Face Recognition Using Skin Texture

by

Dale Christopher Rowe

Image and Information Engineering Group
School of Engineering and Digital Arts
University of Kent

FACE RECOGNITION USING SKIN TEXTURE

by

Dale Christopher Rowe

Thesis submitted to the University of Kent
for the degree of Doctor of Philosophy
in Electronic Engineering

Image and Information Engineering Group

School of Electronics and Digital Arts

University of Kent

July 2009

FACE RECOGNITION

USING SKIN TEXTURE

by

Dale C. Rowe

Thesis submitted to the University of Kent
for the degree of Doctor of Philosophy
in Electronic Engineering

Image and Information Engineering Group
School of Engineering and Digital Arts

Woolf College
July 2009

Abstract

In today's society where information technology is depended upon throughout homes, educational establishments and workplaces the challenge of identity management is ever growing. Advancements in image processing and biometric feature based identification have provided a means for computer software to accurately identify individuals from increasingly vast databases of users. In the quest to improve the performance of such systems in varying environmental conditions skin texture is here proposed as a biometric feature.

This thesis presents and discusses a hypothesis for the use of facial skin texture regions taken from 2-dimensional photographs to accurately identify individuals using three classifiers (neural network, support vector machine and linear discriminant). Gabor wavelet filters are primarily used for feature extraction and are supported in later chapters by the grey-level co-occurrence probability matrix (GLCP) to strengthen the system by providing supplementary high-frequency features. Various fusion techniques for combining these features are presented and their performance is compared including both score and feature fusion and various permutations of each.

Based on preliminary results from the BioSecure Multimodal Database (BMDB) , the work presented indicates that isolated texture regions of the human face taken from under the eye may provide sufficient information to discriminately identify an individual with an equal error rate (EER) of under 1% when operating in greyscale.

An analysis of the performance of the algorithm against image resolution investigates the systems performance when faced with lower resolution training images and discusses optimal resolutions for classifier training. The system also shows a good degree of robustness when the probe image resolution is reduced indicating that the algorithm provides some level of scale invariance. Scope for future work is laid out and a review of the evaluation is also presented.

Acknowledgments

I noticed while writing this thesis that the author has relatively little chance to express himself outside of the factual elements of the subject matter. This is of course intended; to do otherwise would risk presenting an autobiography or novel rather than an academic document. However the authors of such works are permitted the chance to acknowledge and express thanks for the support of peers, friends and family.

As such I would like to express my gratitude firstly to the departmental staff for their ongoing patience with a student's struggles to meet his commitments to his faculty and his family. In retrospect there are times when the balance has been unfairly biased in both directions at which times the administrators, supervisors and staff have shown great understanding and consideration. In particular I would like to thank my supervisor, Dr Farzin Deravi for his support as well as Relly Bowman, Chris Barron and Denyse Menne for their repeated willingness to help in any way they could.

I would also like to thank my colleagues, associates and friends both at the University and at work. In particular, my associates and managers with whom I have been employed during my studies; for their constant words of motivation and confidence - not to mention the many weeks off they have permitted me to take this last year in order to submit. I would like to draw particular attention and thanks to Brendan Hallisey, Kevin Rees, Mike Rees and Ian Schofield for their support and patience at these times.

Heartfelt thanks go also to my parents and sisters who have constantly expressed their belief that I could and would submit, especially to my father for his encouragement, support and time spent helping me balance my work and studies. (I should also mention my gratitude to my youngest sister especially for providing me with copious amounts of ice-cream as I wrote well into the evenings while working away from home!). Although no longer with us, I would like to express gratitude for my father-in-law who showed constant interest and promptings throughout my studies.

But by far my greatest thanks and appreciation goes to my wife and children - without whom this work would be but a distant dream. For a man is only as good as the woman who stands at his side. If this work is of any merit then equal recognition should go to my wife. At periods when my studies seemed like a black cloud with no end suspended over us all, my family has been the beacon in the night. This work is dedicated in its entirety to my wife Alison, and each of our children.

In closing I would like to thank God for the inner strength, conviction, faith and inspiration he has given me to complete this thesis.

Table of Contents

Table of Figures

Table of Tables

1 INTRODUCTION

Biometrics is in essence a combination of the words biological and metrics meaning quite literally the measurement of biological traits. Typically in today's society when discussing biometrics as a field of study one is referring to the capability of computer systems to isolate and recognise a subset of suitable biological features in order to identify, or verify the identity of an individual.

The need for computer systems to be able to perform the tasks of identification and verification has grown rapidly in recent years fuelled in part by the increased power and capabilities of modern computer systems. The typical common home and workplace personal computer now provides multiple processing cores, clock speeds measured in billions of operations per second, the capability to deal with large high precision numbers and ever increasing amounts of memory. This provides an excellent platform upon which to provide recognition systems to a large user base that previously was simply impractical or impossible.

Figure 1.1 - Modern Compact Camera performing Face Tracking (Nikon S6)

In the commercial sector, many consumer cameras now come with the ability to perform basic face recognition to provide better photo composition, lighting balance and focus area selection (Figure 1.1). Such techniques have been documented in journals [Lai and Yuen 2006] in recent years. More advanced systems such as the recent iPhoto application for the Apple Macintosh series of computers have even offered complete face recognition systems that recognise faces in a users photos and then attempts to identify them based on past user provided information (Figure 1.2) with varying levels of success.

In reference to the aforementioned iPhoto example (Figure 1.2) the two faces have in fact been incorrectly identified although previous training information for both individuals has been provided. Although this represents a worst case scenario it does highlight the fact that there remains always room for improvement in finding algorithms and techniques to enhance the technology.

Figure 1.2 - Apple iPhoto Application Face Recognition

Within the security sector and forensic sciences, face recognition systems are widely used (and in some cases depended on) for both identification and identity confirmation (verification). While the technologies that support identification and verification systems may be very similar, they are used for different purposes. Identification in this context is the ability to recognise an unknown individual from a dataset of individuals and may be performed offline on historical data. In contrast, verification is the ability of a system to confirm an individual's identity based on his or her biometric features. In the latter instance, the system is already provided with a claimed identity at the time of verification and is often required to provide a near instant response confirming or rejecting the users proposed identity.

Face recognition systems have rapidly become a prevalent area of research with many practical applications. This study presents facial skin texture from two specific regions as a frequency based feature and introduces a framework for an experimental analysis of its performance.

1.1 PURPOSE OF RESEARCH

The purpose of the work is to examine the feasibility of and methods by which facial skin texture may be employed for biometric subject identification. This thesis will focus on two specific skin regions located underneath the subjects eyes as highlighted in Figure 1.3 and Figure 1.4. (Section 5.4 page 68 provides greater insight into the selection of these areas and their labelling as 2A and 2B).

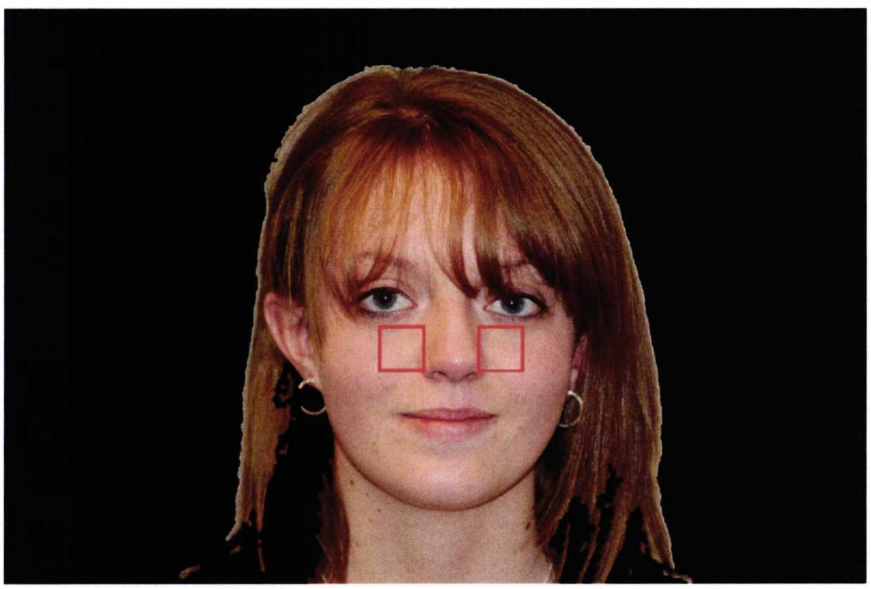


Figure 1.3 - Regions of Interest 2A and 2B

In recent years, the use of computers to assist the recognition of subjects from digital photographs or video recordings has evolved at a rapid rate to the extent where commercial off the shelf (COTS) products are readily available with biometric security capabilities.

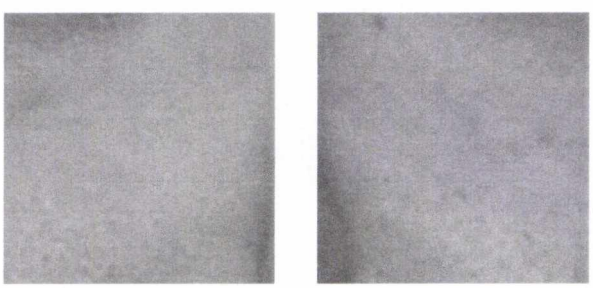


Figure 1.4 - Greyscale Images of Regions 2A and 2B (135x135 pixels)

The focal point of many modern biometric systems is the geometric patterns or shapes contained within imagery. The algorithms and techniques presented in this thesis differ from this norm by focussing on these skin texture regions as the sole feature for identification. The work presented highlights the evolution and fusion of texture recognition methods to evaluate whether they may be a reliable feature for the task of subject identification.

To fully exploit the unique features present in skin texture, operating in the frequency domain rather than or in addition to the spatial domain presents a significant advantage. Specifically the ability to recognise and discern between low and medium frequency patterns that might be difficult to recognise within the spatial domain. Recently some studies within the frequency domain have been used to enhance spoof detection mechanisms [Nixon and Rowe 2005] and subject verification [Maio and Nanni 2005] using fingerprint texture. This study differs by testing the algorithms against the more challenging task of subject identification as opposed to verification and utilises skin texture as a standalone biometric feature. (Note that this is not to say the system presented is unsuitable in a verification environment. See Section 4.6.3 for more information).

While accepted for some time that fingerprints are statistically unique between individuals, the same is not held for other skin surface textures. Although authors such as [Gomez 2002] and [Choi, Tse et al. 2008] present methods for facial tracking and detection using facial textures, when compared with non texture based methods the use of facial skin texture as a sole identification biometric is a novel and thus far untested process.

This study is intended to explore the hypothesis that skin textures within the facial region provide sufficient means to identify a subject in a biometric system. It also presents a critical analysis of the implementation and requirements of such a system.

1.2 Scope of Research

The scope of this research is to evaluate the performance of an innovative technique which uses Gabor wavelets as the prime means of subject identification. Later sections of the study also assess the capabilities of spatial frequency feature techniques to enhance the system's ability to isolate and describe high frequency patterns which may be difficult for a frequency domain based system to differentiate. This thesis will discuss the development of such a method, measure its performance and compare it to alternative techniques for facial recognition.

The study will also assess the performance of the technique when faced with common challenges such as variations in lighting and image resolution. Finally a critical review of the algorithms and their findings will be presented.

1.3 CONTRIBUTION

This study introduces facial skin texture from two specific regions below the eye as a biometric feature and proposes their use for subject identification. To extract features, both frequency and spatial based techniques are evaluated, their performance compared and the benefits and mechanisms of fusing the data are considered.

Although studies exist which use texture based features techniques in a biometric setting, the use of this specific skin segment as a basis for identification is a novel proposal. It is plausible that initial preconceptions for this genre of biometric would discourage its use. This may be due the idea that skin alone does not offer a sufficiently unique feature space. An exploration of this type of system suggests that in relatively small datasets (79 individuals) this type of system demonstrates encouraging results indicating that it may be suitable as a biometric candidate.

The study performs an analysis of various permutations of presenting biometric data to the classifiers in an attempt to identify an optimal arrangement. Once the optimal arrangement is identified, an investigation into the effects of scaling and cross region matching is presented. The study has shown that this skin region does contain sufficient information to enable it to be proposed as a field for further study. The study has also indicated that image resolution may not have as strong a relationship to system performance as initially supposed.

The practical use of such a system could offer certain benefits. Firstly, even though this preliminary study has suggested suitability as a singular biometric, a case may well exist for its use in a multi-biometric system to add confidence to other feature biometrics. In addition, despite our best efforts, there are instances in biometric systems where the ideal information for subject identification is not contained within an image. The reasons for this may be many - perhaps poor pose, insufficient lighting or even attempting to interpret biometric data from a photograph that was not captured following a set biometric data collection protocol. As a second benefit it is then plausible that a system which exploits skin texture to establish identity could prove useful in circumstances where pose is not ideal.

1.4 Early Hypothesis on Image Resolution

The objective of the research was to ascertain if an individual's skin texture from a specific facial region provides a sufficient feature set to enable recognition. The assumption was made that in order to provide the best feature set, high resolution photographs (in the order of 6 megapixels) would provide best results. It was predicted that by increasing the resolution, skin texture features would become more visible and hence increase the systems performance.

For these reasons, it was decided to perform the analysis on the newly introduced Biosecure Multimodal Database (BMDB) [Ortega-Garcia, Fierrez et al. 2009]. Produced under a rigid framework to high specifications by a consortium of higher educational establishments the database provided a completely new image library with which to work. The key benefit of using the BMDB were the high resolution facial images (8.3Mpixels) with which to work.

During the evaluation in Chapter 7, this assumption was scrutinised. It was found that in fact, resolution did not directly impact performance in the manner predicted. Although at extremely low resolutions the accuracy of the system diminished, there were scaling reduction factors that produced superior results than the source resolution. This indicates that the system may be more flexible to working with lower resolution image databases than initially supposed.

