for a multimodal system is defined as the situation where a subject fails to enrol on both devices in the chosen combination, then for each possible combination of two biometric modalities (Fingerprint/Voice, Fingerprint/Face, Voice/Face) the failure to enrol rates are as shown in Table 4.2.

**Table 4.2:** Failure to enrol for dual modalities

| **Modality** | **Failure –to- enrol rate (%)** | | |
|---|---|---|---|
| | Fingerprint / Voice | Fingerprint / Face | Face / Voice |
| **Error rates** | 0.9 ± 1.2 | 0.0 | 0.0 |

The table shows an improvement in the performance that occurs as a result of the availability of more than one modality, and it shows that the weakness in one modality can be compensated by the strength of another.

**Experiment 3:** Failure to enrol in three modalities

Table 4.3 shows the percentage of subjects who failed to enrol successfully in all three biometric devices.

**Table 4.3:** Failure to enrol rates for three modalities

| **Modality** | **Failure to enrol rate (%)** |
|---|---|
| | Fingerprint, Voice & Face |
| **Error rates** | 0.0 |

Indeed, moving to a three-modality system showed the failure to enrol rate drop to zero. This is a powerful indicator of the value of a multimodal system in overcoming the deficiencies in any single modality for any particular potential enrolee.

### 4.3.2 Results Obtained from Analysing the Sessions

As previously mentioned, each volunteer took part in two separate data collection sessions, the first involving enrolment on each of the devices under test together with a post-enrolment verification check. Each volunteer undertook three verification attempts at this session. A second session was undertaken at least one month later where three additional verification attempts were carried out using the enrolment templates generated at the first session.

In this section, two different situations that were raised from these sessions are presented and these are:

#### 4.3.2.1 The Time-based Changes in Biometric Data

The passage of time may give rise to a situation where, while having successfully enrolled and verified at an initial session, re-testing at a later date results in a verification failure (for example, because of damage to the finger in the intervening period, or a change in the way in which the finger was subsequently presented to the capture device, and so on). The time-based changes in biometric data was investigated by considering the proportion of subjects who, after a satisfactory enrolment and verification in the first session, failed to verify identity in all three attempts at the second.

Three different experiments were carried out to investigate the effect of time-based changes in biometrics when using multiple modalities.

**Experiment 1:** Failure rates in single modalities
Table 4.4 shows the failure rates with respect to each of the three biometric devices.

**Table 4.4:** Failure rates for each modality

| **Modality** | **Failure to verify rates (%)** | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Error rates** | 14.0 ± 4.6 | 1.4 ± 1.5 | 26.7 ± 5.8 |

These results also show the considerable variability across the available devices. The voice biometric provides, by a significant margin, the most stable performance, while the face biometric performs relatively poorly, generating a failure rate almost twice that of the fingerprint modality.

These results suggest that, while the voice system might present some difficulties at enrolment (cf. Table 4.1), the performance returned is much more stable than for the other modalities once a satisfactory enrolment has been achieved.

**Experiment 2:** Failure rates in dual modalities

The verification failure rate when a combination of two modalities is adopted refers to the situation where successful verification based on at least one of the modalities in the chosen combination has not been achieved, then for each possible combination of two biometric modalities (Fingerprint/Voice, Fingerprint/Face, Voice/Face) the failure rates are as shown in Table 4.5.

**Table 4.5:** Failure rates for dual modalities

| **Modality** | **Failure to verify rates (%)** | | |
|---|---|---|---|
| | Fingerprint / Voice | Fingerprint / Face | Face / Voice |
| **Error rates** | 0.5 ± 0.9 | 3.6 ± 2.5 | 0.0 |

Clearly, a significant improvement in performance is evident in the multiple modality scenario. The failure rates were reduced dramatically (cf. Table 4.4) by using a dual modality system, with the error rate falling to zero for the Voice/Face combination. This sharply contrasts with failure rates of as much as 27 % when a single modality is adopted.

**Experiment 3:** Failure rates in the three modalities

The verification rate is defined as a failure to satisfactorily verify identity in all three modalities tested. Table 4.6 shows the failure rate in all three modalities

**Table 4.6:** Failure rates for three modalities

| **Modality** | **Failure to verify rates (%)** |
|---|---|
| | Fingerprint, Voice & Face |
| **Error rates** | 0.0 |

Indeed, using three modalities reduced the error rate to zero. The experiments above suggest that although one modality may change dramatically over time, it is unlikely for two or more modalities to do the same.

#### 4.3.2.2 The Goat Phenomenon

The "goats" are the proportion of people who generate inherently unstable data with respect to a particular biometric in a system and who consequently may have a high risk being falsely rejected by the system [Plamondon89]. Goats may therefore need to be excluded from a system or treated as special cases in some way. The goat phenomenon was investigated by determining the proportion of subjects who after a satisfactory enrolment failed to verify identity in all attempts both at the first and the second session. Three experiments were carried out to investigate whether an individual could be a "goat" in more than one biometric system.

**Experiment 1:** Goats in single modalities

Table 4.7 shows the goats probability with respect to each of the three biometric devices.

**Table 4.7:** Goats in each modality

| **Modality** | **Goats (%)** | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Error rates** | 4.1 ± 2.6 | 0.0 | 1.8 ± 1.8 |

The voice system seems less vulnerable to instability (no goats identified) than either of the other two modalities.

**Experiment 2:** Goats in dual modalities

Goats in a combination of two modalities refers to the proportion of people who had a satisfactory enrolment, but failed to verify identity on each of the modalities in the chosen combination, then for each possible combination of two biometric modalities (Fingerprint/Voice, Fingerprint/Face, Voice/Face) the goats are as shown in Table 4.8

**Table 4.8:** Goats in dual modalities

| Modality | Goats (%) | | |
|---|---|---|---|
| | Fingerprint / Voice | Fingerprint / Face | Face / Voice |
| **Error rates** | 0.0 | 0.0 | 0.0 |

It was noticed that the "goats" that get through one type of biometric system, are not the same as those who cause a problem with another system. Table 4.8 showed that it was extremely unlikely that an individual, whose fingerprints are read inaccurately, for example, would also have a voiceprint that is hard to recognize.

**Experiment 3:** Goats in the three modalities

Table 4.9 shows the goats probability in the three biometric devices.

**Table 4.9:** Goats in the three modalities

| Modality | Goats (%) |
|---|---|
| | Fingerprint, Voice & Face |
| **Error rates** | 0.0 |

The experiments demonstrated that "goats" exist in every biometric system, but an individual who is a "goat" in one biometric is unlikely to be also a "goat" in a different one.

### 4.3.3 Exploitation of Re-try Strategies and Learning Effects

To exploit the effects of re-try strategies, two types of experiments were carried out. In both experiments the issues related to multiple enrolment attempts and their effects on error rates were considered, since it is usual to allow more than one such attempt in initiating the exploitation of a biometric-based system.

#### 4.3.3.1 False Rejection Rate as a Function of the Number of Enrolment Attempts

For each type of biometric technology there is an associated learning curve. The more often a user accesses a particular biometric device and the more practised the user becomes, the less likely it will be that the machine will fail to recognise that person [BWG02]. This is because the user has grown more consistent in presenting his/her biometric feature. In this experiment the effects of multiple enrolment attempts on the false rejection rate were explored. This was done by determining the proportion of subjects who failed to verify identity either in the first or in the second session with respect to the enrolment attempts. In Table 4.10, 1st enrolment attempt refers to the portion of subjects who were successfully enrolled in the first attempt, 2nd enrolment attempt refers to the portion of subjects who after having problems enrolling in the 1st attempt was successfully enrolled in the 2nd attempt and 3rd enrolment attempt refers to the portion of subjects who had problems enrolling in the 1st and 2nd attempt, but were successfully enrolled in the 3rd attempt. Table 4.10 shows the results obtained from this experiment.

**Table 4.10**: False rejection rate as a function of enrolment attempts

| Attempt | False rejection rate (%) | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| 1st enrolment attempt | 10.9 ± 4.1 | 0.9 ± 1.2 | 31.2 ± 6.1 |
| 2nd enrolment attempt | 9.5 ± 3.9 | 0.5 ± 0.9 | 1.4 ± 1.5 |
| 3rd enrolment attempt | 2.7 ± 2.1 | 0.0 | 0.5 ± 0.9 |

It was noticed that as the number of attempts to achieve a successful enrolment increases, the false rejection rate decreases. This shows the positive effect of "training during use" associated with this type of activity.

#### 4.3.3.2 Failure to Enrol as a Function of the Number of Enrolment Attempts

Although "failure to enrol" in the previous sections was regarded as failure to enrol in all three attempts, in this experiment we considered the "failure to enrol" as the failure to enrol with respect to enrolment attempts. In Table 4.11,4.12 and 4.13, 1st enrolment attempt refers to the portion of subjects who had problems enrolling in the first attempt, but were successfully enrolled in the 2nd, the 2nd enrolment attempt refers to the portion of subjects who had problems enrolling in both the 1st attempt and the 2nd attempt, but were successfully enrolled in the 3rd and 3rd enrolment attempt refers to the portion of subjects who had problems enrolling in all three attempts and no further attempts were made. Table 4.11 shows the results obtained from this experiment.

**Experiment 1:** Failure to enrol in single modality

Table 4.11 shows the failure to enrol rate on each of the biometric systems with respect to the enrolment attempts.

**Table 4.11**: Failure to enrol rate as a function of enrolment attempts in single modalities

| Attempt | Failure to enrol rate (%) | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| 1st enrolment attempt | 28.5 ± 6 | 29.4 ± 6 | 4.1 ± 2.6 |
| 2nd enrolment attempt | 7.7 ± 3.5 | 16.3 ± 4.9 | 2.3 ± 2 |
| 3rd enrolment attempt | 2.7 ± 2.1 | 10.0 ± 4 | 1.4 ± 1.5 |

The results demonstrate why multiple attempts are generally necessary in practice and, especially, show how a much poorer performance would be recorded if only a single enrolment attempt was allowed. A clear message here is the illustration of the positive effect of "training during use" associated with this type of activity.

**Experiment 2:** Failure to enrol in dual modalities

The failure to enrol rate when a combination of two modalities is adopted is defined as a failure to complete a satisfactory enrolment process on both of the available devices. Table 4.12 shows the performance characteristics for dual modalities.

**Table 4.12**: Failure to enrol rate as a function of enrolment attempts in dual modalities

| Attempt | Failure to enrol rate (%) | | |
|---|---|---|---|
| | Fingerprint / Voice | Fingerprint / Face | Face / Voice |
| 1st enrolment attempt | 10.4 ± 4.0 | 0.9 ± 1.2 | 0.9 ± 1.2 |
| 2nd enrolment attempt | 2.3 ± 2.0 | 0.0 | 0.0 |
| 3rd enrolment attempt | 0.9 ± 1.2 | 0.0 | 0.0 |

**Experiment 3:** Failure to enrol in the three modalities

Table 4.13 shows the failure to enrol in the three modalities with respect to enrolment attempts.

**Table 4.13**: Failure to enrol rate as a function of enrolment attempts in all three modalities

| Attempt | Failure to enrol rate (%) |
|---|---|
| | Fingerprint / Voice / Face |
| 1st enrolment attempt | 0.5 ± 0.9 |
| 2nd enrolment attempt | 0.0 |
| 3rd enrolment attempt | 0.0 |

It was noticed that as the number of modalities increases more attempts are required to achieve a successful enrolment on all of them together.

### 4.3.4 Effects of Biometrics on Each Other

The experiments carried out in the previous sections supported the intuitive assumption that the combination of multiple modalities improves performance by providing more information for making identity decisions. On the other hand, a different intuition suggests that if a strong modality is combined with a weaker one, the resulting decision environment is in a sense averaged, and the combined performance will lie somewhere between that of the two modalities conducted individually (and hence will be degraded from the performance that would be obtained by relying solely on the strongest one) [Daugman00]. To investigate the second suggestion and to see the effect of different modalities on each other and their effect on the performance of the system, three experiments were undertaken. In these experiments the proportion of subjects who had a successful enrolment and verified identity on both sessions were computed.

**Experiment 1:** Successful subjects in single modality
Table 4.14 shows the performance of each of the modalities

**Table 4.14**: Successful subjects in single modality

| **Modalityr** | **Successful subjects (%)** | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Error rates** | 78.3 ± 5.4 | 88.7 ± 4.2 | 67.4 ± 6.2 |

Although the voice system has presented some difficulties at enrolment (cf. Table 4.4), its overall performance is much better than the other two modalities. The face biometric performed poorly during verification (c.f. Table 4.5), which affected its overall performance.

**Experiment 2:** Successful subjects in dual modalities
Table 4.15 shows the performance of dual modalities

**Table 4.15**: Successful subjects in dual modality

| Modality | Successful subjects (%) | | |
|---|---|---|---|
| | Fingerprint / Voice | Fingerprint / Face | Face / Voice |
| **Error rates** | 70.6 ± 6.0 | 54.8 ± 6.6 | 60.2 ± 6.5 |

It was noticed that when using a relatively strong biometric such as the voice, with a relatively weak biometric (in the sense of their overall performance) such as the face, the resulting performance is even less than the average of both biometrics together.

**Experiment 3**· Successful subjects in three modalities

Table 4.16 shows the successful subjects in all three modalities.

**Table 4.16**: Successful subjects in three modalities

| Modality | Successful subjects (%) |
|---|---|
| | Fingerprint, Voice & Face |
| **Error rates** | 51.1 ± 6.6 |

Since the table above summarises the performance of the system used, it explains the effect that biometric performance has on each other. The voice system performed the poorest during enrolment while the face the poorest during verification. This difference in performance affected the overall performance of the system and hence resulting in the conclusion that sometimes a strong biometric is better alone than in combination with a weaker one.

### 4.3.5 Factors that Influenced the Enrolment Process

The following list is some of the user factors observed by the supervisor during the data collection exercise that affected the performance and resulted in enrolment failure. The factors can be categorised as physiological, behavioural, appearance and job related. It is important to emphasise that these factors are just the supervisor's

observations and that no further investigation has been done on them. The enrolment failure factors for each biometric is also provided in this section

#### 4.3.5.1 Fingerprint Biometric

***User physiology***

- Failure due to dry and cracked fingers.
- Most of the women's failure was due to either their long or narrow fingers. Long fingers made it difficult to position the finger and the same applied to narrow fingers, where it was difficult to place the finger in the centre of the device sensor.
- Men with large fingers had difficulty in positioning their finger, which resulted in enrolment failure.
- Women with long fingernails had difficulty in adjusting their finger; their nails were covering the sensor.
- Left handed people found it difficult to use the Fingerprint device, since that the fingerprint platen of the device was on its left side.

***User behaviour***

- Most of the failures were due to placement; the fingerprint device did not have a frame that limits the positioning of the finger.
- Failures due to sweaty (e.g. tensed person) or cold fingers (coming from cold weather) subjects were advised to dry their fingers with a piece of cloth and to wait till their fingers were warm.

***User job***

- People with jobs that require using mainly their fingers (e.g. cleaners) find it more difficult to enrol.
- Failure due to the unfamiliarity of people with technology.

#### 4.3.5.2 Voice Biometric

***User Physiology***

- Old people with hearing problems and sight problems found it difficult to read the sequences on the screen that resulted in misreading the sequences.
- Failure due to sight problems; people could not see the sequence on the screen and were not able to memorize it, so they ended up mixing it.
- Individuals that suffer from dyslexia had some problems in reading the sequences.

***User behaviour***

- Failure due to frustration of not being accepted by the system, which resulted in change of tone.
- Failure due to positioning; people were insisting in getting closer to the microphone that resulted in producing echo while speaking.
- Failure due to tension; people were sometimes tense throughout the trial.
- Repeating the sequences either quickly or too slowly resulted in a failure. Speaking quickly sounded as if the person was mumbling, while speaking too slow the time slot for each sequence was finished before the person finished the sequence.

***User job***

- Failure due to the unfamiliarity of people with technology.
- Failure due to previous activity in working in radio or TV, people tend to vary their tone while reading the sequences, which resulted in a failure.

#### 4.3.5.3 Face Biometric

***User behaviour***

- As mentioned in the previous chapter, the enrolment process required the user to vary the angle of the face slightly. Failure was due to people tilting their head so quickly that their face was hardly captured by the software.
- Failure due to people looking at the screen instead of the camera.

***User appearance***

- Failure due to wearing shaded glasses, coloured frames or very thick glasses that obscured the eyes.

### 4.3.6 Factors that Influenced the Verification Process

The following list is some of the user factors observed by the supervisor during the trial that affected the performance and resulted in verification failure. The factors are very similar to that affected the enrolment process. The verification failure factors for each biometric is provided in this section

#### 4.3.6.1 Fingerprint Biometric

***User physiology***

- Failure due to dry and cracked fingers.
- Most of the women's failure was due to either their long or narrow fingers. Long fingers made it difficult to position the finger, the same applied to narrow fingers it was difficult to place it in the centre of the device sensor.
- Men with large fingers had difficulty in positioning their finger, which resulted in enrolment failure.
- Women with long fingernails had difficulty in adjusting their finger; their nails were covering the sensor.

- Left handed people found it difficult to use the Fingerprint device, since that the fingerprint platen of the device was on its left side.

***User behaviour***

- Most of the failures were due to placement; the fingerprint device did not have a frame that limits the positioning of the finger.

- Failures due to sweaty (e.g. tensed person) or cold fingers (coming from cold weather) subjects were advised to dry their fingers with a piece of cloth and to wait till their fingers were warm.

#### 4.3.6.2 Voice Biometric

***User Physiology***

- Old people with hearing problems and sight problems found it difficult to read the sequences on the screen that resulted in misreading the sequences.

- Failure due to sight problems; people could not see the sequence on the screen and were not able to memorize it, so they ended up mixing it.

- Failure due to cold that affected the voice.

#### 4.3.6.3 Face Biometric

***User behaviour***

- Majority of the failure was due to the fact that people during enrolment were looking at the screen and during verification were looking at the camera and vice versa.

***User appearance***

- Failure due to different hair style/colour that altered the face appearance

## 4.4 Discussion

A number of interesting points may be drawn from the initial analysis of the performance of the three commercial devices (fingerprint, voice and face) gathered in Chapter 3. The results in this chapter are based on the information extracted from the data collection exercise (a sample is shown in appendix (B)). According to the experimental results of this study, the failure-to-enrol rate reduces when using more than one modality, thus supporting the idea that using multiple biometric modalities increases the performance of the system. The experiments also showed that the performance of a single modality may change significantly over time, but it is unlikely for more than two modalities to do the same. On investigating the "goats" phenomenon, the experiments demonstrated that "goats" exist in every biometric system, but an individual who is a "goat" in one biometric modality may not be also a "goat" in a different one. Exploiting the effects of re-try strategies and learning effects showed that the false rejection decreases as the number of attempts to achieve a successful enrolment increases. It also showed the necessity of multiple enrolment attempts in practice in order to improve the performance of the system. On increasing the number of biometric modalities the results showed that more attempts are required to achieve a successful enrolment on all modalities together thus demonstrating the problem of non-universality in biometrics as not all users are able to enrol in all three biometric modalities. The results also showed the difference in performance of each of the biometrics can affect the overall performance of the system on using a specific combination scheme (AND rule) and that one poorly performing biometric modality can degrade the overall performance of the system. Finally, some common factors, observed by the supervisor during the data collection exercise, influenced both the enrolment and verification process suggesting that if those factors could be reduced an improvement in the system could be achieved.

## 4.5 Summary

This chapter has presented our initial results and observations in respect of the data gathered from our large-scale trial to assess the interaction of a cross-section of the general public with a small set of different biometric modalities.

The data gathered, and the initial observations presented here, provide a preliminary overview of the performance and viability of three different biometric modalities when considered as individual options. It should be noted that the results are based on three specific commercial devices, though it is to be expected that the conclusions drawn are indicative of the general trends of the modalities considered.

The next chapter provides a general overview of information fusion and describes the different architectures and levels for combining multiple classifiers. The purpose of the following chapter is to decide the architecture and level at which the three different modalities collected (fingerprint, voice and face) will be combined.

# Chapter 5

# Decision Fusion for Multi-Modal Biometric Systems

## 5.1 Introduction

This chapter investigates multi-modal biometric systems using the fusion architecture and fusion level that was chosen in Chapter 2. This chapter starts by giving a review of the research in the field of multimodal person recognition using fusion rules at the decision level. An explanation of the experimental set-up used for calculating the error types is then provided and a comparison between the hard decision and the soft decision fusion rules is presented. The hard and soft fusion rules are also used when characterising the system users as sheep, goats, lambs and wolves. The multimodal database collected in Chapter 3 is used for the experiments.

## 5.2 Contributions of the Decision Fusion Rules

The decision fusion rules that were described in chapter 2 have been used in the field of multimodal person recognition. This section provides a review of the commonly used fusion rules at the decision level.

Chibelushi et al have proposed in [Chibelushi93a] to integrate acoustic and visual speech for speaker recognition. The combination scheme used was a simple sum rule. The author has also combined in [Chibelushi93b] information from still face profile images and speech using a form of weighted summation fusion. The results showed that when using optimal weights, the ERR was reduced compared to when using each of the speech or the face profile expert alone.

Jourlin et al used a form of weighted summation fusion to combine the opinions of two experts: a speech expert and a lip expert [Jourlin97]. Using optimal weights, fusion led to better performance than using the underlying experts alone.

Dieckmann et al used the majority-voting scheme to integrate two biometric modalities (face and voice), which were analysed by three different experts: (static) face, (dynamic) lip motion and (dynamic) voice [Dieckmann97].

Kittler et al integrated two modalities (face and lip) for personal identity recognition [kittler97]. Three different combination rules were used such as the product rule, majority voting rule and the sum rule. The results confirmed the benefits of integration and the predicted behaviour of the majority voting and averaging integration strategies, which outperformed the product rule combination.

Kittler et al proposed a multimodal person verification system using three experts: frontal face, face profile and voice [Kittler98]. The outputs of the three experts were soft decisions (scores between zero and one). The best combination results were obtained from a simple sum rule.

Hong et al. presented ways of combining information from two modalities: face and fingerprint images at various levels [Hong99]. Two levels of fusion were considered; score level fusion, where the Bayesian method was used and a decision level fusion, where both the OR and AND rules were used. Experimental results showed that the performance of a biometric system was improved by integrating multiple biometrics.

Ben-Yacoub et al investigated the benefits of classifier combination for a multimodal system for personal identity verification [Ben-Yacoub99]. The system used frontal face images and speech. Results showed that by using the linear weighted scheme and a Support Vector Machine (SVM) classifier there was a significant reduction in the total error rate.

Ross et al combined the matching scores of three traits (Face, Fingerprint and Hand geometry) to enhance the performance of a biometric system [Ross01]. Three different techniques (sum rule, decision tree, linear discriminant analysis) were used to combine the matching scores. Experiments indicated that the sum rule resulted in the best performance.

Shakhnarovich et al proposed person identification based on face and gait cues [Shakhnarovich02]. The different combination rules that were used are max, min, sum and product rules. Experimental results showed that the sum rule outperformed the other rules.

This brief review of the fusion rules used at the decision level reveals that for combining soft decisions the sum rule outperforms other combination rules which supports the idea of adopting and using this rule in this study. Several researchers have also used the majority-voting rule, which is also adopted in this study.

## 5.3 Experimental Setup

As mentioned in Chapter 3, in the data collection process, each volunteer took part in two separate data collection sessions, the first involving enrolment on each of the devices under test together with a post-enrolment verification check where each volunteer undertook three verification attempts. A second session was undertaken at least one month later where three additional verification attempts were carried out using the enrolment templates generated at the first session. The experiments carried out in this chapter focused mainly on the three verification attempts undertaken at the second session, as it was desired to consider any time-based changes that occur in the biometric data.

Before starting the experiments, both the corpus and the database collected were examined and the group of people who had the following errors were discarded: -

- The group that failed to enrol in any of the biometric devices, this is because no templates were generated for them.
- The group that had blank or corrupted images in any of the three attempts of the second session due to entering a PIN but moving on before a proper image is captured.
- The group that had templates but did not have biometric samples due to the fact that they failed to attend the second session.

After discarding these groups the resulting database consisted of 147 subjects, each having a template generated by each of the biometric devices and having all three samples acquired in the second session.

In general and depending on the data available, three different data sets are needed for each classifier. The first data set is called the training set and is used by the classifier to model the different persons. The second data set is called the validation set and is used to fine-tune the classifier, for instance by calculating the decision thresholds. The third data set is called the test set and it is used to test the performance of the classifier. For the experiments carried out in this chapter, a simple experimental protocol was used. In this protocol the first enrolment session and the three verification attempts performed in the second session were used in the following manner:

The first enrolment session was used for training the individual classifiers. This means that each access has been used to model the respective client, yielding 147 different client templates for each modality.

Since it was decided to use the default verification threshold assigned by the vendors of each biometric system then there was no need to have a validation set. The three accessed attempts from each person that was undertaken at the second session were used to test the classifiers. This was done by matching each single client sample

access with his own reference template generating 147 clients, then a cross comparison (all samples compared to all templates except the matching one) was used to establish the impostor distribution [O'Gorman98] generating 147 × 146 = 21462 impostor accesses. This process was applied to every attempt in the second session yielding three testing sets.

Table 5.1 shows an example of the cross comparison matrix with the N "genuine" scores shown in **bold** on the diagonal of the matrix and N (N-1) "impostor" scores above and below the main diagonal. For simplicity, the subjects are represented by the alphabets A, B and C. The genuine scores are generated by matching the verification access sample of each subject with his own template. For example the verification sample of subject A is matched with its own template, the same process is applied for both subjects B and C. On the other hand, the impostor scores are generated by comparing all the verification access samples to all the templates except the matching one, i.e. for example, for subject A, all the verification samples (B and C) are compared with its template, except its own sample. The same process is applied for both subject B and C.

**Table 5.1**: Cross comparison matrix showing classifier scores for N=3

| Samples / Templates | A | B | C |
|---|---|---|---|
| A | **50** | *40* | 30 |
| B | *70* | **60** | 20 |
| C | 10 | 0 | **3** |

To illustrate Table 5.1, lets consider the threshold to be set to 40, the genuine scores on the diagonal show that both subjects A and B are accepted by the system since their scores are equal to or higher than the pre-specified threshold and that subject C is falsely rejected by the system for having a score lower than the threshold. The impostor scores in the upper and lower triangle show that the sample presented by subject B for verification is falsely accepted by the system as being of subject A and

that the sample presented by subject A is falsely accepted by the system as being of subject B since both have scores higher than or equal to the pre-specified threshold.

As already mentioned, for each verification attempt a cross comparison matrix was constructed and used as a test set and since each subject performed three verification attempts, three testing set were produced. These three testing sets were used in the experiments carried out in this chapter, which are explained more in the subsequent sections. It should be noted that the results are presented with a 95 % confidence level calculated as described by [Mansfield02]. The formulas used are presented in Appendix (C)

## 5.4 Combining Classifiers Decisions

Combining classifiers decisions is normally the process of combining soft or hard decisions given by different classifiers. As it was mentioned earlier in Chapter 2 it was decided to classify the hard and soft decisions as two separate sub-level of the decision level [Prabhakar01]. The classifiers in either case can be of the same type but working with different features (e.g. fingerprint and voice), heterogeneous classifiers working with the same features, or a hybrid of the previous two, which is the scope of this work, since the classifiers were heterogeneous working with different features (fingerprint, voice and face).

### 5.4.1 Hard Decision Level

A hard decision is a decision made by the system that returns either a 0 or a 1. In an ensemble of classifier the hard decision from each classifier can be combined using voting techniques.

#### Voting Techniques

Voting techniques are classical empirical techniques where the global decision rule is obtained by fusing the hard decisions made by $m$ biometric modules [Kuncheva02][Alkoot99]. These techniques are sometimes referred to as $k$-out-of $m$ voting techniques, where $k$ relates to the number of classifiers that have to decide on

the identity claimed by a person [Teoh04]. For some values of $k$, particular decision fusion schemes are obtained:

1. $k = 1$. This is called the **OR** rule. The identity claim is accepted if at least one of the $m$ classifiers decides that the person under test is a client.
2. $k = m$. This is called the **AND** rule. The identity claim is accepted only if all the $m$ classifiers decide that the person under test is a client.
3. $k = (m + 1) / 2$. This is called the **Majority Voting** rule. It is a concession between the two previous rules.

### 5.4.2 Soft Decision Level

A soft decision is a decision made by the system that returns a score that lies in the [0, 1] interval. The soft decision from each classifier can be combined using the Summation rule.

#### Sum rule

This method is the simplest combination strategy and it has been widely used as a combination scheme in pattern recognition. In this method, the scores from the classifiers are summed in combined using [Chibelushi93b] [Duc97].

$$F = \sum_{i=1}^{M} w_i S_i \qquad (5.1)$$

Where $S_i$ is the score from the i-th classifier, $w_i$ is the corresponding weight in the [0,1] interval, with the constraint $\sum_{i=1}^{M} w_i = 1$ and $M$ is the number of classifiers used.

This method can either be non-confidence based (simple sum) - same weights for all the classifiers – or confidence based (weighted sum) - different weights for the

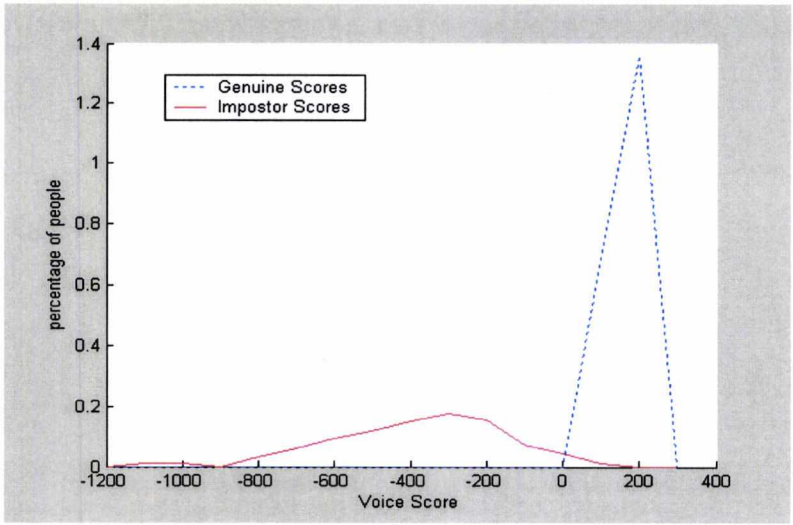
classifiers, assuming a confidence measure is assigned to the classifiers. This approach is also known as *Linear Opinion Pool.*

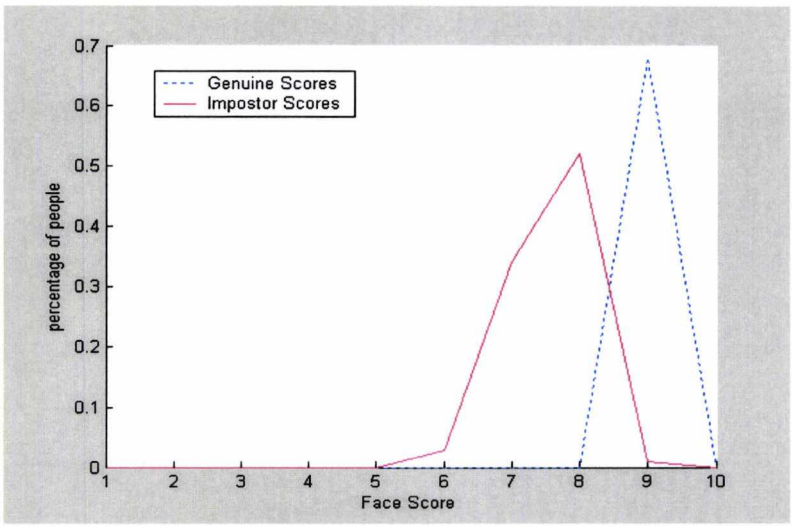
Before applying this fusion rule the raw scores of the different classifiers must be first normalized, where they are mapped into a common range [0, 1]. The score normalization is an essential step because the scores of the individual classifiers may suffer from one or both of the following problems:

- The scores of the individual classifiers may be heterogeneous, that is, one classifier may output a distance measure while another may output a similarity measure.
- The scores of the individual classifiers may have different numerical ranges. For example, one classifier may output scores in the range [0, 1] and another in the range [100, 1000] this will result in the second classifier eliminating the contribution of the first one if the scores are fused without any normalization.

The individual classifiers used in this work suffered from both problems. The scores obtained from the face and the voice modalities were distance scores and those obtained from the fingerprint modality were similarity scores. The individual classifiers also had different numerical ranges, the voice modality score ranges from –1476 to 323, the face modality score ranges from 0 to 10 and finally the fingerprint modality score ranges from 0 to 9. It should be noted that the score ranges of both the fingerprint and face modalities were provided by the vendors of the devices, while the voice modality score range was estimated from using a dataset. This shows the necessity of score normalization into a common domain before combining them. Figure 5.1 shows the conditional distribution of genuine and impostor scores for voice, face and fingerprint modalities.

It should be noted that the y-axis for all the graphs showing the conditional distribution of genuine and impostor scores of the fingerprint modality was adjusted to start from (-10) for better viewing of the graph.



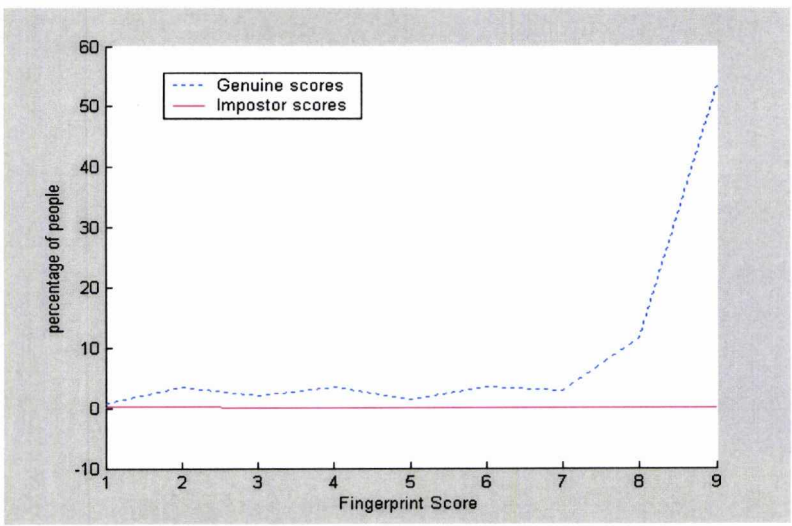




**Figure 5.1**: Conditional distribution of genuine and impostor scores for voice, face and fingerprint respectively

### 5.4.2.1 Normalization Methods

In this section two of the most commonly used normalization methods in the field of biometrics are explored. These are the min-max and the z-score. A new method is also proposed.