1.5 Document Structure

This thesis is divided into two key parts. Chapters 1 to 4 present the reader with background information, surrounding literature and the experimental framework which will be used to conduct the study and experimentations described subsequently in Chapters 5 to 8.

Chapter 2 presents the history of biometric research and its applications in society and includes an introduction into the different requirements for different types of face recognition systems.

Chapter 3 introduces texture as a feature and discusses the evolution of the Gabor Filter. A mathematical model for a filter is derived and examples of its application are applied to two test images.

In Chapter 4, the experimental framework used to conduct the research is established and where appropriate justified. Topics such as hardware and software used in the evaluation are presented along with databases used through various stages of the systems development.

Further on, three classifiers are selected for the evaluation to ensure that a fair analysis of the proposed features is available. The section concludes with a discussion of the evaluation type and procedural measures taken to ensure the integrity and trustworthiness of test results.

Chapter 5 begins the experimental section of this thesis and commences with the introduction of the human face and skin texture are perceived in two dimensions. Although region of interest location was performed manually for this evaluation, a brief examination of image segmentation using skin tonal ranges is covered. The stages of image segmentation and feature extraction are presented and the initial results of using a Gabor filter based system to extract features is assessed against a test database.

Chapter 6 extends the experimental results in Chapter 5 by examining the performance of combining multiple regions. The Grey Level Cooccurrence Probability (GLCP) Matrix is introduced as a supplementary system optimised for the detection of high-frequency texture patterns. The system is then assessed using every possible permutation of both feature and score based fusion for the two skin regions to find the optimal method.

In Chapter 7, the performance of the system is critically assessed by measuring the time taken to both extract features and perform classifier training. The impact of image resolution both on training and testing data is evaluated with further experimental results showing the resolution to performance relationship. A brief investigation into the performance of cross matching skin regions between training and testing is conducted and a statement is made on overall system confidence.

Chapter 8 concludes the evaluation and this thesis by drawing conclusions from the work presented and laying out the path for future research and suggested topics of interest.

Each chapter commences with an explanatory section and terminates with a chapter review allowing the reader to quickly gain perspective on specific areas of this Thesis which may be of interest.

1.6 CHAPTER REVIEW

This chapter has presented a high level overview of the research aims and objectives. A brief introduction to the background of biometric systems and their use in society has been discussed and a statement has been made on an initial underlying hypothesis concerning image resolution which was later found to be unsupportable by the experimental evidence. In summary, this study aims to answer the following questions.

1. Is there a sufficient amount of identifiable information in small regions of facial skin texture to recommend its analysis as a biometric feature?

2. How might the regions in (1) be analysed in order to extract features suitable for classification.

3. How is the performance of (2) impacted by variations in image scaling of the training and probe images?

4. How is the performance of (2) impacted by variations in image scaling of just the probe image?

5. If a system is trained on one region of skin, to what extent is it capable of using skin from another area of the face to recognise an individual?

2 BIOMETRICS AND FACE RECOGNITION

2.1 HISTORY OF BIOMETRICS

Biometrics by definition is the measuring of physiological characteristics of an individual. Today the use of Biometrics is currently an active area of research for the computer assisted identification of persons. There is a common misconception that biometrics is a recent area of research, stemmed by modern requirements such as terrorism threats, crime and user authentication and authorisation, however the use of biometrics to identify individuals dates back to the 14th century as reported by explorer Joao de Barros [Munir 2008] who wrote that the Chinese merchants were stamping children's palm prints and footprints on paper with ink to distinguish the young children from one another.

It is likely at this early a period, the use of manual biometric systems was limited for subject verification from a small known dataset. As communications technology progressed (starting with the patenting of the electric telephone in 1876), the first known use of a multimodal biometric system for subject identification for criminal purposes began.

This differed from previous scenarios in several ways. The information had to be simple to understand and maintain by a large number of database contributors, and needed to be reliable enough to permit criminal identification in a court of law. These constraints changed the fundamental way that biometric systems are required to operate. Now, rather than just assisting with the identification of an individual, biometric systems were being used to prove an individual's identity. With this change came new requirements in terms of confidence and reliability of a biometric system.

In the early 1890's, an anthropologist and police clerk in Paris, Alphonse Bertillon developed a method of multiple body measurements [Bertillon 1896] as shown in Figure 2.1. This system was used by police authorities throughout the world for some time. Its failure was noted and its use rapidly discontinued when it was discovered that some people shared the same measurements and based on the Bertillonage measurements, two people could get confused.

This is a vital lesson in any study of biometrics for subject identification and has given rise to performance measuring techniques which attempt to establish a degree of confidence in a system's performance [Mansfield and Wayman 2002]. The impact of using any automated

system in a real world scenario should be careful measured with users of the system being made aware of both the benefits and risks involved in placing confidence in such a method.

Figure 2.1 - Bertillonage method of Forensic Biometric Identification

After the failure of Bertillonage, the police started using fingerprinting, which was developed by Richard Edward Henry of Scotland Yard, essentially reverting to the same methods used by the Chinese for centuries.

2.2 FACE RECOGNITION

In the past few decades, biometric research has surged forward with modern systems now employed in practical applications where a degree of autonomous functionality, precision and reliability is required. Techniques have diversified from fingerprinting to facial recognition, retinal/iris scanning, limb geometry, vascular pattern recognition and even DNA. While fingerprint biometrics remain the most commonly employed biometric, the industry is devoting significant time and expense into these other areas.

Facial recognition, as one might imagine, involves identifying and extracting facial features for the purpose of individual recognition. Facial recognition has the advantage of being generally non invasive when compared with other techniques. It is noteworthy also to mention that the human brain's primary means of person identification is based on facial features.

There were initially two primary methods for facial recognition which are listed below:

- Geometric
- Pictorial

In geometric facial recognition one considers the spatial configuration of facial features. This usually involves locating the primary facial features such as the eyes, nose and mouth and measuring the geometric distances and angles between these features. These measurements are then used to classify the images for recognition purposes. For example, [Shih and Chuang 2004] describe a series of geometric methods to extract head and face boundaries and facial features.

Pictorial facial recognition in contrast uses a facial template of major features and examines the entire face to perform recognition of frontal face images. One such technique involves computing a face description by approximating the eigenvectors of the face's auto-correlation matrix and is now known as eigenfaces. Developed by Matthew Turk and Alex Pentland in 1987, eigenfaces became the benchmark to which many other techniques are compared and the first facial recognition method that was generally accepted as working [Turk and Pentland 1991].

With eigenfaces, a large set of facial images taken at the same lighting conditions are normalised along the eyes and mouth. All images are then resampled at an identical resolution. Principal component analysis (PCA) is then used to extract the eigenfaces. The technique was so successful that a generic "eigenimage" term is now utilised and the method is now used in other areas including handwriting recognition, lip reading, voice recognition and medical imaging [Allan M. Haggar 1989].

The principle of eigenfaces is fairly straightforward. First, a training set of normalised images is produced as described above. This is called dataset D.

One then subtracts the mean (M), or average image value on a pixel by pixel basis for each image. The resultant image (R) is then stored. From (R) the covariance matrix is calculated

followed by the eigenvectors and eigenvalues. Finally the dataset is thinned to a core selection of principal images using PCA.

When the dataset is thinned, one is left with a selection of standardised face ingredients which have been derived from the statistical analysis of many images. Any new faces are then compared and referenced as percentages of likeness to an eigenface in the dataset.

For example, a new face could be composed of 10% eigenface 1, 60% eigenface 2, 50% eigenface 3 and -20% eigenface 4. By storing the characteristics in this way, quick correlation to existing known averages is possible.

2.3 Objectives of Facial Recognition Systems

2.3.1 Introduction

The idea of recognising a person's face using a computer system is now one of large academic and commercial interest. It should be noted that there are several reasons one might want to recognise a face, be it for identification purposes or assisting an advanced human machine interface system to understand facial expressions.

In computer assisted facial biometrics there are four prime areas that are currently active fields of research (although often the practical implementation of a biometric identification system involves the combination of multiple fields). These can be summarised as:

- Subject Identification
- Subject Verification
- Facial Tracking/localisation
- Facial Expression Analysis

This section will address each of these fields with an explanation and highlights of recent work within that area.

2.3.2 Subject Identification

In subject identification, the objective is to identify an unknown individual from a large database of known individuals. Typically a classifier has been trained with the features of the known database and is then presented with an unknown (probe) image with which to

identify. Various factors may affect the accuracy of the classifier assignment such as classifier performance, quality of the probe image, subject orientation and facial expressions. Typical uses of this type of system are forensics, security closed circuit television systems and criminal record searches.

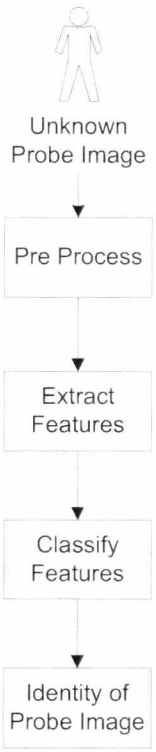


Figure 2.2 - Subject Identification System

Figure 2.2 shows the steps involved in a subject identification biometric system. The system is reliant upon a classifier which has been pre-trained to recognise individuals from a known dataset using the provided features. The accuracy of these form of systems is dependent on both the capability of the features to uniquely describe an individual from a database of N individuals as well as the classifiers ability to use this information to suggest a plausible identity for the subject.

[Jeon, Kwon et al. 2004] use a Gabor based encoding system and genetic algorithm for face recognition. Expanding on this, [Kim, Jeon et al. 2006] show correct face recognition to the FERET database of 97% by using Gabor wavelets assisted by a hardware based genetic algorithm. (See also [Krishna, Black et al. 2006]).

[Liu, Lam et al. 2004] present a paper that applies linear discriminant analysis to extract the principal features of faces for recognition purposes and thus reduces the feature space

dimensionality for faster processing. [Meade, Sivakumar et al. 2005] provide a comparative study between PCA, Gabor Wavelets and Discrete Wavelet Transforms (DWT) for facial recognition. Rather than attempt to reduce feature dimensionality, [Liu 2006] presents a Gabor wavelet based method for face recognition that capitalises on the large dimensional feature vectors produced by varying Gabor orientations and frequencies. (see also [Shen, Bai et al. 2007])

Another unique application of Gabor wavelets is used by [Wang, Li et al. 2006] which involves passing a Gabor wavelet window from left to right, top to bottom across the image to search for specific facial features prior to identification (see also [Nanni and Maio 2007]).

In an effort to accelerate the process of face recognition from large databases, [Chow and Rahman 2006] present a Self Organising Map (SOM) concept which is relatively feature independent. The SOM is capable of creating compressed feature vectors within the database which can be used to rapidly isolate and identify a small groups of potential candidate matches.

In the majority of Gabor feature extraction processes, the focus is on the amplitude of the filtered patterns. [Zhang, Shan et al. 2007] proposes an alternate method of using a phase based histogram to extract features suitable for face identification.

In an attempt to redefine the "flat" space vector-style features commonly used in Gabor facial features, [Cho and Wong 2008] train a neural network to define a tree based structure which is used to represent face identification information.

2.3.3 Subject Verification

In subject verification the objective is somewhat simpler. The individual presents him or herself to the system with a unique identifier. The biometric information is then used to verify that the individual is indeed who he or she claims and not an impostor. In these circumstances, the system is required to give a positive or negative confirmation to a single class within the database. This may be accompanied with a numerical value from the classification score which can indicate to what degree of confidence the system was able to match the subjects features to their known identification record. Typical uses of verification systems include IT Systems authentication, identity confirmation and new biometric passports.

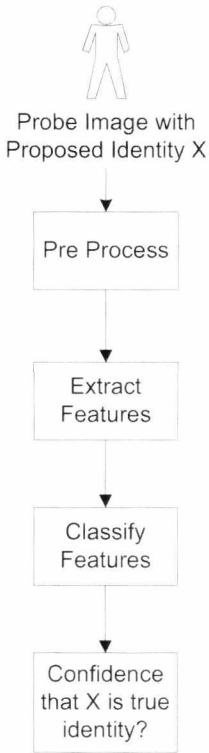


Figure 2.3 - Subject Verification System

[Duc, Fischer et al. 1999] discuss elastic graph matching to increase the classification success rate of eigenface processed images for subject verification (see also [Kotropoulos, Tefas et al. 2000]) although [Gonzalez-Jimenez and Alba-Castro 2007] claim superior performance using valley/ridge feature detection on facial images and performing Gabor feature extraction from the valleys and ridgelines.

2.3.4 FACIAL LOCALISATION AND TRACKING

In facial tracking or localisation, a system is presented with a "noisy" image or sequence of images and required to locate and/or track the face. In this context, image noise implies that the image contains information unrelated to the subjects face and the system is required to filter through this to find facial patterns. Also within this field are image segmentation systems to which one may input a face for feature localisation prior to identification, verification or pose analysis. Typical uses of face tracking and localisation include automatic face tracking cameras, and precursory facial extraction for further identification/verification purposes.

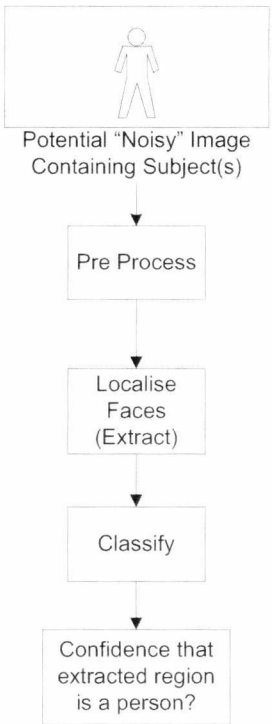


Figure 2.4 - Simplified Localisation System

In Figure 2.4 a simplified localisation system is presented. In this design an image that may contain a face is presented. The system attempts to identify regions which contain facial features and present the user with a confidence value that the regions identified are in fact faces. Note that similar systems may also be used to localise features within a face probe image (such as eyes and mouth). This method is frequently used to automate subject and feature localisation prior to identification or verification.