#### Min-Max method

This method is best used if the maximum and minimum values of the scores produced by the classifier are known [Jain99a][Indovina03][Snelick03] [Marcialis02]. In this case, the minimum and maximum scores are shifted to 0 and 1, respectively.

$$S_{norm} = \frac{S - S_{\min}}{S_{\max} - S_{\min}} \tag{5.2}$$

Where

$S_{norm}$ : is the normalized score

$S$ : is the raw classifier score

$S_{\min}$ : is the minimum score from the set S of all the scores of that classifier

$S_{\max}$ : is the maximum score from the set S of all the scores of that classifier

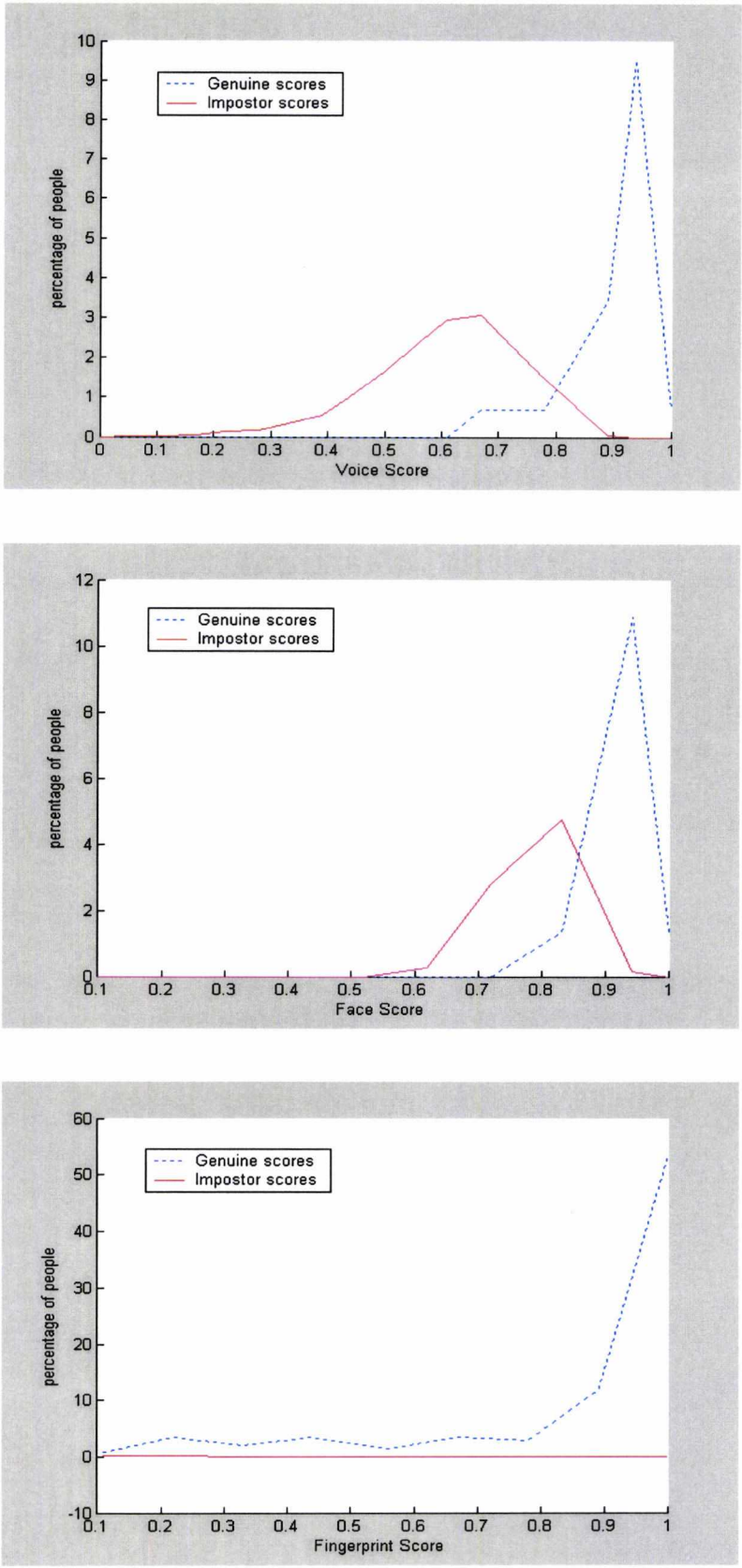



**Figure 5.2**: Conditional distribution of genuine and impostor scores after Min-Max normalization for voice, face and fingerprint respectively

This method is highly sensitive to the boundaries (maximum and minimum values) in the data. Figure 5.5 shows the distribution of fingerprint, voice and face scores after performing this normalization method. The figure shows that the Min-max normalization transforms all the scores into a common range largely and retains the overall original shape of the score distribution except for a scaling factor. Although this method produced the same distribution of scores as the original one, this may not be considered since the maximum and minimum values for the voice modality had to be estimated from a data set, which suggests that these values may change on a different matching set.

**Z-score method**

This is the most commonly used normalization technique. it is calculated by using the arithmetic mean and standard deviation of a given data [Lu04] [Kholmatov03] [Cheung04] [Auckenthaler00]. The normalized scores are given by

$$S_{norm} = \frac{S - \mu}{\sigma} \tag{5.3}$$

Where

$S_{norm}$ : is the normalized score

$S$ : is the raw classifier score

$\mu$ : is the arithmetic mean

$\sigma$ : is the standard deviation

This method is highly sensitive to the arithmetic mean and the standard deviation values. Figure 5.3 shows the distribution of fingerprint, voice and face scores after performing this normalization method. The figure shows that the Z-score normalization largely retained the overall original shape of the score distribution for both the voice and face modalities but not the finger modality. It also fails to map the scores of the different modalities into a common numerical range. This method is not considered to be robust since both the arithmetic mean and the standard deviation are calculated from a certain data set, which might change if calculated on a different one and the fact that it does not map the scores of the different modalities into a common numerical range makes it undesirable for the experiments.

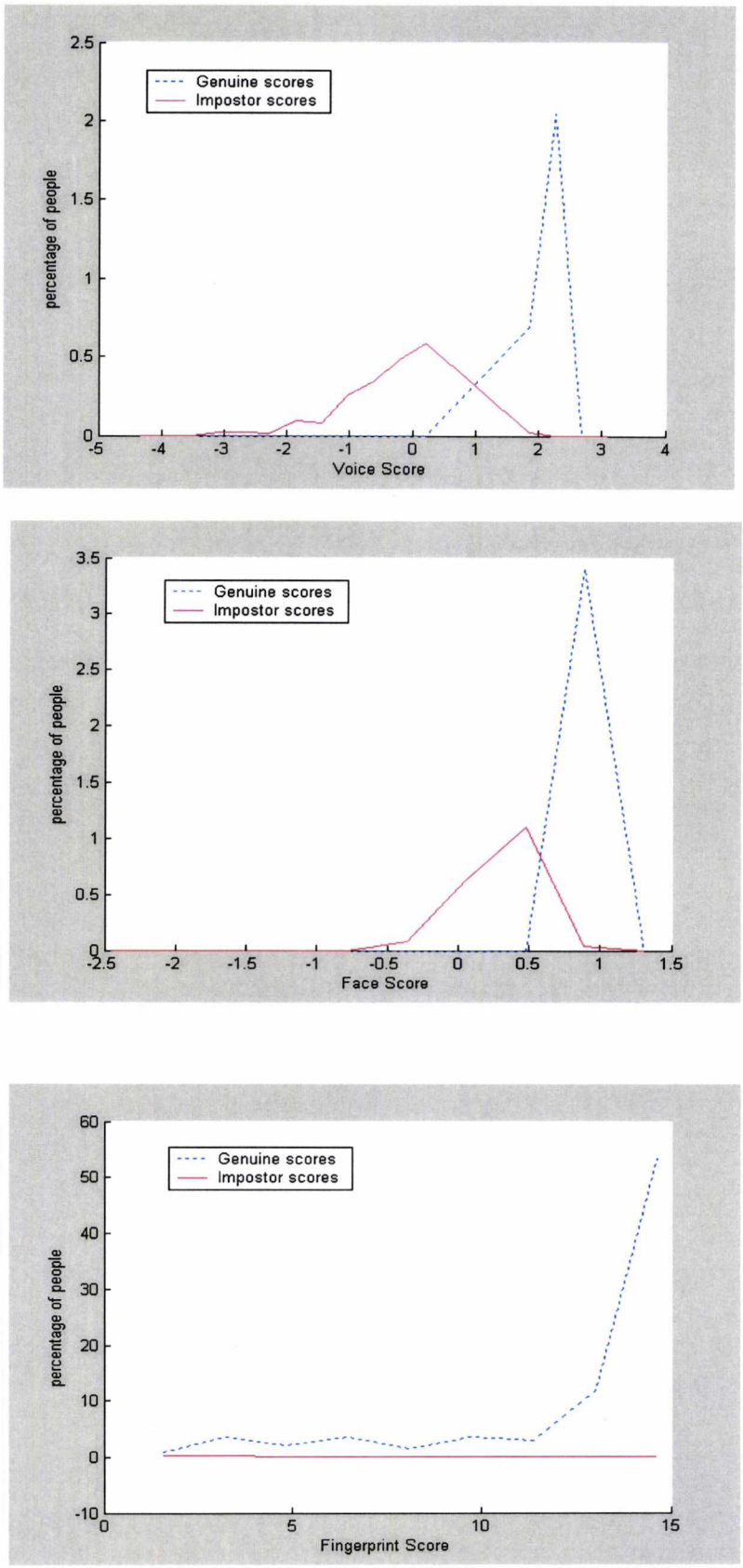



**Figure 5.3**: Conditional distribution of genuine and impostor scores after Z-score normalization for voice, face and fingerprint respectively

**Adaptive Logarithmic method**

The proposed method compares the raw scores to a predefined threshold usually given by the vendors. This method will keep all values below the threshold under 0.5 and all those above the threshold above 0.5. In this way the scores will be mapped in the range [0,1]

$$S_{norm} = \frac{\exp^{(S-Threshold)}}{1+\exp^{(S-Threshold)}} \tag{5.4}$$

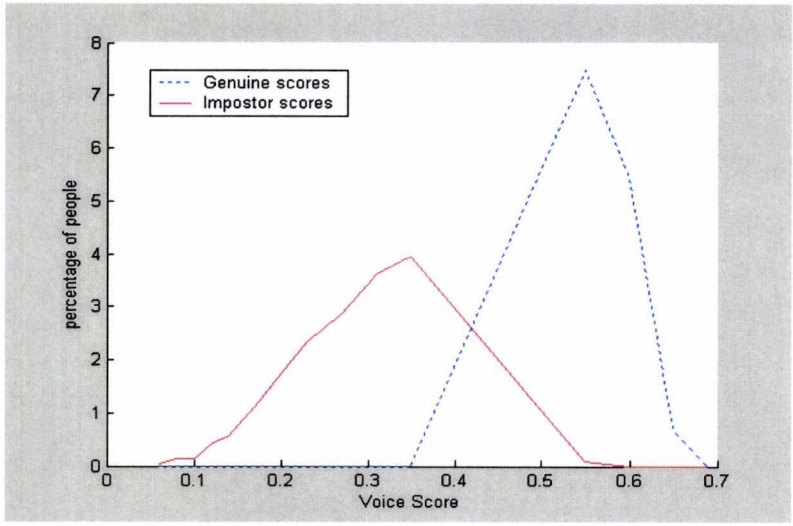
Where

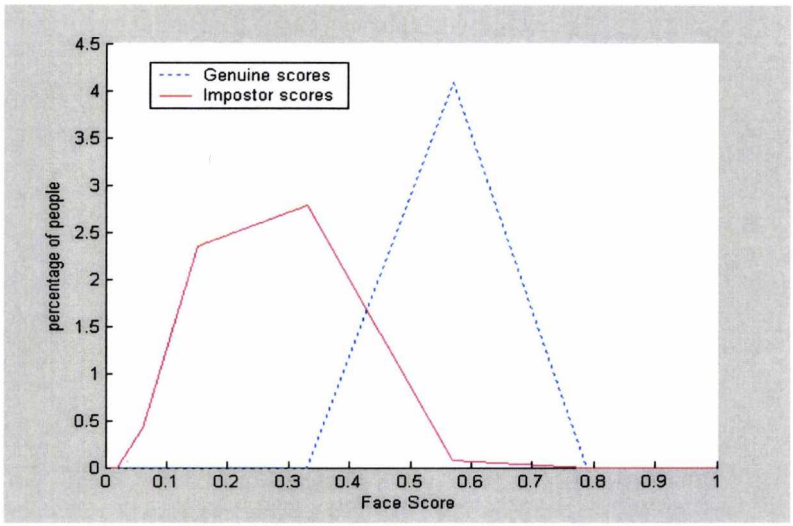
$S_{norm}$ : is the normalized score

$S$ : is the raw matcher score

*Threshold*: is the default threshold of the biometric device.

This method is highly sensitive to the threshold, which is normally provided by the vendors. Figure 5.4 shows the distribution of fingerprint, voice and face scores after performing this normalization method. The figure shows that the proposed normalization method largely retains the original shape of the score distribution with a scaling factor as well as it transforms the scores of the different modalities into a common numerical range. This method is considered to be almost accurate since it does not require the calculation of certain parameters from a data set as the previous two methods.



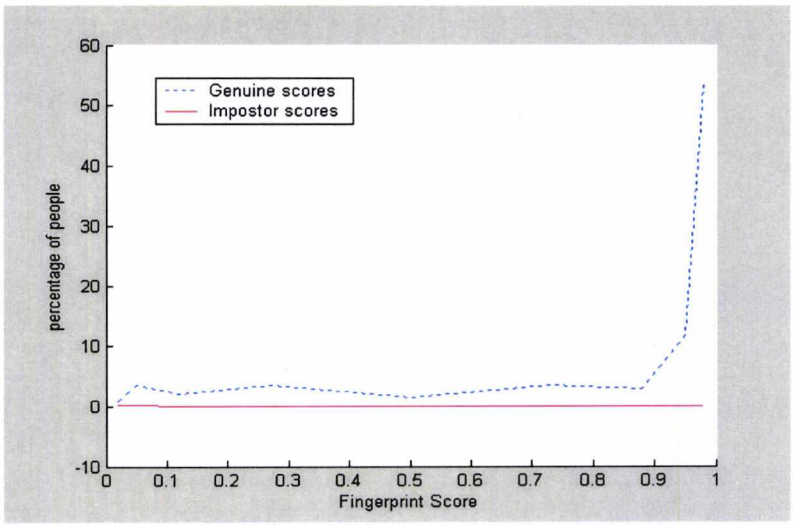




**Figure 5.4**: Conditional distribution of genuine and impostor scores after adaptive logarithmic normalization for voice, face and fingerprint respectively

## 5.5 Scenarios for Accessing a System

A user accessing a system could be either a legitimate user or an impostor attempting to defeat the system. Since, it was desired to calculate the likelihood that a legitimate user (client) is rejected by the system (FRR) and the likelihood that an impostor is accepted by the system during verification as being a legitimate user (FAR). An explanation is presented of how these errors are calculated using the different fusion rules. In this section, two different cases are discussed:

It should be noted that the multimodal database collected in Chapter 3 from the three commercial devices (fingerprint, voice and face) is used for the experiments in this chapter.

### 5.5.1 Genuine Users

This is the case where the user of the system is a legitimate user trying to access the system, this user will either be accepted by the system in the case where the decision fusion rule results in an overall accept decision or rejected by the system in case the overall verification decision made is a reject. The rejection of this user gives rise to one of the two main errors -false rejection error-. In the following two subsections an explanation is provided of how this error is calculated for both the hard and soft decision fusion.

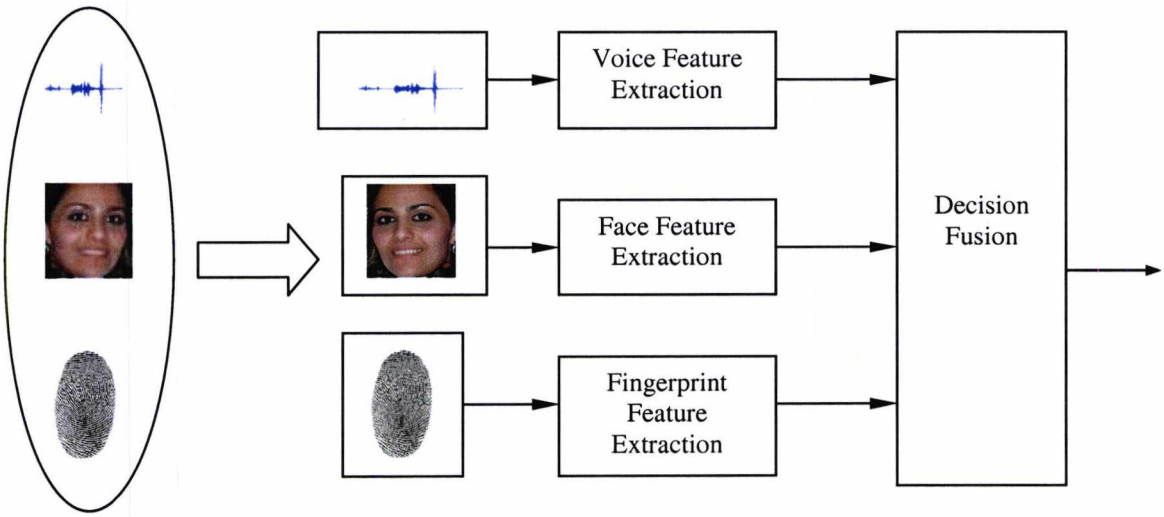



**Figure 5.5:** Genuine user

#### 5.5.1.1 Hard Decision Fusion

A hard decision is a decision made by the classifier that returns either accept or reject, in other words a 1 or a 0. Since, the output of the classifiers of the three modalities (fingerprint, voice and face) used in this project were scores, the default thresholds specified by the vendors of the different biometric devices were assigned to each classifier such that if the genuine matching score is higher than or equal to the pre-specified threshold an accept or a 1 is returned, and if the genuine matching score is lower than the pre-specified threshold a reject or 0 is returned. These decisions are then combined using the following three methods:

##### 5.5.1.1.1 AND Fusion

In AND fusion [Kittler98], the identity claim of a user is accepted only if all the classifiers decide that the person under test is a client, that is, if the outputs returned by the classifiers in all three modalities is a 1 (accept). Hence, a user is falsely rejected if the output returned by the classifiers is a 0 in any single modality.

##### 5.5.1.1.2 Majority Voting

In majority voting [Dieckmann97], the identity claim of a user is accepted if the majority of the classifiers decide that the person under test is a client, that is, if the output returned by the classifiers in any two modalities out of the three is a 1 (accept). Hence, a user is falsely rejected if the output returned by the classifiers is a 0 in any two modalities.

##### 5.5.1.1.3 OR Fusion

In OR fusion [Kittler98], the identity claim of a user is accepted if at least one of the classifiers decides that the person under test is a client, that is, if the output returned by the classifiers in any single modality is a 1 (accept). Hence, a user is falsely rejected if the output returned by the classifiers is a 0 in all three modalities.

#### 5.5.1.2 Soft Decision Fusion

A soft decision is a decision made by the system that generates a score that normally lies in the range [0,1]. As already mentioned, the output of the classifiers of the three modalities (fingerprint, voice and face) used in this project were scores, so they may be combined using the following rule:

##### 5.5.1.2.1 Sum Rule

In the Sum Rule [Ross01], the scores from the classifiers are summed as shown in Equation 5.1, where the summed scores are then compared to a pre-specified threshold to reach the verification decision of whether accepting or rejecting the user. The identity claim of a user is accepted if the summed score of all three classifiers is higher than or equal to the assigned pre-specified threshold. Hence, a user is falsely rejected by the system if the summed score of all three classifiers is lower than the pre-specified threshold.

### 5.5.2 Impostors

This is the case where the user of the system is an impostor attempting to defeat and access the system as being a legitimate user, this user will either be accepted by the system as a legitimate user in the case where the decision fusion rule results in an overall accept decision or rejected by the system in case the overall verification decision made is a reject. The acceptance of this user gives rise to one of the two main errors - false acceptance error-. In the following two subsections an explanation is provided of how this error is calculated for both the hard and soft decision fusion.

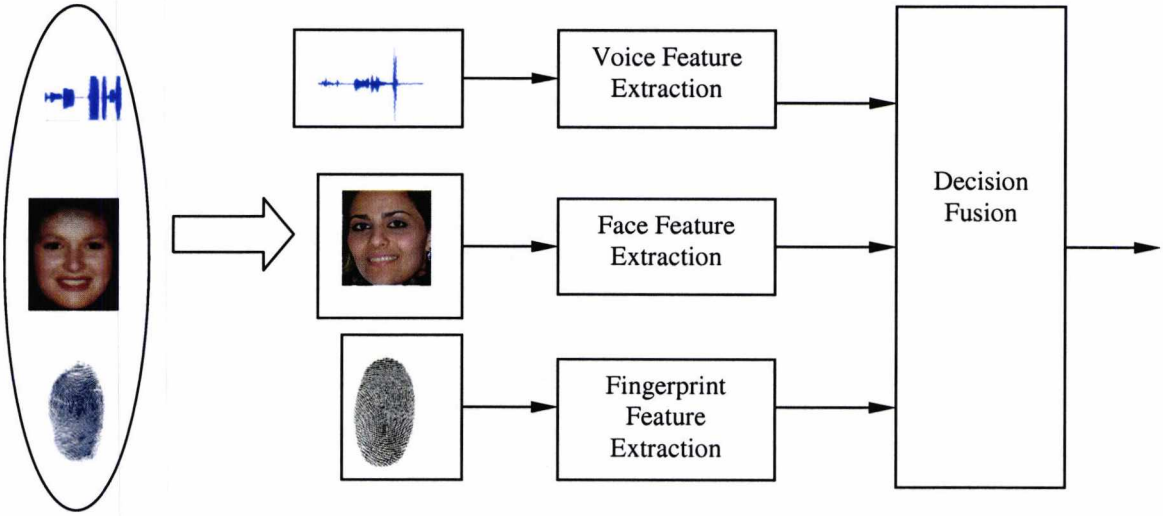



**Figure 5.6:** Impostor user

#### 5.5.2.1 Hard Decision Fusion

As stated previously, a hard decision is a decision made by the system that returns either accept or reject, in other words a 1 or a 0. The output impostor scores of the classifiers of the three modalities (fingerprint, voice and face) used in this project were transformed into hard decisions (0, 1) using the same method described in section 6.4.1.1. The three combination methods used in this section are the same as the ones used in section 6.4.1.1. However, the scenario presented is different.

##### 5.5.2.1.1 AND Fusion

In AND fusion [Kittler98], the identity claim of an impostor is accepted as being that of a legitimate user only if all the classifiers decide that the person under test is a client, that is, if the outputs returned by the classifiers in all three modalities is a 1 (accept). In other words, a false acceptance occurs if an impostor was successful in impersonating a legitimate user in all three modalities.

##### 5.5.2.1.2 Majority Voting

In majority voting [Dieckmann97], the identity claim of an impostor is accepted as being that of a legitimate user if the majority of the classifiers decide that the person

under test is a client, that is, if the output returned by the classifiers in any two modalities out of the three in the proposed system is a 1 (accept). In other words, a false acceptance occurs if an impostor was successful in impersonating a legitimate user in any two modalities.

#### 5.5.2.1.3 OR Fusion

In OR fusion [Kittler98], the identity claim of an impostor is accepted as being that of a legitimate user if at least one of the classifiers decide that the person under test is a client, that is, if the output returned by the classifiers in any single modality is a 1 (accept). In other words, a false acceptance occurs if an impostor was successful in impersonating a legitimate user in any single modality.

### 5.5.2.2 Soft Decision Fusion

As previously stated, a soft decision is a verification decision made by the system that normally generates a score generated in the range [0,1]. Since, the output of the classifiers of the three modalities (fingerprint, voice and face) used in this study were scores, they were combined using the following rule:

#### 5.5.2.2.1 Sum Rule

In the Sum Rule [Ross01], the impostor scores generated from the cross comparison are summed as shown in Equation 5.1, where the summed scores are then compared to a pre-specified threshold to reach the verification decision of whether the system will accept the impostor as being a legitimate user or reject him. The identity claim of an impostor is accepted as being a legitimate user if the summed score of all three classifiers is higher than the assigned pre-specified threshold.

## 5.6 Decision Fusion Error Rates

Both the genuine scenario and the impostor scenario described in the previous sections are used in this section to calculate the two main performance measures of the multimodal system; the false accept rate and the false reject rate. In this section an example of a building access control application is considered where the user

approaches an access point, insert his PIN number and provides live biometric data to the sensors installed. The sensors compare and match the data given by the user to the data present on the database and the system gives a decision of either accept or reject access for that particular individual. All users are given a maximum of three attempts to provide their correct biometric data to gain access. The decision made by the system to accept or reject an individual depends on the fusion method used. For this application two scenarios are considered:

The first scenario considers the user approaching the access point to be a legitimate user. The user provides his live biometric data for gaining access, but for some reason the system fails to recognize him. A second attempt is made by him to provide his correct biometric data and if the user still fails to be recognized by the system, a third and final chance is given to him. If the user fails to be recognized by the system after the third attempt then access is denied for that user.

The second scenario considers the user approaching the access point to be an impostor trying to spoof the system and gain access to the building. If the impostor fails to be recognized by the system, a second attempt is given to him to provide his biometric data and if he fails to be recognized by the system at that attempt, a third and final chance is given to him. Failing to be recognized by the system after the third attempt, the impostor is denied from accessing the building.

Considering the two scenarios mentioned, in the following subsections both the false reject rate (FRR) and the false accept rate (FAR) are calculated for different fusion methods. The three testing sets that were generated in Section 6.3 from the three accessed attempts undertaken at the second session of the data collection exercise are used to calculate the error rates, since each data set is regarded as an attempt to gain access.

It should be noted that whenever FRR in ($2^{nd}$ attempt) is mentioned it refers to the group of genuine users who failed to be recognised by the system at the first attempt and were given a second chance to provide their correct biometric data. While, the FRR in ($3^{rd}$ attempt) refers to the group of genuine users who failed to be recognised by the system at the first and second attempts and were given a third chance to

provide their correct biometric data. On the other hand, the FAR in (2nd attempt) refers to the group of impostors who failed to be recognised by the system at the first attempt and were given a second chance to provide their biometric data. While, the FAR in (3 rd attempt) refers to the group of impostors who failed to be recognised by the system at the first and second attempts and were given a third chance to provide their biometric data.

### 5.6.1 Performance of Hard Decision Fusion Methods

In this section both the false reject rate (FRR) and the false accept rate (FAR) are calculated for the different hard decision fusion methods. The three testing sets are regarded as the three attempts provided by the system to gain access.

#### 5.6.1.1 AND Fusion

In AND fusion, a decision is reached only when all the classifiers agree about it. AND fusion is mainly useful in situations where one would like to detect the presence of an event, with a low false acceptance bias, which means having a high FRR% and low FAR %. Table 5.2 shows the FRR and FAR when using this decision fusion method on the data gathered from the three commercial devices (fingerprint, voice and face).

**Table 5.2**: Error rates for AND fusion

| Attempts | 1st attempt (%) | | 2nd attempt (%) | | 3rd attempt (%) | |
|---|---|---|---|---|---|---|
| | FRR | FAR | FRR | FAR | FRR | FAR |
| Error rates | 59.2 ± 7.9 | 0.0 | 47.6 ± 8.1 | 0.0 | 44.9 ± 8.0 | 0.0 |

#### 5.6.1.2 Majority Voting

One of the simplest methods for combining classifiers is the majority voting strategy. In this method [Dieckmann97], a consensus is reached on the decision by

having a majority of the classifiers declaring the same decision. Table 5.3 shows the FRR and FAR when using this decision fusion method

**Table 5.3**: Error rates for majority voting

| Attempts | 1 attempt (%) | | 2 attempts (%) | | 3 attempts (%) | |
|---|---|---|---|---|---|---|
| | FRR | FAR | FRR | FAR | FRR | FAR |
| Error rates | 20.5 ± 6.5 | 0.05 ± 0.2 | 12.8 ± 5.4 | 0.07 ± 0.27 | 8.1 ± 4.4 | 0.08 ± 0.31 |

The downside to this approach is that an odd number of classifiers is required to prevent ties, which means that this approach would not be used if only two modalities were to be combined.

#### 5.6.1.3 OR Fusion

In OR fusion [Kittler98], a decision is made as soon as one of the classifiers makes a decision. OR fusion is mainly useful where one would like to detect the presence of an event with a low false rejection bias, which means having a low FRR% and high FAR %. Table 5.4 shows the FRR and FAR when using this method.

**Table 5.4:** Error rates for OR fusion

| Attempts | 1 attempt (%) | | 2 attempts (%) | | 3 attempts (%) | |
|---|---|---|---|---|---|---|
| | FRR | FAR | FRR | FAR | FRR | FAR |
| Error rates | 2.1 ± 2.3 | 3.6 ± 14.12 | 0.7 ± 1.3 | 4.9 ± 19.21 | 0.0 | 5.8 ± 22.7 |

It can be concluded that as the number of attempts increases the false reject rate (FRR) decreases and the false accept rate (FAR) increases. It was also noticed that although the OR rule had the best performance over the AND rule and majority

voting rules its main disadvantage is that it can only be used in applications which requires low security since it has the highest FAR.

### 5.6.2 Soft Decision Fusion Methods

As noted previously, a soft decision is a decision made by the system that generates a score. Since the three modalities (fingerprint, voice and face) used in this study had different score ranges, a normalization step was necessary to map all the raw scores from the different matchers into a common range [0, 1] before combining them in the fusion stage. The adaptive logarithmic method proposed in the previous chapter was used to map the different scores of the three modalities into a common range [0,1].

#### 5.6.2.1 Sum Rule

As previously stated, in the sum rule the scores from the different classifiers are summed. The summed score is then compared to a pre-specified threshold to reach a verification decision of whether to accept or reject the user [Duc97] [Kittler98] [Ross01]. Table 5.5 shows the results of the error rates when setting the pre-specified threshold to 0.5 and assigning equal weights to each modality.

**Table 5.5**: Error rates for the sum rule

| Attempts | 1 attempt (%) | | 2 attempts (%) | | 3 attempts (%) | |
|---|---|---|---|---|---|---|
| | FRR | FAR | FRR | FAR | FRR | FAR |
| Error rates | 31.3 ± 7.5 | 0.0 | 23.1 ± 6.8 | 0.0 | 19.0 ± 6.3 | 0.0 |

It was noticed that the majority-voting rule and the OR rule in the hard decision fusion performed better than the sum rule in the soft decision fusion in reducing the false reject rate (FRR) while both the sum rule and the AND rule performed well in reducing the false accept rate (FAR).

## 5.7 Characterising Individual System Users

In biometric systems it is important to know not only what works and to what extent it works, but also to be aware of the causes of errors, i.e. what does not work and why [Pankanti02]. The characterization of individual users that contribute to the overall biometric recognition system errors has received little attention.

Bolle et al. [Bolle00] suggested in his evaluation techniques for biometrics-based authentication systems that some measures to characterize the target population should be given.

Doddington et al [Doddington98] showed that the error rates vary across the population. It has led to the jocular characterization of the target population as being composed of "*sheep*" and "*goats*". In this characterization, the *sheep* for whom authentication systems perform reasonably well, are well behaved and dominate the population, whereas the *goats*, though in a minority, tend to determine the performance of the system through their disproportionate contribution of false reject errors. Like targets, impostors also have barnyard appellations, which follow from in homogeneities in impostor performance across the population. Specifically there are some impostors who have unusually good success at impersonating many different targets. These are called "*wolves*". There are also some targets that are easy to imitate and thus seem unusually susceptible to many different impostors. These are called "*lambs*".

The overall performance of any biometric system can be improved if some of the most difficult individuals (e.g. the "goats", the hard to match subjects) were to be excluded. Detecting theses individuals for whom the system performs poorly and dealing with them will result in an increase in the system performance.

In this section the four different terms that characterize the system users are measured using the data gathered in Chapter 3 to investigate their effects on the performance of a multimodal biometric system.

Table 5.6 shows a sample of the cross comparison matrix that was used to measure the four terms that characterize the system users. The subjects of the system are represented by the letters of the alphabet A, B, C, D. Templates refer to the subjects' enrolment templates and Samples refer to the verification sample provided by the subject. $S_{ij}$ represent the scores obtained from matching the samples against the templates, where i represent the rows and j the columns. The scores $S_{ij}$ (where i=j) on the diagonal determine whether a user is a sheep or goat. This is determined by means of an appropriate threshold. On the other hand, the scores $S_{ij}$ (where i≠j) above and below the diagonal may indicate the presence of a wolf or a lamb, which are determined by means of an appropriate threshold.

**Table 5.6**: Characterising Individual User Matrix

| Samples / Templates | A | B | C | D |
|---|---|---|---|---|
| A | $S_{11}$ | $S_{12}$ | $S_{13}$ | $S_{14}$ |
| B | $S_{21}$ | $S_{22}$ | $S_{23}$ | $S_{24}$ |
| C | $S_{31}$ | $S_{32}$ | $S_{33}$ | $S_{34}$ |
| D | $S_{41}$ | $S_{42}$ | $S_{43}$ | $S_{44}$ |

For clarification consider Table 5.7. The scores presented in the cross comparison matrix were generated in the same way as in Table 5.1, that is, the scores on the diagonal are generated by matching the verification sample of each subject with his own template, on the other hand, the scores above and below the main diagonal are generated by comparing all the verification access samples to all the templates except the matching one.

**Table 5.7**: Example of Characterising Individual User Matrix

| Samples / Templates | A | B | C | D | E |
|---|---|---|---|---|---|
| A | **70** | 10 | *2* | 20 | *6* |
| B | *54* | **60** | 5 | 15 | *52* |
| C | *52* | 32 | **80** | 2 | 10 |
| D | 35 | 8 | 19 | **30** | *56* |
| E | *7* | 5 | 3 | 1 | **20** |

In this example consider the threshold to be set to 50, considering the scores on the diagonal, the scores for subjects A, B and C are higher than the pre-specified threshold which means that they performed reasonably well and were accepted by the system. These subjects are referred to as the *sheep* of the system. While, both subjects D and E are referred to as the *goats* of the system each having a score lower than the pre-specified threshold. On the other hand, the scores above and below the main diagonal show that the sample of subject A (when compared to the templates of the other subjects) was accepted by the system as being of both subjects B and C since the scores obtained were higher than the pre-specified threshold with values 54 and 52 respectively and that the sample of subject E (when compared to the templates of the other subjects) was accepted as subjects B and D since the scores obtained were higher than the pre-specified threshold with values 52 and 56 respectively. Both subjects A and E are referred to as the *wolves* of the system since their samples are strong enough to successfully impersonate other subjects. Subjects B, C and D are referred to as *lambs* since their templates were easily imitated by different impostors such as A and E.

As it was previously suggested, knowing the causes that affect the performance of the system and dealing with them could result in an improvement in the system performance.

In the next subsections an investigation is provided based on the data gathered in Chapter 3 on the performance of the well-behaved majority, which are the *sheep* of the system and the troublesome minorities, which are the *goats*, *wolves* and *lambs* of

the system. Some proposed ways of reducing the troublesome users is also investigated.

### 5.7.1 The Sheep

*Sheep* are the group of subjects that dominate the population and for which authentication systems perform reasonably well. The performance of the system depends on the proportion of *sheep* in the system since, the higher is the proportion of *sheep*, the lower is the proportion of *goats* and hence, the lower is the false reject rate (FRR) which is an important factor in the system performance. Table 5.8 shows the proportion of users who represent the *sheep* measured in each modality. Table 5.9 shows the proportion of users who represent the *sheep* of the system measured under the different decision fusion rules.

The proportion of *sheep* measured under the different decision fusion rules in Table 5.9 were evaluated as follow:

In the AND rule the group of subjects who were referred to as *sheep* in all the three modalities (fingerprint, voice and face) were calculated, in the majority voting rule the group of subjects who were referred to as *sheep* in any two modalities were measured and in the OR rule the group of subjects who were referred to as *sheep* in any single modality were calculated.

**Table 5.8**: Proportion of sheep in each modality

| **Modality** | **Sheep of the system (%)** | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Total** | 72.1 ± 7.2 | 81.6 ± 6.3 | 63.9 ± 7.8 |

**Table 5.9**: Proportion of sheep under the decision fusion rules

| **Fusion Rule** | **Sheep of the system (%)** | | | |
|---|---|---|---|---|
| | AND rule | Majority Voting | OR rule | Sum rule |
| **Total** | 40.1 ± 7.9 | 83.7 ± 6.0 | 95.9 ± 3.2 | 68.7 ± 7.5 |

From Table 5.8, the voice modality seems to have the highest proportion of *sheep* among the three modalities. It was also noticed that when comparing the proportion of *sheep* found in each single modality with the proportion of *sheep* found under the different decision fusion rules there were more *sheep* when using either the majority voting rule or the OR rule than there were when using any single modality. This suggests that using more than one modality increases the proportion of *sheep* and hence decreases the false reject rate (FRR) of the system. It was also realised that when comparing the results of the hard decision fusion with that of the soft decision fusion, both the majority-voting rule and the OR rule in the hard decision fusion performed better in increasing the proportion of *sheep* in the system than the sum rule in the soft decision fusion. The results also demonstrates that among the fusion rules the AND rule seem to decrease the proportion of *sheep* in the system thus increasing the false reject rate (FRR) of the system.

### 5.7.2 The Goats

*Goats* are the group of subjects whose pattern of activity when interfacing with the system varies beyond the specified range allowed by the system, and who consequently may be falsely rejected by the system. The *goats* decrease the performance of the system; the higher is the proportion of *goats* in the system the higher is the false reject rate. Since, it was desired to calculate the proportion of users who are consequently falsely rejected by the system, three attempts were considered for each modality such that the user is regarded as a *goat* if he is falsely rejected by the system in all three attempts. Table 5.10 shows the proportion of users who represents the *goats*, measured in each modality. Table 5.11 shows the proportion of users who represents the *goats* of the system measured under the different decision fusion rules.

The proportion of *goats* measured under the different decision fusion rules in Table 5.11 were evaluated as follow:

The AND rule measured the group of subjects who were *goats* in all the three modalities (fingerprint, voice and face), the majority voting rule measured the group

of subjects who were *goats* in any two modalities and the OR rule measured the group of subjects who were *goats* in any single modality.

**Table 5.10**: Proportion of goats in each modality

| Gender | Goats of the system (%) | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Total** | 15.6 ± 5.9 | 3.4 ± 2.9 | 36.1 ± 7.8 |

**Table 5.11**: Proportion of goats under the decision fusion rules

| Fusion Rule | Goats of the system (%) | | | |
|---|---|---|---|---|
| | AND rule | Majority Voting | OR rule | Sum rule |
| **Total** | 44.9 ± 8.0 | 8.1 ± 4.4 | 0.0 | 19.0 ± 6.3 |

It was observed from Table 5.10 that the voice modality had the lowest proportion of *goats* among the three modalities. Comparing Table 5.10 and Table 5.11 suggests that using multiple modalities decreases the proportion of *goats* in the system and hence increases the performance of the system since the proportion of *goats* found under the different decision fusion rules were less than the proportion of *goats* found in each modality alone with the exception of the AND rule. Comparing the results of the different fusion rules demonstrated that both the majority-voting rule and the OR rule in the hard decision fusion performed better in decreasing the proportion of *goats* in the system than the sum rule in the soft decision fusion. Finally, the results showed that the AND rule seems to increase the proportion of *goats* in the system thus decreasing the performance of the system.

### 5.7.3 The Lambs

*Lambs* are the group of subjects who are exceptionally vulnerable to impersonation. The *lambs* affect the performance of the system, the higher is the proportion of *lambs* the less secure is the system since it means that either the users have a

relatively weak biometric data that can be impersonated by others or the impostors impersonating them have a strong biometric data. Both cases affects the performance of the system by increasing the false accepts rate (FAR) which is another important factor in the system performance. Table 5.12 shows the proportion of users who represents the *lambs*, measured in each modality. Table 5.13 shows the proportion of users who represents the *lambs* of the system measured under the different decision fusion rules.

The proportion of *lambs* measured under the different decision fusion rules in Table 5.13 were evaluated as follow:

In the AND rule the group of subjects who were *lambs* in all the three modalities (fingerprint, voice and face) were calculated, in the majority voting rule the group of subjects who were *lambs* in any two modalities were measured and in the OR rule the group of subjects who were *lambs* in any single modality were measured.

**Table 5.12**: Proportion of lambs in each modality

| Gender | Lambs of the system (%) | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Total** | 0.0 | 70.7 ± 7.4 | 48.3 ± 8.1 |

**Table 5.13**: Proportion of lambs under the decision fusion rules

| Gender | Lambs of the system (%) | | | |
|---|---|---|---|---|
| | AND rule | Majority Voting | OR rule | Sum rule |
| **Total** | 0.0 | 5.4 ± 3.7 | 86.4 ± 5.5 | 0.0 |

Table 5.12 shows that the fingerprint modality is the most secure system among the three modalities used in this work with a no *lambs* in the system. It also shows the voice modality being is the most vulnerable modality to impersonation among the three. Comparing Table 5.12 and Table 5.13 shows that the proportion of *lambs*

present in each single modality is more than the proportion of *lambs* present when using the different fusion rules (with the exception of the OR rule) which supports the idea that combining multiple modalities improve the performance of the system. Table 5.13 illustrates that a user can be a *lamb* in a single modality (shown in the OR rule) or a *lamb* in two modalities (shown in the majority voting rule) but cannot be a *lamb* in three modalities (shown in the AND rule). It was realised that the sum rule in soft decision fusion resulted in an elimination of the *lambs* from the system.

### 5.7.4 The Wolves

*Wolves* are the group of subjects that are successful at impersonating others. The *wolves* decrease the performance of the system, the higher is the proportion of *wolves* in the system, the higher is the false acceptance rate (FAR) and the less secure is the system. A user impersonating others can have two possible explanations either the user have a strong biometric data or the impersonated subjects have a weak biometric data. Table 5.14 shows the proportion of users who represents the *wolves*, measured in each modality. Table 5.15 shows the proportion of users who represents the *wolves* of the system measured under the different decision fusion rules.

The proportion of *wolves* measured under the different decision fusion rules in Table 5.14 were evaluated as follow:

The group of subjects who were *wolves* in all the three modalities (fingerprint, voice and face) were calculated by the AND rule, the group of subjects who were *wolves* in any two modalities were calculated using the majority voting rule and the group of subjects who were *wolves* in any single modality were calculated using the OR rule.

**Table 5.14**: Proportion of wolves in each modality

| Gender | Wolves of the system (%) | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| **Total** | 0.0 | 53.0±8.0 | 44.0± 8.0 |

**Table 5.15**: Proportion of wolves under the decision fusion rules

| Gender | Wolves of the system (%) | | | |
|---|---|---|---|---|
| | AND rule | Majority Voting | OR rule | Sum rule |
| **Total** | 0.0 | 6.1 ± 3.8 | 71 ± 7.3 | 0.0 |

Table 5.14 shows that the fingerprint modality is the most secure modality among all three, the fact that there were no *wolves* at all means that it was difficult to impersonate the biometric data of any user. Combining multiple modalities increase the performance of the system since it is quite difficult to impersonate a user in more than one modality. Table 5.14 and 5.15 show that the proportion of *wolves* under different decision rules (with the exception of the OR rule) is less than the proportion of *wolves* in each single modality. It can be seen that both the AND rule in the hard decision fusion and the sum rule in the soft decision fusion eliminated the *wolves* from the system.

### 5.7.4.1 Types of Wolves

As previously mentioned, the *wolves* decrease the system performance and cause the existence of *lambs* in the system. Knowing the *wolves* and their types can help in dealing with them and hence increase the system performance. In this section four types of *wolves* are proposed, which are divided into the following categories: -

1. **Type A**
   A user impersonating only ***one subject*** in a ***single modality***
2. **Type B**
   A user impersonating ***two or more subjects*** in a ***single modality***

3. **Type C**
   A user impersonating only ***one subject*** in ***two modalities*** *simultaneously*

4. **Type D**
   A user impersonating ***two or more subjects*** in ***two modalities*** *simultaneously*

Type A and Type B are grouped together since both of them are dealing with a user impersonating others in a single modality while Type C and Type D are grouped together both dealing with a user impersonating others in two modalities simultaneously. Investigating if a user could impersonate others in more than two modalities was not considered since the results from Table 5.15 showed that no *wolves* were found in all three modalities. Table 5.16 shows the wolves of Type A and Type B. Table 5.17 show the wolves of Type C and Type D.

**Table 5.16**: Proportion of wolves of Type A and B

| Type of wolves | Wolves in the system (%) | | |
|---|---|---|---|
| | Fingerprint | Voice | Face |
| Type A | 0.0 | 17.7 ± 6.2 | 7.5 ± 4.3 |
| Type B | 0.0 | 34.0 ± 7.7 | 36.0 ± 7.8 |

It was striking to realize that the proportion of users who had the ability to impersonate two or more subjects (Type B) is more than the proportion of users who had the ability to impersonate only one subject (Type A). There are two possible explanations for this, either the subjects who have been impersonated (*lambs*) have a weak biometric data (for example due to template ageing) or the users who are impersonating others (*wolves)* have a very strong biometric data which enables them to impersonate more than one person. Table 6.16 also showed that the fingerprint modality is more secure than the other modalities with no *wolves* of Type A or Type B being present.

**Table 5.17**: Proportion of wolves of Type C and D

| Type of wolves | Wolves in the system (%) | | |
|---|---|---|---|
| | Fingerprint & Voice | Voice & Face | Face & Finger |
| Type C | 0.0 | 5.4 ± 3.7 | 0.0 |
| Type D | 0.0 | 0.7 ± 1.3 | 0.0 |

*Wolves* Type C and Type D are "stronger" *wolves* since they have the ability of personating other subjects in two modalities simultaneously. In Table 5.17 it can be seen that in any combination of two modalities including the fingerprint modality, the result was an elimination of the *wolves*. This shows that the fingerprint modality is very robust and combining it with any other modality will increase the performance and security of the system. This table also shows that there were *wolves* that were able to impersonate two different subjects in two modalities (face and voice) simultaneously.

## 5.8 Discussion

A number of interesting points may be drawn from the above analysis regarding the combination of multiple modalities using the decision fusion rules. The results in this chapter are based on the comparison between the performance of the hard decision fusion rules (AND rule, OR rule and majority voting rule) and the performance of the soft decision fusion rule (sum rule). It is also based on investigating the effect of characterising the individual users as *sheep*, *goats, lambs* and *wolves*. According to the experimental results of this study, the false reject rate (FRR) of the system is reduced more by using the majority-voting rule and the OR rule in the hard decision fusion than by using the sum rule in the soft decision fusion, whereas the false accept rate (FAR) of the system is reduced to zero by using only the AND rule in the hard decision fusion and the sum rule in the soft decision fusion. A general conclusion can be drawn that the sum rule in the soft decision fusion performed better than the hard decision fusion rules for the present system

since it reduced the false accept rate (FAR) to zero and the false reject rate (FRR) to a level which is acceptable for many applications.

Considering the characterization of the individual users as *sheep*, *goats*, *lambs* and *wolves* and their effect on the performance of the system, several conclusions may be made. Considering initially the users who were characterized as being the *sheep* of the system, the results illustrated that the higher is the proportion of *sheep* in the system the better is the performance since it results in a reduction in the FRR. On the other hand, the results showed that the *goats* decrease the performance of the system since they lead to false reject and the less is the proportion of *goats* the better is the performance of the system. *Lambs* and *wolves* decrease the performance of the system since they lead to false accept. The less is the proportion of *lambs* and *wolves* the better is the performance of the system. Bearing in mind theses result, the three modalities (fingerprint, voice and face) adopted in this study were analysed. The fingerprint modality seemed to outperform the other two modalities by having a relatively low proportion of *goats* (15.6 %) and a zero number of *lambs* and *wolves* in its system. Although the voice modality had the lowest proportion of *goats* among the three modalities it also had the highest proportion of *lambs* and *wolves*, which suggest that the voice modality is the most vulnerable to impersonation among the three modalities. The face modality showed the worst performance among the three modalities by having the highest proportion of *goats* and almost 50 % of the users were *lambs* and *wolves*

The study of the effects of the hard decision fusion rules (AND rule, OR rule and majority voting) and the soft decision fusion rule (sum rule) on the characterization of the users as *sheep, goats, lambs* and *wolves* suggests that the AND rule eliminates the *lambs* and *wolves*, but increase the proportion of *goats* in the system. On the hand, the OR rule appear to decrease the proportion of *goats*, but increases the *lambs* and *wolves* in the system. The majority voting seems to outperform the AND rule and the OR rule in providing an almost acceptable proportion of *goats, lambs* and *wolves.* The sum rule seems to have both the benefit of the AND rule in diminishing the *lambs* and *wolves* and that of the majority voting in reducing the *goats* to an acceptable number.

The intuitive assumption that combining multiple modalities increases the performance of the system, was supported by the fact that the results obtained by fusing the different modalities, under the different decision fusion rules were better than when using single modalities. The fusion of multiple modalities proved to reduce the proportion of *lambs*, *goats* and *wolves* in the system thus increasing the performance of the system.

## 5.9 Summary

In this chapter a comparison between the performance of hard decision fusion and soft decision fusion in multimodal biometric systems was made. The results showed that the hard decision fusion outperformed the soft decision fusion in reducing the false reject rate, while the soft decision performed equally well as the AND rule in reducing the false accept rate (FAR) to zero. The effect of characterizing the individual users as *sheep, goats, lamb* and *wolves* on the performance of the system was investigated and different types of *wolves* were proposed. The experimental results suggested that the performance of the system could be improved if the proportion of *lambs*, *goats* and *wolves* are reduced.

The next chapter provides a general overview of genetic algorithms (GAs) and proposes exploiting it for optimising the performance of multimodal biometric recognition systems.

# Chapter 6

# Introduction to Genetic Algorithms

## 6.1 Introduction

This chapter starts by giving a general overview of the genetic algorithms (GAs) and their different parameters. The reasons behind using genetic algorithms as an optimisation technique instead of other alternatives is then provided. Finally, a description of some application areas for GAs is provided and a proposal for exploiting them in the field of biometric system optimisation is presented.

## 6.2 What are Genetic Algorithms?

Genetic Algorithms were first proposed by John Holland in the 1960s and further developed by Holland and his students and colleagues at the University of Michigan in the 1960s and 1970s [Holland75]. GAs are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetics. The basic techniques of GAs are designed to simulate processes in natural systems necessary for evolution, especially those that follow the principles first laid down by Charles Darwin in his concept of "Survival of the Fittest" since, in nature, competition among individuals for scanty resources results in the fittest individuals dominating

over the weaker ones. GAs typically maintain a constant-sized population of individuals ('chromosomes'), which represent samples from the space to be searched [Davis91]. Each individual is evaluated on the basis of its overall fitness with respect to some pre-specified functionality and across some particular application domain. New individuals (samples from the search space) are produced by selecting high performing individuals to produce 'offspring' which retain many of the features of their 'parents'. The result is an evolving population, which exhibits progressively improved fitness with respect to the given functionality ('goal'). Figure 6.1 outlines the key features of a typical genetic algorithm. A population of individual structures is initialised and then evolved from generation *t* to generation *t*+1 by repeated applications of fitness evaluation, selection, crossover and mutation [Dixon78]. In the following sections a description of each of these parameters is given in details [Goldberg89].

```
t = 0;  /* Initial Generation */
Population_Initialise (t);
Fitness Evaluation (t);
Repeat
        t = t + 1; /* Next Generation */
        Selection (t);
        Crossover (t);
        Mutation (t);
        Fitness Evaluation (t);
        Reinsertion (t);
Until best individuals meets criterion;
```

**Figure 6.1**: A Simple Genetic Algorithm

### 6.2.1 Population Representation

Genetic algorithms operate on a population of strings [Back93]. Each string "also called *chromosome*" represents one possible solution in the searching space for a particular problem [Rawlins91]. Each chromosome represents a set of parameters called *genes* where each gene corresponds to a feature of the problem and has its own position in the chromosome, which is called *locus*. Each gene is encoded by a given number of *allele*. The allele can be represented by binary, real number or other forms and its range is usually defined by the problem specified. Chromosomes are represented by different encoding types depending on the problem being explored. In this section we will discuss the different types of encoding of theses chromosomes.

#### 6.2.1.1 Binary Encoding

Binary encoding is the most commonly used representation of chromosomes in Genetic algorithms [Bramlette91]. In this type of encoding, the chromosomes consist of a string of 0's and 1's. Each chromosome consists of "genes", with each gene being represented by a number of alleles (i.e. 0,1). Figure 6.2 shows the individual structures in the population

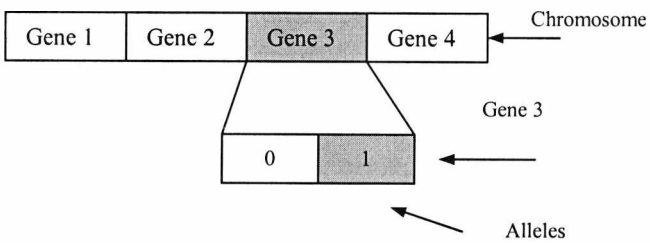



**Figure 6.2**: Chromosome with binary encoding

Each bit in the string can represent some characteristic of the solution or it could represent whether or not some particular characteristic was present.

Although binary encoding has several advantages, such as its relative simplicity and its ability of generating many possible chromosomes, even with a small number of genes, its main drawback is that it is often not a natural way of representation for many problems and sometimes corrections must be made after crossover and/or mutation.

#### 6.2.1.2 Permutation Encoding

Permutation encoding is normally used in ordering problems, such as the travelling salesman problem (TSP), or task ordering problems [Lucasius92]. In this type of encoding, chromosomes are represented by strings of numbers that represent a position in a sequence. Figure 6.3 shows the chromosome with permutation encoding.

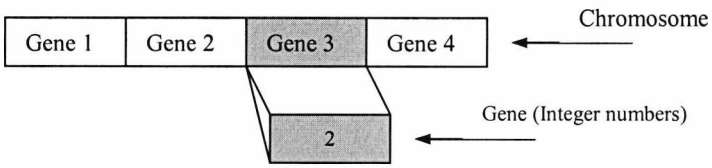



**Figure 6.3**: Chromosome with permutation encoding

In the TSP each number would represent a city to be visited.

The main drawback of this type of encoding is that sometimes corrections must be made after crossover and/or mutation to leave the chromosome consistent (e.g. having a real sequence of the cities to be visited in the travelling salesman problem).

#### 6.2.1.3 Value Encoding

Direct value encoding is used in problems where some more complicated values such as real numbers are used. The use of real-valued genes in GAs is claimed by Wright [Wright91] to offer a number of advantages in numerical function optimisation over binary encoding. Efficiency of the GAs is increased, as there is no

need to convert chromosomes into a binary representation before each function evaluation; and less memory is required as efficient floating-point internal computer representations can be used directly [Michalewicz92]. In this type of encoding, every chromosome is a sequence of some values. These values, apart from being real numbers, can sometime be characters such as A, B or any labels such as "back, left". Figure 6.4 shows the chromosome with value encoding.

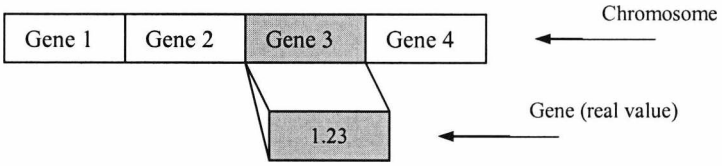



**Figure 6.4**: Chromosome with value encoding

Value coding is a good choice for some special problems where the use of binary coding for these problems would be difficult, such as finding weights for a neural network where the real values in the chromosomes represent weights in the neural network. However, for this encoding it is often necessary to develop some new crossover and mutation operations specific for the problem.

### 6.2.2 The Objective and Fitness Function

The objective function is used to provide a measure of how individuals have performed in the problem domain [Whitley93]. In the case of a minimization problem, for example, the fit individuals will be those, which have the lowest numerical values of the associated objective function. This raw measure of fitness is usually used as an intermediate stage in determining the relative performance of individuals in a genetic algorithm. Another function, the *fitness function*, is used to transform the objective function value into a measure of relative fitness [De Jong75], thus:

$$F(x) = g(f(x)) \qquad (6.1)$$

where $f$ is the objective function, $g$ transforms the value of the objective function to a non-negative number and $F$ is the resulting relative fitness. The two common transformation processes will be described in the following subsection.

### 6.2.3 Selection

Selection is the stage where the individuals of the population compete among each other to become parents of the next generation [Blickle95]. The fitter the member of the population the more likely it is to produce an offspring. There are many different types of selection operators. One common approach always selects the "fittest" solution and discards the worst, but there are hundreds of variants of this scheme [Goldberg89] [Baker85]. None is right or wrong in absolute terms. In fact, some will perform better than others depending on the problem domain being explored.

The first step in the selection stage is to transform the objective function value into a measure of relative fitness as mentioned in the previous subsection. This is performed either by:

- Fitness Scaling or
- Fitness Ranking

**Fitness Scaling: -**

This transformation method was suggested by Goldberg in [Goldberg89]. In this method the objective values of a population is scaled into a fitness measure by using the following linear transformation

$$f' = af + b \qquad (6.2)$$

where $f$ is the objective value of an individual , $a$ is a positive scaling factor if the optimisation is for maximizing and negative if it is for minimizing. The offset $b$ is

used to ensure that the resulting fitness values are non-negative and $f'$ is the resulting scaled fitness value of an individual.

To maintain a certain relationship between the maximum fitness individual in the population and the average population fitness, the following constraint equations are used:

$$f'_{\max} = f_{avg} * C_s \qquad (6.3)$$

$$f'_{avg} = f_{avg} \qquad (6.4)$$

where $f'$ is the scaled maximum fitness, $f'_{avg}$ is the scaled average fitness of the population, $f_{avg}$ is the average objective value of the population and $C_s$ is a scaling constant that specifies the expected number of copies of the best individual in the next generation. Increasing $C_s$ will increase the selection pressure (bias towards best individual and quicker convergence), decreasing $C_s$ will decrease the selection pressure. The linear coefficients $a$ and $b$ are calculated from the given constraint equations.

Using linear scaling, the expected number of offspring is approximately proportional to that individual performance As there is no constraint on an individual's performance in a given generation, highly fit individuals in early generations can dominate the reproduction causing rapid convergence to possibly sub-optimal solutions. Similarly, if there is a little deviation in the population, then scaling provides only a small bias towards the most fit individual.

**Fitness Ranking: -**

This transformation method was suggested by Baker in [Baker85]. This method overcomes the reliance on an extreme individual. Ranking introduces a uniform scaling across the population and provides a simple and effective way of controlling selective pressure [Whitley89]. (Selective pressure indicates the probability of the best individual being selected compared to the average probability of selection of all individuals). In this method individuals are sorted in order of their objective values and then reproductive fitness values are assigned according to rank [Back91]. The

fitness assigned to each individual depends only on its position in the individuals rank and not on the actual objective value. The fitness of individuals in the population is calculated as:

$$Fitness(Pos) = 2 - SP + 2(SP - 1)(Pos - 1)/(N - 1) \qquad (6.3)$$

Where $N$ is the number of individuals in the population, $Pos$ is the position of an individual in this population (least fit individual has $Pos$ =1, the fittest individual $Pos = N$) and $SP$ is the selective pressure normally in the range [1.0-2.0].

Lets take the following example, where a single chromosome has an objective value far in excess of the others, which means that the other chromosomes will have very few chances to be selected. The fitness ranking is better in these cases than the fitness scaling approach, it will operate by ranking the population and then assigning each chromosome a fitness value from this ranking. The worst will have fitness 1, second worst 2 etc. and the best will have fitness N (number of chromosomes in population). This provides a chance for all the chromosomes to be selected. Table 6.1 and 6.2 shows the objective values and fitness values (after applying the fitness ranking method) of four chromosomes respectively

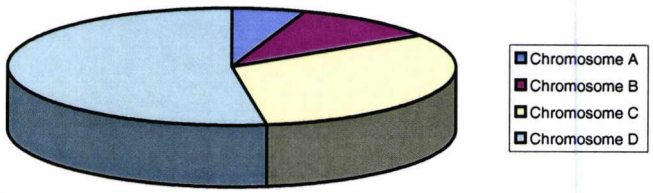
**Table 6.1** :Objective values of individuals

| | |
|---|---|
| Chromosome A | 1 |
| Chromosome B | 2 |
| Chromosome C | 6 |
| Chromosome D | 10 |

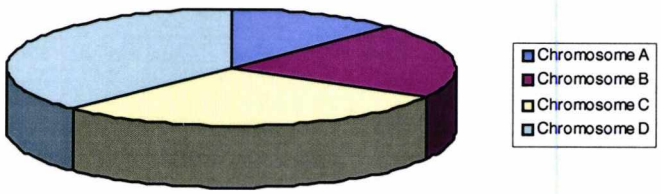
**Table 6.2:** Fitness values of individuals with SP =1.5

| | |
|---|---|
| Chromosome A | 0.5 |
| Chromosome B | 0.83 |
| Chromosome C | 1.16 |
| Chromosome D | 1.5 |

Figure 6.5 shows how the situation changes after applying the ranking method.




Situation before ranking




Situation after ranking

**Figure 6.5**: Rank-based fitness assignment

The drawback of this method is that it can lead to slower convergence, because the best chromosomes do not differ so much from other ones.

The actual selection is performed in the next step where parents are selected according to their fitness. The two main selection methods are described in the following subsections.

#### 6.2.3.1 Roulette Wheel Selection

Roulette Wheel Selection is the most commonly used selection technique [Goldberg89]. It can be regarded as allocating each of the population members a Pie-shaped slice on a roulette wheel, with each slice proportional to the member's fitness value. Selection of a population member to be a parent can then be viewed as a spin of the wheel, with the winning population member being the one in whose slice the roulette spinner ends up. Although this selection procedure is random, each parent's chance of being selected is directly proportional to its fitness. The least fit members will gradually be driven out of the population [Bin Azhar02]. Figure 6.6 illustrates the idea of the roulette wheel and it is obvious from this example that Chromosome 3 has a good chance of being selected more than once, and this shows that the stronger chromosomes will begin to dominate, eradicating the weaker ones from the population.

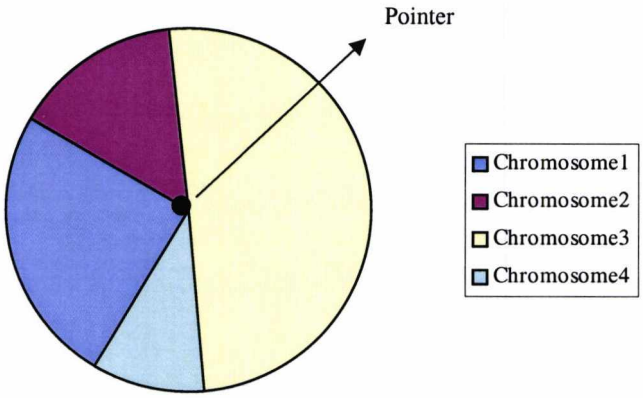



**Figure 6.6**: Roulette Wheel Selection

#### 6.2.3.2 Universal Stochastic Sampling

Universal Stochastic Sampling (USS) is a single-phase sampling algorithm with minimum spread and zero bias [Baker87]. Bias is defined as the absolute difference between an individual's actual and expected selection probability. Zero bias indicates that an individual's selection probability equals its expected number of trials. Spread is the range in the possible number of trials that an individual may

achieve. Instead of the single selection pointer employed in roulette wheel method, USS uses N equally spaced pointers, where N is the number of selections required. The population is then shuffled randomly and a single number is generated, *num*. This indicates the position of the first pointer. The N individuals are then chosen by generating the N pointers spaced by 1/N, [*num*, *num*+1/N,...,*num*+ (N-1)/N] and selecting the individuals whose fitness span the positions of the pointers. The number of copies that an individual gets is equal to the number of pointers that lie within the corresponding slot. As individuals are selected entirely on their position in the population, USS has a zero bias [Mitchell96]. Figure 7.7 illustrates the idea of the USS. In this example there are four chromosomes, so there will be four pointers, after spinning the wheel chromosome D will have two copies since there are two pointers within its slot, both chromosome B and chromosome C will have one copy each and chromosome A will have none.

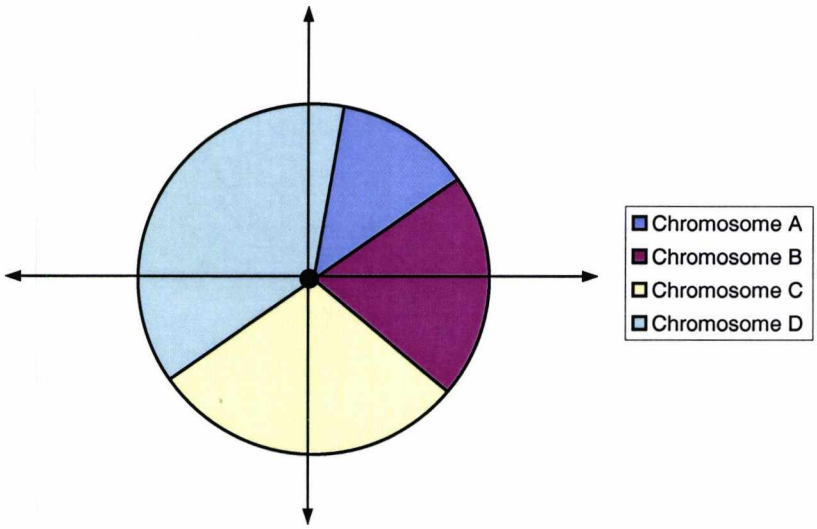



**Figure 6.7**: Universal Stochastic Sampling

### 6.2.4 Genetic Operators

Selection alone cannot introduce any new individuals into the population, i.e., it cannot introduce new points in the search space. These are generated by genetically inspired operators, of which the most well known are *crossover* and *mutation* [Spears98].

The crossover and mutation operators are the most important part of a genetic algorithm and are the main influence on the performance of the algorithm [Muhlenbein95]. Usually, there is a predefined probability of procreation associated with each of these operators. Traditionally, these probability values are selected such that crossover is the most frequently used, with mutation being resorted to only relatively rarely. This is because the mutation operator is a random operator and serves to introduce diversity into the population. The kind of operator to be applied to each member of the gene pool is determined by random choice based on these probabilities. Of the two operators, mutation involves only a single parent and results in the creation of a single offspring. The crossover operator involves two parents and generates two offsprings.

#### 6.2.4.1 Crossover

Crossover is not usually applied to all pairs of individuals selected for mating [Bremermann62]. A random choice is made, where the likelihood of crossover is applied. If crossover is not applied, offspring are produced simply by duplicating the parents. In this subsection we will describe the different types of crossover operators.

**Single point Crossover**

Single point crossover is the simplest form of crossover. It operates by randomly selecting a single cutting point in the two selected parents' chromosomes, resulting in the production of two "head" segments and two "tail" segments. The tail segments are then swapped over to produce two new full-length chromosomes [Muhlenbein95]. The two offspring each inherit some genes from each parent. Figure 6.8 illustrates the single point crossover.