In the late nineties, a method was presented that utilised Gabor wavelets to detect human faces from colour images [Garcia, Zikos et al. 1999]. [Imaoka and Okajima 2004] use a massive number (196,800) of Gabor features for face detection within images containing large amounts of background noise. [Li, Ou et al. 2005] uses a similar method to recognise human faces using Gabor features and support vector machine classifiers.

[Kyrki, Kamarainen et al. 2004] discuss a method of searching for objects of interest within a Gabor feature space and [Shih and Chuang 2004] use geometric face models and Gabor filters to identify known features (such as eyes) in order to facilitate head, face and facial feature extraction. [Ersi and Zelek 2005] later propose a rotation invariant Gabor-based system for face feature detection which is claimed also exhibit brightness, contrast and some scaling invariance (See also [Gneushev 2007]). Face detection is also fast becoming a capability of mobile devices such as telephones and cameras for assisted focussing or

"tagging" systems [Lai and Yuen 2006] (Tagging refers to computer assisted facial recognition which then prompts the user for names of faces recognised).

A recent study [Lian and Lu 2007] proposes a gender classification system based from multi-resolution, multi-angle images that uses Gabor wavelets for feature extraction.

2.3.5 Facial Expression Analysis

Facial expression analysis is the use of biometric systems to attempt to interpret facial expression or mood and is particularly useful in human to machine interfaces and lie detection systems.

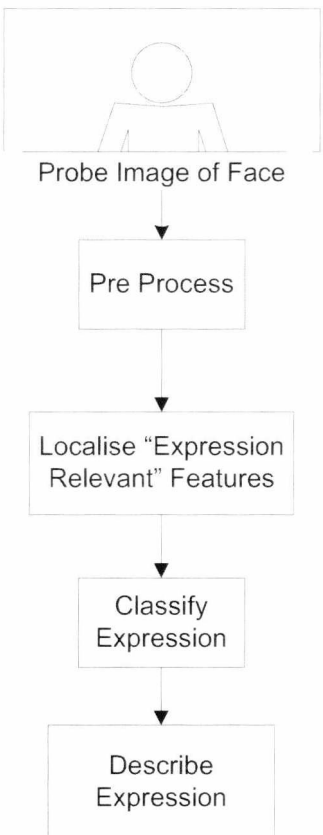


Figure 2.5 - Simplified Expression Analysis System

With the recent advancements in digital imaging and computer power, autonomous biometric systems have made considerable advances. [Donato, Bartlett et al. 1999] describe a facial gesture recognition system which uses Gabor Wavelets to identify 12 unique facial actions with 96% accuracy, [Chen, Lai et al. 2004] use a similar method against the MIT Face database for pose discrimination and achieve a success rate of 93.5%. This study is

later extended with the application of a hybrid boost learning algorithm for enhanced classification. [Chen, Huang et al. 2008]

In an attempt to identify not just the facial expression, but the strength of the expression [Amin, Afzulpurkar et al. 2005] apply a fuzzy c-mean clustering algorithm alongside Gabor wavelets resulting in a ranking of face emotion.

2.3.6 3D Facial Recognition

Although more a "technique" supportive of any of the prior biometric objectives, the topic of 3D face recognition is unique enough to warrant independent mention.

Three dimensional face recognition can be used to further enhance the performance of 2D algorithms by providing compensation for head orientation and lighting changes. Frequently a 3D mesh of the face is generated which allows native 3D recognition algorithms to work alongside proven 2D systems for a greater degree of accuracy. To its downside, 3D recognition requires the use of specialist camera equipment that is capable of range measurement or projecting a mesh onto the subjects face.

[Wang, Chua et al. 2001] [Wang, Chua et al. 2002] propose a hybrid 2D/3D Gabor feature based system to identify subjects from a database of 20 individuals from different visual perspectives. Although not directed towards biometric identification, [Tarini, Yamauchi et al. 2002] describe a technique to fuse multiple 2D images to create a 3D texture mesh of a person's head for modelling purposes. [Abate, Nappi et al. 2006] use 3D face meshes and 2D textures to generate a composite feature vector which is then reduced using principal component analysis (PCA) for identification purposes.

2.4 Scope for Future Work

Biometrics is a rapidly progressing field of study and presents many diverse avenues for future expansion. Since the 9/11 attacks on the United States, government institutions have diverted substantial funding and research efforts into computer aided recognition systems as part of a global counter-terrorism effort. As part of this effort, the USA now fingerprints and takes facial photographs of every international visitor at all major airports creating perhaps the largest ever database of human biometric information [US-DHS 2009].

Outside of governmental departments, biometrics also has a large base in enterprise organisations where fingerprint or iris recognition devices are used for user authorisation and authentication in both computer systems and physical location access.

In the consumer market, recent advances in low powered microchips have permitted advanced facial recognition software to be built into compact digital cameras and recording devices allowing facial tracking for optimal subject focus.

It would be a fair assessment to say that biometric research is a broad field of study with a substantial history. In terms of technological implementation and new approaches and implementations, biometric recognition systems is still in relative infancy and one can expect the capability of electronic equipment to recognise imagery and video in new and interesting ways in the coming years.

2.5 CHAPTER REVIEW

This chapter has provided a brief introduction into biometric recognition systems in both the pre and post information technology eras. In particular, face recognition has been discussed and it has been presented that there are a minimum of four distinct facial recognition system key fields which include:

- Subject Identification
- Subject Verification
- Facial Localisation and Tracking
- Facial Pose Analysis

The prime objective of this thesis is to present a system capable of subject identification. Identification of an unknown individual from a database of many is generally seen as the most challenging requirement of a biometric identification system - thus a critical evaluation against this requirement should if successful provide sufficient information to enable its use in other areas.

The reader should be able to review this section to gain a perspective on what requirements are placed on a system of this type in terms of reliability, confidence and role.

3 TEXTURE AND THE FREQUENCY DOMAIN

3.1 PERCEPTION OF TEXTURE

Visual texture provides an abundant source of information about the nature of surfaces and objects all around us. As human beings, much of the detail of what we see is taken for granted. Often oblivious to the intrinsic elements that surround us, details such as texture are often noticed and processed only by our subconscious in an effort to "see the bigger picture".

Many surfaces may be defined or identified primarily by their surface texture. In addition to recognising familiar textures, the human visual system is capable of recognising three dimensional paths and objects by correlating the perspective changes in texture.

Before progressing too far into the discussion of texture, a clear definition of texture for the purposes of this study should be established. A texture is by definition 'the feel, appearance or consistency of a surface, substance or fabric' [Hawker and Soanes 2005]. Texture is a discernable item and one which human beings are adept at recognising. By merely viewing an object, an individual may be able to judge what materials the object is constructed from, whether the surface is wet or dry, hard or soft, and many other features describing the object. Judgements are made from the perception of texture based on the consistency and appearance of the object. Repeating or like patterns are identified and recognised based on prior experiences. In some instances, the person is able to describe the touch-sensation of an object merely by viewing it and identifying the visual appearance of texture.

3.2 DEFINING TEXTURE

If one were to look at texture as "a set of re-occurring frequencies of spatial patterns that may be suitable for identifying an image or part of an image" then almost any image, image subset or collection of images may be defined in its entirety by its texture. For example, this thesis is a composition of letters, numbers, images and symbols printed primarily in black on a white background. The pattern of letters on this page may be defined as a texture which is unique to this page alone within the document. However, one will also note that the page is sufficiently similar to other pages of text as to be identified at a glance as a "page of text". From this particular example, it could be concluded that:

1. The composition of numbers, symbols and letters on this page may be viewed as a texture of the page. The texture of any specific page allows the page to be identified generically as a page of text.

2. The texture of any specific page allows each page to be identified as individual and could potentially be used to isolate exact replicas (ie. photocopies).

This gives rise to two important but opposing properties of texture - similarity and uniqueness. Both may be defined relative to other textures with which we are familiar. By providing exact definitions to these two quantities it is possible to use ones experience to classify new textures either as an association with items previously viewed or as something new or as is commonly performed, as some function of the two.

It is proposed that this relationship could be represented as:

$$x = f(x) + g(x) \tag{3.1}$$

Where x is the texture being studied, $f(x)$ represents a function identifying how distinct a texture is from previous examples and $g(x)$ represents a mapping function showing the textures similarity to previous textures.

By identifying and describing texture in relative terms, every new texture examined becomes a function of previous textures allowing the experience of an individual to assist them. On studying a medieval castle wall for the first time, a viewer will notice patterns of stone and masonry unique to the construction of the castle, yet the human mind is capable of instantly recognising this form of construction and linking it to modern day brickwork and construction.

Similarly one may view clouds in the sky which are in a constant state of change and motion – every day completely unique and random, yet the capability to identify the visual stimuli as clouds is present no matter what shape and form they take – in fact young children frequently observe that the patterns in clouds take forms of known objects. This ability to associate is what makes the human visual system so powerful and interesting.

Having an understanding of the capabilities of the human visual cortex is important in any effort to reproduce computer based recognition systems. The capability to correlate two or more unique items for recognition purposes is an ability that one strives to replicate in the field of texture recognition.

Research into the visual cortex suggests that much of our capability to identify textures comes from isolating characteristics in the frequency patterns within texture. [Hanazawa 2004]

3.3 TEXTURE RECOGNITION

Researchers have shown an active interest in texture recognition before its use in face recognition. This section addresses briefly some relevant research into texture processing that falls outside of the biometric category. (For texture research within the biometric field, please refer to Section 2.3).

The remote sensing sector has long pursued the capability to autonomously classify aerial and satellite imaging textures to various categories such as grass, sand, water, forest and other land type surfaces [Cao, Peng et al. 2004]. In the early nineties, [Barber and Ledrew 1991] present a method for determining sea ice from synthetic aperture radar remotely (SAR) sensed images based on its textural properties (see also [Soh and Tsatsoulis 1999; Clausi 2001] and [Clausi and Yue 2004]). [Kaplan 1999] present a paper on a scale invariant system which uses the Hurst parameter of the fractional Brownian motion for texture recognition and successfully apply the technique to identify textures in remotely sensed SAR images.

A generic texture recognition system that measures energy in the frequency domain is used by [Aujol, Aubert et al. 2003] to perform autonomous texture segmentation and classification. An alternative system attempts to define all possible micro-texture features as textons, similar to elements in a period table where unknown textures are identified as compounds of textons [Zhu, Guo et al. 2005], this is similar in principle to the discussion of Eigenfaces in Section 2.2 .

Texture is also of significant interest in the medical research sector particularly for the detection of abnormalities such as lung cancer [G.S.Cox 1992], tumour detection [Santoro, Prevete et al. 2006], [Rangayyan and Ayres 2006] and ultrasound image interpretation [Chen, Lu et al. 2001], [Xie, Jiang et al. 2005].

In industry, texture based systems can be used to detect imperfections in manufactured items such as textiles [Escofet, Navarro et al. 1998], [Latif-Amet, Ertuzun et al. 2000], [Bodnarova, Bennamoun et al. 2002], [Mak and Peng 2008], semiconductors [Bourgeat, Meriaudeau et al. 2006] and protein crystallisation measurement [Pan, Shavit et al. 2006].

One difficulty in working in the frequency domain can be caused by the apparent loss of the geometric shape information. Studies indicate however that some element of shape information is preserved in this domain and demonstrates how this may be recovered in a post-filtered environment [Massot and Herault 2008] [Wen L. Hwang 1996] or even provide 3D based information from a 2 dimensional image based on texture properties [Wen L. Hwang 1996].

3.4 Frequency Domain Features

Fourier analysis has a long and complicated history involving several individuals who investigated varying physical phenomena. History is aware of civilisations as early as the Babylonians having an awareness of using trigonometric sums – the use of sums of harmonically related sines, cosines or periodic complex exponentials - to describe and predict periodic astronomical phenomena [McKean 1972].

The first detailed work however begins with L. Euler in 1748 who studied the motion of "normal modes" of a vibrating string. Euler hypothesised that if the vertical deflection of a string at time t and distance x along the string, then at any given point in time the position of the string may be represented as harmonically related sinusoidal functions of x. Euler noted from this that if he was able to accurately describe a string at a point in time, he was also able to describe the coefficients for the same function for a future point in time [Euler 1749; Euler 1750].

This discovery by Euler led to significant study and debate throughout the mid 18th century. A few years later in 1753, D. Bernoulli argued that all physically possible motions of a string could be represented by such combinations of normal modes but did not support this mathematically and perhaps for this reason, his ideas were widely disregarded [Bernoulli 1753]. In fact in 1759, J. L. Lagrange criticised the use of such trigonometric series as in his own view, it was impossible to represent signals with corners (i.e. discontinuous slopes) using this method.

In this environment, Jean Baptiste Joseph Fourier presented his ideas some fifty years later. His contributions are even more significant when one takes into account the scepticism and hostility towards trigonometric series at the period. In addition to his mathematical genius, Fourier played an active role in politics (at times placing his life in jeopardy) in the years following the French revolution and became a personal travelling associate of Napoleon Bonaparte accompanying him on an expedition to Egypt and later in 1802 being appointed by Napoleon to a prefect in the region of Grenoble.

By 1807, Fourier completed a study on mathematical physics which used series of harmonically related sinusoids to represent the temperature distribution through a body. In addition to this, he made the claim that such a series could represent "any" periodic signal. It should be noted that although his studies and insight are what in time provided a strong basis for his works, many of his ideas and concepts had been discovered by others. In fact, Fourier did not actually contribute to the mathematical theory of Fourier series but through his claims and understanding of the topic, and work in other areas spurred the development and research of the area.

Unfortunately for Fourier, his paper of 1807 was rejected due to J.L Lagrange's vehement objections to the proposal for the same reasons he had criticised Bernoulli's work some fifty years earlier. For this reason, the paper was never published. Undeterred however, Fourier persevered and eventually had his research published by the Insitut de France under the text *Théorie analytique de la chaleur* [Fourier 1955] which was successfully published in 1822.

Much later in his life, Fourier's work began to be recognised for its significance in mathematics, science and engineering. Due to the important role sinusoidal signals play in so many fields, Fourier's research has been successfully applied to many varied and different fields.

3.4.1 THE GABOR FILTER

Texture is a topic of much interest and discussion in the field of computer based image recognition. If one were to classify an image as an array of intensity values ranging between two arbitrary numbers, then the texture of that image would be defined as repeating or similar patterns between two specific points within the array. Identifying these patterns is second nature to human beings who seem to possess an almost innate ability to identify common textures such as brickwork, carpet, grass and other surfaces with ease.

Replicating these abilities in computer software involves mapping textures to simple information "features" that may be used to discern between various surfaces, presenting a technique which successfully performs this "feature extraction" will form the basis of this thesis. By representing textures as a sequence of "features", a system may be trained to "classify" future images presented into a known category. This process is known as "classification".

The history of texture processing using computer systems had its debut in the early nineties when computer systems first became commonplace in the laboratory, business and home

environments. For the first time, the ability and technology to perform intensive computational algorithms on images was possible.

Early fields of study in texture processing were focussed on two major areas: Identifying features and characteristics of the human body and the inspection of remote sensing data from overhead and satellite photographs. Although a few studies were conducted outside these fields, the majority of research focuses on one of these two domains with a practical application in mind.

For its ability to extract frequency information from a signal, Fourier analysis has proven to be a powerful technique in signal processing. However, when applied to image processing, Fourier analysis suffers from confusion due to spectral features from different parts of the image being fused and mixed together.

In the field of image processing, the ability to extract spatially localized features is of importance. Gabor filters are a popular tool in this field due to their ability to perform exactly this function. Dénes Gábor was a Hungarian physicist born in 1900. His most notable works include the invention of holography [Gabor 1948]. The basic principle leading to his discovery was to incorporate phase as well as amplitude in optical imaging information. Extraordinarily his research was published over a decade before the first invention of the laser which allowed holography to become a reality.

Perhaps key to his success as an inventor and researcher was Gabor's ability to observe and reproduce biological functions such as sight and communication in theoretical mathematics. Of particular interest is his 1946 paper entitled "Theory of Communication" [Gabor 1946] where Gabor introduces the frequency domain as storing the 'essence' of information.

Gabor's work on communications led to the utilization of "Gabor Filters" in both communication compression and feature extraction for image processing.

By modifying different input parameters of a Gabor filter, it is possible to isolate specific texture patterns within an image. Optimising a filter bank for particular texture patterns has been a subject of much discussion (see also Chapter 2).

Laine and Fan [Andrew Laine 1993] introduced a multi-channel approach to texture recognition which featured Discrete Wavelet Transformation (DWT) [R.R. Coifman 1992; Wickerhauser 1992; Unser 1993] and found the algorithms were also more successful than spatial analysis at identifying images with low-frequency changes.

Wavelets are frequently found to be advantageous in signal processing due to their ability to examine a signal within a specified temporal and scale domain. By contrast, Fourier

transforms analyse signals over an infinite period and often miss "detail" or information spikes in signals because of this.

In a study on shape from texture, [Wen L. Hwang Circa 1996] uses a continuous wavelet transform to establish the orientation of a texture from ridge patterns. By extracting this information, it is possible to identify three dimensional planar shapes. Ribeiro and Hancock [Eraldo Ribeiro 1998] continue this research and establish a successful algorithm to use texture for planar vanishing point detection, thus allowing three dimensional estimation from a two dimensional image.

In 1998, a team at the University of Barcelona proposed a multi-orientation and multi-scale method of Gabor filter feature extraction in emulation of the human vision process for defect detection in textures [Escofet, Navarro et al. 1998]. Multi-scale techniques were also employed in a Gabor wavelet transform technique used for texture analysis in remote sensing applications [Zhu and Yang 1998].

The approximation of a function using a superposition of functions has existed since Joseph Fourier's work in the 1800's where he realised that sines and cosines could be superimposed to represent other functions. As an extension of this work, wavelet analysis utilises different scales and resolutions to analyse data in an attempt to see both "the forest and the trees" of a signal. A large window will allow the "forest" to be seen clearly, while a smaller window will allow small discrepancies between the "trees" to be identified.

In image processing, Fourier analysis can be limited in its accuracy as by definition the functions produced are non-local (stretch to infinity) and have difficult handling spikes in a function. By contrast, discrete wavelet transforms allow approximating functions that are contained within finite domains and are ideal for handling such localised spikes in data.

The study of image spatial frequency and orientation is one of much interest; in particular the ability to quantify features such as texture and pattern repetition in the frequency domain has led to significant research in this field.

3.4.2 THE FAST FOURIER TRANSFORM

Suppose for example, a simple sinusoid $y = \sin(x)$ and its corresponding fast Fourier transform (FFT) are considered as shown in Figure 3.1 and Figure 3.2.

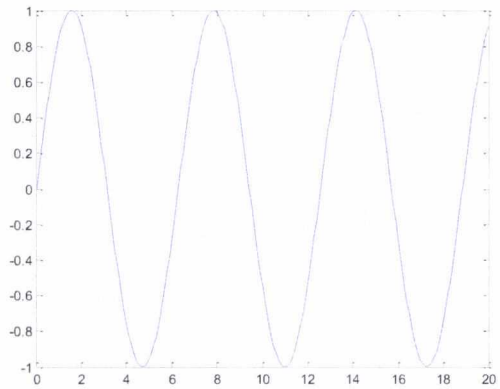


Figure 3.1 - Plot of function $y = sin(x)$

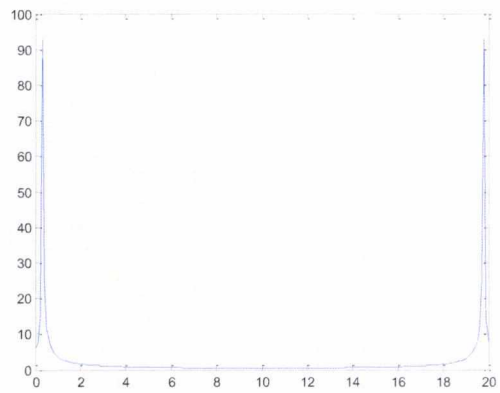


Figure 3.2: Fast Fourier Transform of $y = sin(x)$

It is relatively simple to discern the frequency of the sinusoid by studying its spatial representation in Figure 3.1. One can see that the signal repeats with period $x = 2\pi$. This information is also preserved in the FFT by a corresponding peak in the magnitude.

By introducing a compression of the wave along the x axis a similar wave can be shown with period $x = \frac{\pi}{2}$ for which the function $y = \sin(4x)$ is depicted in Figure 3.4.

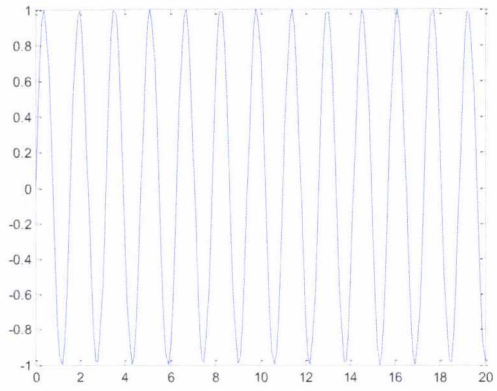


Figure 3.3 - Plot of function $y = sin(4x)$

Again from the FFT the related magnitude spike corresponding with the wave period in the x axis (Figure 3.4) clearly highlights the preservation of information within the frequency domain.

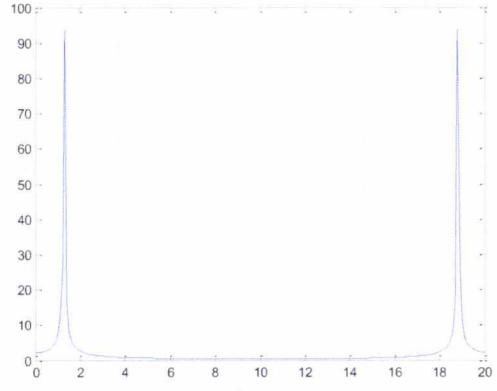


Figure 3.4: Fast Fourier Transform of $y = sin(4x)$

When these two functions are combined, the resultant sinusoidal signal becomes less clear – however the FFT of the wave retains the clarity and facility to extract the frequencies of each curve with relative ease (see Figure 3.3).

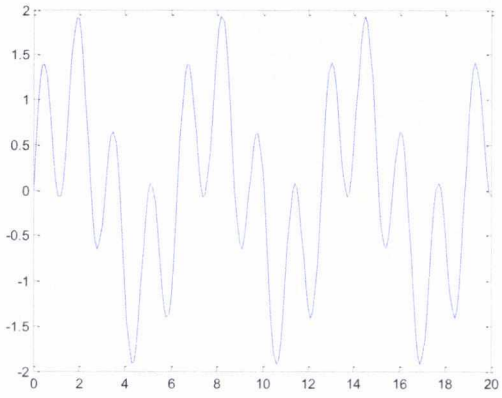


Figure 3.5 - Plot of function of $y = sin\ x + sin(4x)$

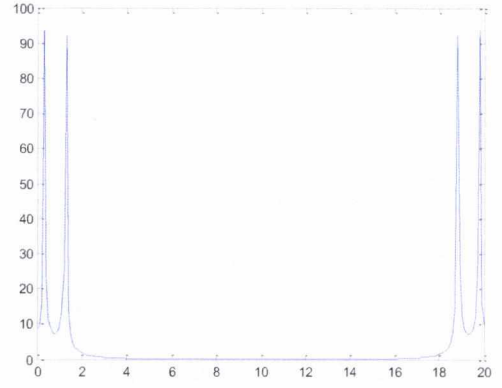


Figure 3.6: Fast Fourier Transform of $y = sin(x) + sin(4x)$

The study of signals within the frequency domain is a widely studied area and significant other information pertaining to the background and development of frequency domain systems can be found in many reference books and papers. This thesis will focus solely on the application of frequency domain techniques to face recognition.

3.4.3 DERIVING GABOR FILTERS

A Gabor filter is formed by multiplying a defined harmonic function $s(x, y)$ by a Gaussian function $w(x, y)$ such that:

$$f(x, y) = s(x, y)w(x, y) \tag{3.2}$$

The output function $f(x, y)$ is known as a Gabor filter and is convolved with the input image to obtain an output. By performing this function directly in the frequency domain using the convolution-multiplication property, the computational overhead can be substantially reduced.

The harmonic function is also known as a complex sinusoidal wave and can be defined as:

$$s(x, y) = e^{j(2\pi(u_0 x + v_0 y) + \varphi))} \tag{3.3}$$

where φ defines the phase and (u_0, v_0) the frequency of the sinusoidal.

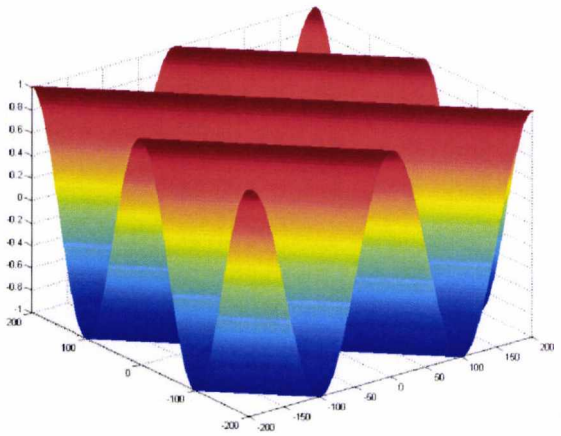


Figure 3.7 - 3D Projection of the Real Component of a Complex 2D Sinusoidal Wave

Johann Carl Frederich Gauss introduced the concept of a "normal distribution" later known as Gaussian distribution in the early eighteen hundreds. Gaussian distributions have been used in statistical analysis to represent a "natural" distribution of elements from astronomical phenomenon to psychological measurements.