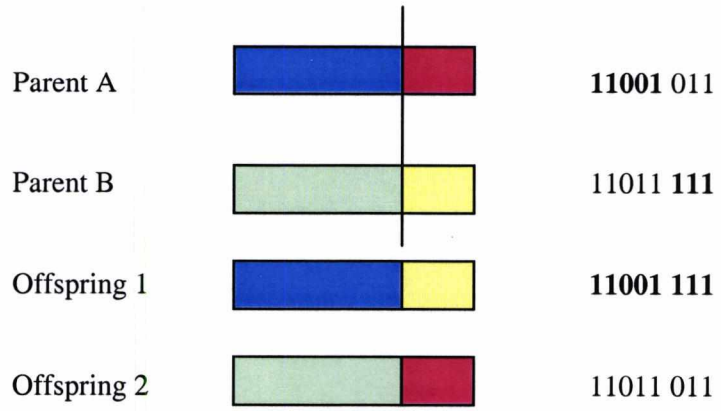



**Figure 6.8**: Single point crossover

**Multi-point Crossover**

This crossover operator was first introduced by De Jong in [De Jong75]. It involves the division of the original string parents into m cut-points, and then the bits between successive crossover points are exchanged between the two parents to produce two new offspring. This process is illustrated in Figure 6.9.

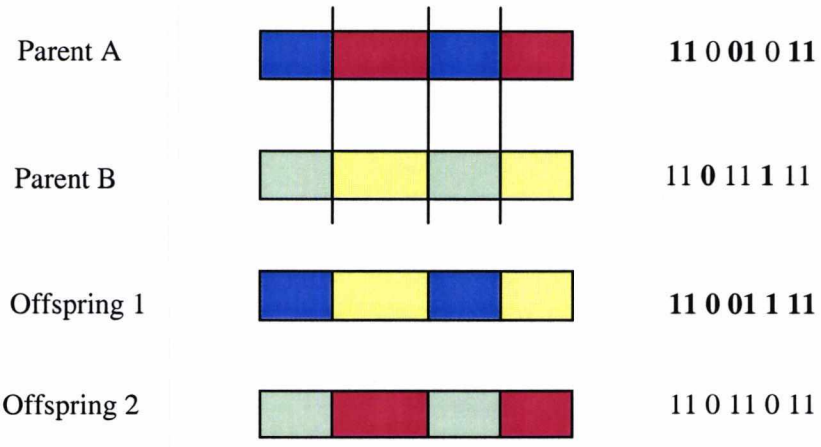



**Figure 6.9**: Multi-point crossover (m=3)

The idea behind multi-point crossover, is that the parts of the chromosome representation that contribute most to the performance of a particular individual may

not necessarily be contained in adjacent sub strings [Booker87]. Further, the disruptive nature of multi-point crossover appears to encourage the exploration of the search space, rather than favoring the convergence to highly fit individuals early in the search, thus making the search more robust [Spears91].

**Uniform Crossover**

This crossover operator was introduced by Syswerda in [Syswerda89]. It does not use cut-points but instead creates offspring by using a crossover mask, which is created at random. The parity of the bits in the mask indicates which parent will supply the offspring with which bits. The following example illustrates the process, Consider the following two parents, crossover mask and resulting offspring:

| | |
|---|---|
| P1 = | 1 0 1 1 0 0 0 1 1 1 |
| P2 = | 0 0 0 1 1 1 1 0 0 0 |
| Mask = | 0 0 1 1 0 0 1 1 0 0 |
| O1 = | 0 0 1 1 1 1 0 1 0 0 |
| O2 = | 1 0 0 1 0 0 1 0 1 1 |

Here, the first offspring, O1, is produced by taking the bit from P1 if the corresponding mask bit is 1 or the bit from P2 if the corresponding mask bit is 0. Offspring O2 is created using the inverse of the mask or, equivalently, swapping P1 and P2.

Uniform crossover, like multi-point crossover, has been claimed to reduce the bias associated with the length of the binary representation used and the particular coding for a given parameter set. This helps to overcome the bias in single-point crossover towards short substrings without requiring precise understanding of the significance of individual bits in the chromosome representation.

**Intermediate Recombination**

This crossover operator is used when given a real-valued encoding of the chromosome structure [Mühlenbein93]. It is a method of producing new chromosomes around and between the values of the parent chromosomes. Offspring are produced according to the rule:

$$Offspring = \alpha.Parent_1 + Parent_2(1-\alpha) \tag{6.3}$$

where $\alpha$ is a scaling factor chosen uniformly at random over some intervals, typically [-0.25, 1.25]. Each variable in the offspring is the result of combining the variables in the parents according to the above expression with a new $\alpha$ chosen for each pair of parents genes. In geometric terms, intermediate recombination capable of producing new variables within a slightly larger hypercube than that defined by the parents but constrained by the range of $\alpha$ as shown in Figure 6.10

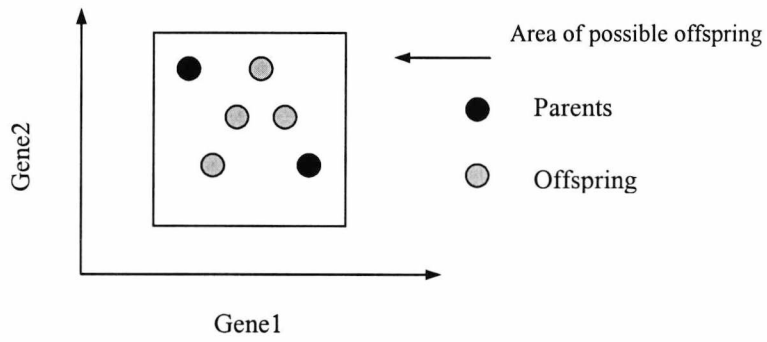


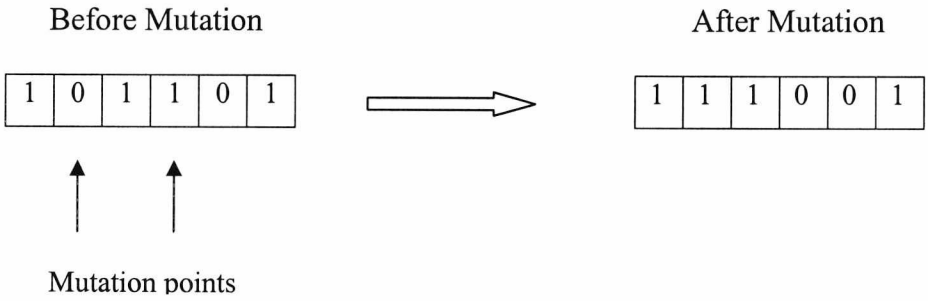

**Figure 6.10**: Geometric effect of Intermediate Recombination

### 6.2.4.2 Mutation

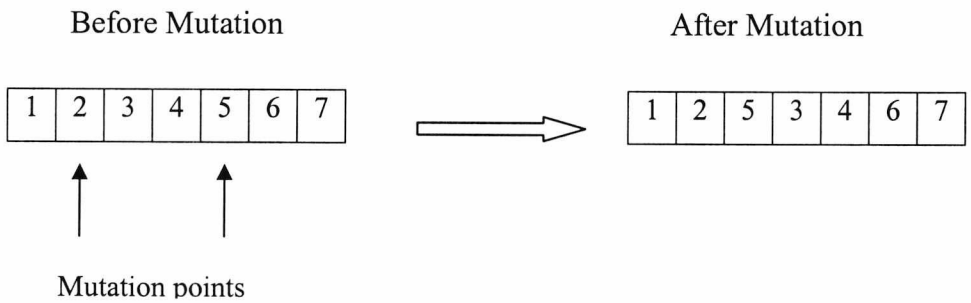
Mutation operates by randomly changing one or more alleles of a selected individual and it acts as a perturbation operator to allow for inserting new information into the population [Whitley95]. Mutation is considered as a background operator with a very low probability of application. The role of mutation is often seen as providing a

guarantee that the probability of searching any given string will never be zero and acting as a safety net to recover good genetic material that may be lost through the action of selection and crossover. The example below illustrates the effect of mutation on different chromosome representations:

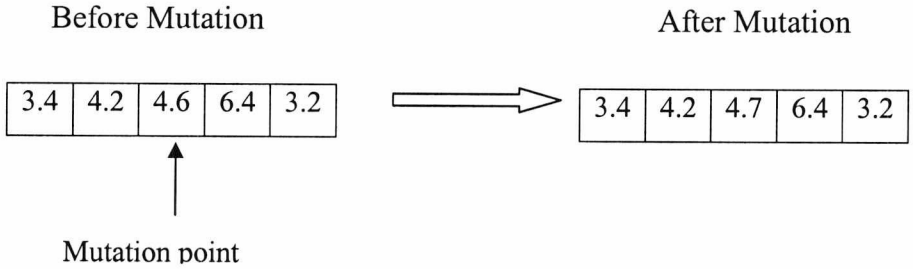
**Bit inversion** For binary representation, bits positions are chosen randomly and corresponding bit values negated (0 becomes 1 and 1 becomes 0).




**Order changing** For permutation representation, mutation is done by picking two alleles at random and moving one so that it is next to the other.




**Value representation** For value representation, a small number is added or subtracted from selected values



With non-binary representations, mutation is achieved by either perturbing the gene values or random selection of new values within the allowed range. Wright in [Wright91] and Janikow in [Janikow91] demonstrated how real-coded GAs might take advantage of higher mutation rates than binary-coded GAs, increasing the level of possible exploration of the search space without adversely affecting the convergence characteristics.

### 6.2.5 Reinsertion

Once a new population has been produced by selection and recombination of individuals from the old population, the fitness of the individuals in the new population is determined [Belew97]. If fewer offspring are produced than the size of the original population, then the offspring have to be reinserted into the old population. Similarly, if not all offspring are to be used at each generation or if more offspring are generated than needed a reinsertion scheme must be used to determine which individuals should be inserted into the new population. There are different schemes of reinsertion such as: -

**Pure reinsertion**: In this scheme, the number of offspring produced is as many as the parents and all parents are replaced by the offspring.

**Elitist reinsertion**: In this scheme, the offspring produced is less than the parents and the worst parents are replaced.

**Fitness-based reinsertion**: In this scheme, more offspring are produced than needed for reinsertion and only the best offspring are reinserted.

Pure reinsertion is the simplest reinsertion scheme. Every individual lives one generation only. This scheme is used in the simple genetic algorithm. However, it is very likely, that very good individuals are replaced without producing better offspring and thus, good information is lost.

Elitism reinsertion dictates that the old parent individuals will be pooled together with the new offspring individuals and then the ranking of all individuals will be performed according to their fitness value. The best-fitted individuals, selected from

the pool, will substitute the old parent population. This technique guarantees survival of the best adapted individuals but also hinders evolution if these apparently well-adapted individuals approach a local optimum instead of the global one.

The fitness-based reinsertion dictates the ranking to be performed only on the offspring population of individuals and the best out of these to substitute the least fit parent. However, with every generation some new individuals are inserted. It is not checked whether the parents are replaced by better or worse offspring. Because parents may be replaced by offspring with a lower fitness, the average fitness of the population can decrease. However, if the inserted offspring are extremely bad, they will be replaced with new offspring in the next generation. Thus, this selection procedure might lose well adapted parent individuals but it provides also the power to leave local optima in search for the global optimum.

### 6.2.6 Termination of the Genetic Algorithms

Since genetic algorithms are stochastic iterative processes that are not guaranteed to converge, a termination condition must either be specified as some fixed, maximal number of generations or as the attainment of an acceptable fitness level [Banzhaf99].

## 6.3 Comparison of Genetic Algorithms with Other Techniques

Most research into GAs has concentrated on finding empirical rules for getting them to perform well. There is no accepted "general theory" which explains exactly why GAs have the properties they do. Nevertheless, several hypotheses have been put forward which can partially explain the success of GAs. Holland's Schema theorem [Holland75] was the first rigorous explanation of how GAs work. According to Goldberg [Goldberg89], the power of the GAs lies in their ability to find good building blocks.

Any efficient optimisation algorithm must use two techniques to find a global maximum: *exploration* to investigate new and unknown areas in the search space, and *exploitation* to make use of knowledge found at points previously visited to help find better points. These two requirements are contradictory, and a good search algorithm must find a tradeoff between the two. The general optimization algorithms fall under three categories: Enumerative schemes, deterministic algorithms and stochastic algorithms. A brief description of some of the most commonly used deterministic and stochastic algorithms is provided as follow:

**Random Search**

The brute force approach for difficult functions is a random search. These techniques do not use any knowledge gained from previous results [Holland75]. Points in the search space are selected randomly, or in some systematic way, and their fitness is evaluated. The best optimum values are recorded when discovered while performing random walks on the problem space. This is a very unintelligent strategy and is rarely used by itself

**Gradient methods**

A number of different methods for optimising well-behaved continuous functions have been developed which rely on using information about the gradient of the function to guide the direction of search [Bunday94]. If the derivative of the function cannot be computed, because it is discontinuous, for example, these methods often fail. Such methods are generally referred to as *hillclimbing*. They can perform well on functions with only one peak, but on functions with many peaks, they suffer from the problem that the first peak found will be climbed, and this may not be the highest peak. Having reached the top of a local maximum, no further progress can be made.

**Iterated Search**

Random search and gradient search may be combined to give an *iterated hillclimbing* search. Once one peak has been located, the hillclimb is started again, but with another, randomly chosen, starting point. This technique has the advantage

of simplicity, and can perform well if the function does not have too many local maximum points. However, since each random trial is carried out in isolation, no overall picture of the domain is obtained. As the random search progresses, it continues to allocate its trials evenly over the search space. This means that it will still evaluate just as many points in regions found to be of low fitness as in regions found to be of high fitness.

**Enumerative**

These techniques are the simplest, they work within a finite search space, or at least a discretized infinite search space [Goldberg89]. The algorithm then starts looking at objective function values at every point in the space, one at the time.

**Simulated annealing**

This is essentially a modified version of hill climbing. Starting from a random point in the search space, a random move is made [Rutenbar89]. If this move takes us to a higher point, it is accepted. If it takes us to a lower point, it is accepted only with probability p(t), where t is time. The function p(t) begins close to 1, but gradually reduces towards zero, the analogy being with the cooling of a solid. Initially therefore, any moves are accepted, but as the "temperature" reduces, the probability of accepting a negative move is lowered.. Like the random search, however, simulated annealing only deals with one candidate solution at a time, and so does not build up an overall picture of the search space. No information is saved from previous moves to guide the selection of new moves.

It can be noticed that both the enumerative and random search methods are not efficient when the search space is significantly large or the problem is significantly difficult. The gradient methods are inadequate if the search space is noisy (one with numerous peaks). Gradient methods also depend upon the existence of derivatives or well-defined slope values. But, the real world of search is fraught with discontinuities, vast multimodal noisy search spaces. The iterative search does not perform well if the function has too many local maximum points. The simulated annealing deals only with one candidate solution at a time, and so does not build up an overall picture of the search space.

Although genetic algorithms take the systematic convergent properties of gradient searches and combine them with the generalization and simplicity of randomised, iterative and enumerative searches, this approach differs from these search methods in that:

- Genetic algorithms work with a coding of the parameter set, not the parameters themselves.

  The natural parameter set of the optimisation problem must be coded as a finite length string of symbols over a finite alphabet. GAs exploit coding similarities in a very general way; as a result they are largely unconstrained by the limitation of other methods (e.g. continuity of a function, or the existence of a derivative function).

- Genetic algorithms use probabilistic transition rules based on fitness rather than using deterministic rules.

  Genetic algorithms do not use simple random search but rather use probability as a guide toward likely improvement.

- Genetic algorithms use an objective function information, not derivatives or rather auxiliary knowledge

  Gradient search, for example, require derivatives (calculated analytically or numerically) in order to climb the current peak. GAs are blind. They only require payoff values associated with individual strings. GAs attempt to develop broadly based schemes by ignoring auxiliary information.

- Genetic algorithms search from a population of points, not a single point.

  Moving point to point in search spaces that are multimodal (that have many optimum points) is a perfect prescription for locating false peaks. GAs on the other hand work form a rich database of points simultaneously, climbing many peaks in parallel, thereby reducing the probability of finding a false peak (weaker local minimum/maximum points) as compared to point–to-point methods.

## 6.4 Application Areas of Genetic Algorithms

Genetic Algorithms (GAs) in various forms have been applied to many scientific and engineering problems, including the following:

- **Automatic Programming:** GAs have been used to develop computer programs for specific tasks [Koza93] and to design other computational structures such as cellular automata [Mitchell93] and sorting networks [Hillis90].

- **Economic Models:** GAs have been used to model processes of innovation, the development of bidding strategies and the emergence of economic markets [Brian Aurther93] [Holland91]

- **Immune System Models:** GAs have been used to model various aspects of the natural immune system, including somatic mutation during an individual's lifetime and the discovery of multi-gene families during evolutionary time [Cellada92] [Farmer86].

- **Ecological Models:** GAs have been used to model ecological phenomena such as biological arms races, host-parasite co-evolution, symbiosis and resource flow in ecologies [Lindgren93] [Taylor89]

- **Population Genetics Models:** GAs have been used to study questions in population genetics, such as " under what conditions will a gene for recombination be evolutionary viable?" [Bergman92] [Fogel90]

- **Interactions between evolution and learning:** GAs have been used to study how individual learning and species evolution affect one another [Ackley92] [Belew90] [Fontanari90].

- **Models of Social Systems:** GAs have been used to study evolutionary aspects of social systems, such as the evolution of cooperation, the evolution of communication and trail-following behaviour in ants [Axelrod86] [Werner92] [Collins92]

- **Optimisation:** GAs have been used in a wide variety of optimisation tasks, including numerical optimisation and combinatorial optimisation problems such as circuit design and job shop scheduling [De Jong75]

- **Machine and Robot Learning:** GAs have been used for many machine-learning applications, including classification and prediction tasks. GAs have also been used to design neural networks, to evolve rules for learning classifier systems or symbolic production systems and to design and control robots [Belew92] [Holland86] [Davidor91].

Genetic algorithms, apart from their generally high computational cost, have been shown to be able to out-perform conventional optimisation techniques of difficult, discontinuous, multimodal and noisy functions, which makes the GA an attractive choice to be used in the field of biometric recognition, since that the search space in this field is fraught with discontinuities and vast multimodal noisy spaces.

Several research studies have already used genetic algorithms in the field of biometric recognition.

In 1991, Caldwell and Johnson created a system that was used to help witnesses reconstruct facial depictions of criminals [Caldwell91]. The system had a large library of basic facial features, which contained images of noses, foreheads, ears, etc. The system uses a 35 bit binary string to encode the features and creates an initial population of 20 strings (faces). The witness then rank each face (from 0 to 9), and these scores serve as the fitness value. Then a new generation is created using selection, crossover and mutation.

In 1996, Bala *et al* addressed the problem of crafting visual routines for eye detection from real grey-level facial imagery using a hybrid method that integrates genetic algorithms and decision trees [Bala96]. The experimental results reported demonstrated the feasibility of the approach in terms of feature selection and the corresponding eye detection.

In 1997, Bala *et al* introduced a hybrid method that integrates genetic algorithms and decision tree learning in order to evolve useful subsets of discriminatory features for recognizing complex visual concepts [Bala97]. A Genetic Algorithm was used to search the space of all possible subsets of a large set of candidate discrimination features, which were then evaluated by using the decision-tree learning algorithm. The experimental results reported, using both satellite and facial image data, indicated that learning does indeed help evolution in several important ways. The error rates on the underlying classification tasks were observed to decrease significantly when learning and evolution were allowed to dynamically interact.

In 1998, Liu and Wechsler integrated GAs for capturing the non-accidental spatiotemporal properties ('regularities') called Optimal Projection Axes (OPA) for face recognition by searching through all the rotations defined over whitened PCA subspaces [Liu98]. Evolution was driven by a fitness function defined in terms of performance accuracy and class separation ('scatter index'). Accuracy indicates the extent to which learning has been successful so far, while the scatter index gives an indication of the expected fitness on future trials. Experimental results showed that when using a large data set (1107 facial images from the US army FERET database) it resulted in a recognition of 92 % when compared with other methods (eigenfaces (87 %) and MDF (86 %).

It is apparent that all the previous contributions in the field of biometric recognition have focused on using GAs in a mono-modal biometric system and have all focused on using the genetic algorithm techniques at the feature level.

## 6.5 Summary

In this chapter an introduction to genetic algorithms and a discussion of the different operators that influence their performance was provided. A comparison between GAs and other optimisation techniques was also presented declaring the advantages that GAs have over the alternatives. Finally a description was provided of some application areas of GAs, especially in the field of biometric-based recognition of

individuals. The literature showed that most researchers have used GAs in mono-modal biometric systems and at the feature level.

The next chapter proposes the use of genetic algorithms (GAs) in multi-modal biometric system and demonstrates how they might have a valuable role to play at the decision level.

# Chapter 7

# Optimising Multimodal Person Recognition

## 7.1 Introduction

In this chapter a proposed approach for the optimisation of a multimodal person recognition system based on the use of genetic algorithm is described. Several experiments are performed involving the effects of both hard decision fusion rules and soft decision fusion rules on the performance of the system. The experiments also illustrate the important role that the proposed approach plays in system optimisation by solving problems associated with score normalization and weights/ threshold settings.

## 7.2 Performance Measurements of Biometric Systems

As previously mentioned in Chapter 1, the overall performance of a system can be evaluated in terms of its *accuracy*. Measuring the accuracy is critical for determining whether the system meets its requirements and, in practice, how the system will respond in a variety of situations [Golfarelli97]. It is traditionally characterised by two error statistics: false reject rate (FRR) and false accept rate (FAR). A false reject

occurs when a system rejects a valid identity; a false accept occurs when a system incorrectly accepts a claimed identity as valid when it is not. In a perfect biometric system, both error rates would be zero. Unfortunately, biometric systems are not perfect. However, as shown by the study of binary hypothesis testing [Van Trees68], either of the two (FAR, FRR) can be reduced to an arbitrarily small value by changing the decision threshold, but with the drawback of increasing the other one. A unique measure can be obtained by combining these two errors into the total error rate (TER) or its complementary, the total success rate (TSR) [Martin00]. In this chapter, the total error rate (TER) is the key factor in all the experimental results.

$$TER = FRR + FAR \tag{7.1}$$

$$TSR = 1\text{-}TER \tag{7.2}$$

Most biometric person verification systems return a score indicating the likelihood that the user is a genuine client or an impostor. Selecting a threshold over which scores are considered to indicate genuine clients instead of impostors can modify the relative performance of FAR and FRR. A typical threshold chosen is the one that reaches the equal error rate (EER), where FAR = FRR [Pierrot98]. Note that EER and TER, while similar, are different concepts: EER is often used to select a threshold but cannot be used to measure the performance of a system on unknown data, while TER can be used to measure this performance. This can be explained by the fact that the threshold selected on one dataset for FAR=FRR will not give the same ERR on an unknown data while the same threshold can be applied to an unknown dataset to measure its TER.

In order to achieve an effective comparison of the performance of different biometrics systems each with a different threshold range, a description independent of threshold scaling is required [Bengio04]. The ROC Curve and the DET Curve are used for evaluating the overall performance of a system while eliminating the threshold parameter. The ROC Curve has been taken to denote either the Receiver Operating Characteristic [Centor91] [Hanley89] [Egan75] or alternatively, the

Relative Operating Characteristic [Swets73]. It is usually used for pattern or signal detection systems where the false accept rate (FAR) is plotted -on the x-axis- against the Genuine Accept Rate (GAR) –on the y-axis-. The Detection Error Trade-off (DET) Curve on the other hand [Mansfield01b], plots the FAR –on the x-axis- against the FRR –on the y-axis- giving uniform treatment to both types of error, a logarithmic scale for both axes is then used to spread out the plot to better distinguish the performance of the different systems [Mansfield 02].

In order to compare the performance of the different biometric modalities used in this project, which are fingerprint, voice, and face, the DET curve was used. Figure 7.1 shows the DET curves for the each of the modalities used in this project

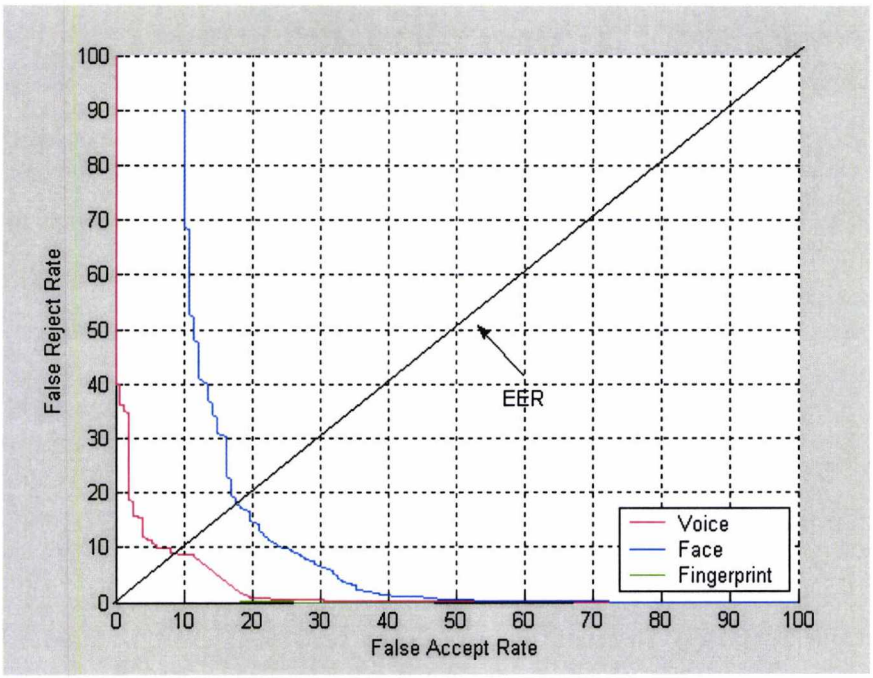



**Figure 7.1**: Detection error trade-off: FAR vs. FRR

The lower and further left on the graph that an operating point occurs, the better the performance can be considered [Martin97], with the origin representing the "perfect" performance. Figure 7.1 shows that the voice system had the best

accuracy, followed by the fingerprint and finally the face system. The graph shows that the fingerprint is the most secure modality against spoofing among the other modalities with a false accept rate of zero.

The DET curves can also be plotted for the different decision fusion rules. Figure 7.2 (a) and 7.2 (b) shows the DET curves plotted for the soft decision fusion methods and the hard decision fusion method respectively. The DET curves plotted were performed through an exhaustive search.

It should be noted that the scores used in plotting the DET curves for the different decision fusion rules were the raw un-normalised scores of each modality, except for the sum rule, where the scores used were normalized by using the normalization method proposed in Chapter 5, as the raw un-normalised scores of each modality could not be used in this case.

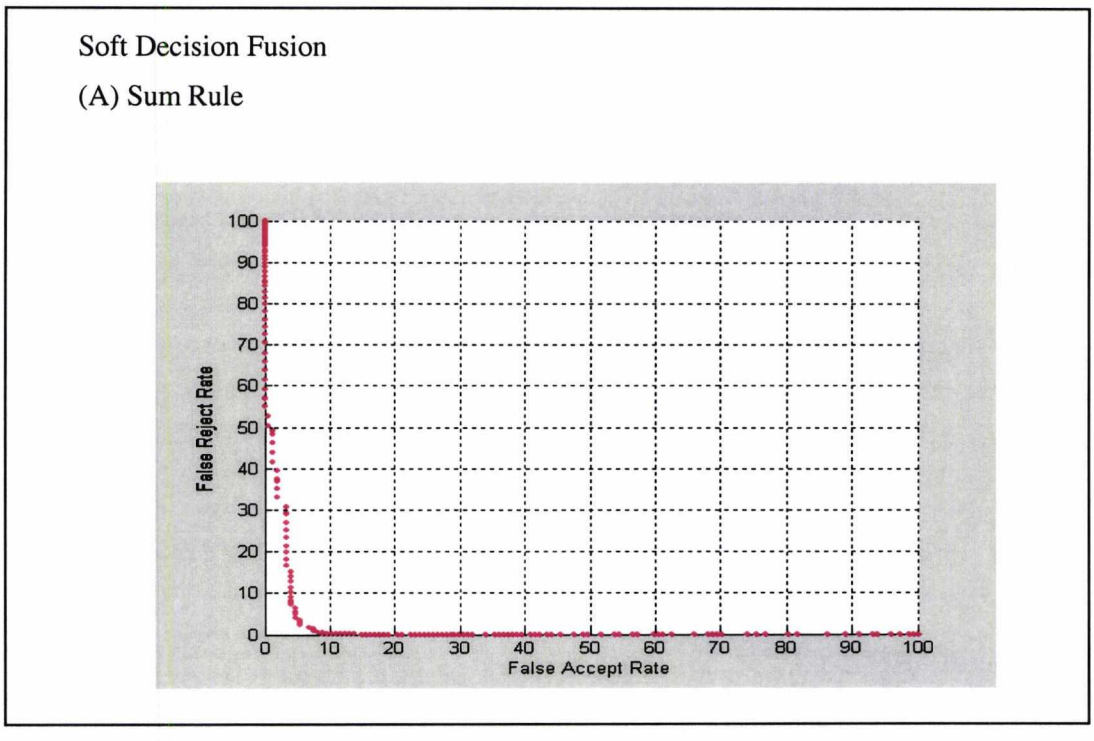
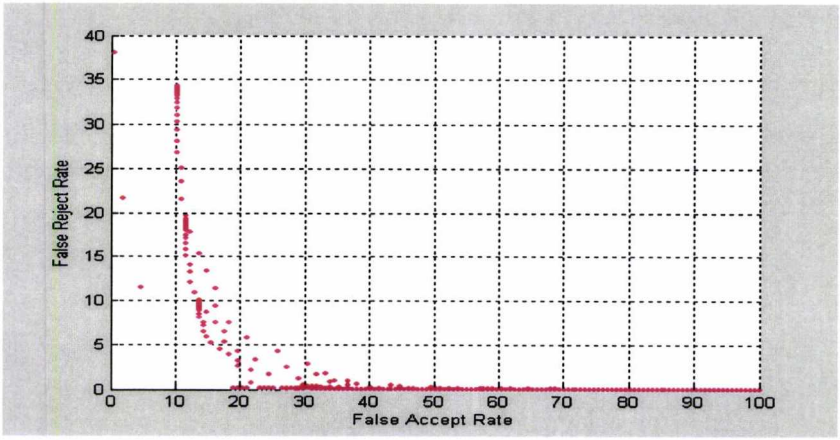


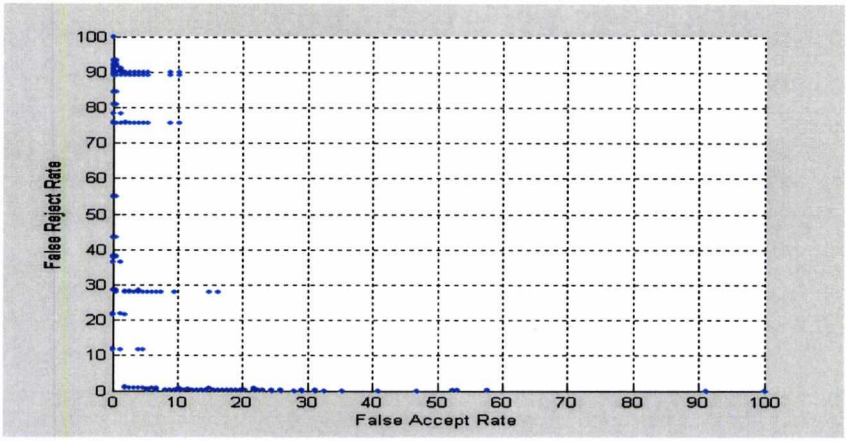

**Figure 7.2 (a)**: Soft decision fusion methods

Hard Decision Fusion

(A) AND Rule




(B) OR Rule




(C) Majority Voting

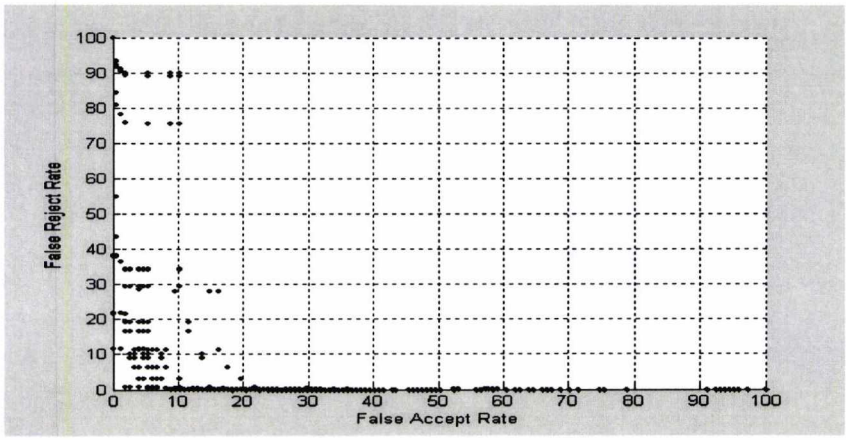



**Figure 7.2 (b)**: Hard decision fusion methods

From the plotted DET curves it can be concluded that improving the performance of a single biometric system by searching for the optimal verification threshold that lowers both its FAR and FRR and thus resulting in a low TER might not be a difficult task, but as the number of biometric modalities increases the search space becomes more complex and determining the optimal verification threshold of the individual modalities under different decision fusion rules becomes almost impossible. Furthermore, having to normalise the scores before combining them using the sum rule reduces the dimensionality of the space thus resulting in inaccurate recognition rates [Altinacy03].

In the next section a more formal approach for the optimisation of a multi-modal biometric identity verification scenario is proposed that solves the problem associated with threshold settings.

## 7.3 An Approach to Optimising Multimodal Configurations

The specific approach proposed here adopts an optimisation technique based on the use of genetic algorithms (GAs) [Kuncheva93]. The empirical approach adopted uses the raw un-normalised "scores" generated by each of the biometric devices in the system, offering an effective "plug-and-play" design philosophy to system implementation. Figure 7.3 shows the proposed optimising architecture that integrates genetic algorithms and the different decision fusion rules for solving the problem associated with threshold settings.

In the next section a detailed description of the proposed optimising architecture is provided.

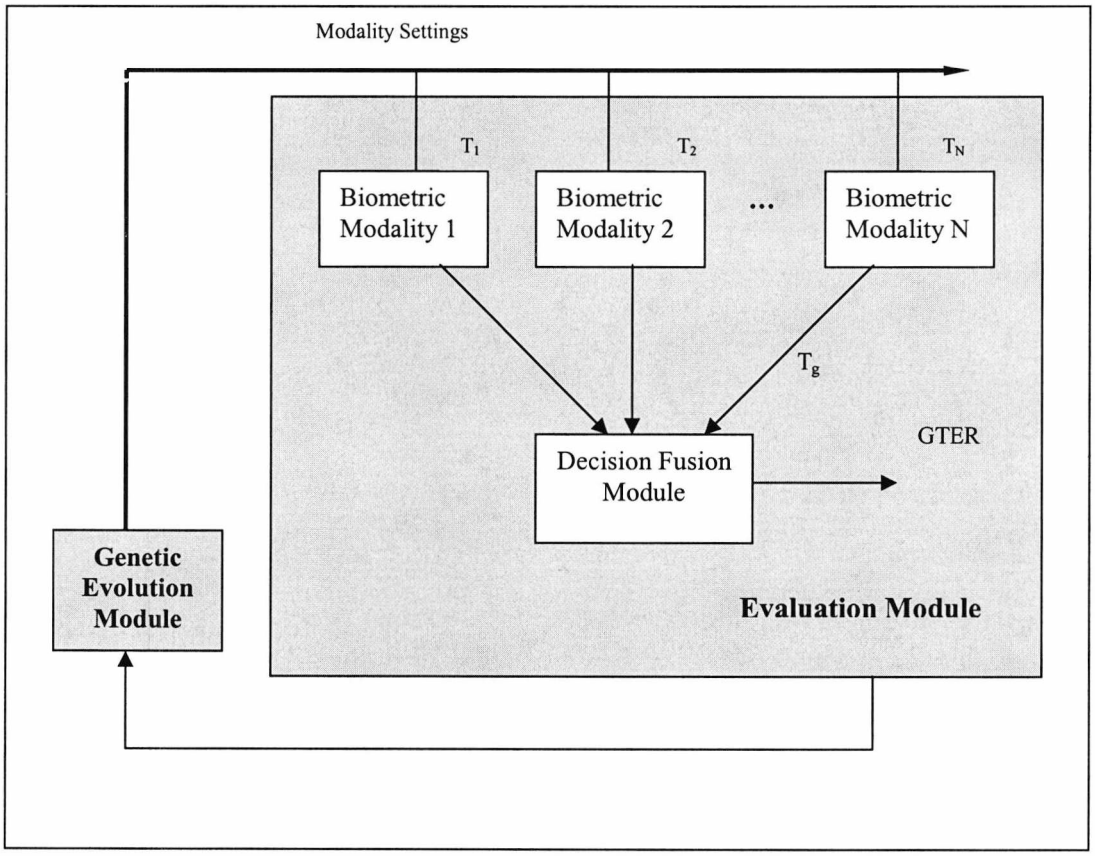



**Figure 7.3**: The optimising architecture

## 7.4 Description of the Optimising Architecture

The architecture consists of two main components, the genetic evolution module and the evaluation module. In the following paragraphs a description of each component is given in detail. This architecture was implemented by using the genetic algorithm toolbox provided in [GA Toolbox].

### 7.4.1 The Genetic Evolution Module

The genetic evolution module is responsible for generating a population of individuals (Chromosomes) and applying the genetic operators to them depending on their fitness, as will be described later in this chapter.

#### 7.4.1.1 Chromosome Representation and Genetic Operators

Each chromosome generated has 'N+1' genes, where N is the number of modalities used in the system. Each of the N genes represents a local threshold to which that particular modality is set for the verification process. These local thresholds are represented in Figure 7.4 by $T_i$, where i =1 to N. The 'N+1' *th* gene represents the global threshold associate with the decision fusion rule. This global threshold is represented in Figure 7.4 by $T_g$. As previously explained in Chapter 6, the chromosomes are represented by different encoding types depending on the problem being explored. In this work the *value encoding* type was used, since the thresholds of the different modalities used in this project were all real values. Figure 7.4 shows the chromosome representation.

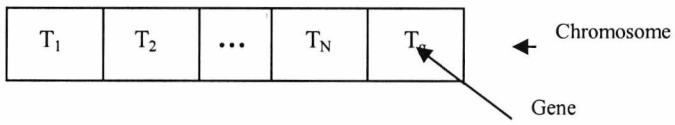



**Figure7.4**: General chromosome representation

The different genetic operators that were described in the previous chapter were used in this study as follows:

For selection, where the individuals of the population compete among each other to become parents of the next generation, the *roulette wheel selection* technique was used. This technique was chosen because the selection procedure is biased to the member with the highest fitness value. For crossover, where new individuals are introduced into the population, *intermediate recombination* was used. This method was used because of its ability in dealing with chromosomes represented by real-valued encoding. For mutation, where new information is inserted into the population by randomly changing one or more genes of a selected individual, the function for the mutation of real-valued population was used as described in the genetic algorithm toolbox used [GA Toolbox]. For reinsertion, the *fitness-based reinsertion* scheme was used, where the offspring are selected for reinsertion

according to their fitness. The offspring with the higher fitness replace the less fit ones.

### 7.4.2 The Evaluation Module

The evaluation module is responsible for providing fitness feedback for each chromosome produced by the genetic evolution module. This fitness feedback is expressed as a function to be minimized where:

$$\text{Fitness Function} = \text{Global Total Error Rate} \quad (7.3)$$

$$\text{Global Total Error Rate} = \text{Global False Accept Rate} + \text{Global False Reject Rate} \quad (7.4)$$

The Global Total Error Rate (GTER), is calculated by fusing the local decisions of the biometric modalities under a specific decision fusion rule and thus creating a global decision from which the GFAR and GFRR, and hence the GTER, are calculated.

The decision fusion module uses either the hard decision fusion rules or the soft decision fusion rules.

## 7.5 Operation of the Optimising Architecture

In the previous section the components of the optimising architecture were described. In this section the operation of the proposed algorithm is presented, as illustrated in Figure 8.5. The operation commences by the genetic evaluation module generating an initial population of random chromosomes. Each chromosome represents the local threshold of each modality and the global threshold associated with the decision fusion rule used. Each chromosome generated is then passed to the evaluation module where the fitness function is calculated producing the GTER, given the thresholds and the fusion rule. When all chromosomes have been evaluated they are sent back to the genetic evolution module (GEM) where they are

ranked in a descending order using the fitness ranking method (described in the previous chapter), with the most fit chromosome having the lowest GTER and the least fit chromosome having the highest GTER. The genetic evolution module (GEM) then performs the genetic operations: selection, crossover, mutation and reinsertion, producing a new generation of chromosomes that are passed to the evaluation module for fitness evaluation. This process is repeated until the number of generations (100) is achieved.

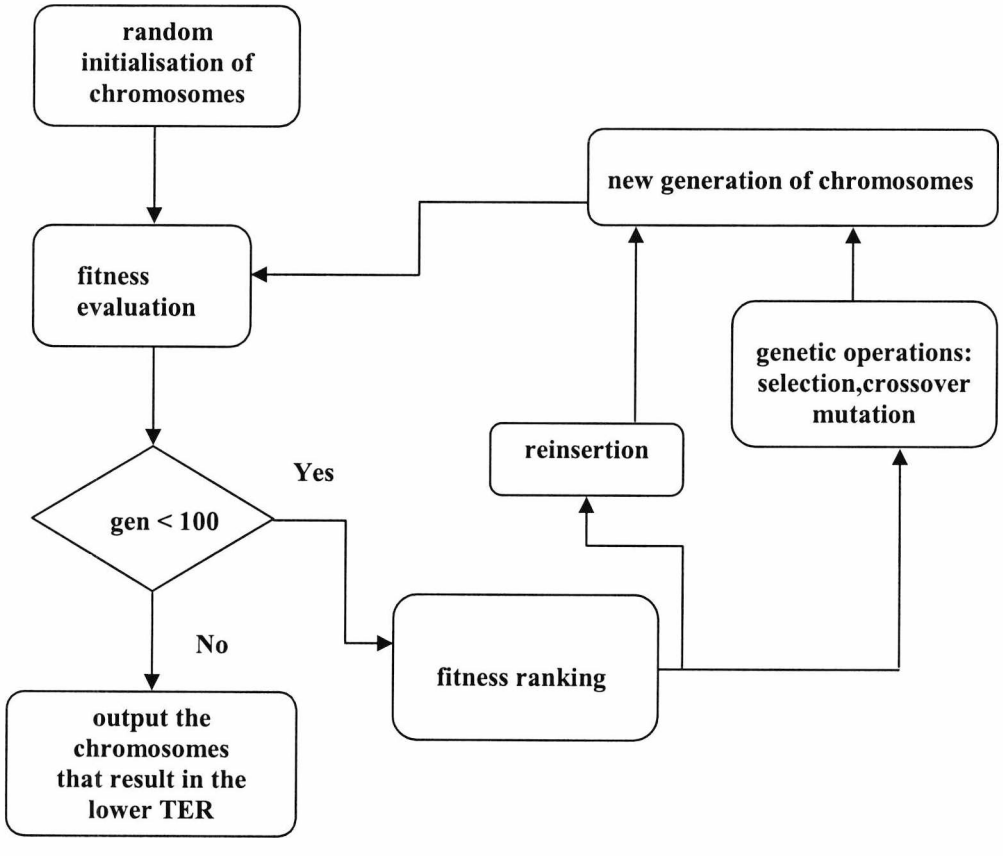



**Figure 7.5**: Flowchart of the optimising architecture

## 7.6 Experimental Methodology

The optimising architecture described in the previous sections was implemented using the genetic algorithm toolbox in [GA Toolbox]. For the genetic evolution module (GEM) a constant population size of 100, a crossover rate of 0.6 and a mutation rate of 1/(number of genes) were used, as described in the genetic algorithm toolbox [GA Toolbox]. Each chromosome consisted of four genes, three representing the local thresholds of the three modalities used in this work (fingerprint, voice and face) and the fourth representing the global threshold associated with the decision fusion rule. All chromosomes representations used real values as previously stated.

The testing methodology operates in the following manner. Since GAs are stochastic algorithms, it is difficult to formally specify convergence criteria. As the fitness of a population may remain static for a number of generations before a superior individual is found, the application of conventional termination criteria becomes problematic. A common practice was used where the GA is terminated after a pre-specified number of generations. The experiments, which are reported here, were terminated after 100 generations in a single run. The choice of having a single run with a large population of 100 chromosomes instead of using multiple independent runs with a smaller population was based on the conclusion made by Cantu-Paz in [Cantu-Paz03] that states "A single run with the largest population possible reaches a better solution than multiple independent runs. Similarly, a single large run reaches the global solution faster than multiple independent runs". After the GA is terminated the chromosomes with the lowest TER were selected. It is possible for the chromosomes to present more than one optimal solution in this search space. If multiple solutions yield the same TER, any one of the resulting solutions can be selected as a final solution. In such a case, the chromosome having the threshold values closer to the default ones was preferred since it was desired to know to what extent must the thresholds be tuned from their original value to achieve the minimum TER.

## 7.7 Experimental Results

The ultimate goal for any biometric system is to make authentication decisions. However, the actual decision making process, specifically the setting of decision thresholds, has often been neglected in the field of biometrics. Making these decisions has often been dismissed as an unchallenging problem to be addressed during application development. However, in real operational systems, the problem has been found to be very challenging. The thresholds, for any realistic field deployment and eventual independent operation, have to be set up ahead of time during enrolment, (i.e., *a priori,*). The alternative is setting the thresholds *a posteriori* using the information available from the aggregate similarity scores recorded during the matching process. In this work, the default threshold set by the vendors of the biometric modalities used in this project was regarded as the *a priori* threshold since they were set up during the enrolment process in the data collection exercise. Alternatively, the *a posteriori* threshold is estimated using the proposed optimising architecture.

For estimating the *a posteriori* threshold and carrying out different experiments, three different data sets from the collected database were used for training, validating and testing the proposed approach; according to the following criteria.

The enrolment session was used for training the individual classifiers. This means that each access has been used to model the respective client, yielding 147 different client models, as previously mentioned.

The first access from each person in the second enrolment session was used for validation. This was done by matching each single client sample access with his own reference model, generating 147 clients. Then a cross comparison was used to establish the impostor distribution [O'Gorman98] generating $147 \times 146 = 21462$ impostor accesses, as explained in Chapter 5. This data set was used by the optimising architecture to search for the optimal local and global thresholds.

The second access from each person in the second enrolment session was used for testing the thresholds calculated from the validation set.

Two configurations were considered for calculating the total error rate (TER). Figure 7.6(a) and 7.6 (b) shows these, designated configuration I and configuration II respectively.

**Configuration I**

In this configuration, the validation set is used by the proposed optimising architecture to tune the thresholds of the biometric modalities, as will be described later in this chapter. The minimum total error rate (TER) generated by the tuning of the thresholds (*a posteriori* thresholds) on this dataset is recorded and compared with the TER computed when using the *a posteriori* thresholds on the testing set.

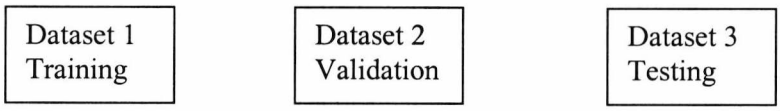



**Figure 7.6 (a)**: Configuration I for computing the TER

**Configuration II**

In this configuration the *a priori* threshold, set by the vendors of the biometric modalities was used in the testing set to compute the total error rate (TER).

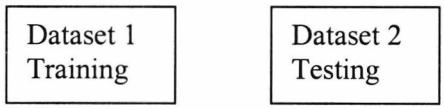



**Figure 7.6 (b)**: Configuration II for computing the TER

In this section the results obtained from using the *a priori* and the *a posteriori* thresholds are compared. It should be noted that all the experiments carried out in this section used the scenarios described in Chapter 5 to calculate both the false reject rate (FRR) and the false accept rate (FAR) for the different fusion rules.

### 7.7.1 Hard Decision Fusion Rules

As previously mentioned the hard decision fusion rule are combined using the three rules; AND, OR and majority voting. For these experiments each chromosome consists only of three genes, each representing the local threshold of the three modalities used in this project (fingerprint, voice and face) as shown in Figure 7.7

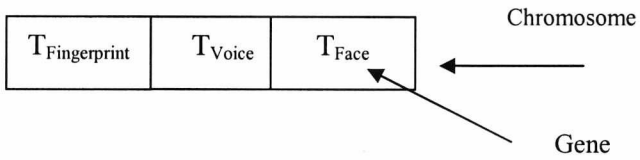



**Figure 7.7**: Chromosome representation for hard decision fusion

This is because when using the hard decision fusion rules such as AND, OR or majority voting, no global threshold is involved; the FAR, FRR and TER are calculated only from the local thresholds of the biometric modalities as demonstrated in Figure 7.8. The results from the different fusion rules are presented in the following subsections.

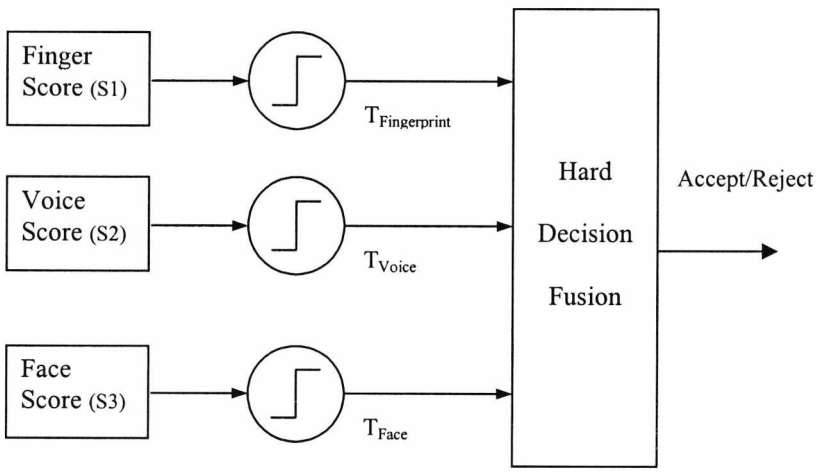



**Figure 7.8**: Evaluation of TER using hard decision fusion

#### 7.7.1.1 AND Fusion

In AND fusion, for a user to be accepted as a genuine client, all the classifiers must agree. That is, the score of each single modality must be greater than its pre-specified local threshold. The AND fusion rule is mainly used in high security access applications where a main concern is break-ins, and hence the system operates on a low FAR at the expense of a high FRR [Jain99]. Figure 7.9 shows the minimum TER plotted after every generation for the AND fusion.

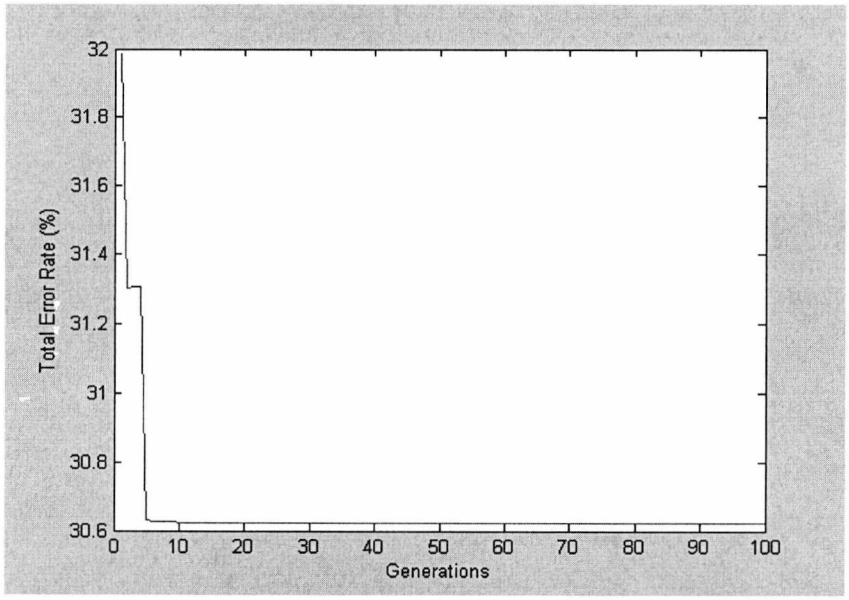



**Figure 7.9**: Minimum TER vs. Generation for AND fusion

Figure 7.9 illustrates the reduction of the total error rate (TER) over the 100 generations with the AND rule being used. As can be observed from Figure 7.9, a minimum TER of 30.62 % (computed when the thresholds were chosen on the validation set), was obtained after tuning the local threshold of each modality. It can also be seen that the TER remained constant after the $10^{th}$ generation. The thresholds computed from the validation set were then used on the test set to compute the total error rate (TER).

Table 7.1 compares several results; it compares the TER obtained when using the *a priori* thresholds, the resulting TER after setting the *a posteriori* thresholds on the validation set and the TER obtained when using the *a posteriori* thresholds computed on the test set. In addition to that the table includes the results obtained from carrying an exhaustive search. Although this method is not sufficient for complex problems, it was done only on this experiment to compare its performance with the optimizing method proposed.

**Table 7.1**: Comparison between the a *priori* threshold and the *a posteriori* threshold results for the AND rule

| Rule used | Settings used | Local Thresholds | | | Total Error Rate (%) | |
|---|---|---|---|---|---|---|
| | | Finger | Voice | Face | FRR | FAR |
| AND Fusion | *a priori* Threshold | 5 | 0 | 8.7 | 57.1 | 0.0 |
| | *a posteriori* Threshold (Validation set) | 3 | -262 | 6.8 | 30.62 | 0.0 |
| | *a posteriori* Threshold (test set) | 3 | -262 | 6.8 | 34.0 | 0.0 |

The table shows an expected reduction of 40.5% in the TER if the local threshold of each biometric modality is tuned from its default setting (*a priori* threshold) by using the optimisation architecture proposed. It is also seen that the TER from the validation set (30.62 %) was better than the TER (34.0 %) obtained when using the *a posteriori* threshold on the test set.

A simple experiment was carried out using the exhaustive search method with an increment step of 0.01 since it was decided that the results would be up to 2 decimal points. Although this method is not sufficient for complex problems, it was done only on this experiment to compare its performance with the optimizing method proposed. Figure 7.2 shows the results obtained from the exhaustive search.

**Table 7.2**: The exhaustive search results

| Method used | Local Thresholds | | | Total Error Rate (%) | |
|---|---|---|---|---|---|
| | Finger | Voice | Face | FRR | FAR |
| *Exhaustive serach* | 3 | -262 | 6.8 | 30.62 | 0.0 |

The table showed that the exhaustive search provided similar results to the optimising method proving that the optimising method is efficient. It should be noted that the exhaustive search was not done on other experiment.

#### 7.7.1.2 OR Fusion

In OR fusion, for a user to be accepted at least one of the classifiers must indicate that the person is a genuine client. That is, the score of at least one single modality must be greater than the pre-specified local threshold.

Figure 7.10 shows the minimum TER plotted after every generation for the OR fusion.

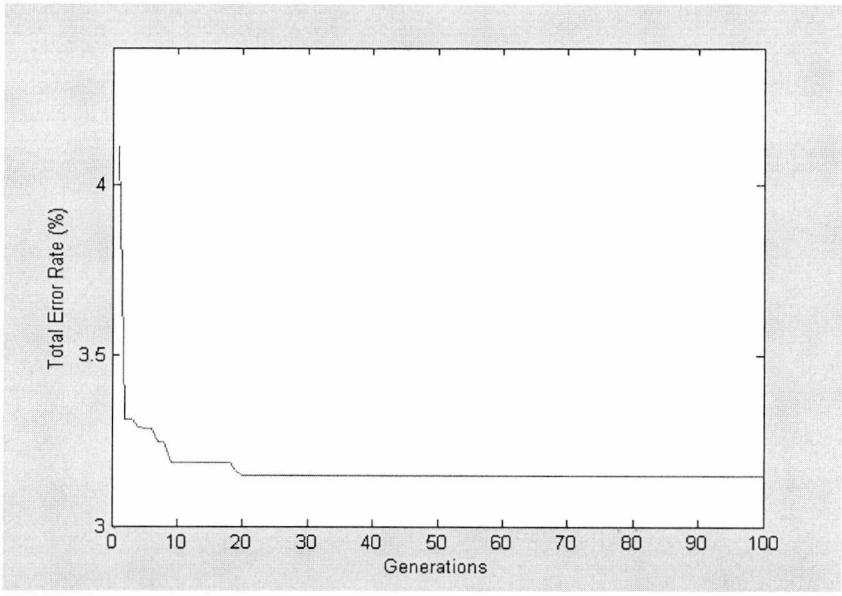



**Figure 7.10**: Minimum TER vs. Generation for OR fusion

Figure 7.10 shows a decreasing total error rate (TER) over the 100 generations with the OR rule being used. From Figure 8.10 it can be seen that the minimum total error rate obtained after tuning the local threshold of each modality on the validation set was 3.2 %. It can also be observed that the TER remained constant after the 20$^{th}$ generation. The thresholds computed from the validation set were then used on the test set to compute the TER.

Table 7.3 presents three different results; it presents the TER obtained when the *a priori* threshold is used on the testing set, it also presents the TER obtained when the *a posteriori* thresholds is set on the validation set and finally it presents the TER obtained when using the *a posteriori* thresholds computed on the test set.

**Table 7.3**: Comparison between the a *priori* threshold and the *a posteriori* threshold results for the OR rule

| Rule used | Settings used | Local Thresholds | | | Total Error Rate (%) | |
|---|---|---|---|---|---|---|
| | | Finger | Voice | Face | FRR | FAR |
| OR Fusion | *a priori* Threshold | 5 | 0 | 8.7 | 4.1 | 3.8 |
| | *a posteriori* Threshold (Validation set) | 3 | -262 | 6.8 | 2.0 | 1.2 |
| | *a posteriori* Threshold (test set) | 3 | -262 | 6.8 | 4.8 | 1.3 |

Table 7.3 illustrates that if the local thresholds of the biometric modalities are re-calibrated from their default settings (*a priori* threshold) by using the optimisation architecture proposed, a reduction of 22.8 % in the TER occurs. From Table 7.2 it is seen that using the *a posteriori* threshold on the validation set give better results than when using the same *a posteriori* threshold on the test set.

### 7.7.1.3 Majority Voting

In majority voting fusion, for a user to be accepted, the majority of the classifiers must agree that the person is a genuine client. That is, the scores of any two modalities out of the three (in our current scenario) must be greater than their pre-specified local thresholds. Figure 7.11 shows the minimum TER plotted after every generation for the majority voting rule.

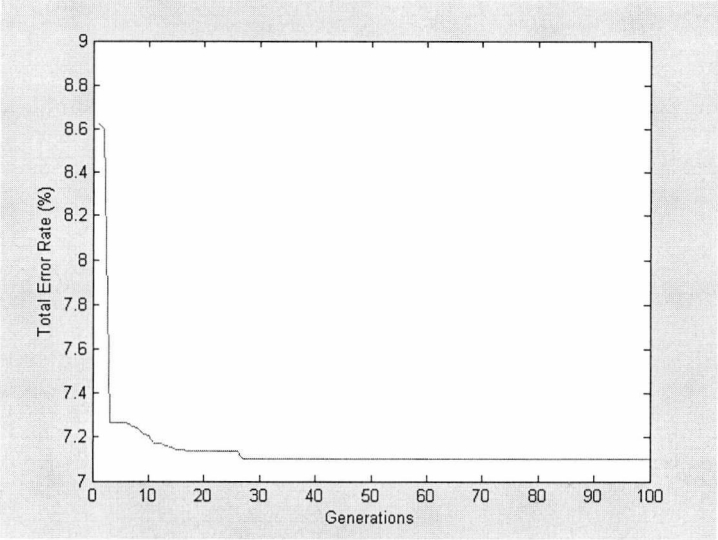



**Figure 7.11**: Minimum TER vs. Generation for majority voting fusion

Figure 7.11 shows a decreasing total error rate (TER) over the 100 generations with the majority voting rule being used. Figure 7.11 illustrates that the minimum TER that can be attained after various tuning of the local thresholds of the modalities (using the proposed method) is 7.1 %. It is seen that the TER remained constant after the $30^{th}$ generation. The *a posteriori* thresholds that were computed on the validation set were used on the test set to compute the TER.

Table 7.4 provides three results computed from the experiments. It provides the TER obtained when using the *a priori* thresholds, the TER attained after setting the *a posteriori* thresholds on the validation set, and the TER obtained when using the computed *a posteriori* thresholds on the test set.

**Table 7.4**: Comparison between the *a priori* threshold and the *a posteriori* threshold results for the majority rule

| Rule used | Settings used | Local Thresholds | | | Total Error Rate (%) | |
|---|---|---|---|---|---|---|
| | | Finger | Voice | Face | FRR | FAR |
| Majority Voting Fusion | *a priori* Threshold | 5 | 0 | 8.7 | 15.6 | 0.02 |
| | *a posteriori* Threshold (Validation set) | 3 | 3 | 7.18 | 6.1 | 1.0 |
| | *a posteriori* Threshold (test set) | 3 | -262 | 6.8 | 7.5 | 1.1 |

Table 7.4 shows that a reduction of almost 50 % in the total error rate (TER) is expected to occur if the local thresholds of the biometric modalities used here are tuned slightly from their *a priori* thresholds by using the proposed optimisation architecture. The results also show that the TER computed on the validation set was better than the TER computed on the test set.

It can be concluded that the majority voting approach had the highest expected reduction in the TER with a slight tuning of the thresholds compared with both the AND rule and the OR rule.

Table 7.5 summarises the results obtained from the different hard decision fusion rules

**Table 7.5**: Comparative performance of the different hard decision fusion methods

| Fusion Method | *a priori* Threshold | | | *a posteriori* Threshold (Validation set) | | | *a posteriori* Threshold (Test set) | | |
|---|---|---|---|---|---|---|---|---|---|
| | FRR | FAR | TER | FRR | FAR | TER | FRR | FAR | TER |
| AND | 57.1 | 0.0 | 57.1 | 30.6 | 0.0 | 30.6 | 34.0 | 0.0 | 34.0 |
| OR | 4.1 | 3.8 | 7.9 | 2.0 | 1.2 | 3.2 | 4.8 | 1.3 | 6.1 |
| Majority Voting | 15.6 | 0.02 | 15.26 | 6.1 | 1.0 | 7.1 | 7.5 | 1.1 | 8.6 |

Table 7.5 shows that the total error rate (TER) computed on the validation set was always better than the TER computed on the test set in all the hard decision fusion methods. This is due to the fact that the validation set was used as a training set in tuning the local thresholds of the biometric modalities as well as a testing set in computing the TER, which results in an overestimation of the performance. This is due to the fact that the proposed optimising architecture will generate the best results for the same data set it have been trained on.

### 7.7.2 Soft Decision Fusion

As previously defined, a soft decision is a decision made by the system that generates a score that normally lies in the range [0,1]. Since the biometric modalities used in this project had different score ranges, the proposed optimising architecture was used to solve the problem of combining the different scores using the sum rule without having to actually normalise the scores. Figure 7.12 shows the method used for computing the FAR, FRR and hence TER when using the sum rule. The $W_i$ (where i = 1 to 3) represent the weights assigned to each biometric modality.

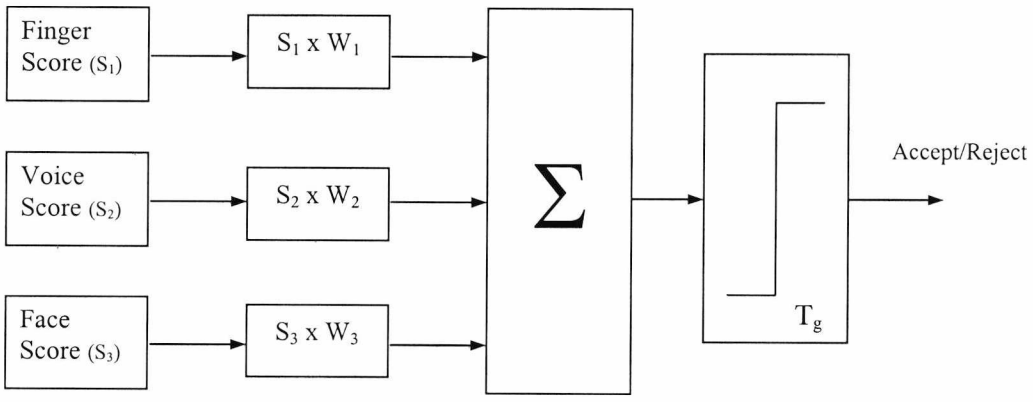



**Figure 7.12**: Evaluation of TER using soft decision fusion

#### 7.7.2.1 Sum Rule

In the sum rule, a user is accepted if the summed score of the three biometric modalities is greater than the pre-specified global threshold ($T_g$). Two experiments were undertaken to evaluate this approach and both are based on the structure shown in Figure 7.12.

**Experiment 1**

In this experiment the raw scores of the biometric modalities were used and equal weights were assigned to each modality. For this experiment each chromosome consists of one gene only, representing the global threshold ($T_g$) as shown in Figure 7.13 . Although this experiment is a simple one it does show the flexibility of the optimising method in dealing with small problems and extending it to larger ones.

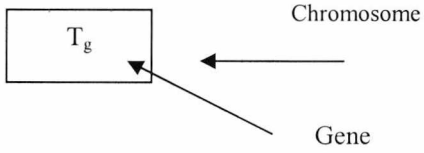



**Figure 7.13**: Chromosome representation for summation rule

Figure 7.14 shows the minimum TER plotted after every generation for the sum rule. It shows the variation in the total error rate (TER) over the 100 generations as the sum rule was used. From Figure 7.14 the results show that a minimum TER of 11.76 % was attained on the validation set after tuning the global threshold (Tg) by using the proposed optimising architecture. It was also shown that the TER remained constant after the $5^{th}$ generation. The global thresholds computed on the validation set were used on the test set to compute the TER on that data set.

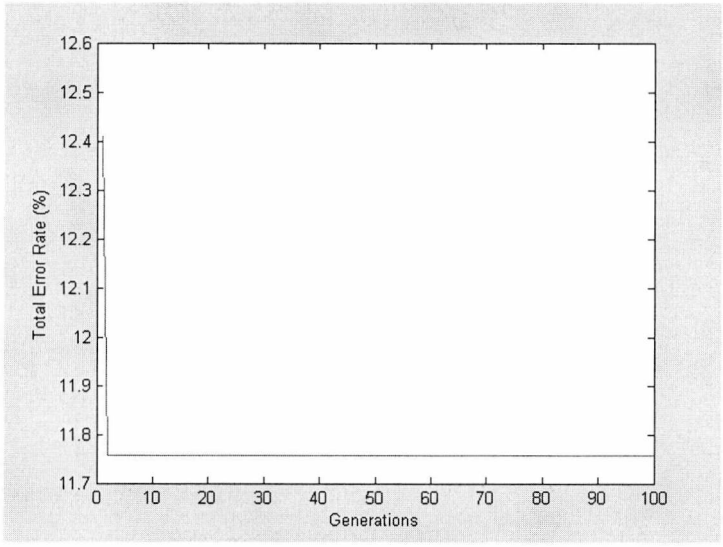



**Figure 7.14**: Minimum TER vs. Generation for summation rule

Table 7.6 shows the results obtained when using the *a posteriori* threshold on the test set.

**Table 7.6**: Results obtained when using the optimisation method

| Threshold | False Reject | False Accept |
|---|---|---|
| 2.58 | 8.16 % | 1.66 % |

It was observed that in this experiment the TER (9.82 %) computed on the test set was better than the TER computed on the validation set (11.76 %) when using the sum rule.