A 2-Dimensional Gaussian curve can be described as (see [Movellan 2005]):

$$w(x, y) = Ke^{(-\pi(a^2(x-x_0)_r{}^2 + b^2(y-y_0)_r{}^2))} \tag{3.4}$$

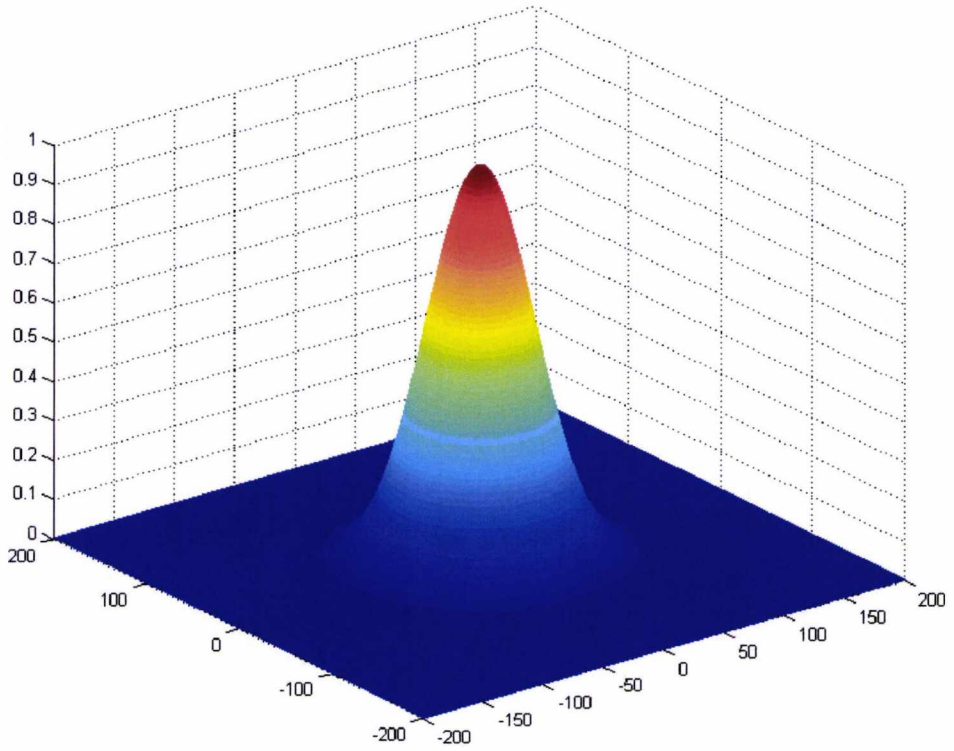


Figure 3.8 - 3D Projection of a 2 Dimensional Gaussian Function

Where a and b scale the Gaussian, x_0 and $y_0$are the peak of the function, K modifies the output amplitude and the r subscript represents the following rotation operations:

$$(x - x_0)_r = (x - x_0)cos\theta + (y - y_0)sin\theta \quad \textbf{(3.5)}$$

$$(y - y_0)_r = -(x - x_0)sin\theta + (y - y_0)cos\theta \quad \textbf{(3.6)}$$

Multiplying the Gaussian function (3.4) against the sinusoidal carrier (3.3) gives:

$$f(x,y) = Ke^{(-\pi(a^2(x-x_0)_r{}^2 + b^2(y-y_0)_r{}^2))}e^{j(2\pi(u_0x+v_0y)+P))} \quad \textbf{(3.7)}$$

This function is known as the Gabor filter. The real and imaginary components of which are shown in Figure 3.9 and Figure 3.10.

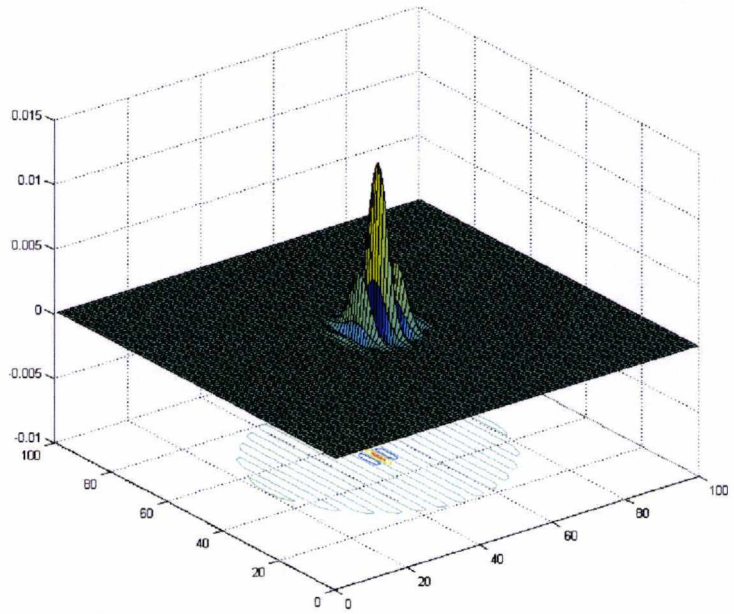


Figure 3.9 - Real Component of a Gabor Filter

The combined effect of the sinusoidal carrier and Gaussian can clearly be seen in these illustrations as the sine wave becomes spatially localised within the bounds set by the Gaussian.

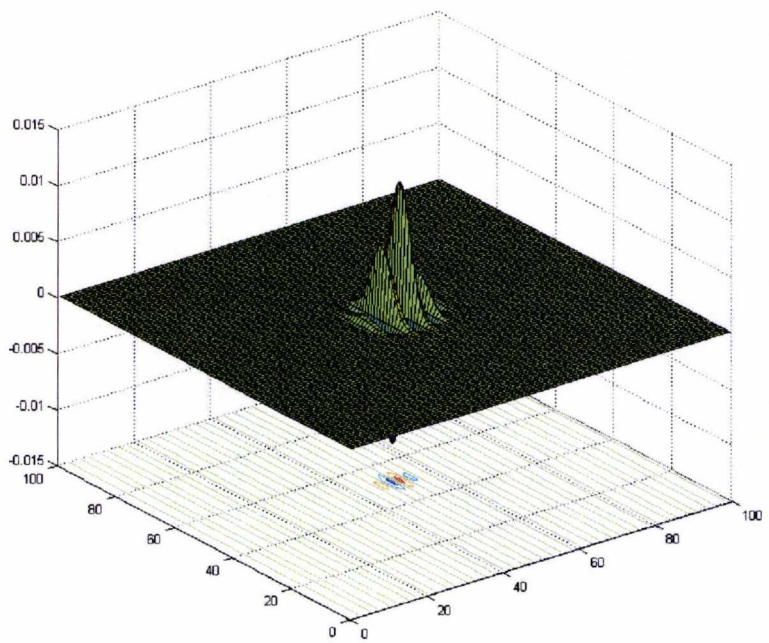


Figure 3.10 - Imaginary Component of a Gabor Filter

3.4.4 APPLIED EXAMPLE OF A GABOR FILTERED IMAGE

Take for example the filter from (3.7) :

$$f(x,y) = \frac{\alpha\beta}{\pi} e^{(-\alpha^2(x\cos\theta + y\sin\theta)^2 - \beta^2(x\sin\theta + y\cos\theta)^2 + 2j\pi f_0(x\cos\theta + y\sin\theta))} \quad \textbf{(3.8)}$$

Where α represents the frequency to width along the wave, β represents the frequency/width ratio orthogonal to the wave.

Taking $\theta = 0°$ at a frequency $f_0 = 0.02$ cycles/pixel (or $\omega = 50$ pixels).

Figure 3.11 shows a simple 64 x 64 monochromatic test image depicting a white square on a black background.

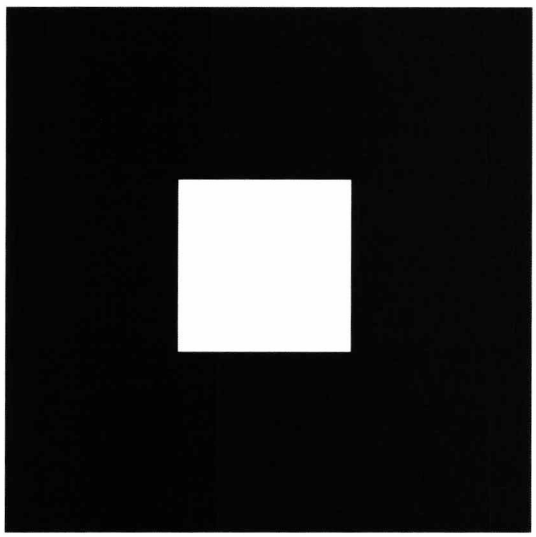


Figure 3.11 - Example Image before Processing

A Fast Fourier Transform (FFT) of the magnitude of the image in Figure 3.11 is represented in Figure 3.12.

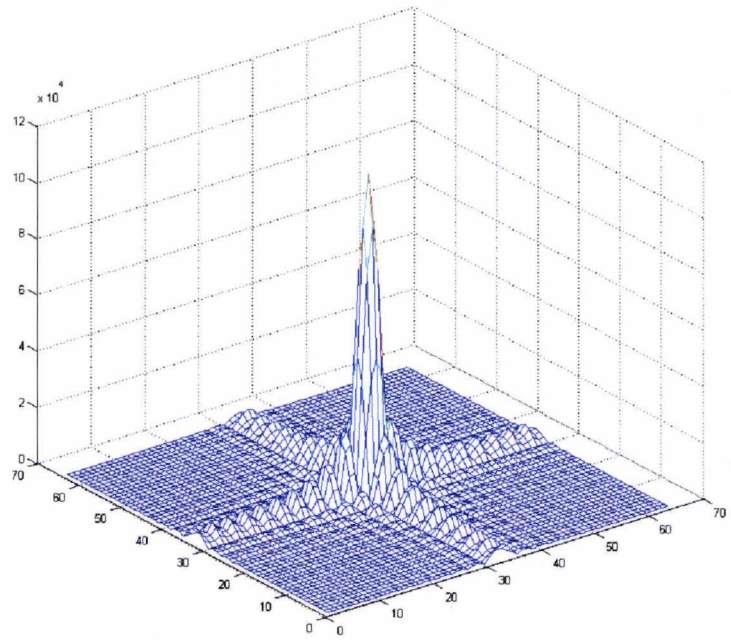


Figure 3.12 - FFT of Figure 3.11

Taking a sample filter from (3.8) with values $\theta = 0°$ and f_0 of 0.02 cycles/pixel with a width to orthogonal factor ($\alpha\beta$) of 0.1 would provide the Gabor filter depicted in Figure 3.13 and Figure 3.14.

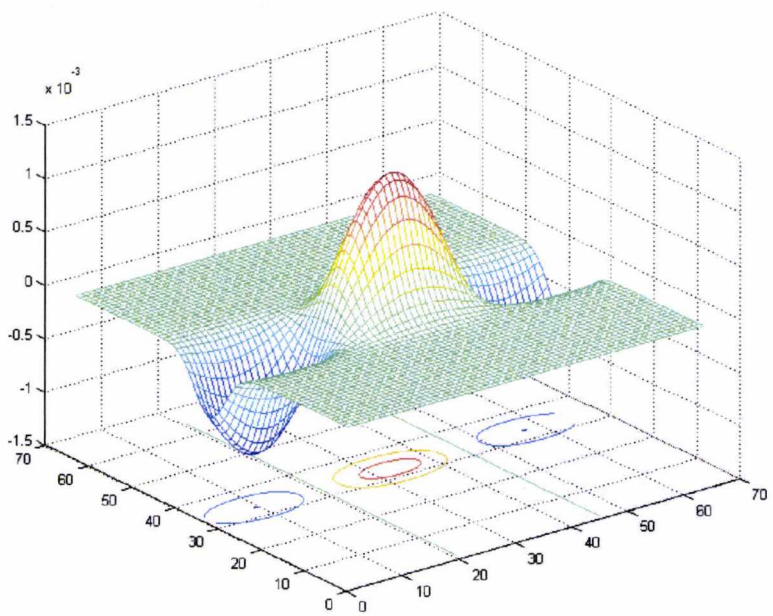


Figure 3.13 - 2 Dimensional Gabor Filter $f_0 = 0.02$, $\theta = 0$

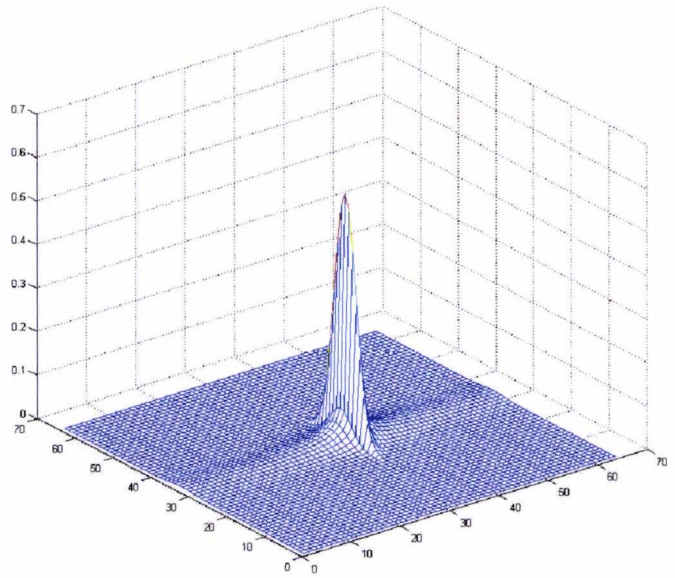


Figure 3.14 - FFT of a Gabor Filter (Magnitude)

The filter is then transformed to the frequency domain using an FFT which allows a rapid point by point multiplication of the images (thus providing a convolution of the spatial image and filter). For clarity in visualising the output on screen, the magnitude of the filter has been shown.

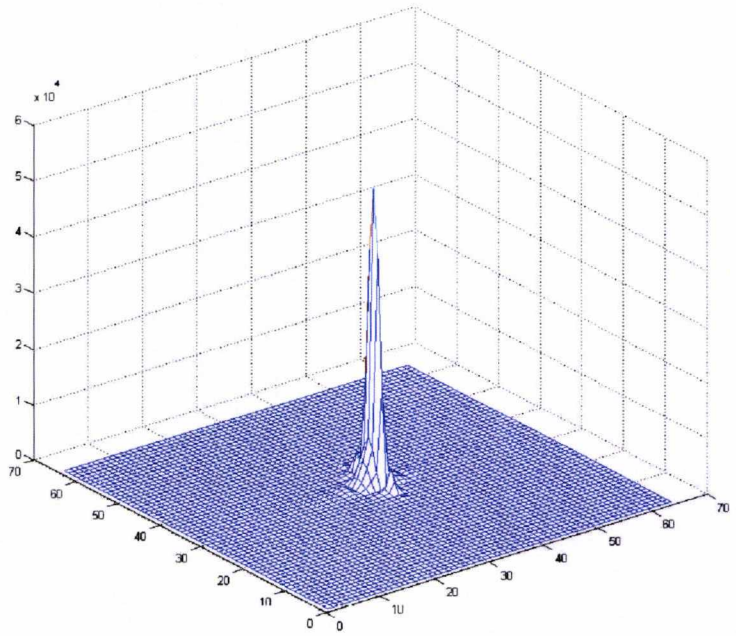


Figure 3.15 - Convolved Gabor Filter and Image (Magnitude)

Various functions can then be performed on the frequency domain representation of the filtered image to measure energy, standard deviation and any other such features which may be used as features for classification in image recognition.