**Experiment 2**

In this experiment the raw scores of the different modalities were weighted in order to vary the importance of the matching scores of each biometric modality [Jain02]. For this experiment each chromosome consists of four genes, three representing the weights for each biometric modality and one representing the global threshold for

the sum rule as shown in Figure 7.15. This fusion rule is often called *weighted summation rule.*

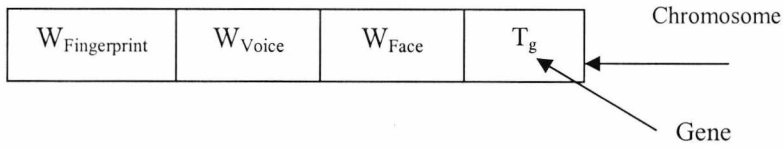



**Figure 7.15**: Chromosome representation for weighted summation rule

Figure 7.15 shows the minimum TER plotted after every generation for the weighted summation rule. It shows that a minimum TER of 4.56 % was reached on the validation set after tuning the global threshold (Tg) by using the proposed optimising architecture. The graph shows that the TER remained constant after the $15^{th}$ generation. The weights and the global threshold computed on the validation set were used on the test set to compute the TER.

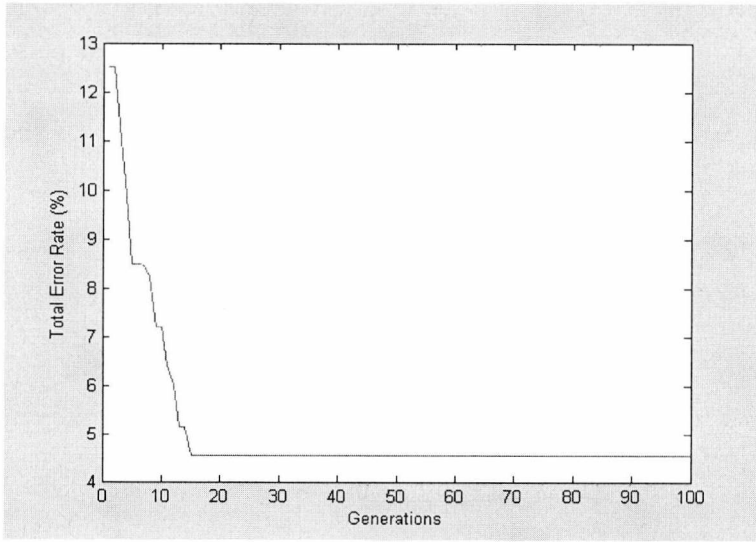



**Figure 7.16**: Minimum TER vs. Generation for the weighted summation rule

Table 7.7 shows the results obtained when using the *a posteriori* threshold computed by using the proposed optimising architecture on the test set.

**Table 7.7**: Results obtained for the weighted summation rule

| Weights | | | Threshold | FRR | FAR |
|---|---|---|---|---|---|
| W1 | W2 | W3 | | | |
| 0.48 | 0.03 | 0.49 | 4.14 | 7.5 % | 1.2 % |

Table 7.7 shows that the voice modality was the least weighted; this can be explained by the fact that the problem of computing the total error rate (TER) is a minimization problem, (that is, computing the minimum TER) and since the voice modality had the largest range of scores among the other modalities a small weight had to be assigned to it to minimize the TER during computation. It was also observed that the total error rate (TER) computed on the validation set was better than the TER computed on the test set.

### 7.7.3 Hybrid Decision Fusion

It was decided to perform some further experiments using a hybrid fusion method combining the majority voting rule in the hard decision fusion scenario and the sum rule in the soft decision fusion. The choice of the majority voting rule was based on its highest performance in reducing the TER among the other hard decision fusion methods. The approach adopted used the raw scores generated by each of the three biometric modalities of interest here. Figure 7.17 shows the hybrid architecture used to compute the TER. Three experiments were undertaken in this part of the investigation, based on the structure shown in Figure 7.17.

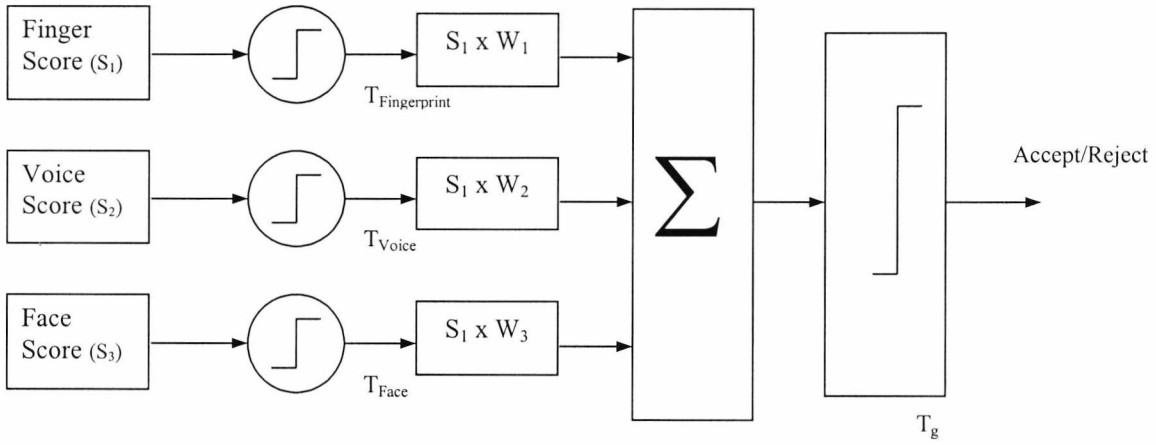



**Figure 7.17**: Evaluation of TER using Hybrid decision fusion

The experiments carried out for both the hard decision fusion and the soft decision fusion used the scenarios described in Chapter 6. The experiments in this section (the hybrid decision fusion) applied the following criteria in calculating the FAR and the FRR. If the user accessing a system is a legitimate user, then the identity claim of this user is accepted: first, if the scores of any two of the biometric modalities are greater than their pre-specified local threshold and second if the summed score is greater than the pre-specified global threshold. Hence, a legitimate user is falsely rejected by the system in two cases: first if the scores of any two of the biometric modalities are lower than their pre-specified local threshold and second if the scores of any two of the biometric modalities are greater than their pre-specified local threshold, but the summed score is less than the pre-specified global threshold. On the other hand, if the user accessing a system is an impostor trying to spoof the system, then the claimed identity of this user is accepted as being that of a legitimate user if the scores of any two of the biometric modalities are greater than their pre-specified local threshold and if the summed score is greater than the pre-specified global threshold.

**Experiment 1**

In this experiment the local thresholds of the biometric modalities were set to the default values (*a priori* thresholds) given by the vendors of the biometric devices, 5 for the fingerprint, 0 for the voice and 8.7 for the face. The majority voting rule was

applied, where the scores from at least two modalities must pass their local thresholds and then the proposed optimising architecture is used to compute both the best weights and the global threshold that minimises the TER. For this experiment each chromosome consists of four genes, three representing the weights for each biometric modality and one representing the global threshold for the sum rule as shown in Figure 7.18.

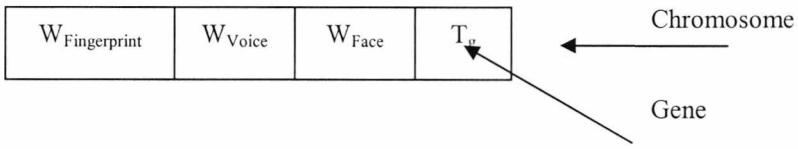



**Figure 7.18**: Chromosome representation for the hybrid method

Figure 7.19 shows the minimum TER plotted after every generation for the hybrid fusion. The graph shows that the minimum TER that could be attained by adjusting the weights and tuning the global threshold was 20.5 %. It also shows that this TER remained constant after the 15$^{th}$ generation. The weights and the global threshold computed from the validation set were then used to compute the TER on the test set.

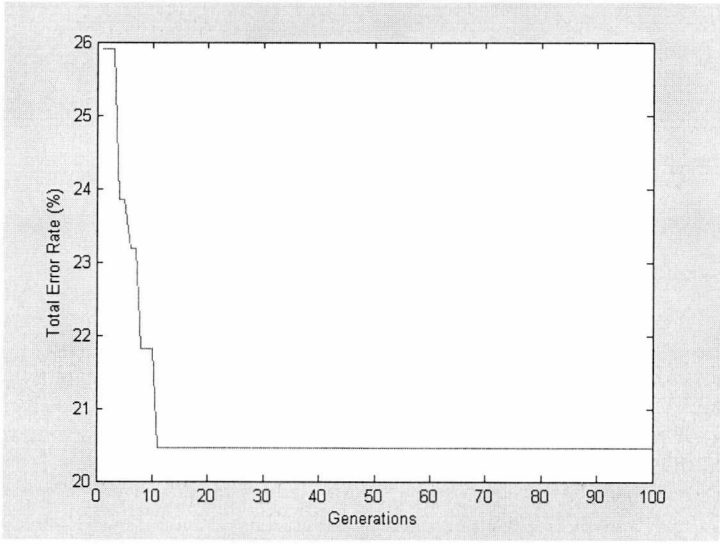



**Figure 7.19**: Minimum TER vs. Generation for the hybrid fusion

Table 7.8 shows the results obtained when using the computed *a posteriori* threshold on the test set.

**Table 7.8**: Results from using the *a posteriori* threshold on the test set after optimisation

| Weights | | | Threshold2 | FRR | FAR |
|---|---|---|---|---|---|
| W1 | W2 | W3 | | | |
| 0.28 | 0.02 | 0.70 | 3.6 | 17.7 % | 0.02 % |

The results showed that the TER computed using the *a posetriori* threshold on the test set are better than the TER computed on the validation set.

**Experiment 2**

In this experiment equal weights were assigned to each biometric modality. The proposed optimising architecture was used to compute both the local thresholds of each of the biometric modalities and the global threshold. For this experiment, each chromosome consists of four genes, three representing the local threshold of the biometric modalities and one representing the global threshold for the sum rule as shown in Figure 7.20.

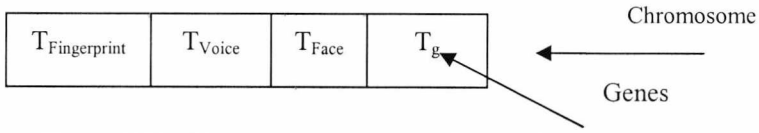



**Figure 7.20**: Chromosome representation for the hybrid method

Figure 7.21 shows the minimum TER plotted over the 100 generations for this experiment. From the graph it was observed that the minimum total error rate computed was 13 % and that this value remained constant after the $30^{th}$ generation. The computed local and global thresholds from the validation set were used to compute the TER on the test set.

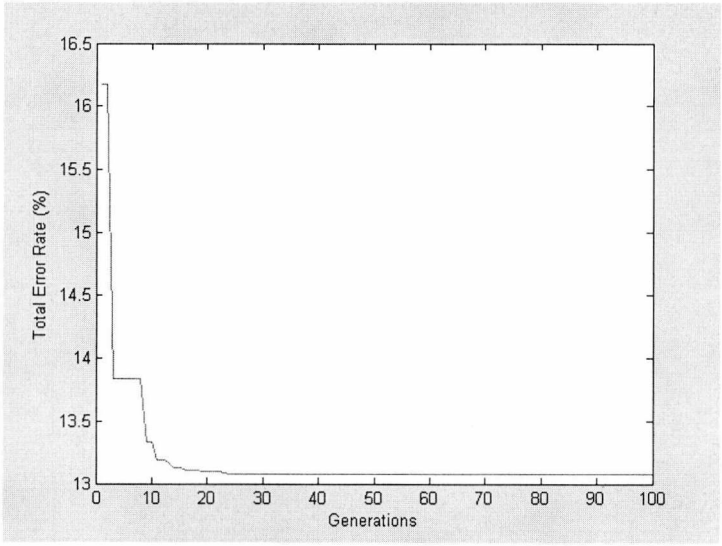



**Figure 7.21**: Minimum TER vs. Generation for the hybrid fusion

Table 7.9 shows the results of using the *a posteriori* threshold on the test set

**Table 7.9**: Results obtained after optimising the thresholds

| Biometrics | | | Threshold2 | FRR | FAR |
|---|---|---|---|---|---|
| Finger | Voice | Face | | | |
| 3 | -2 | 7.54 | 2.58 | 10.2 % | 1.0 % |

The TER computed on the test set is seen to be better than the TER computed on the validation set.

**Experiment 3**

In this experiment the proposed optimising method was used to compute the best weights, the local thresholds of the biometric modalities and the global threshold that minimises the TER. For this experiment, each chromosome consists of seven genes, three representing the weights for each modality, another three representing the local thresholds of the biometric modalities, and one representing the global threshold for the sum rule as shown in Figure 7.22.

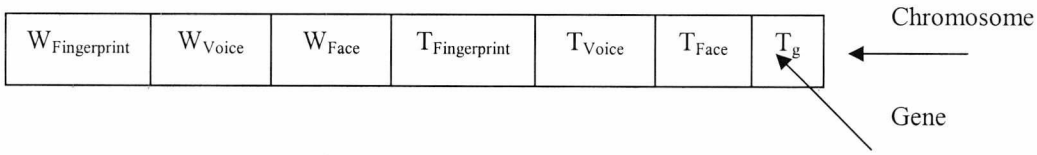



**Figure 7.22**: Chromosome representation for the hybrid method

The experiment showed that the minimum TER that was computed by tuning the thresholds and adjusting the different weights was 7.1 % as shown in Figure 7.23 and that this value remained constant after the 30$^{th}$ generation. The thresholds and the weights computed on the validation set were used to compute the TER tested on the test set. The TER obtained by using the computed *a posteriori* threshold on the test set is shown in Table 7.10.

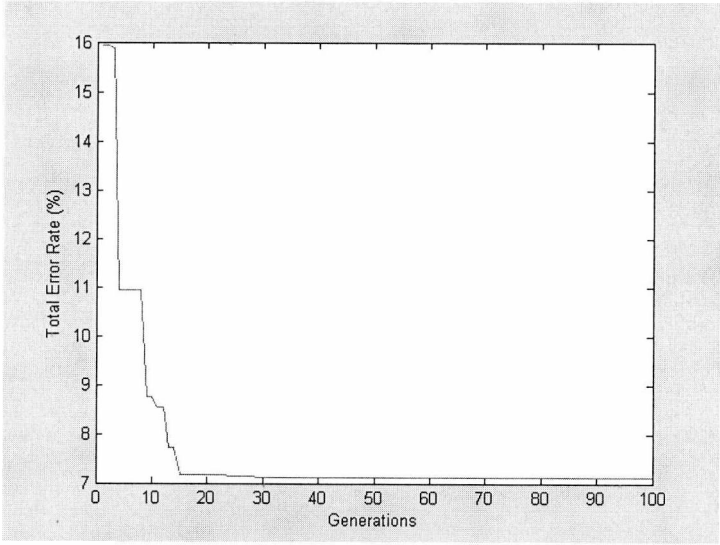



**Figure 7.23**: Minimum TER vs. Generation for the hybrid fusion

**Table 7.10**: Results obtained for the hybrid method

| Weights | | | Threshold1 | | | Threshold2 | FRR | FAR |
|---|---|---|---|---|---|---|---|---|
| W1 | W2 | W3 | Finger | Voice | Face | | | |
| 0.38 | 0.02 | 0.60 | 3 | 3 | 7.2 | 3.26 | 9.5 % | 1.2 % |

In this experiment the TER computed on the validation set was better than the TER computed using the *a posetriori* threshold on the test set.

Table 7.11 summarises the results obtained from the different experiments.

**Table 7.11**: Comparative performance of the soft and hybrid decision fusion methods

| Rule Used | Experiment undertaken | *a posteriori* Threshold (%) (Validation set) | | | *a posteriori* Measure (%) (Test set) | | |
|---|---|---|---|---|---|---|---|
| | | FRR | FAR | TER | FRR | FAR | TER |
| Sum Rule | Experiment 1 | 10.2 | 1.6 | 11.8 | 8.2 | 1.7 | 9.9 |
| | Experiment 2 | 3.4 | 1.2 | 4.6 | 7.5 | 1.2 | 8.7 |
| Hybrid Rule | Experiment 1 | 20.4 | 0.1 | 20.5 | 17.7 | 0.02 | 17.72 |
| | Experiment 2 | 12.2 | 0.8 | 13.0 | 10.2 | 1.0 | 11.2 |
| | Experiment 3 | 6.1 | 1.0 | 7.1 | 9.5 | 1.2 | 10.7 |

Table 7.11 shows that in the sum rule the results obtained from experiment 2 are better than the results computed from experiment 1. This is due to the fact that weighting was used in experiment 2. Weighting varies the importance of matching scores of each biometric modality, thus increasing the system performance. It was also observed that when using the hybrid rule, the performance is better when both the local thresholds of the biometric modalities and the global threshold are tuned (as in experiment 2) rather than when adjusting the weights and tuning the global threshold (as in experiment 1). The system performs even better if the weights are adjusted and both the local and the global threshold are tuned as in (experiment 3). The results also showed that in some experiments the TER in the test set was better than the TER computed on the validation set. This often happens if the test set is a well-behaved dataset, meaning that the verification samples provided by the users on that dataset were good samples.

The adoption of the hybrid method (as shown in Table 7.11) generates a higher TER than both the sum rule and the majority voting rule (as shown in Table 7.5)

### 7.7.4 Normalized-Sum Fusion

A further more complicated experiment was carried out in this section to explore the flexibility of the optimizing method in searching a larger space. It was decided to use the optimizing method to search for the parameters that normalising the scores before summing them. Figure 7.27 shows the normalised-sum method used to compute the TER. Two experiments were carried out in this section.

Using a general form of the Adaptive logarithmic method described in Chapter 5

$$S_{norm} = \frac{\exp^{(\frac{S-Threshold}{C})}}{1+\exp^{(\frac{S-Threshold}{C})}} \qquad (7.1)$$

Where

$S_{norm}$ : is the normalized score

$S$ : is the raw matcher score

*Threshold*: is the threshold of the biometric device.

$C$ : denotes the left and right edges of the region in which the function is linear, i.e. it exhibits linear characteristics in the interval (Threshold ± C). Figure 7.24 shows an example of the normalizing method, where the scores in the [0, 9] range are mapped to the [0, 1] range using Threshold = 5, C = 1.

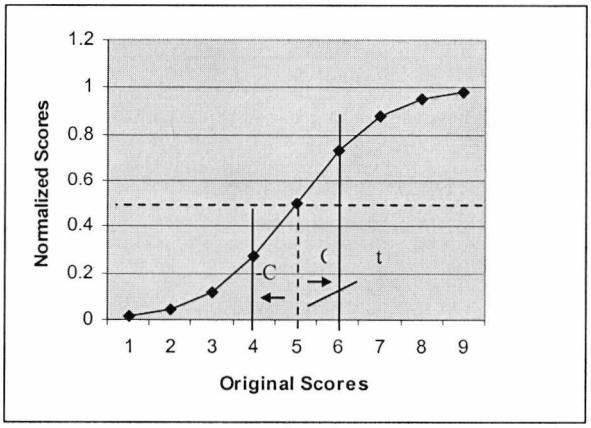



**Figure 7.24**: Normalisation method (threshold (t) =5, C= 1)

This method transforms the scores into the [0, 1] interval. But, it requires careful tuning of the parameters (Threshold (t) and C) to obtain good efficiency. In general, the threshold is chosen to be some value falling in the region of overlap between the genuine and impostor score distribution, and C is made equal to the extent of overlap between the two distributions toward the left and right of the threshold (t), respectively as shown in Figure 7.25. This normalization scheme provides a linear transformation of the scores in the region of overlap, while the scores outside this region are transformed non-linearly.

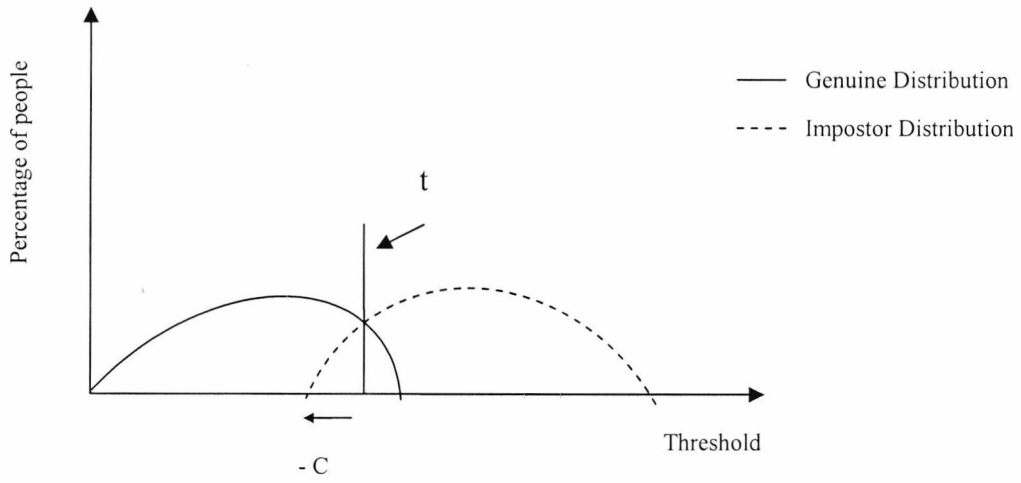
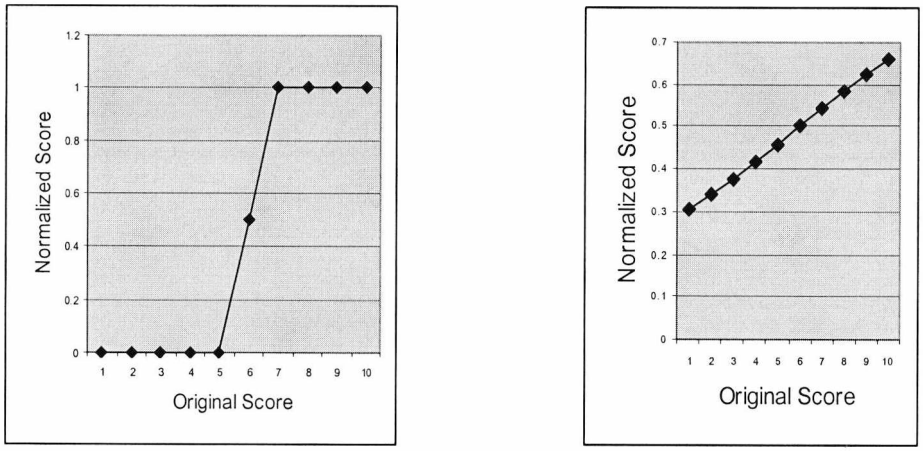



**Figure 7.25**: Distribution of genuine and impostor scores.

Changing the threshold (t) and the parameter (C) affects the way the scores are transformed into the region [0, 1] and hence changes the shape of the genuine and impostor distribution and trying to tune them for efficient results becomes difficult. When C becomes larger most of the scores are linearly transformed while the smaller C becomes less score are linearly transformed and the steeper is the slope as shown in Figure 7.26.




C is small C is large

**Figure 7.26**: transformation of scores into the range [0, 1]

Figure 7.26 shows the transformation of scores in the range [0, 1] when tuning the parameter C and fixing the threshold (t). The figure shows that increasing the parameter C does not normalize the scores into the range [0, 1] and hence a fine tuning of C is needed for efficient results.

Figure 7.27 shows the normalised-sum method used to compute the TER, two experiments were carried out to tune the thresholds and the parameter C for each of the biometric modalities.

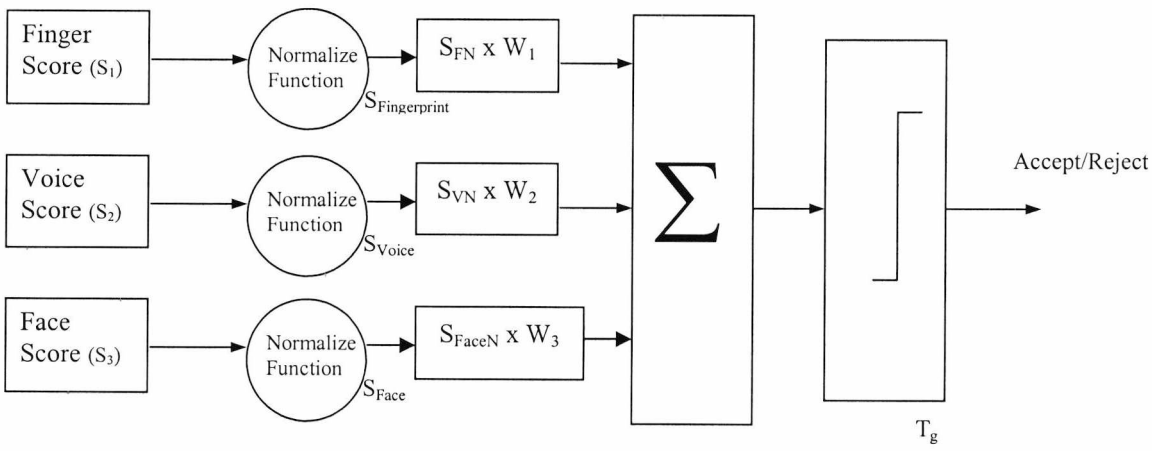



**Figure 7.27**: Evaluation of TER using normalized-sum fusion

**Experiment 1**

For this experiment, each chromosome consists of seven genes, three representing the local thresholds of the biometric modalities, another three representing the *C* parameter for each modality and one representing the global threshold for the sum rule as shown in Figure 7.28. For this experiment equal weights were assigned for each modality.

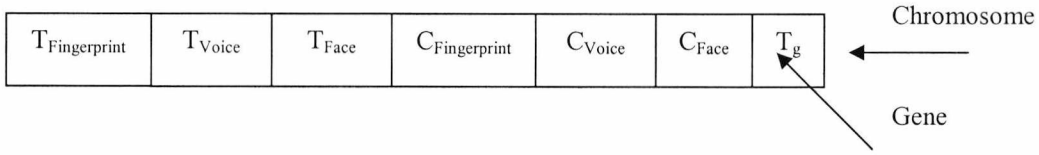



**Figure 7.28**: Chromosome representation for the normalized-sum fusion

Figure 7.29 showed that the minimum TER computed by tuning the thresholds and the parameter C was 6.68 % and that this value remained constant after the 20$^{th}$ generation. The thresholds and the weights computed on the validation set were used to compute the TER tested on the test set. The TER obtained by using the computed *a posteriori* threshold on the test set is shown in Table 7.11.

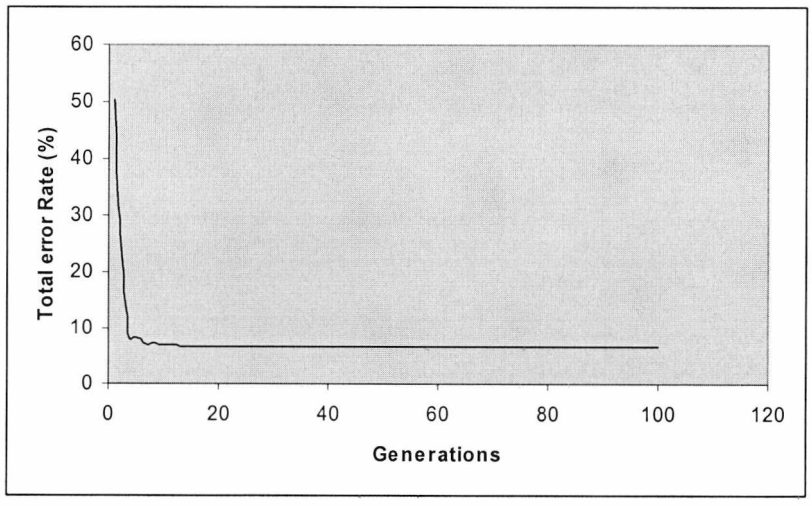



**Figure 7.29**: Minimum TER vs. Generation for the normalized-sum fusion

Table 7.12 shows the results obtained by using the optimising method to tune the different parameters

**Table 7.12**: Results obtained for the normalised summed fusion

| Threshold1 | | | C parameter | | | Threshold2 | TER |
|---|---|---|---|---|---|---|---|
| Finger | Voice | Face | $C_{Finger}$ | $C_{Voice}$ | $C_{Face}$ | | |
| 7 | 280 | 9 | 3.3 | 76.5 | 6.7 | 0.31 | 4.92 % |

Table 7.11 shows that by tuning the thresholds of each modality and adjusting the parameter C, the performance of the system improves compared to when using it without normalizing the scores as shown in Table 7.5.

**Experiment 2**

In this experiment, each chromosome consists of ten genes, three representing the local thresholds of the biometric modalities, another three representing the *C* parameter for each modality, three more representing the weights assigned for each modality and one representing the global threshold for the sum rule as shown in Figure 7.30.

| Fingerprint | Voice | Face | $C_{Fingerprint}$ | $C_{Voice}$ | $C_{Face}$ | $W_{Fingerprint}$ | $W_{Voice}$ | $W_{Face}$ | $T_g$ |
|---|---|---|---|---|---|---|---|---|---|

Gene

**Figure 7.30**: Chromosome representation for the normalized-sum fusion

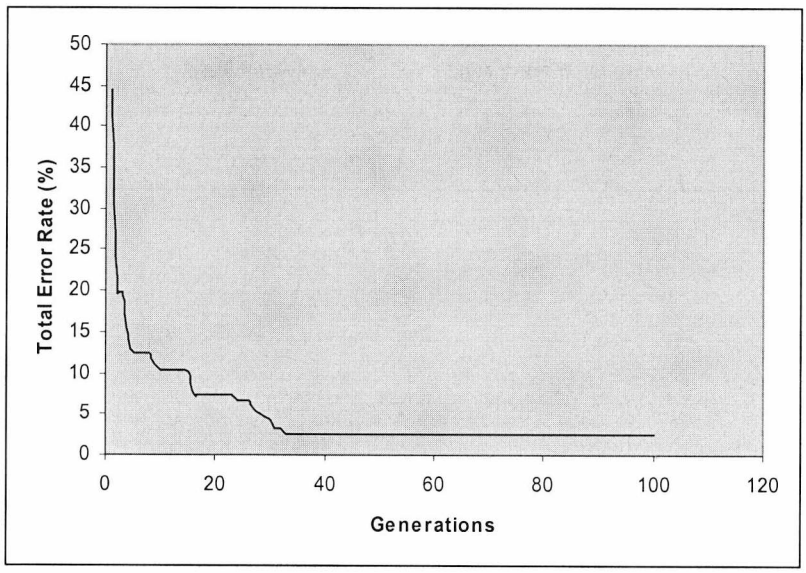



**Figure 7.31**: Minimum TER vs. Generation for the normalized-sum fusion

Figure 7.31 shows the minimum TER plotted over the 100 generations for this experiment. The graph show that the minimum total error rate computed was 2.47 % and that this value remained constant after the 35$^{th}$ generation. The computed parameters from the validation set were used to compute the TER on the test set.

Table 7.13 shows the results obtained by using the optimising method to tune the different parameters

**Table 7.13**: Results obtained after normalising the different parameters

| Threshold1 | | | C parameter | | | Weights | | | Threshold2 | TER |
|---|---|---|---|---|---|---|---|---|---|---|
| Finger | Voice | Face | $C_{Finger}$ | $C_{Voice}$ | $C_{Face}$ | $W_{finger}$ | $W_{Voice}$ | $W_{Face}$ | 0.40 | 3.0 % |
| 7 | -151.2 | 6.26 | 4.2 | -42.9 | 3.26 | 0.2 | 0.2 | 0.6 | | |

Table 7.13 shows that by adjusting the different parameters, the performance of the system improves even better compared with the results obtained from Table 7.7

The experiments that have been carried out in this chapter explored the power of using genetic algorithm. Genetic algorithm provided a flexible structure where it can be used to search a small space of only one gene as shown in Figure 7.13 as well as a large space with up to 10 genes as shown in Figure 7.30. In addition to that, the experimental results of this study showed that in all the decision fusion methods used, the minimum TER was attained between the $5^{th}$ and the $35^{th}$ generation which means that the genetic algorithm provides the optimum results in a less computational time when compared with other search method such as the exhaustive search which will take between 100 iteration for one gene to $4*10^{20}$ iterations for 10 genes before reaching the optimal solution. Another advantage that genetic algorithms have over other search methods is its ability of finding several solutions which provides the same result; this gives flexibility in choosing the best result for the problem. The experiments showed that genetic algorithm solved two of the main problems in the multimodal biometric field which are the setting of the thresholds and the normalization of scores.

## 7.8 Discussion

A number of interesting points may be drawn from the different experiments that were undertaken. The results reported in this chapter are based on testing the proposed optimising architecture and investigating its performance when using the different fusion rules. According to the experimental results of this study, the majority voting rule in the hard decision fusion methods had the highest expected reduction in the total error rate (TER) among the other rules and this was achieved by only tuning the local thresholds slightly from their *a priori* adjustment. The experimental results also showed that in all the hard decision fusion methods the total error rate (TER) computed on the validation set was better than the total error rate (TER) computed by using the *a posteriori* thresholds on the test set. This can be explained by the fact that using the same data set for both training and testing results in overestimating the performance of the system. Comparing the hard decision fusion results obtained when using the *a priori* thresholds with the one generated by using the proposed optimising method showed that the *a posteriori* thresholds obtained by using the proposed optimising method resulted in an improvement in the

optimising architecture solved the problem associated with threshold setting by tuning the local thresholds of the biometric modalities.

The experimental results also showed that the proposed optimising architecture not only solved the problem of threshold setting, but also the problem of score normalization by combining the raw scores of the biometric modalities using the sum rule without having to normalise the scores beforehand. In addition, the optimising architecture solved the problem associated with weight setting as it was able to adjust the weights when using the weighted summation rule and increase the system performance. The results also showed that by tuning the different parameters of the normalizing function increases the system performance. The results showed that the weighted sum rule performed better than the sum rule, proving that varying the relative importance of the matching scores of each biometric modality increases the system performance.

Comparing the hybrid decision fusion results with that of the majority voting and the sum rule showed that both the majority voting and the sum rule performed better alone than when combining them.

Considering the number of generations required to reach the optimal solution (minimum TER), the experimental results of this study showed that in all the decision fusion methods used, the minimum TER was attained between the $5^{th}$ and the $30^{th}$ generation. The experiments also showed that the optimal solution is found much sooner if the space to be searched is smaller, such as in the case of finding the optimal weights, as they lie in the range [0, 1]. However, if the space is large, as in the case of finding the thresholds, more generations are required to reach the optimal solution.

It can be concluded that the proposed optimising architecture solved the problems associated with threshold settings. This clearly demonstrates a valuable potential strategy that can be used to set the *a priori* thresholds of the different biometric devices before using them. The proposed optimising architecture solved the problem of score normalisation, which makes it an effective "plug-and-play" design philosophy to system implementation. It also solved problems associated with

weight settings, which is used in many applications for varying the relative importance of the different parameters used.

## 7.9 Summary

In this chapter a proposed approach for the optimisation of a multimodal person recognition system based on the use of genetic algorithm was described. Several experiments were undertaken to test the proposed optimising architecture and to investigate its performance when using the different possible fusion rules. A hybrid decision fusion rule was also explored which combines the majority voting and the sum rule. The experimental results showed that the proposed approach could play an important role in system optimisation by, for example, determining system parameters, which reduce the total error rate. A further benefit of this approach is its ability in resolving some of the problems associated with score normalization in setting weights/thresholds.

The next chapter presents a summary of the work carried out in this thesis and suggests some further work for the future.

# Chapter 8

# Conclusions and Further Work

## 8.1 Introduction

This chapter summarises the work presented in this thesis and presents the main conclusions that have been drawn from the work. It also provides some suggestions for future research. First, a summary of each chapter is provided as a separate section.

### 8.1.1 Chapter 2: Multimodal Biometric System Concepts

This chapter presented a review of research studies in the field of multimodal biometric systems concerning possible fusion approaches and the multimodal databases. The literature review emphasised combining multiple modalities and showed that the overall performance of the system can improve by integrating multiple biometric systems rather than by using a single biometric alone. In this chapter the different architectures for combining classifiers were described from which the parallel architecture was chosen to be adopted in this work. Four different levels on which the data can be fused were discussed: the sensor data level, the feature level, the matching score level and the decision level. A brief overview describing these fusion approaches in combining multimodal biometric system

revealed the obvious preference of combining multiple biometrics at the decision level (both the matching score level and the decision level) compared with combining them at the sensor and the feature level, and this made the fusion at the decision level the main focus of this work In this chapter the decision fusion level was divided into two sub-levels; a hard decision level, where the scores combined used voting techniques, and a soft decision level, where the scores combined used the summation rule.

The literature also revealed some of the challenges that play an important role in enhancing the performance of multimodal biometric systems, such as the choice of a normalization method that maps the output scores of different biometric modalities into a common interval [0,1] before fusing them (see, for example, Snelick in [Snelick03]).

The literature also showed that one of the important factors in evaluating the performance of automatic recognition systems and in fusing multiple modalities is the availability of a large multimodal biometric database acquired under real conditions for testing the algorithms. An overview of the publicly available medium and large-scale multimodal databases was provided. These databases were not used in this study for various reasons such as their unavailability at the start of this research, their modest size and the fact that some of them consist only of two biometric modalities where it was preferred in this study to evaluate the performance of a system with more than two biometric modalities.

The focus of this study was therefore to determine methods and ways of solving some of the important problems and challenges, which the literature survey had identified.

### 8.1.2 Chapter 3: Data Collection Trial

In this chapter the problem of the unavailability of a large multimodal biometric database acquired under real conditions to be used for testing different algorithms was solved by developing a multimodal biometric database to be used in this work. This chapter described the data collection exercise that was undertaken at the

University of Kent as part of this project. Three biometric modalities were collected in this project: fingerprint, voice and face, using commercially available devices (with verification thresholds set mid range according to manufacturers' specifications). Fingerprint was chosen for its high reliability, voice and face for their low-cost hardware and high user acceptability. The data collection protocol that was followed was linked to a scenario-testing regime and was developed within the emerging guidelines for best practice in biometric testing.

The data collection exercise involved collecting biometric samples from 221 volunteers (45 % females and 55% males) with an age range from 18 years to 65 years and above. The data collection protocol adopted required that each volunteer undertake two separate data collection sessions, the first involving enrolment on each of the systems adopted (up to three enrolment attempts per device were permitted), together with a post-enrolment verification check. Each volunteer undertook three verification attempts at this session. A second session was undertaken at least one month later where three additional verification attempts were carried out using the enrolment templates generated at the first session. This data collection exercise continued for almost 6 months, starting in November 2001 and finishing in May 2002.

### 8.1.3 Chapter 4: The Multimodal Database and Some Preliminary Analysis

In this chapter a preliminary analysis on the collected multimodal database was provided. The analysis focused mainly on performance comparisons between verification based on single individual biometric modalities and approaches, which exploit the opportunity to combine data from multiple biometric modalities. The analysis focused mainly on Type I verification errors (FRR). Several experiments were reported in this chapter to estimate the performance of the system achieved with typical users and using commercially available devices.

The experiments reported in this chapter investigated several factors that affected the performance of the system, such as the failure to enrol rate, the time-based changes in biometric data and the "goat" phenomenon. The experiments also

exploited the effects of re-try strategies and learning effects. Finally, a list of some of the users-related factors which were observed by the supervisor during the trial that influenced both the enrolment process and the verification process was provided

A number of interesting points were drawn from the different experiments that were carried out. According to the experimental results of this study, the failure to enrol rate reduces when more than one biometric modality is used. The experimental results also showed that the performance of a single modality might change dramatically over time, although it is unlikely for two or more modalities to do the same. It was also demonstrated that "goats" exist in every biometric system and that an individual who is a "goat" in one biometric modality is unlikely to be also a "goat" in a different one. Finally, exploiting the effects of re-try strategies and learning effects showed how training reduces the error rates and improves the system performance.

### 8.1.4 Chapter 5: Combining Multimodal Biometric System using Decision Fusion.

In this chapter a brief review of research studies of the most commonly used fusion rules at the decision level revealed that for combining soft decisions the sum rule outperformed other combination rules, which supported the idea of adopting and using this rule in this study. Regarding hard decisions, the AND and OR rule seem to be the dominating rules. The review also revealed that most of the research that used majority voting rule used it in combining soft decisions, though in this study it was used also to combine hard decisions. The chapter also presented the two most commonly used score normalisation methods and proposed a novel method for score normalisation.

In this chapter a set of experiments were carried out to compare the performance of the hard decision fusion methods (AND, OR and majority voting) with that of the soft decision fusion method (sum rule). The experiments were also based on investigating the effect of characterising the individual users as *sheep*, *goats*, *lambs* and *wolves* on the performance of the system.

The comparative analysis showed that the hard decision fusion rules especially the majority voting and the OR rule, outperformed the soft decision fusion rules in reducing the false reject rate (FRR) whereas the soft decision fusion method (sum rule) outperformed the hard decision fusion methods with the exception of the AND rule in reducing the false accept rate to zero. From the results, a general conclusion was drawn that the sum rule in the soft decision fusion performs better than the hard decision fusion method for the present system since it reduced the FAR to zero and the FRR to a level, which is acceptable for many applications.

The characterisation of individual users as *sheep*, *goats*, *lambs* and *wolves* and their effect on the performance of the system has been considered by several researchers such as [Doddington98] [Pankanti02]. However, none of these has considered analysing what effects the different fusion rules have on *sheep*, *goats*, *lambs* and *wolves*. In this chapter the comparative analysis that was carried out on the effects of the hard decision fusion methods (AND, OR and majority voting) and the soft decision fusion methods (sum rule) on the characterisation of the users as *sheep*, *goats*, *lambs* and *wolves* determined that the AND rule eliminates the *lambs* and *wolves*, but increase the proportion of *goats* in the system. This is in contrast with the OR rule, which decreases the proportion of *goats* and increases both the number of *lambs* and *wolves* in the system. The majority voting approach performed better than both the AND rule and the OR rule in providing a relatively low proportion of *goats*, *lambs* and *wolves*. On the other hand, the soft decision fusion method (sum rule) diminished the *lambs* and *wolves* from the system and provided a relatively low proportion of *goats*.

In this chapter four different types of *wolves* were proposed and different experiments were carried out to investigate their effect on the system performance

### 9.1.5 Chapter 6: Introduction to Genetic Algorithm

In this chapter an introduction to genetic algorithms and the different operators that influence their performance was provided. A comparison between genetic algorithms and other optimisation techniques was presented in order to illustrate the

advantages that genetic algorithms have over other approaches. The chapter also presented some application areas of genetic algorithms particularly in biometric-based recognition of individuals. An overview of the research studies showed that most of the research applied genetic algorithms in mono-modal biometric systems and at the feature fusion level.

### 8.1.7 Chapter 7: Optimising Multimodal Person Recognition

In this chapter an approach to the optimisation of a multimodal person recognition system based on the use of genetic algorithms was proposed. The proposed optimising architecture played an important role in resolving some of the problems associated with score normalisation in setting weights/thresholds. The experiments performed in this chapter aimed at setting the threshold and weights that provide a minimum total error rate (TER) when using the different decision fusion rules. The experiments in this chapter were carried out using a constant population size of 100, a crossover of 0.6 and a mutation rate of (1/ number of genes) as described by the genetic algorithm toolbox.

Several points were drawn from the different experiments that were carried out. The experimental results of this study showed that in the hard decision fusion methods the majority voting had the highest expected reduction in the TER, which was achieved by slightly tuning the local threshold from their *a priori* settings. The results also showed that in the hard decision fusion methods the *a posteriori* thresholds computed by using the proposed optimising method improved the performance of the system more than when using the *a priori* thresholds. This shows that the proposed optimising method particularly addressed the problem associated with threshold settings in the hard decision fusion methods.

The proposed optimising method evenly addressed the problem of score normalisation by combining the raw scores using the sum without having to normalise them beforehand. The problem associated with weight setting was also solved by the proposed optimising method by adjusting the weights when using the weighted summation rule.

## 8.2 Suggestions for Future Work

Some suggestions are proposed here for future work. Although the experiments were carried out on a database that contains three biometric modalities (fingerprint, voice and face), it is suggested that an investigation should be carried out using a larger database with more than three modalities such as the BIOMET database.

In Chapter 5 different types of *wolves* were proposed. A useful area for further study is to investigate issues concerning these *wolves* such as the reasons of their existence and whether their existence is a result of the biometric devices used or their possession of some distinctive features, their average age and gender and whether a male could be a wolf and spoof a female template and vice versa. A suggestion is to test *wolves* on different biometric devices.

Finally, the proposed optimising method was used at both the score and the decision level. An important area of further study is to explore the use of similar optimisation techniques for the fusion of multiple modalities at the feature level and to investigate the effects on the overall system performance.

## 8.3 Summary

The main objective of this project was to investigate the fusion of multimodal biometric verification systems and to evaluate their performance. A multimodal biometric database that consists of fingerprint, voice and face modalities was gathered for this purpose. The multiple modalities were fused both at the score and the decision level to support a system that can meet challenging and varying requirements. Some issues related to the implementation of multimodal biometric system was also addresses such as the setting of the verification thresholds adopted by each biometric device. The main achievements of this work are summarized as follow:

1. The development of a multi-modal database for use in person verification experiments.

2. The proposal of a novel method for score normalization.

3. A comparative analysis of the performance of different fusion rules when characterizing the system users as *sheep, goats, lambs* and *wolves*

4. The proposal of a novel approach for the optimization of multimodal biometric systems based on the use of genetic algorithms for solving problems associated with weights/thresholds settings and score normalization.

# Part 2

## Bibliography and Appendices

# Bibliography


[**Ackley92**] D. H. Ackley and M. L. Littman, " Interactions Between Learning and Evolution", In C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmusen, editors, *Artificial Life II*, pp.487-507, 1992.

[**Alkoot99**] F. M. Alkoot, J. Kittler, " Experimental Evaluation of Expert Fusion Strategies", *Pattern Recognition Letters*, Vol. 20, pp. 1361-1369, 1999.

[**Allgrove99**] C. Allgrove, "A Study of Automatic Signature Verification Techniques in the Context of a Practical Document Processing Application", 1999

[**ATM99**] "ATM to Use Face Recognition Technology"
http://www.networkusa.org/fingerprint/page5a/fp-atm-facial-scans.html

[**Auckenthaler00**] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score Normalization for Text-independent Speaker Verification Systems," *Digital Signal Processing*, vol. 10, pp. 42–54, 2000.

[**Axelrod86**] R. Axelrod, " An Evolutionary Approach to Norms", *The American Politics Science Review*, 80, 1986.

[**Back91**] T. Bäck and F. Hoffmeister, " Extended Selection Mechanisms in Genetic Algorithms" *In Proceedings of the Fourth International Conference on genetic Algorithms*, San Mateo, California, USA, pp 92-99,1991

[**Back93**] T. Bäck and H. P Schwefel, "An Overview of Evolutionary Algorithms for Parameter Optimization", *Evolutionary Computation*, Vol.1, No.1, pp. 1-23, 1993. http://lumpi.informatik.uni-dortmund.de/people/baeck/papers/ec93.ps.Z

[**Bailly-Baillire03**] E. Bailly-Baillire, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, J.-Ph. Thiran, "The BANCA Database and Evaluation Protocol", *AVBPA 2003, LNCS 2688*, pp. 625-638, 2003.

**[Baker85]** J. E. Baker, " Adaptive Selection Methods for Genetic Algorithms", *Proceedings of an International Conference on Genetic Algorithms and their Applications*, pp. 101-111, 1985.

[**Baker87**] J. E. Baker, " Reducing Bias and Inefficiency in the Selection Algorithm", *Proc. ICGA 2*, pp. 14-21, 1987.

[**Bala96**] Bala J., K. DeJong, J. Huang, H. Vafaie, and H. Wechsler, " Visual Routine for Eye Detection Using Hybrid Genetic Architectures", *Proc. of 13th International Conference on Pattern Recognition (ICPR)*, pp.50-54, Vienna, Austria, 1996.

[**Bala97**] J. Bala, K. DeJong, J. Huang, H. Vafaie, and H. Wechsler, " Using Learning to Facilitate the Evolution of Features for Recognizing Visual Concepts", *Evolutionary Computation*, Vol .4, No.3, pp.1-14, 1997.

[**BANCA**] http://www.ee.surrey.ac.uk/banca/

[**Baltzakis01**] H. Baltzakis and N. Papamarkos, " A New Signature Verification Technique Based on a Two- Stage Neural Network Classifier", *Engineering applications of Artificial Intelligence*, Vol.14, No.1, pp.95-103, 2001

[**Banzhaf99**] W. Banzhaf, and C. Reeves," Foundations of Genetic Algorithms 5", San Francisco, California, USA: Morgan Kaufmann Publishers, 1999.

[**Barniv81**] Y. Barniv and D. Casasent, " Multisensor Image Registration: Experimental Verification", *Proceedings of the SPIE,* Vol. 292, pp. 160-171, 1981.

[**BBC02**] BBC News, " Hi-tech Signatures to Fight Fraud", 8th November, 2002
http://news.bbc.co.uk/1/hi/technology/2420143.stm

[**Beiser99a**] V. Beiser, "Casinos Fight Back with Tech", *Wired News*, May 4, 1999

[**Beiser99b**] V. Beiser, "Biometrics Breaks into Prisons", *Wired News*, August 21, 1999
http://www.wired.com/news/technology/0,1282,21362,00.html

[**Belew90**] R. K. Belew, " Evolution, Learning and Culture: Computational Metaphors for Adaptive Algorithms", *Complex Systems*, Vol. 4, pp 11-49, 1990.

[**Belew92**] R. K. Belew, J. McInerney and N. N. Schraudolph, " Evolving Networks: Using the Genetic Algorithm with Connectionist Learning", In C. G. Langton, C. Taylor, J. D

Farmer and S. Rasmussen, editors, *Artificial Life II*, Santa Fe Institute studies in the Science of Complexity, pp. 511-547,1992.

[**Belew97**] R. K Belew and M. D Vose, " Foundations of Genetic Algorithms 4", San Francisco, California, USA: Morgan Kaufmann Publishers, 1997.

[**Bengio01**] S. Bengio, J. Marithoz, and S. Marcel, "Evaluation of Biometric Technology on XM2VTS", IDIAPRR 21, IDIAP, 2001.
http://citeseer.ist.psu.edu/bengio01evaluation.html

[**Bengio02**] S. Bengio, Christine Marcel, Sébastien Marcel and J. Mariéthoz, " Confidence Measures for Multimodal Identity Verification", *Information Fusion*, Vol.3, No.4 pp. 267-276, 2002.

[**Bengio04**] S. Bengio and J. Mariéthoz, " The Expected Performance Curve: A New Assessment Measure for Person Authentication", In *Proceedings of Odyssey 2004: The Speaker and Language Recognition Workshop*, 2004.

[**Ben-Yacoub98**] S. Ben-Yacoub, "Multi-Modal Data Fusion for Person Authentication using SVM", *IDIAP-RR 7*, IDIAP, 1998.

[**Bergman92**] A. Bergman and M. W. Feldman, " Recombination Dynamics and the Fitness Landscape", *Physica D*, Vol.56, pp 57-67, 1992

[**Bigun97**] E. Bigun, J. Bigun, B. Duc and S. Fischer, "Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics", *In Proceedings of First International Conference on AVBPA*, (Crass-Montana, Switzerland), pp. 291-300, March 1997.

[**Bimbot97**] F. Bimbot and G. Chollet, "Assessment of Speaker Verification Systems". In *Handbook of Standards and Resources for spoken language systems*, Mouton de Gruyter, 1997.

[**Bin Azhar02**] M. A. Hannan Bin Azhar, " Design of an FPGA Based Adaptive Controller for Collision-free Robot Navigation", MSc Thesis, Department of Electronics, University of Kent, Canterbury, UK, 2002.

[**Blickle95**] T.Blickle and L. Thiele, ": A Comparison of Selection Schemes used in Genetic Algorithms (2. Edition). TIK Report No. 11, Computer Engineering and Communication

Networks Lab (TIK), Swiss Federal Institute of Technology (ETH) Zürich, Switzerland, 1995.
http://www.tik.ee.ethz.ch/Publications/TIK-Reports/TIK-Report11abstract.html

[**Bolle00**] Ruud. M. Bolle, Sharath Pankanti and Nalini. K. Ratha, " Evaluation Techniques for Biometrics-Based Authentication Systems (FRR)", *Proceedings International Conference on Pattern recognition (ICPR)*, pp. 2831-2837, Barcelona, Spain, September 03-08, 2000.

[**Booker87**] L. Booker, "Improving Search in Genetic Algorithms", *In Genetic Algorithm and Simulated Annealing*, L. Davis (Ed.), pp 61-73,Morgan Kaufmann Publishers, 1987.

[**Bramlette91**] M. F. Bramlette, "Initialization, Mutation and Selection Methods in Genetic Algorithms for Function Optimization", *Proc ICGA 4*, pp. 100-107, 1991.

[**Bremermann62**] J. H. Bremermann, " Optimization through Evolution and Recombination". *In Yovits, M. C. et al.: Self-organizing systems*. Washington, DC: Spartan Books, pp. 93-106, 1962.

[**Brian Arthur93**] W. Brian Arthur, " On Designing Economic Agents that Behave Like Human agents", *Evolutionary Economics*, Vol 3, pp 1-22, 1993.

[**Brooke94**] N. M. Brooke, J. M. Tomlinson and K .R. Moore, " Automatic Speech Recognition that includes Visual Speech Cues ", *Proceedings of the Institute of Acoustics-1994 Autumn Conference of Speech and Hearing*, vol.16, No.5, pp. 15-22, 1994.

[**Brunelli95**] R. Brunelli and D.Falavigna, " Person Identification Using Multiple Cues", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.17, No.10, pp.955-966, October1995.

[**Bunday84**] B. D. Bunday, " Basic Optimisation Methods", *Edward Arnold*, 1984.

[**BWG00**] Biometrics Working Group, Best Practices in Testing and Reporting Performance of Biometric Devices, January 2000.

[**BWG02**] Biometrics Working Group, Biometrics for Identification and Authentication-Advice on Product Selection, March 2002.

[**Caldwell91**] C. Caldwell and V. S. Johnson, " Tracking a Criminal Suspect through Face-Space with a Genetic Algorithm", *In Proceedings of the Fourth International Conference on Genetic Algorithms*, pp 416-421, 1991.

[**Campbell97**] J. P. Campbell, " Speaker Recognition: A Tutorial", *Proceedings of the IEEE*, Vol.85, No.9, pp. 1437-1462, 1997

[**Cantu-Paz03**] E. Cantu-Paz and D. E. Goldberg, "Are Multiple Runs of Genetic Algorithms Better than One", *LNCS 2723*, pp.801-812, 2003.

[**Canuto00**] A Canuto, G Howells, And M Fairhurst, " The Use of Confidence Measures to Enhance Combination Strategies in Multi-Network Neuro-Fuzzy Systems", *Connection Science*, 12(3/4), pp.315-331, 2000.

[**Caruana89**] R. A. Caruana, L. A. Eshelman and J. D. Schaffer, "Representation and Hidden bias II: Eliminating defining length bias in genetic search via shuffle crossover", *In Eleventh International joint Conference on Artificial Intelligence*, N. S. Sridharan (ED.), Vol.1, pp 750-7555, Morgan Kaufmann Publishers, 1989.

[**Celada92**] F. Celada and P. E. Seiden, " A computer Model of Cellular Interactions in the Immune System", *Immunology Today*, Vol 13, No 2, pp 56-62, 1992.

[**Centor91**] R. M. Centor, " Signal Delectability: TheUse of ROC Curves and their Analysis", *Medical Decision Making*, Vol 11, pp. 102-106, 1991.

[**Cheung04**] M.C. Cheung, M.W. Mak and S.Y. Kung, "Intramodal and Intermodal Fusion for Audio-Visual Biometric Authentication", *Proc. International Symposium on Intelligent Multimedia, Video, and Speech Processing,* 2004.

[**Chibelushi93a**] C. C. Chibelushi, J. S. Mason and F. Deravi, " Integration of Acoustic and Visual Speech for Speaker Recognition", *EUROSPEECH'93*, pp.157-160, 1993.

[**Chibelushi93b**] C. C. Chibelushi, F. Deravi and J. S. Mason, " Voice and Facial Image Integration for Speaker Recognition", *IEEE International Symposium and Multimedia Technologies and Future Applications*, Southampton, UK, 1993.

[**Chibelushi96**] C. C. Chibelushi, S Gandon, J. S. Mason, F. Deravi, and D Johnston, " Design Issues for a Digital Integrated Audio-Visual Database". *IEE Colloquium on Integrated Audio-Visual Processing for Recognition, Synthesis and Communication.* London, Digest Reference Number 1996/213, pages 7/1-7/7, November 1996.

[**Chibelushi97a**] C. C. Chibelushi, J. S. Mason and F. Deravi, " Audio-Visual Person Recognition: An Evaluation of Data Fusion Strategies", *Proceedings of the European Conference on Security, IEE*, London, pp 26-30, April 1997.

[**Chibelushi97b**] C. C. Chibelushi, J. S. Mason and F. Deravi, "Feature Level Data Fusion for Bimodal Person Recognition", *Sixth International Conference on Image Processing and its Applications, IEE*, pp 339-403, July 1997.

[**Chibelushi99**] C. C. Chibelushi, F. Deravi and J. S. Mason, " Adaptive Decision Fusion for Robust Pattern Recognition", *IEEE Transactions of Systems, Man and Cybernetics- Part B: Cybernetics*, Vol.29, No.6, pp. 902-907,1999.

[**CNN02**] CNN World News, "Schiphol Backs Eye Scan Security", March 27, 2002 http://www.cnn.com/2002/WORLD/europe/03/27/schiphol.security

[**Cole95**] G. Cole, "Giving Voice to Security", *Financial Times*, September 15, 1995 http://www.weverify.com/vervcsec.htm

[**Collins92**] R. J. Collins and D. R. Jefferson, " AntFarm: Towards Simulated Evolution", In C. G. Langton, C. Taylor and S. Rasmussen, editors, *Artificial Life II*, pp 579-601,1992

[**Coots02**] F. T. Cootes, H. Kang, G. Wheeler, L. Butcher and J. C. Taylor, " Face Recognition Using active Appearance Models" *BMVA Symposium on Advancing Biometric Technologies*, Royal Statistical society, London 200

[**Court03**] W. Court, " Biometrics: Evaluation Criteria and Scenario Based Performance Testing", *GSEC*, June 2003.

[**Czyz04**] J. Czyz, J. Kittler and L. Vandendorpe, " Multiple Classifier Combination for Face-Based Identity Verification", *In Pattern Recognition*, Vol 37, pp 1459-1469, 2004

[**Das**] R. Das, "An Application of Biometric Technology: Signature Recognition", Technology Executives Club.
http://technologyexecutivesclub.com/artbiomterissignature.htm

[**Das1**] R. Das, "An Application of Biometric Technology: Voice Recognition", Technology Executives Club.
http://technologyexecutivesclub.com/artbiomterissignature.htm

[**Dasarathy94**] B. V. Dasarathy, " Decision Fusion", *IEEE Computer Society Press*, 1994.

[**Daugman00**] J. Daugman, " Biometric Decision Landscape", Technical report No. TR482, University of Cambridge, Computer Laboratory, 2000.

[**Davies94**] S. G. Davies, "Touching Big Brother: How biometric technology will fuse flesh and machine", *Information Technology and People*, vol.7, no. 4, 1994.

[**Davidor91**] Y. Davidor, " Genetic algorithms and robotics", *Robotics and Automated Systems, World Scientific*, Singapore, 1991.

[**Davis91**] L.D. Davis, Handbook of Genetic Algorithms. Van Nostrand Reinhold, 1991.

[**De Jong75**] K. A. De Jong," An Analysis of the Behaviour of a Class of Genetic Adaptive Systems", PhD thesis, The University of Michigan, ANN Arbor, MI, 1975.

[**Dieckmann97**] U. Dieckmann, P. Plankensteiner and T. Wagner, " Sesam: A Biometric Person Identification System using Sensor Fusion", *Pattern Recognition Letters*, Vol.18, No.9, pp. 827-833, September 1997.

[**Dixon78**] L. C. W Dixon and G. P Szego, " The Optimization Problem: An Introduction" In Dixon, L. C. W. and Szego, G. P. (ed.): Towards Global Optimization II, New York: North Holland, 1978.

**[Doddington98]** G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, " Sheep, Goats, Lambs and Wolves: a Statistical Analysis of Speaker Performance in the NIST1998 Speaker Recognition Evaluation", *Proc. of ICSLD 98*, Sydney, Australia, November 1998.

[**Duc97**] B. Duc, G. Maytre, S. Fischer and J. Bigun, " Person Authentication by Fusing Face and Speech Information", *Ist Int'l Conference on Audio- and Video-based Biometric Person Authentication*, LNCS 1206, pp. 311-318,1997.

[**Duc97a**] B. Duc, E. S. Bigun, J. Bigun, G. Maitre, and S. Fischer, " Fusion of Audio and Video Information for Multi modal Person Authentication", *In Pattern Recognition Letters,* Vol.18, No.9, pp. 835-843, 1997.

[**Egan75**] J. P. Egan, "Signal Detection Theory and ROC Analysis", Academic Press, 1975

[**EWA01**] Electronic warfare Associates, Biometric Technology Security Evaluation under the Common Criteria, September 2001.

[**Fairhurst97**] M. Fairhurst and A. Rahman, " A Generalised Approach to the Recognition of Structurally Similar Handwritten Characters", *IEE Vision, Image and Signal processing*, Vol.144, No.1, pp.15-22, 1997.

[**Farmer86**] J. D. Farmer, N. H. Packard and A. S. Perelson, " The Immune System, Adaptation and Machine Learning", *Physica D*, Vol.22, pp 187-204, 1986.

[**FingerScan**] http://www.finger-scan.com

[**Fitzegerald89**] K. Fitzgerald," The Quest for Intruder-Proof Computer Systems", *IEEE Spectrum*, Vol.26, No.8, pp.22-26, 1989.

[**Fogel90**] D. B. Fogel and J. W. Atmar, " Comparing Genetic Operators with Gaussian Mutations in Simulated Evolutionary Processes using Linear Search", *Biological Cybernetics*, Vol.63, pp 111-114, 1990.

[**Fontanari90**] J. F. Fontanari and R. Meir, " The Effect of Learning on the Evolution of Asexual Populations ", *Complex Systems*, Vol.4, pp. 401-414,1990.

[**Furui97**] S. Furui, "Recent Advances in Speaker Recognition", *Pattern Recognition Letters*, Vol. 18, No. 9, pp. 859-872, 1997.

[**GA Toolbox**] http://www.shef.ac.uk/~gaipp/ga-toolbox

[**Garcia-Salicetti03**] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L. les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacretaz, " BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities", *Lecture Notes in Computer Science*. Springer-Verlag Heidelberg, August, vol. 2688, pp. 845-853, 2003

[**George96**] T. George, "Biometric Encryption- New Developments in Biometrics", *The 18th International Privacy and Data Conference*, September 19, 1996, http://infoweb.magi.com/~privcan/conf96/se_tomko.html.

[**Goldberg89**] D. E. Goldberg, " Genetic Algorithms on Search, Optimisation and Machine Learning", *Addison-Wesley Pub Co*, 1989.

[**Goldberg91**] D.E. Goldberg and K. Deb, " A Comparative Analysis of Selection Schemes Used in Genetic Algorithms", In G. J. E. Rawlins (ed.), *Foundations of Genetic Algorithms*, pp 69-93. San Mateo, CA: Morgan Kaufmann, 1991.

[**Golfarelli97**] M. Golfarelli and D. Maio, " On the Error-Reject Trade-Off in Biometric Verification Systems", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, 1997.

[**Hall01**] D.L. Hall, J. Llinas, "Multisensor Data Fusion", in: D.L. Hall, J. Llinas (Eds.), Handbook of Multisensor Data Fusion, *CRC Press*, pp. 1–10, 2001,

[**Hanley89**] J.A.Hanley, "Receiver Operating Characteristics (ROC) Methodology: The State of the Art", *Critical Reviews in Diagnostic Imaging*, Vol.29, pp. 307-335, 1989.

[**Hill78**] R. B. Hill, Apparatus and Method for Identifying individuals Through their Retinal Vasculature Patterns, US Patent No. 4109237, 1978

[**Hillis90**] W. D. Hillis, " Co-evolving Parasites Improve Simulated Evolution as an Optimisation Procedure", *Physica D*, Vol.42, pp.228-234, 1990.

[**Holland75**] H. J. Holland, " Adaptation in Neural and Artificial Systems", University of Michigan Press, Ann Arbor, MI, 1975.

[**Holland86**] J. H. Holland, K. J. Holyoak, R .E. Nisbett and P. Thagard, " Induction: Processes of Inference, Learning and Discovery", *MIT Press*, 1986.

[**Holland91**] J. H. Holland and J. H. Miller, " Artificial Adaptive Agents in Economic Theory", Technical Report 91-05-025, Santa Fe Institute, Santa Fe, New Mexico, 1991.

[**Hong88**] L. Hong, Y. Wan and A.K. Jain, "Fingerprint Image Enhancement: Algorithms and Performance Evaluation", *IEEE Transactions on PAMI* ,Vol. 20, No. 8, pp.777-789, August 1998.

[**Hong98**] L. Hong and A. Jain, " Integrating Faces and Fingerprints for Personal Identification", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.16, No.12, pp.1295-1306, 1998.

**[Hong99]** L. Hong, A. Jain and S. Pankanti, " Can Multibiometrics Improve Performance?" *Proceedings AutoID'99*, Summit, NJ, pp. 59-64, Oct 1999.

[**Hsu02**] R.-L. Hsu and A. K. Jain, "Semantic Face Matching"', *Proc. IEEE Int'l Conf. Multimedia and Expo (ICME)* , Lausanne, Switzerland, Aug. 2002.

[**Huang97**] K. Huang and H. Yan, " Off-line Signature Verification Based on Geometric Feature Extraction and Neural Network Classification", *Pattern Recognition*, Vol.30, No.1, pp. 9-17, 1997.

[**IBG99**] International Biometric Group, "Overview of Biometrics-Face Geometry" http://www.biometricgroup.com

[**Indovina03**] M. Indovina, U. Uludag, R. Snelick, A. Mink and A. Jain, "Multimodal Biometric Authentication Methods: A COTS Approach", *Proc. MMUA 2003, Workshop on Multimodal User Authentication*, pp. 99-106, Santa Barbara, CA, December 11-12, 2003.

[**Iris99**] "Bank will ID Customers by Pattern of Eye's Iris ", May 13, 1999 http://www.mercurycenter.com/resources/search/

[**Jain97**] A. K. Jain, L. Hong, S. Pankanti and R.Bolle, "An Identity Authentication System using Fingerprints". *In Proceedings of the IEEE,* Vol 85, No 9, pp. 1365-1388, 1997.

[**Jain99**] A. K. Jain, R. Bolle and S. Pankanti (eds), " Biometrics: Personal Identification in Network Society", *Kluwer Academic Publisher*, 1999.

[**Jain99a**] A. K. Jain, L.Hong, Y. Kulkarni, "A Multimodal Biometric System using Fingerprints, Face and Speech", *2nd Int'l Conference on Audio- and Video-based Biometric Person Authentication*, Washington D.C., pp. 182-187, March 22-24, 1999.

[**Jain99b**] A.K.Jain, A.Ross and S.Pankanti, "A Prototype Hand Geometry-based Verification System" In Proc. of 2nd Int'l Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), Washington, pp.166-171, 1999.

[**Jain00**] Anil Jain, Lin Hong and Sharath Pankanti, "Biometric Identification", *Communications of the ACM*, Vol.43, Issue 2, pp. 90-98, February 2000.

[**Jain01**] A. K. Jain and S. Pankanti, "Biometrics Systems: Anatomy of Performance", *IEICE Trans. Fundamentals*, Vol. E84-D, No. 7, pp. 788-799, 2001.