To complete the above illustration, Figure 3.16 shows the filtered image output after being returned to the spatial domain and shown as a bitmap, this clearly shows the impact of the low frequency Gabor filter applied to the image with the resultant loss of boundaries.

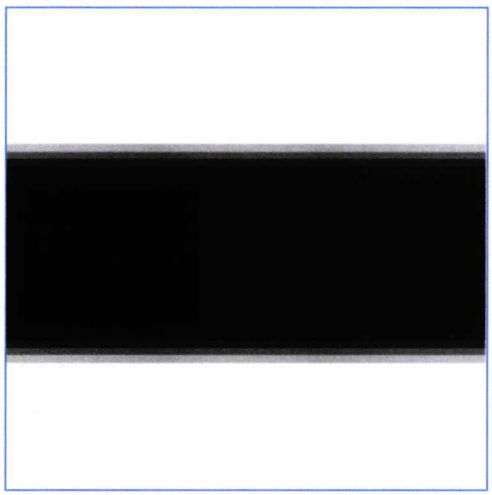


Figure 3.16 - Resultant Image after Filtering

In the second example, an image if vertical lines is taken. (64 Pixels, pattern frequency - 0.5 cycles per pixel). This is a high frequency pattern that should be difficult for a Gabor filter to isolate at the previous frequency settings.

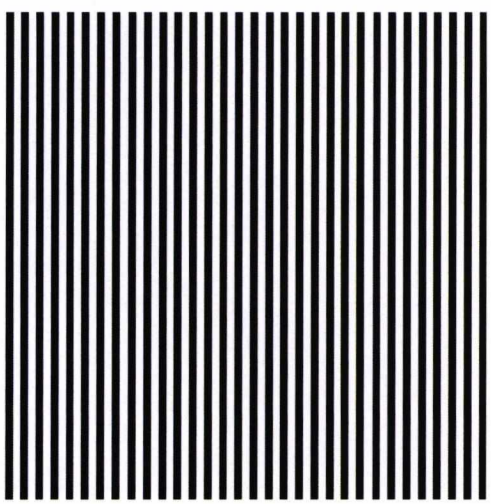


Figure 3.17 - Sample Image 2

Examining the FFT of this image provides the magnitude as shown in Figure 3.18. It is clear from this the difference in the frequency conversion of the diagram. Two spikes are evident representing the high frequency pattern observed in the source image along the horizontal axis.

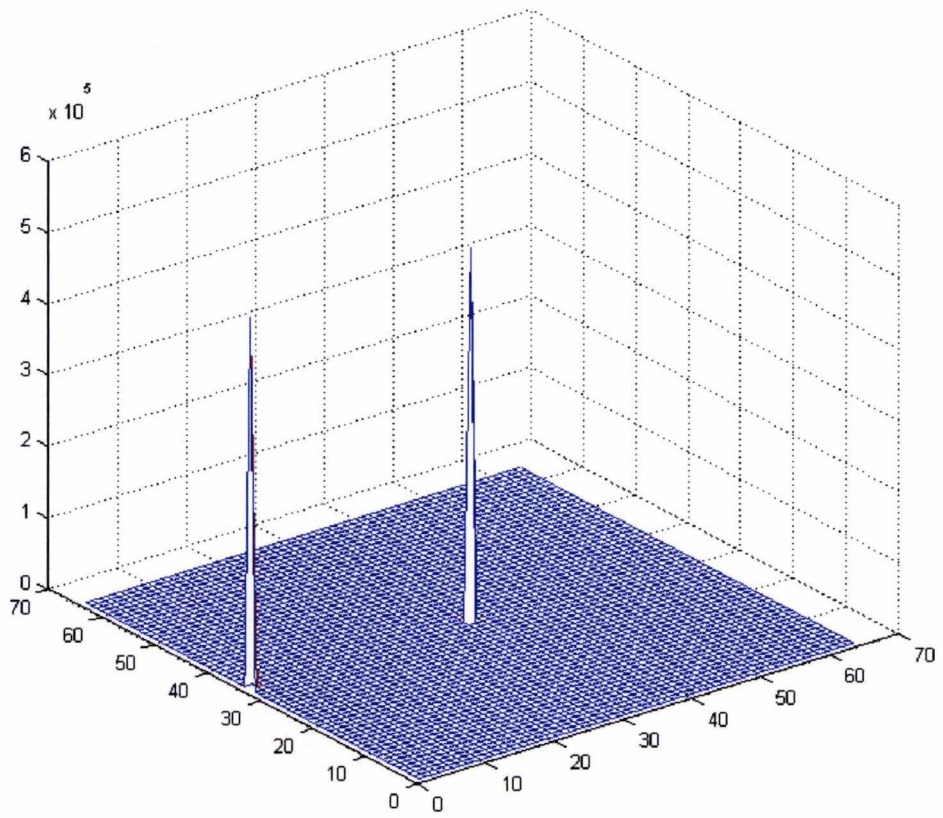


Figure 3.18 - FFT of Image 2

Convolving this image with the filter with F0 = 0.02 cycles per pixel (or 50 pixels per cycle) virtually eliminates all but the DC component of the filter as would be expected. The returned image (or lack thereof) is shown in Figure 3.20.

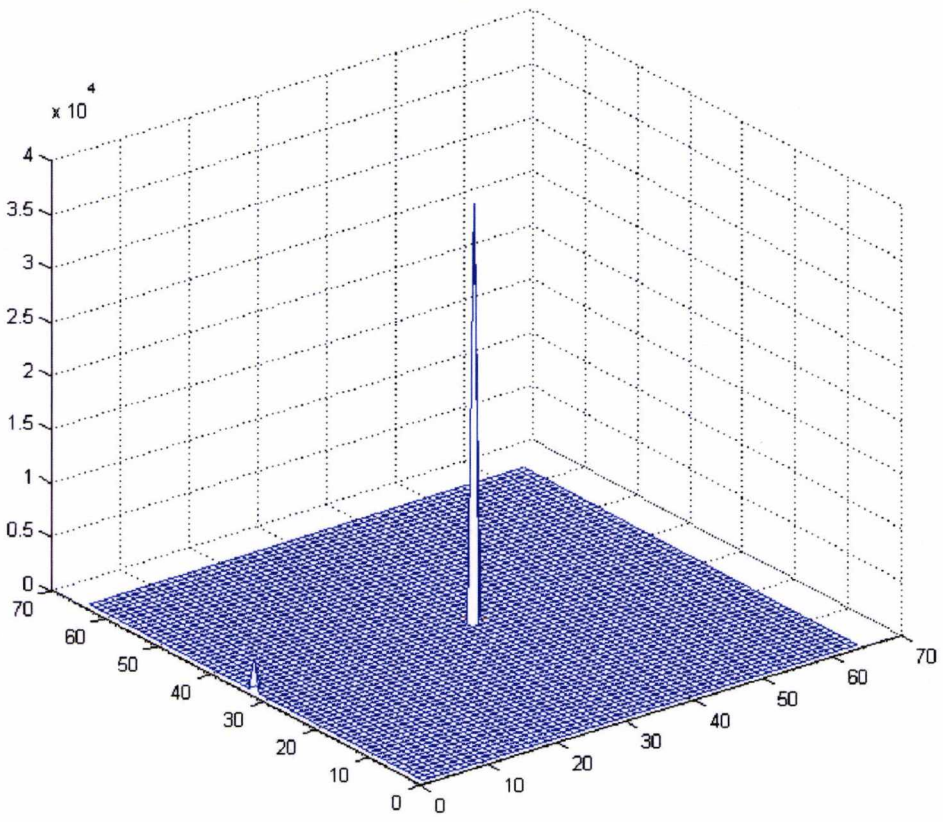


Figure 3.19 - Magnitude of Convolved Image

Figure 3.20 - Resultant Filtered Image (2)

3.5 CHAPTER REVIEW

This chapter has discussed how we perceive texture and an attempt to define and quantify how texture may be described by its properties. A review of recent activities in texture recognition is provided with descriptions of non-biometric applications to supplement the information provided in Chapter 2 . The frequency domain has been introduced with a discussion of its benefits in recognising texture.

The Gabor filter has been introduced in detail from both a historical and mathematical perspective. This has supported the findings of other researchers that Gabor filters provide significant advantage over spatial domain techniques in the filtering of the medium to low frequency range.

In conclusion, an applied practical example of the effect of Gabor filters on two test images has been shown.

4 EXPERIMENTAL FRAMEWORK

4.1 INTRODUCTION

The aim of this chapter is to present in a clear and concise fashion the configuration of equipments, tools and standards used in the preparation of results for this study. By defining these parameters for the entire study, consistency and controlled results may be achieved which facilitates the extension of this research in the same controlled environment at a future point in time.

The ability to reproduce test results via accurate configuration and documentation of test parameters is key to the acceptance and validity of test results. In order to conduct both a consistent and reproducible evaluation, this chapter has been written to provide the framework used to carry out all experimentation. The objective of this is that in future activities, the performance of both enhanced and/or alternative methods can be accurately compared with the methods presented in this document.

This chapter is divided into sections detailing specific elements of the experimental configuration such as hardware, software and policies maintained during the tests. At the beginning of each section the reader will find a brief statement which will where appropriate indicate the estimated impact that to the configuration may have on the results obtained. Obviously an exact reproduction of the test conditions will be difficult given the variances in both hardware and software available and the rate of change within the market, but by reading this chapter the reader should be able to ascertain which steps, datasets and conditions need to be reproduced accurately for comparative results to be valid.

4.2 SOFTWARE FRAMEWORK

Although unavailable at the start of research, all results have been re-executed with the versions of software listed in this section. One should note that although this research has been carried out on 64 bit capable hardware and software, the results obtained should be identical if performed on 32 bit hardware with a similar configuration. To validate this statement, some tests were reproduced on 32 bit hardware with identical results. It was noted however, that when dealing with extremely large image arrays, the 32 bit version of MatLab and Windows reported out of memory errors and failed to complete certain tasks.

Software Package	Version and Details
Operating System	Microsoft Windows Vista x64 Business
Development Environment	MatLab 2008a 64 bit edition
External Libraries	PRTool 4.1 (2009)

Table 4.1 - Software Framework

4.3 HARDWARE FRAMEWORK

The hardware used is offered as reference material but should not impact the results, only the time taken to obtain the results. Photographic hardware used to capture database images may have a significant impact if of substantially lower or higher quality. In chapter 1.1a)i)6the effects of image scale variance is assessed and this may be beneficial to review when selecting photographic equipments.

4.3.1 COMPUTER EQUIPMENT

Sufficient memory and CPU scalability (64 bit) appear to be required when dealing larger images in MatLab.

32 bit systems are limited to address 4Gb of memory (including paging further operations to disk) which can prove limiting. Most 64 bit systems are capable of addressing a minimum of 16Tb of memory.

Tests have shown that image sizes greater than 7 million pixels benefit significantly from more than 4Gb of available RAM with early results on 32 bit hardware frequently failing with out of memory errors.

Processor	Intel Core 2 Series, Q6600 Quad Core
Memory	8Gb Dual Channel DDR2 (1066MHz)
Hard Drive	700Gb RAID-0 Striped
Video Card	GeForce 8800GTX with 768Mb Dedicated DDR3
Visual Display	3840x1200 usable pixels (Dual 24" 1920x1200 Panels)

Table 4.2 - Computer Specification

4.3.2 PHOTOGRAPHIC EQUIPMENT

To generate the high resolution database the following equipment and configuration was used.

Equipment	Details
Camera	Nikon D50 6 MPixel Nikon D200 10.1 MPixel
Lens	Nikor 135mm Portrait
Flash	Nikon SB600 Speedlight
ISO Setting	ISO Auto (Max 400)
Quality Setting	RAW
Colour Mode	SRGB
Tripod	Manfrotto

Table 4.3 - Photographic Equipment used for HRDB Dataset

4.4 DATABASE FRAMEWORK

4.4.1 INTRODUCTION

This section presents the databases used in the research and justifies their selection. In order to give as broad an analysis as possible three types of database have been selected. These are described below.

At the commencement of this study, the availability of high resolution face databases was extremely limited. To rapidly acquire such a database at this period proved difficult and for this reason, the "High Resolution Database" or "HRDB" was produced. The later release of the "Biosecure Multimodal Database" or BMDB provided an excellent independent benchmark upon which to test the developed algorithms.

4.4.2 ALGORITHM DEVELOPMENT AND CUSTOMISATION

As mentioned in Chapter 3, in the feature extraction process, the frequencies isolated by Gabor filters are tuned to a specific orientation and frequency by the adjustment of variables

within the filter equation. The optimisation of these variables may be likened to 'tuning' the filter to provide a feature capable of producing optimum post-classification results.

Chapter 5 describes how the Gabor filter size is calculated to provide scale invariance based on the individuals eye separation value in pixels. It is believed that this measure will increase the ability of the system to cope with variations in input image resolution (indeed, this is evaluated in Chapter 7).

Due to the relatively late release of the BMDB database in this study and the assumption that the texture recognition rate would benefit to high resolution images, the HRDB was collected as described later in this section. As this dataset is purpose collected for the study, conducting an impartial assessment using it could risk to present biased results. Given also the nature of the algorithm in requiring some measure of "tuning", there is a risk for the algorithm to become tailored specifically to a small dataset thus showing results that may not be representative of testing on a larger database.