[**Jain01**] A. K. Jain, S. Pankanti, S. Prabhakar, and A. Ross, "Recent Advances in Fingerprint Verification", *Invited Paper for 3rd International Conference on Audio- and Video-Based Person Authentication (AVBPA),*, pp. 182-191, Sweden, June 6-8, 2001.

[**Jain02**] A. K. Jain and A. Ross, "Learning User-specific Parameters in a Multibiometric System", *Proc. of IEEE International Conference on Image Processing (ICIP)*, (Rochester, NY), pp. 57-60, September 2002.

[**Jain04a**] A. K. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, Vol. 14, No. 1, pp. 4-20, January 2004.

[**Jain04b**] A. K. Jain and A. Ross, "Multibiometric Systems", *Communications of the ACM, Special Issue on Multimodal Interfaces*, Vol. 47, No. 1, pp. 34-40, January 2004.

[**Jain04c**] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman, "Biometrics: A Grand Challenge", *Proc. International Conference on Pattern Recognition (ICPR)* , vol. II, pp. 935-942, Cambridge, UK, Aug. 2004.

[**Janikow91**] C. Z. Janikow and Z. Michalewicz, "An Experimental Comparison of Binary and Floating Point Representations in Genetic Algorithms", *Proc. ICGA 4*, pp.31-36, 1991.

[**Jankowski90**] C.Jankowski, A. Kalyanswamy, S.Basson and J. Spitz, " NTIMIT: A Phonetically Balanced, Continuous Speech Telephone Bandwidth Speech Database", *Proc. International Conf. Acoustics, Speech and Signal Processing*, Vol 1,pp.109-112, 1990

[**Jourlin97**] P. Jourlin, J. Luettin, D. Genoud and H. Wassner, " Acoustic-labial Speaker Verification", *Pattern Recognition Letters*, Vol. 18, No. 9, pp. 853-858, 1997.

[**Kamel03**] Mohamed S. Kamel and Nayer M. Wanas, " Data Dependence in Combining Classifiers", *Proc. 4th Int. Workshop on Multiple Classifier Systems (MCS 2003)*, Guilford, U.K, June 2003, F. Roli and Terry Windeatt Eds., LNCS 2709, pp. 1-14.

[**Kholmatov03**] A. A. Kholmatov, "Biometric Identity Verification using On-Line & Off-Line Signature Verification", MSc thesis, Sabanci university, 2003.

[**Kittler98**] J. Kittler, M. Hatef, R. Duin and J. Matas, " On Combining Classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20, pp. 226–239,1998.

[**Koza93**] J. R. Koza, " Genetic Programming: On the Programming of Computers by Means of Natural Selection", *MIT Press*, Cambridge, MA, 1993.

[**Kumar03**] A. Kumar, D. C. M. Wong, H. C. Shen and A. K. Jain, "Personal Verification Using Palmprint and Hand Geometry Biometric", *In Proc of 4th Int'l Conf. On Audio and Video-Based Biometric Person Authentication*, pp.668-678, June 2003

[**Kuncheva93**] L .I. Kuncheva, "Genetic Algorithm for Feature Selection for Parallel Classifiers", *Information Processing Letters*, Vol.46, pp.163-168, 1993

[**Kuncheva02**] Ludmila I. Kuncheva, " A Theoretical Study on Six Classifier Fusion Strategies", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.2, pp. 281-286, Feb 2002.

[**Lindgren93**] K. Lindgren and M. G. Nordhal, " Artificial food webs", In C. G. Langton, editor, *Artificial Life III*, 1993.

[**Liu98**] C. Liu and H. Wechsler, " Evolution of Optimal Projection Axes (OPA) for Face Recognition", *3rd. Int.conf.on Automatic Face and Gesture Recognition*, Nara, Japan, 1998.

[**Lu04**] X. Lu, D. Colbry and A. K. Jain, "Three-Dimensional Model Based Face Recognition", *Proc. International Conference on Pattern Recognition (ICPR)*, Vol.I, pp. 362-366, Cambridge, UK, August 2004.

[**Lucasius92**] C. B. Lucasius and G. Kateman, "Towards Solving Subset Selection Problems with the Aid of the Genetic Algorithm", In *Parallel Problem Solving from Nature 2*, R. Männer and B. Manderick, (Eds.), pp. 239-247, Amsterdam: North-Holland, 1992.

[**Luettin97**] J. Luettin, Visual Speech and Speaker Recognition, PhD Thesis, Department of Computer Science, University of Sheffield, 1997.

[**M2VTS**] http://www.tele.ucl.ac.be/PROJECTS/M2VTS/

[**Mai03**]. L. Mai, T. Tan, Y. Wang, and D. Zhang, "Personel Identification Based on Iris Texture Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol 25,No.12, pp.1519-1533, 2003.

[**Maltoni03**] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition, Springer Verlag, June 2003.

[**Mansfield01a**] T. Mansfield, G. Kelly, D. Chandler and J. Kane, Biometric Product Testing Final Report, March 2001.
http://www.cesg.gov.uk/site/ast/biometrics/media/BiometricTestReportpt1.pdf

[**Mansfield01b**] T. Mansfield, Detection Error Trade-Off and confidence Interval Analysis Report, December 2001.

[**Mansfield02**] A. J. Mansfield and J. L. Wayman, Best Practices in Testing and Reporting Performance of Biometric Devices, version 2, August 2002. http://www.cesg.gov.uk/site/ast/biometrics/media/BestPractice.pdf

[**Marcialis02**] G.L. Marcialis and F. Roli, "Fusion of LDA and PCA for Face Recognition", *Proceedings of the Workshop on Machine Vision and Perception*, 8 th Workshop of the Italian Association for Artificial Intelligence (AIIA'02), available at the URL: http://wwwdii. ing.unisi.it/aiia2002.

[**Marcel02**] S. Marcel and S. Bengio, "Improving Face Verification using Skin Color Information", *Proceedings of the 16th International Conference on Pattern Recognition*, 2002.

[**Miros99**] Miros, " Evolving Approaches to Recognizing a Friendly Face", Retrieved September 11, 1999 from the World Wide Web: http://www.miros.com/Neural_networks_description.htm.

[**Martin97**] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET Curve in Assessment of Detection Task Performance," *In Proceedings of Eurospeech' 97, Rhodes, Greece*, pp. 1895-1898, 1997.

[**Martin00**] A. Martin, M. Przybocki, "The NIST 1999 Speaker Recognition Evaluation- An Overview", *Digital Signal Processing*, Vol.10, pp.1-18, 2000.

[**MCYT**] MCYT Consortium site, http://www.infor.uva.es/biometria

[**Messer99**] K Messer, J Matas, J Kittler, J Luettin and G Maitre, "XM2VTS: The Extended M2VTS Database", *In Proceedings of the 2nd Conference on Audio and Video-based Biometric Person Authentication AVBPA'99*, Springer Verlag, New York, http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb, 1999.

[**Michalewicz92**] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer Verlag, 1992.

[**Miller94**] B. Miller, " Vital Signs of Identity", *IEEE Spectrum*, Vol.31, No.2, pp.22-30, 1994.

[**Mitchell93**] M.Mitchell, J.P.Crutchfield and P.T.Hraber, " Evolving Cellular Automata to Perform Computations: Mechanisms and Impediments", *Physica D*, 1993.

[**Mitchell96**] Mitchell, M.: An Introduction to Genetic Algorithms. Cambridge, Massachusetts: MIT Press, 1996.

[**Moenssens71**] A. Moenssens, Fingerprint Techniques. Chilton Book Company, London, 1971

[**Mühlenbein93**] H. Mühlenbein and D. Schlierkamp-Voosen, "Predictive Models for the Breeder Genetic Algorithm", *Evolutionary Computation*, Vol.1, No.1, pp.25-49, 1993.

[**Mühlenbein95**] H. Mühlenbein and D. Schlierkamp-Voosen, " Analysis of Selection, Mutation and Recombination in Genetic Algorithms". In Banzhaf, W. and Eeckman, F. H.: Evolution as a Computational Process. Lecture Notes in Computer Science 899, pp. 142-168, Berlin: Springer-Verlag, 1995.
ftp://borneo.gmd.de/pub/as/ga/gmd_as_ga-95_03.ps

[**MCYT02**] J. Ortega, D. Simon, M. Faundez, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, and Q.-I. Moro, " MCYT: A Multimodal Biometric Database". *In Proc. COST 275 Workshop on the Advent of Biometrics on the Internet*, pages 123-126, Rome, 2002.

[**Naik89**] J. M. Naik, L. P. Netsch, and G. R. Doddington, " Speaker Verification over Long Distance Telephone Lines", In *Proceedings of the 1989 International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, Scotland, pp 524-527, May 1989.

[**Negin00**] M. Negin and T. Camus, " An iris Biometric System for Public and Personal Use", *IEEE computer*, special issue, February 2000.

[**Newham95**] E. Newham, " The Biometric Report", SBJ Services, 1995.

[**NIST00**] NIST report to the United States Congress, "Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability", November 13, 2000.
http://www.itl.nist.gov/iad/894.03/NISTAPP_Nov02.pdf

[**Ortega-Garcia02**] Ortega, J., Simon, D., Faundez, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., and Moro, Q.-I., "MCYT: A Multimodal Biometric Database", *Proc. COST-275 Biometric Recognition Workshop*, Rome, Italy, Nov. 2002, pp. 123–126.

http://www.fub.it/cost275

[**Ortega-Garcia03**] Ortega, J., Simon, D., Faundez, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., and Moro, Q.-I., "MCYT Baseline Corpus: A Bimodal Biometric Database", *IEE Proc.Image Signal Process*, Vol.150, No.6, pp. 395-401, 2003.

[**O'Gorman98**] L. O'Gorman, " Fingerprint Verification", In A. Jain, etal (eds), Biometrics: Personal Identification in a Network Society, (Kluwer Academic Press, 1998).

[**Pankanti01**] S. Pankanti, S Prabhakary A K. Jain, "On the Individuality of Fingerprints", *In the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 805-812, Hawaii, December 11-13, 2001.

[**Penny02**] W. Penny, " Biometrics: A Double Edged sword- Security and Privacy", *SANS Institute*, 2002
http://www.sans.org/rr/papers/6/137.pdf

[**Phillips00**] P. J. Phillips, A. Martin, C. Wilson, M. Przybocki, " Introduction to Evaluating Biometric Systems", *IEEE Computer Magazine,* pp. 56-63, January 2000.

[**Pierrot98**] J. B. Pierrot, J. Lindberg, J. Koolwaaij, H. P. Hutter, D. Genoud, M. Blomberg, and F. Bimbot, " A Comparison of a Priori Thresholds Settings Procedures for Speaker Verification in the CAVE project", *In proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, Vol. I, pp.125-128, 1998.

[**Pigeon98**] S. Pigeon and L. Vandendorpe, " Multiple Experts for Robust Face Authentication", *SPIE, editor, Optical Security and Counterfeit Deterrence II*, 3314, pp.166-177, 1998.

[**Plamondon89**] R. Plamondon and G. Lorette, " Automatic Signature Verification and Writer Identification-The State of the Art", *Pattern Recognition*, Vol.22, pp.119-128, 1989.

[**Ponti99**] R. Ponti, " Facial Features Identification". Retrieved June 3, 1999 from the World Wide Web:
http://www.tech.purdue.edu/it/resources/aidc/BioWebPages/Biometrics_Face.html.

[**Prabhakar01**] S. Prabhakar, A.K.Jain, "Decision-level Fusion in Biometric Verification", *In Proceedings of the 2nd International Workshop on Multiple Classifier Systems (MCS),* pp. 345-359, 2001

[**Prabhakar03**] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric Recognition: Security & Privacy Concerns", *IEEE Security & Privacy Magazine*, Vol. 1, No. 2, pp. 33-42, March-April 2003.

[**Prabhakar03a**] S. Prabhakar and A. K. Jain, "Fingerprint Matching", In *Automatic Fingerprint Recognition Systems*, N. Ratha and R. Bolle (eds.), Springer Verlag, New York, 2003.

[**Prabhakar03b**] S. Prabhakar, A. K. Jain, and S. Pankanti, "Learning Fingerprint Minutiae Location and Type", *Pattern Recognition*, Vol. 36, No. 8, pp. 1847-1857, 2003.

[**Press99**] "Bank of America Offers Fingerprint Access to Online Banking", January 6, 1999, Press Release,

http://www.interrnetnews.com/ec-news/1999/01/0601-bank.html

[**Prins98**] C.Prins, " Biometric Technology Law: Making our Body Identify for us: Legal Implications of Biometric Technology", *Computer Law and Security Report*, Vol. 14, No.3, p.160, 1998

[**Rahman99**] A. Rahman and M. Fairhurst, " Enhancing Multiple Expert Decision Combination Strategies Through Exploitation of a Priori Information Sources", *IEE Vision, Image and Signal Processing*, Vol.146, No.1, pp.1-10, 1999.

[**Rawlins91**] G. J. E Rawlins, "Foundations of Genetic Algorithms". San Mateo, California, USA: Morgan Kaufmann Publishers, 1991.

[**Rhodes56**] H. T. F. Rhodes, "Alphonse Bertillon: Father of Scientific Detection", Abelard-Schuman, New York, 1956

[**Roli02**] F. Roli, J. Kittler, G. Fumera and D. Muntoni, " An Experimental Comparison of Classifier Fusion Rules for Multimodal Personal Identity Verification Systems", *Proc. 3rd Int. Workshop on Multiple Classifier Systems (MCS 2002)*, Cagliari, Italy, June 2002,F. Roli and J. Kittler Eds., LNCS 2364, pp. 325-335.

[**Roli02**] F. Roli, Tutorial, " Fusion of Multiple Classifiers", University of Cagliari, 2002

[**Ross94**] P. E. Ross, " I can read your Face", *Forbes*, pp.304-305, 1994

[**Ross01**] A.Ross, A.K. Jain and J. Qian, " Information Fusion in Biometrics", *Proc. of 3rd Int. Conf. On Audio- and Video-Based Biometric Person Authentication,* pp. 354-359, 2001.

[**Ross03**] A. Ross, A. K. Jain, and J. Reisman, "A Hybrid Fingerprint Matcher", *Pattern Recognition*, Vol. 36, No. 7, pp. 1661-1673, 2003.

[**Rutender89**] R. A. Rutender, "Simulated Annealing Algorithms: An Overview", *IEEE Circuits and Devices Magazine*, pp. 19-26, 1989.

[**SAFLINK99**] SAFLINK corporation, " SAFLINK develops Way to Secure Internet Banking/ Brokerage Account Balances, Bill Payment and Funds Transfer Using Biometrics", Press Release, June 24, 1999.
http://biz.yahoo.com/prnews/990624/fl_saflink_1.html

[**Sanderson02**] C. Sanderson, K.K.Paliwal, "Information Fusion and Person Verification using Speech and Face Information", Research paper *IDIAP-RR 02-33, IDIAP*, 2002.

[**Sanderson03**] C.Sanderson and K.K.Paliwal, "Noise Resistance Audio-Visual Verification Via Structural Constaints, *Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing*, Vol. V, pp. 716-719, Hong Kong, April 2003.

[**Sanderson04**] C. Sanderson and K. K. Paliwal, "Identity Verification using Speech and Face Information" *Digital Signal Processing, Vol* 14, pp.449-480, 2004

[**Schneier99**] B. Schneier, "Inside Risks: The Uses and Abuses of Biometrics," *Comm. ACM,* vol. 42, no. 8, p. 136. Aug. 1999,

**[SecuGen]** http://www.secugen.com

[**Shen97**] W. Shen, M. Surette and R. Khanna, "Evaluation of Automated Biometrics-Based Identification and Verification Systems", *Proceedings of IEEE*, Vol.85, No.9, pp.1465-1478, 1997

[**Sidlauskas88**] D.R. Sidlauskas, "3D Hand Profile Identification Apparatus", US Patent No.4736203, 1988.

[**Silsbee96**] P. Silsbee, A. Bovik, "Computer Lipreading for Improved Accuracy in Automatic Speech Recognition", *IEEE Trans. Speech Audio Process*, Vol 4, pp.337–351, 1996.

[**Snedecor67**] W. G Snedecor and W. G. Cochran, Statistical Methods, 6th edition. Iowa State University Press, 1967.

[**Snelick03**] R. Snelick, M. Indovina, J. Yen, and A. Mink, "Multimodal Biometrics: Issues in Design and Testing", *In Proceedings of Fifth International Conference on Multimodal Interfaces*, (Vancouver, Canada), November 2003.

[**Spears91**] W. M. Spears and K. A. De Jong, "An Analysis of Multi-Point Crossover", In *Foundations of Genetic Algorithms*, J. E. Rawlins (Ed.), pp. 301-315, 1991.

[**Spears97**] W M. Spears. Recombination Parameters. In *The Handbook of Evolutionary Computation*, T. Baeck, D. Fogel and Z. Michalewicz (editors), 1997, IOP Publishing and Oxford University Press.

[**Spears98**] W M. Spears, The Role of Mutation and Recombination Evolutionary Algorithms Ph.D. thesis, George Mason University, Fairfax, VA, 1998.

[**Swets73**] J. A Swets," The Relative Operating Characteristic in Psychology", *Science*, Vol. 182, pp.990-1000, 1973.

[**Syswerda98**] G. Syswerda, "Uniform Crossover in Genetic Algorithm", *Proc ICGA3*, pp 2-9, 1989.

[**Taylor89**] C. E. Taylor, D. R. Jefferson, S. R. Turner and S. R. Goldman, " RAM: Artificial Life for the Exploration of Complex Biological Systems", In C. G. Langton, editor, *Artificial Life*, pp 275-295,1989.

[**Teoh04**] Andrew Teoh, S. A. Samad, A. Hussain, " Nearest Neighbourhood Classifiers in a Bimodal Biometric Verification System Fusion Decision Scheme", *Journal of research and practice in Information Technology*, Vol.36, No.1, pp.97-112, February 2004.

[**Turk91**] M. Turk, and A. Pentland, " Eigenfaces for Recognition", *Journal of Cognitive Neuroscience,* Vol. 3, pp.71-86, 1991

[**Uludag04**] U. Uludag, A. Ross and A. K. Jain, "Biometric Template Selection and Update: A Case Study in Fingerprints", *Pattern Recognition*, Vol. 37, No. 7, pp. 1533-1542, July 2004.

[**Van Trees68**] H. L. Van Trees, " Detection, Estimation and Modulation Theory", Vol.1 John Wiley & Sons, New York, 1968.

[**Verivoice**] http://www.verivoice.com

[**Visionics**] http://www.visionics.com

[**Visionics99**] Visionics, " The Underlying Algorithm: Local Feature Analysis", Retrieved October 11, 1999 from the World Wide Web: http://www.visionics.com/Faceit/What/LFA.htm.

[**Wang03**]Y. Wang, T. Tan and A. K. Jain, "Combining Face and Iris Biometrics for Identity Verification", *In Proceedings of Fourth International Conference on AVBPA,* (Guildford, U. K.), pp. 805-813, June 2003.

[**Wark00**] T. Wark, Multi-modal Speech Processing for Automatic Speaker Recognition, PhD thesis, School of Electrical & Electronic Systems Engineering, Queensland University of Technology, Brisbane, 2000.

[**Wayman98**] J. L. Wayman, " Technical Testing and Evaluation of Biometric Identification Devices" *in A. Jain, etal (eds), Biometrics: Personal Identification in a Networked Society,* (Kluwer Academic Press, 1998), pp. 345-367

[**Wayman99**] J. L. Wayman, " Fundamentals of Biometric Authentication Technologies", *Proc CTST'99*, Chicago, May 1999, pp. 390-410.

[**Wayman99a**] J. L. Wayman, " Confidence Interval and Test Size Estimation for Biometric Data", National Biometric Test Centre Collected Works 1997-2000, San Jose State University, 2000, pp. 91-95.

[**Wayman99b**] J.L. Wayman, "Technical Testing and Evaluation of Biometric Identification Devices" in A. Jain, etal (eds), Biometrics: Personal Identification in a Networked Society, (Boston, Kluwer Academic Press, 1999)

[**Wayman00**] J. L. Wayman, " Picking the Best Biometric for your Application", *Proc.CTST00*, May2000

[**Werner92**] G. M. Werner and M. G. Dyer, " Evolution of Communication in Artificial Organisms", In C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen, editors, *Artificial Life II*, pp 659-687. 1992.

[**Whitley89**] D. Whitley, "The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best", *Proc. ICGA 3*, pp. 116-121,Morgan Kaufmann Publishers, 1989.

[**Whitley93**] L. D Whitley, " Foundations of Genetic Algorithms 2", San Mateo, California, USA: Morgan Kaufmann Publishers, 1993.

[**Whitley95**] L. D Whitley and M. D Vose, " Foundations of Genetic Algorithms 3", San Francisco, California, USA: Morgan Kaufmann Publishers, 1995.

[**Wildes97**] R. P. Wildes, " Iris Recognition: An Emerging Biometric Technology", in *Proceedings of the IEEE*, vol. 85, No. 9 1997.

[**Wright91**] A. H. Wright, "Genetic Algorithms for Real Parameter Optimization", In *Foundations of Genetic Algorithms*, J. E. Rawlins (Ed.), Morgan Kaufmann, pp.205-218, 1991.

[**Xu92**] L. Xu, A. Krzyakk and C. Y. Suen, " Several Methods for Combining Multiple Classifiers and their Applications in Handwritten Character Recognition", *IEEE Transactions on System, Man and Cybernetics*, Vol.22, No.3, pp.418-435, 1992.

[**Zhu02**] Y. Zhu, T. Tan, and Y. Wang, "Biometric Personal Identification Based on Iris Pattern", *ICPR2000: the 15th International Conference on Pattern Recognition*, Barcelona, Spain, pp.805-808, 2002.

# Appendix A

# Data Collection Information

## A.1 Introduction

This appendix presents the data sheets used during the data collection exercise. As previously mentioned in Chapter 3 each volunteer took part in two separate data collection sessions, the first involving enrolment on each of the biometric devices together with a post-enrolment verification check of three attempts. In the second session three additional verification attempts were carried out using the enrolment templates generated at the first session. In this appendix the data sheets that were used in both sessions are provided. The data sheets include a brief personal information of each volunteer, a written consent form for participation, some comments that were written by the supervisor during the data collection and the results of enrolment and verification in each session.

# A Personal Information Data Sheet

**Name: .....**

**Occupation: .....**

**Phone: .....** **Consent Form** ☐

**E-mail: .....**

---

**Trial Details**

**ID/PIN Number:.....**

**Gender:**

☐ **Male** ☐ **Female**

**Age:**

☐ **18-24** ☐ **25-34** ☐ **35-44**

☐ **45-54** ☐ **55-64** ☐ **65+**

**Features:**

☐ **Glasses** ☐ **Contact Lenses** ☐ **Beard**

☐ **Moustache** ☐ **Other feature**

## B Consent Form

**Department of Electronics**
**University of Kent**

CONSENT FORM

Please tick

1. I confirm that I have read and understood the Information for Project Volunteers for the above project ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reasons. ☐

3. I consent to my biometric data being collected during the trial and stored electronically. ☐

4. I understand that this data will only be used for the purposes of evaluating performance of biometric techniques and devices, by the Department of Electronics at the University of Kent. ☐

5. I agree to take part in the project as a volunteer. ☐

6. I understand that the advertised honorarium will only be paid on completion of the two measurment sessions. The payment will be made by cheque. ☐

| Name of Volunteer | Date | Signature |
|---|---|---|

PLEASE BRING YOUR COMPLETED CONSENT FORM WITH YOU ON YOUR FIRST VISIT.

C Data Sheet for Enrolment and Verification in the First Session

PIN .....

**NAME: .....**

**ENROLMENT** **VERIFICATION**

| **SYSTEM** | **Failure to match** | **Failure to Acquire** | **System rejection** | **OK** | **Attempts** | | |
|---|---|---|---|---|---|---|---|
| | | | | | **1** | **2** | **3** |
| **FINGERPRINT** | | | | | | | |
| **VOICE** | | | | | | | |
| **FACE** | | | | | | | |

## D The Supervisor Observations on the First Session

**Comments on**

**Problems with Fingerprint:**

**Enrolment:**

**Verification:**

**Problems with Voice:**

**Enrolment:**

**Verification:**

**Problems with Face:**

**Enrolment:**

**Verification:**

E Data Sheet for Verification in the Second Session

**Date : .....**

**PIN : .....**

**Name: .....**

| **SYSTEM** | **Verification Attempts** | | | Comments |
|---|---|---|---|---|
| | 1 | 2 | 3 | |
| **FINGERPRINT** | | | | |
| **VOICE** | | | | |
| **FACE** | | | | |

# Appendix B

# Database Entities

## B.1 Introduction

This appendix presents a sample of the supplementary data collected at the time of the data collection exercise that was carried out to obtain biometric samples and scores. The preliminary analysis provided in Chapter 4 is based on this database. In this appendix a sample of the database is provided with a description of each parameter.

## B.2 Description of the Database

In this section a description of the different parameters of the database is provided. The sample of the database provided consists of 10 subjects as shown in Table B.1.

PIN: represents the personal identification number given to each subject.

Gender: refers to the gender of the subject.

Age: refers to the age range of the subject.

Date of 1$^{st}$ visit: refers to the date of the first session attended by the subject.

Finger (FTE): refers to the number of attempts the subject failed to enrol in the fingerprint modality (max of 3 attempts allowed).

| PIN | Gender | Age | Date of 1st visit | Finger (FTE) | Voice (FTE) | Face (FTE) | Finger (FTV) | Voice (FTV) | Face (FTV) | Date of 2nd visit | Finger (FTV2) | Voice (FTV2) | Face (FTV2) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1082 | Male | 55-64 | 26/11/2001 | 2 attempts | 3 attempts | | 2 attempts | | 1 attempt | 24/01/02 | 2 attempts | not enrolled | 3 attempts |
| 1236 | Female | 55-64 | | | | | 2 attempts | | | 24/01/02 | 3 attempts | | |
| 1147 | Male | 45-54 | | | | | | | | 23/01/02 | | | |
| 1481 | Male | 25-34 | | 1 attempt | | | | | | 8/3/2002 | 1 attempt | | 3 attempts |
| 1503 | Female | 35-44 | | | | | | | 2 attempts | 26/02/02 | | | |
| 1597 | Male | 18-24 | | | | | | | | 4/3/2002 | | | |
| 1090 | Female | 65+ | | | 1 attempt | | | | 3 attempts | 24/01/02 | 1 attempt | | 2 attempts |
| 1600 | Male | 55-64 | | | | | | | | 24/01/02 | | | 2 attempts |
| 1902 | Female | 35-44 | | | 2 attempts | | | | | 23/01/02 | 1 attempt | | |
| 1112 | Male | 65+ | 27/11/2001 | | 1 attempt | 1 attempt | 3 attempts | | 1 attempt | 15/01/02 | 2 attempt | 1 attempt | 2 attempts |

**Table B.1:** Table of supplementary data used in generating the preliminary analysis provided in Chapter 4

Voice (FTE): refers to the number of attempts the subject failed to enrol in the voice modality (max of 3 attempts allowed).
Face (FTE): refers to the number of attempts the subject failed to enrol in the face modality (max of 3 attempts allowed).

"1 attempt" in (FTE): means that the subject had problems enrolling in the specified modality in the first attempt but was successfully enrolled in the second attempt.

"2 attempts" in (FTE): means that the subject had problems enrolling in the specified modality in both the first and the second attempt but was successfully enrolled in the second attempt.

"3 attempts" in (FTE): means that the subject had problems enrolling in the specified modality in all three attempts.

"blank space" in the FTE means that the subject had no problems enrolling in the specified modality.

Finger (FTV): refers to the number of attempts the subject failed to verify in the fingerprint modality.

Voice (FTV): refers to the number of attempts the subject failed to verify in the voice modality.
Face (FTV): refers to the number of attempts the subject failed to verify in the face modality.

"1 attempt" in (FTV): means that the subject had problems verifying in the specified modality in the first attempt but was successfully verified in the second attempt.

"2 attempts" in (FTE): means that the subject had problems verifying in the specified modality in both the first and the second attempt but was successfully verified in the second attempt.

"3 attempts" in (FTE): means that the subject had problems verifying in the specified modality in all three attempts.

"blank space" in the FTV means that the subject had no problems verifying in the specified modality.

Date of 2nd visit: refers to the date of the second session attended by the subject.

"1 attempt" in (FTV2): means that the subject in the second session had problems verifying in the specified modality in the first attempt but was successfully verified in the second attempt.

"2 attempts" in (FTV2): means that the subject in the second session had problems verifying in the specified modality in both the first and the second attempt but was successfully verified in the second attempt.

"3 attempts" in (FTV2): means that the subject in the second session had problems verifying in the specified modality in all three attempts.

"blank space2 in the FTV2 means that the subject in the second session had no problems verifying in the specified modality.

"not enrolled" means that the subject was not enrolled in the first session and as a result no enrolment template was generated that could be used for verification in the second session.

# Appendix C

# Confidence Interval Estimation for Biometric Data

## C.1 Introduction

This appendix presents the method adopted for calculating the confidence intervals of the error rates determined in this thesis.

## C.2 Estimation of the Uncertainty in Measured Error Rates

The variance is a statistical measure of uncertainty and it is used in estimating the confidence intervals. In this section the formulas and methods for estimating the variance of performance measure is provided as suggested by [BWG02].

The estimation of the variances in the measured error rates followed the equation given by Bickel in [Bickel98] since it is used in the cases where a cross comparison was used to establish the impostor distribution and where the error rate vary across the population, that is, when different subjects have different individual false reject rates and different subject pairs have different individual false accept rates.

### C.2.1 Variance Estimation of False Reject Rate

The formulas presented in this section were used to estimate the variance of the false reject rate as well as the failure to enrol rate. The variance was estimated using the following equation, the derivation of this estimate can be found in [Snedecor67].

$$\sigma = \frac{p(1-p)}{n-1} \tag{C.1}$$

where

$$p = \frac{1}{n}\sum a_i \tag{C.2}$$

$n$ : Number of enrolled volunteers.

$a_i$ : Count of false reject for $i^{th}$ volunteer.

$p$: Observed false reject rate.

$\sigma$: Estimated variance of observed false reject rate.

### C.2.2 Variance Estimation of False Accept Rate

The formulas presented in this section were used to estimate the variance of the false accept rate. The variance was estimated using the following formula as given by Bickel in [Wayman99a] [Wayman99b].

$$\sigma^2 = \frac{1}{n^2(n-1)^2}\left[\sum(c_j + d_i)^2 - \frac{4}{n}q^2\right] \tag{C.3}$$

where

$$q = \frac{b_{i,j}}{n(n-1)} \tag{C.4}$$

$n$ : Number of enrolled volunteers.

$b_{i,j}$: Count of false accept for $i^{th}$ volunteer against $j^{th}$ template.

$c_j$: Count of false accepts against $j^{th}$ template by any volunteer.

$d_i$: Count of false accepts by $i^{th}$ volunteer against any template

$q$: Observed false accept rate.

$\sigma^2$ :Estimated variance of observed false reject rate.

## C.3 Confidence Intervals Estimation for FRR and FAR

Confidence interval is commonly known as the margin of errors. The statistical term " confidence interval" is defined as the probability that the true parameter is within the interval that surrounds the estimate of the parameters FRR and FAR.

Under the assumption of normality, 100(1-α) % confidence bound on the observed error rates are given by

$$E \pm z(1-\frac{\alpha}{2})\sqrt{\sigma^2} \qquad \text{(C.5)}$$

where

$z(1-\frac{\alpha}{2})$ : indicates the number of standard deviations from the origin required to encompass $(1-\frac{\alpha}{2})$ % of the area under the standard normal distribution. For α=5 % that is for a 95 % confidence limits the value is 1.96.

E: represents either the false reject rate or the false accept rate depending on what is calculated.

$\sigma^2$ is either the estimated variance of the observed false reject rate or the estimated variance of the observed false accept rate depending on what is calculated.

Often when this formula is applied, the confidence interval reaches into negative values for the observed error rates. This is due to the non-normality of the distribution of the observed error rates.

## C.4 Confidence Intervals for Proportions

The formulas presented in this section were used to estimate the confidence intervals for the proportion of sheep, lambs, goats and wolves. The confidence interval is estimated using the following formula as suggested by Spiegel in [Spiegel90].

$$P \pm z(1-\frac{\alpha}{2})\sqrt{\frac{P(1-P)}{N}} \qquad \text{(C.6)}$$

where

$z(1-\frac{\alpha}{2})$ : indicates the number of standard deviations from the origin required to encompass $(1-\frac{\alpha}{2})$ % of the area under the standard normal distribution. For $\alpha$=5 % that is for a 95 % confidence limits the value is 1.96.

$N$ : Number of enrolled volunteers.

$P$: Observed proportion of sheep, lambs, goats or wolves.

# Appendix D

# Publications

A list of the publications, relating to some of the work presented in this thesis, is given below.

M. C. Fairhurst, J. George, F. Deravi, "Scenario based data collection trials for the evaluation of multi-modal biometric processing: a preliminary report", *Proceedings of Knowledge-Based Intelligent information Engineering Systems & Allied Technologies (KES 2002)*, pp1217-1221, Sept 3-5, 2002.

M. C. Fairhurst, R. M. Guest, F. Deravi, J. George, "Using Biometrics as an Enabling Technology in Balancing Universality and Selectivity for Management of Information Access", *Proceedings of Universal Access. Theoretical Perspectives, Practice, and Experience: 7th ERCIM International Workshop on User Interfaces for All,* LNCS 2615, pp 249-259, Paris, October 24-25, 2002.

M. C. Fairhurst, F. Deravi, N. J. Mavity, J. George, K. Sirlantzis, " Intelligent management of Multimodal Biometric Transactions", *Proceedings of Knowledge Based Intelligent information Engineering Systems (KES 2003),* LNAI 2774, pp 1255-1260, Oxford, Sept 3-5, 2003.

M. C. Fairhurst, F. Deravi, and J. George, " Towards Optimised Implementation of Multimodal Biometric Configurations", *CIHSPS 2004, IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, Italy, July 2004.