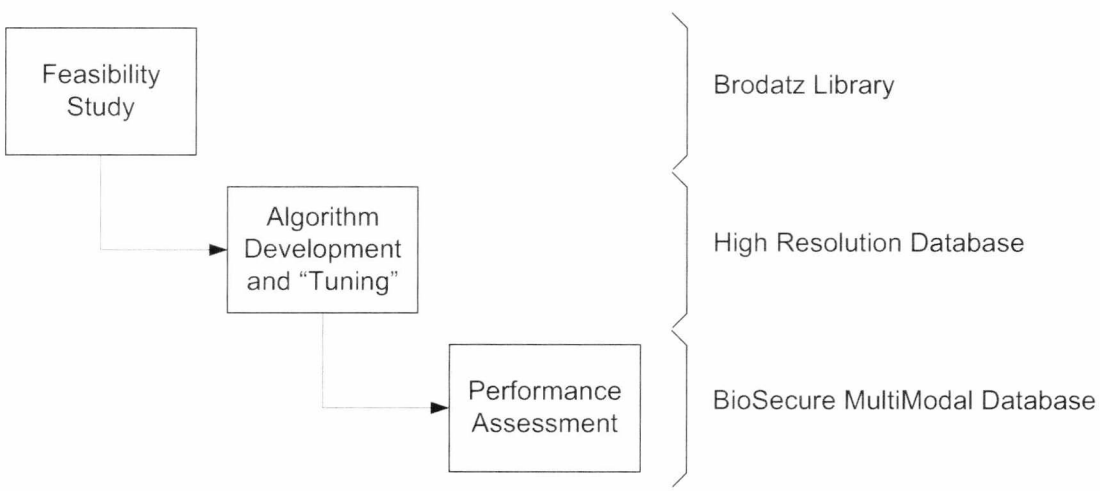


Figure 4.1 - Use of Image Libraries

For these reasons it was practical to develop and tune the algorithms using the HRDB database and perform testing using the BMDB database which became accessible in 2009. This also facilitates increased confidence in the performance of the system by using independently collected data in the evaluation.

4.4.3 THE BRODATZ ALBUM

4.4.3.1 INTRODUCTION

The Brodatz texture database [Brodatz 1966] is often used as a baseline for measuring texture recognition performance and was primarily used during initial algorithm testing and performance measurement before biometric databases became available. The Brodatz subset used included 111 greyscale source images.

This database was used initially in early investigations and algorithm development in the absence of a suitable biometric face database. When the HRDB (High Resolution Database) was produced as part of this research, the Brodatz Album was no longer used. It is mentioned in this section to provide a complete background into the development of the system presented in later chapters.

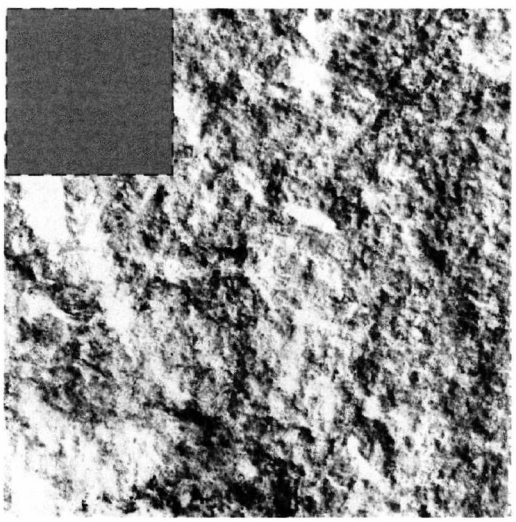


Figure 4.2 - Creation of Training Sets from Brodatz Database (1)

Three database sets were created with varying resolutions of 32x32, 64x64 and 128x128 pixel crops from the source images as shown in Figure 4.2. A horizontal and vertical overlap of 50% was permitted in the creation of these databases as shown in Figure 4.3.

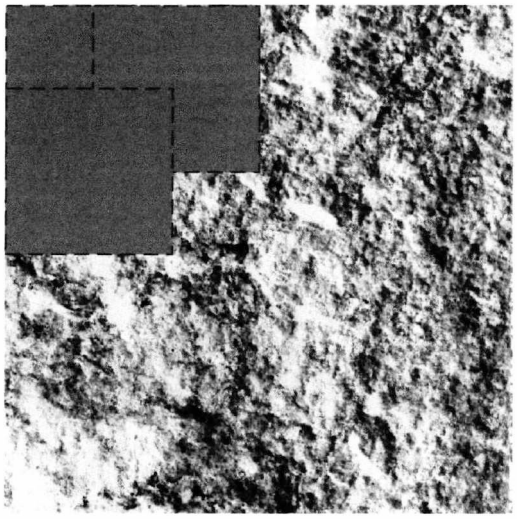


Figure 4.3 - Creation of Training Sets from Brodatz Database (2)

As the Brodatz database was only used before the availability of Biometric databases and in algorithm development, performance results have not been included in this thesis.

4.4.4 HIGH RESOLUTION DATABASE (HRDB)

4.4.4.1 INTRODUCTION

The High Resolution Database (HRDB) was created as part of this evaluation in response to an early requirement to design and develop algorithms with probe images of sufficient resolution to identify skin texture. At this stage of the research, there was no access to high resolution multi mega-pixel probe images. The HRDB database was created with the cooperation and support of the author's family members.

The high resolution database is a gallery of 14 subjects with a total of 26 probe images per subject. Each subject was photographed at one of two sittings. (9 subjects at sitting 1, 5 subjects at sitting 2). The apparatus used at each sitting differed hence the database was divided into two distinct datasets to enable fair usage.

4.4.4.2 DATABASE CREATION AND PROTOCOL

The HRDB was created using a Digital SLR camera (Section 4.3.2 using the protocol described below.

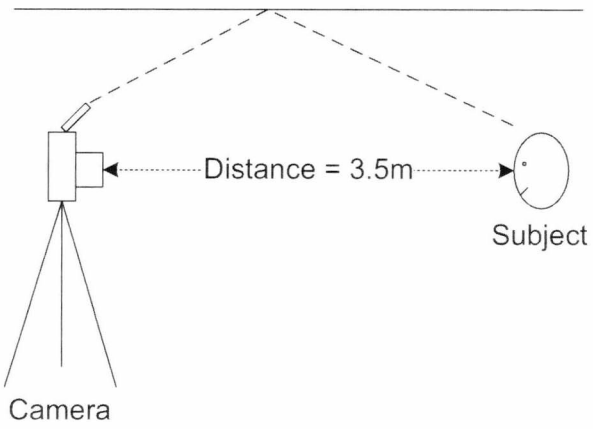


Figure 4.4 - HRDB Creation Process

The subject was asked to stand on a line measured 3.5m from the lens position with their body placed squarely facing the camera apparatus.

The subject was asked to vary their facial expression by naturally smiling or frowning as they would normally. The subject was instructed to maintain a constant head position facing the camera.

The photographs were taken indoors using natural lighting supplemented with a bounce flash at a distance of 3.5 metres from the subject. A total of 35 photos of each subject were taken in a single sitting from which were then evaluated for the following criteria:

1. Quality of photograph (Focus, Lighting)
2. Pose (Poses where the subject had inadvertently moved the angle of their head to the camera were discarded)
3. Expression (Any pose where the subject's expression had caused distortion in the upper cheeks were removed).

After this thinning process, 26 images of each subject remained which formed the database. The database was then split into two sets for classifier training and evaluation. The images were split using a simple odd/even assignment. (Odd numbered images assigned for training and even numbered assigned for testing).

Two datasets were created from two different subject pools. Dataset 1 comprised the authors children, parents, siblings and their family (9 subjects). Dataset 2 comprised the authors spouses siblings and their family (5 subjects).

4.4.4.3 LIMITATIONS OF THE HRDB DATABASE

The HRDB Dataset does suffer from some limitations which are:

1. There is no ethnic and limited genetic variation between the subjects.
2. There photographs were not obtained using any established biometric data collection protocols.

Even given these limitations, the HRDB database offers an advantage during evaluation. By developing the algorithm using the this database, the confidence of the evaluation using the established BMDB database may be increased. The feature extraction algorithms tuning parameters (frequency, orientation and scaling factors) were established using a separate database following different collection protocols. This ensures a degree of independence when evaluating performance using the BMDB database and indicates that the performance is not hinged on tuning to an evaluation database.

4.4.4.4 POST PROCESSING

The gallery has undergone post processing as explained below.

When a typical image is considered such as the one shown in Figure 4.5 it is useful to be able to isolate the subject's face from background distractions within the image. Particularly with reference to texture recognition, background can create undesirable features and hinder accurate recognition results.

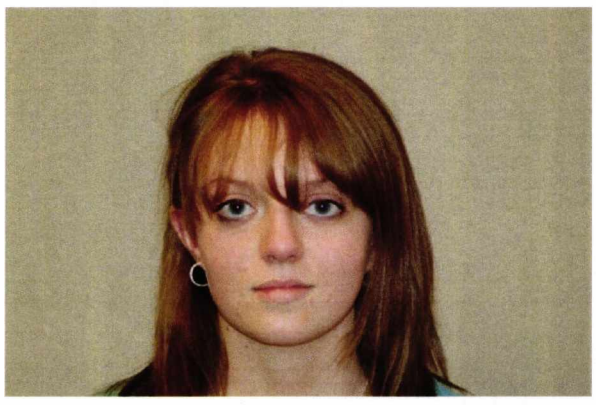


Figure 4.5 - Facial Image from High Resolution Dataset

A study in colour component selection for skin recognition showed that over 97% of skin samples could be accurately represented by a simple selection process in a specific colour space [Gomez 2002]. By evaluating different colour spaces, Gomez was able to identify

several colour components were suitable for skin detection. The process of detecting skin is based on Gomez's work and begins by examining the image in the HSV colour space and generating a skin "mask" of typical skin HSV values. This mask is then overlaid onto the source image and the image is then dilated to ensure that only large consistent areas of skin are flagged as skin surface.

Publicly available MatLab functions are available which implement this process using the Hue (base colour similarity), Saturation (colourfulness relative to its own brightness) and Value (brightness) colour model and enable the raid isolation of skin colouration from background noise. Figure 4.6 demonstrates the output of the previously shown image after passing a through a skin selection filter. This mechanism is used to pre-process all images utilised in this chapter.

Figure 4.6 - Skin Identification for Post Processing

It was decided to perform post processing for this database due to its use in the early stages of algorithm design. By removing background "noise" in this fashion, the images became simpler to manage and use in the preliminary stages of development.

4.4.4.5 DATABASE SPECIFICATIONS

Parameter	Value
Image Resolution	6 MP /10 MP (Median eye separation 440 pixels)
Colour Depth	24 Bits per Pixel
Colour Space	sRGB
Camera	Nikon D50 - Dataset 1/D200 - Dataset 2

Table 4.4 - HRDB Specifications

4.4.5 BMDB BioSecure Database

4.4.5.1 Introduction

The Multi-Scenario Multi-Environment BioSecure Multimodal database [Ortega-Garcia, Fierrez et al. 2009] is a new multimodal biometric database designed and acquired within the framework of the European BioSecure Network of Excellence. Acquisition of the data was achieved in two sessions over a balanced gender and age distribution of individuals. Data collected includes 2 Dimensional faces, Fingerprint, Handwriting, Iris, Signature and Speech samples and was collected with the collaboration of over 11 European institutions from more than 600 individuals.

4.4.5.2 Data Selection

A subset of the database of 79 individuals acquired from dataset 2 (DS2) of the database which was created under controlled conditions in an office environment. Specifically, the 2-D face images from DS2 are used in the performance assessment of the algorithms presented in this study.

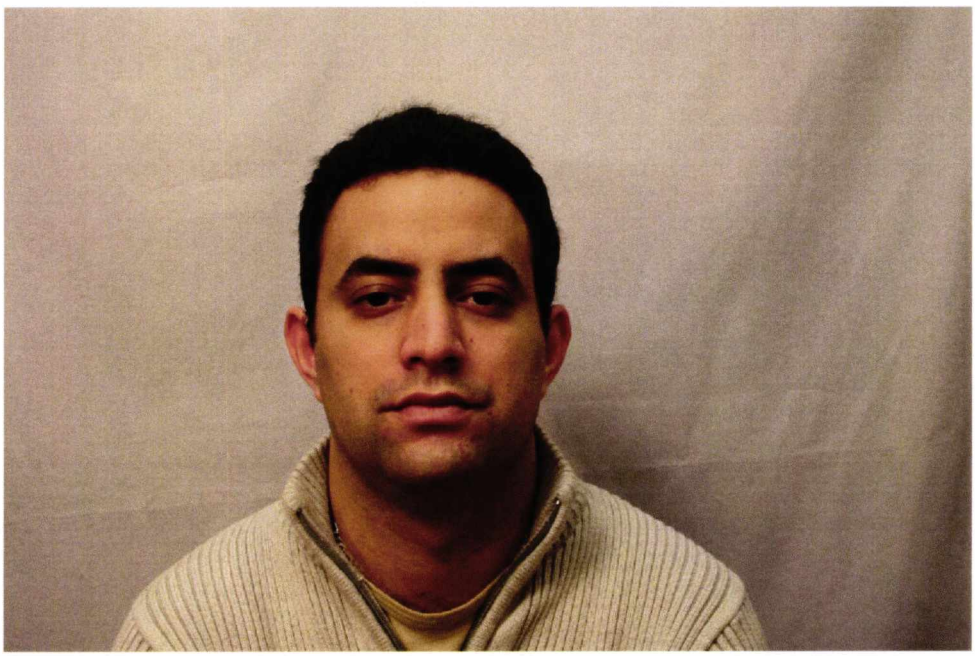


Figure 4.7 - Sample Image from BioSecure Database (From BioSecure Website)

DS2 2-D faces were acquired using a digital camera over two sittings. At each sitting a total of four images were acquired two of which were taken with a flash and two without. A total of two sittings per individual were held providing a total dataset of 8 images per person. This produces a dataset of 79 classes with a total of 632 probe images.

For legal reasons, samples from the BMDB BioSecure gallery cannot be provided in this document other than the image shown in Figure 4.7 (which was obtained from the sample images provided at http://biosecure.it-sudparis.eu/AB/).

Typical eye separation distance in this database ranges from 330 - 486 pixels.

4.4.5.3 TRAINING AND PROBE GALLERIES

Given the relatively small number of probe images per class, data has been evenly distributed to the training and probe galleries providing four images per class for each purpose. This also ensures that classifiers are tested on different probes than images used during the training phase.

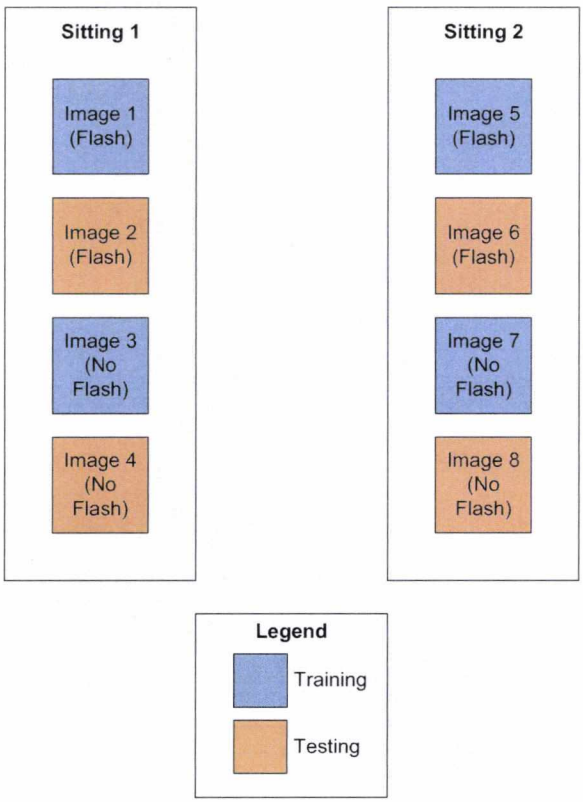


Figure 4.8 - BioSecure DS2 Subset Training and Testing Assignment Process

The classes are distributed with every other image going to the training set with the remaining probe images held for testing. This ensures that an even distribution of the flash illuminated photos and each sitting are present in both the training and testing galleries.

4.4.5.4 DATABASE SPECIFICATIONS

Parameter	Value
Image Resolution	7.8 MP (Median Eye Separation 408 pixels)
Colour Depth	24 Bits per Pixel
Colour Space	sRGB
Camera	Canon EOS 30D

Table 4.5 - BMDB Specifications

4.4.6 DATABASE COMPARISON AND SELECTION SETS

Three unique databases have been described in the preceding sections and are summarised in Table 4.6.

Database	Description	Purpose in Evaluation
Brodatz	Generic Published Texture Database	Initial Investigations
HRDB	Controlled Unpublished Face Database	Feature Extraction Algorithm Development
BMDB	Controlled Published Face Database	System Performance Testing

Table 4.6 - Summary of Image Databases

The scope of each database is established at this stage. It is generally not considered good practice to design image recognition algorithms for the same image gallery that will be used in performance assessment as this can lead to biased results. For this reason, all algorithm development was performed on the HRDB database as a development database while the performance analysis was performed on the published (and hence comparable) BMDB database.

To summarise, the test results in chapters 5 - 7 were performed on a unique dataset which was not used during the development of the feature extraction process. By using one database for the systems developmental testing and training, and a separate distinct database for the performance evaluation the impartiality of the system to the test data can be assured.

4.5 CLASSIFIER FRAMEWORK

4.5.1 INTRODUCTION

The process by which extracted features becomes understandable information is known as classification as shown in Figure 4.9.

Figure 4.9 - Classification Process

This section provides an introduction into classifier types, their benefits and limitations and classifier terminology. The role of a classifier is to assign raw feature vectors into understandable classes. Typically, a trained classifier will have a list of possible outcome assignments known as classes. Several variants, such as whether or not one knows how feature data will become distributed to classes (ie. class probabilities) may be an important consideration in classifier selection.

There are several forms of classifiers currently in use in some variant form. Typical examples of classifier types are:

- Linear
- Quadratic
- K Nearest Neighbour
- Boosting
- Decision Tree
- Neural Networks
- Bayesian Networks
- Hidden Markov Models

The objective of the research in this thesis is heavily weighted in the direction of proof of concept and establishing a suitable feature set for the classification of individuals. For this

reason classifier design carries somewhat less importance - it is required, but not a focal point of this stage of research.

Given this objective, three commonly used classifiers are used for all tests to provide a base for comparison. A description of each classifier, its parametric configuration and the reason for its choice are provided in this section. The selection of three distinct classifier types to ensure that the evidence presented maintains a degree of classifier independence and impartiality. By testing the proposed algorithms and techniques with a variety of classifiers, each with its own level of computational requirements a fair assessment may be made of the potential accuracy of the system as well as an estimate of the associated cost in terms of computational power associated with each classifier.

All classifiers described here have been practically implemented using the PRTools framework as presented in Section 4.2 .

4.5.2 Note on Classifier Tuning

The aim of the study is to identify if there exists any potential for facial skin texture as a biometric feature suitable for identification. The classifiers presented have been chosen to allow a general estimate of the systems performance and are likely operating at sub optimal levels for the data being analysed. In essence, classification is treated to a large extent as an "off the shelf" solution based on PRTools and Matlab. Further analysis remains to be done on the impact of fine tuning the classifiers.

Later chapters in the thesis comment on the overall statistical confidence of the system and explain that for any increase in accuracy to be substantiated, the tests should be performed on a much larger dataset. For this reason there is little to be gained by attempting to improve the classifiers performance on the image databases used in this study.

4.5.3 Back Propagation Feed Forward Neural Network

Artificial Neural Networks are a commonly used classifier in pattern recognition which have been used in tumour detection [G.S.Cox 1992], texture classification [Tang, Tian et al. 2003] and adult image identification [Lee, Kuo et al. 2006] to name but a few.

In an attempt at modelling a simplistic version of the biological brain, a neural network is composed of a series of interconnected neurons. Although many times less intricate, the fundamental principles of operation are similar to a biological neural network.

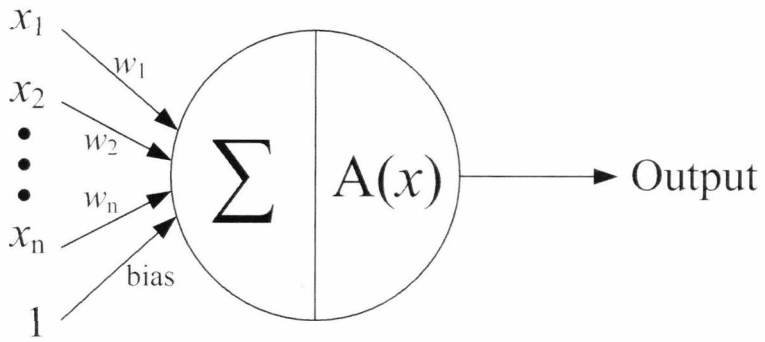


Figure 4.10 - Perceptron with Activation Function (A)

The feed forward neural network is a multi-layer perceptron network with no feedback loops [Bishop 1995], [Hellstrom and Kanal 1989] where a single perceptron takes the form similar shown in Figure 4.10 and Equation 4.1:

$$O = \sum_{i=1}^{d} x_i w_i + w_n \tag{4.1}$$

Where O represents the perceptron summation operation, x is a supplied input feature, w is the weight and w_n represents the threshold or bias. The perceptron takes a series of inputs and multiples each of these by their respective weights (this is initially seeded to a random value between 0 and 1).

$$A(x) = \frac{1}{1 + e^{-x}} \tag{4.2}$$

The perceptron's output is passed to an activation function which formulates the output as a result of defined function. In a back propagation neural network, this output is typically a sigmoid function as shown in (4.2). This assists the system in tuning the biases and weights by establishing the minimum error from the deriving the gradient of an error function.

Back propagation refers to the networks ability to self-tune under supervised learning in order to adjust weights throughout the network based on output error deltas. The method was first proposed by [Polyak 1964] and subsequently implemented by several authors. [Rumelhart, Hinton et al. 1986] provides a noteworthy study on the topic.

Figure 4.11 demonstrates an example neural network. The network shown has an input and output layer plus an internal hidden layer. In the displayed figure, there are three input features, 4 hidden nodes and 5 output nodes that may be used to provide class assignments. Although some studies have attempted to generate algorithms for the automatic determination of optimal hidden nodes [Lam 1995] there is little evidence to support the use of these algorithms in large dimensional pattern recognition systems.

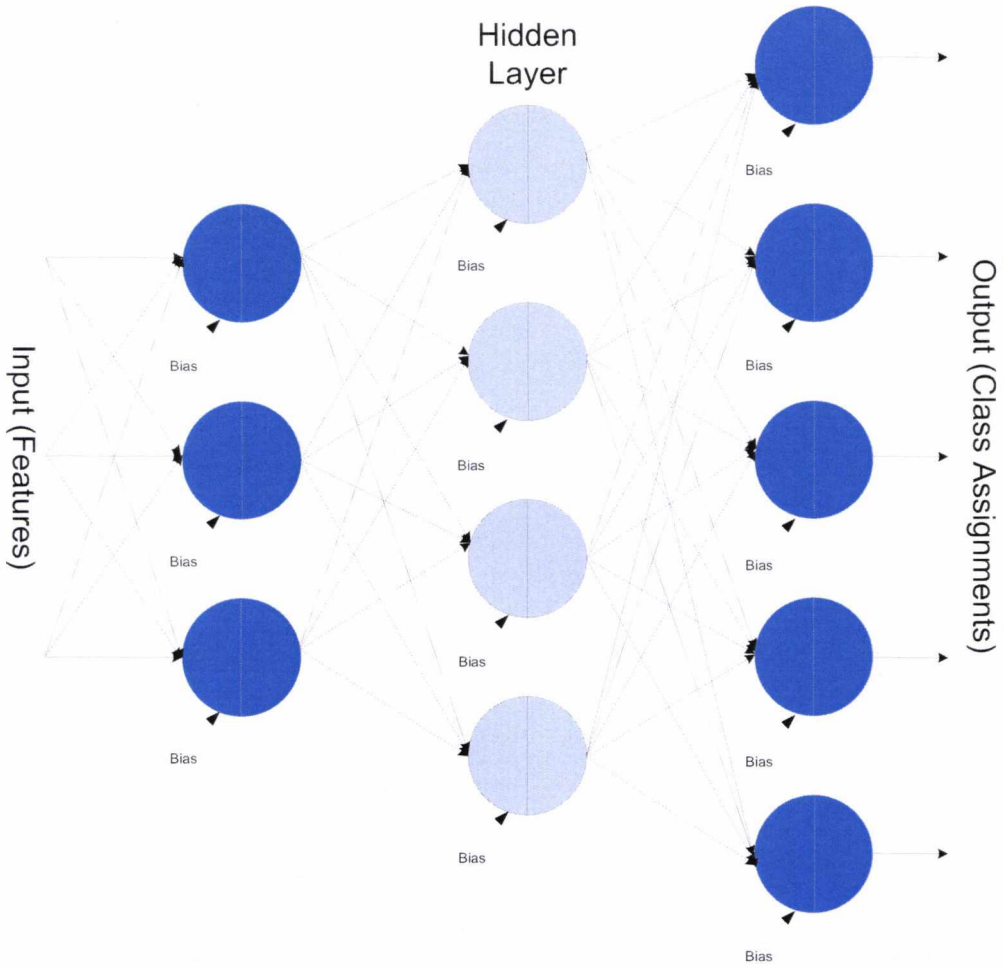


Figure 4.11 - Example Neural Network (Weights not shown)

The back propagation feed forward network is used as the baseline neural network classifier in this study with 3 layers (input, hidden and output). The hidden layer contains 48 nodes (twice that of a typical 24-dimensional input vector as suggested by [Bishop 1995]).

Initial practical experimentation of 24 hidden and 96 hidden node systems showed that with 24 hidden nodes, the system struggled to converge particularly on larger input vectors whereas a 96 hidden node network showed little tangible benefit other than substantially increased training times.

The PRTools `bpxnc` function is used to implement the neural networks used for testing in this study.

Due to the time taken to successfully train a neural network of this complexity being measured in half hour periods on the hardware specified earlier in this section, neural networks are rated as costly in terms computational power.

4.5.4 BAYESIAN LINEAR DISCRIMINANT CLASSIFIER

A Bayesian classifier [Minksy 1961] is a simple classifier which uses independent feature matching to determine class probabilities [Duda, Hart et al. 2001]. Put simply, a car may be recognised as a car if it has four wheels, a steering wheel, doors and an engine. Although many of these features may be shared with other objects, each feature is independently assessed without bias to other potential matches.

Linear discriminant classifiers are formed by searching for linear functions which provide the best separation between the output classes by assuming normal densities with equal covariance matrices. This is an attempt to reduce to the feature dimensionality in a manner similar to principal component analysis (PCA).

A classifier is said to be linear if the function of the feature vectors describing the class is linear. Equation (4.3) shows the mapping relationship of a linear classifier taking $\vec{w}$ as the weight vector learnt from a set of labelled training examples and $\vec{x}$ as the input feature vector. f is hence the function that either provides a direct linear mapping outcome based on the input vector, or the probability of an input vector belonging to a specific class.

$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j (w_j x_j)\right) \tag{4.3}$$

Despite its name, linear discriminant analysis belongs to the generative rather than the discriminative classifier family .Classifiers are typically classed as either generative models

or discriminative models. A generative linear classifier operates by modelling the conditional probability of an outcome. Examples of generative linear classifiers are Linear Discriminant Analysis and the Naive Bayes Classifier. [Andrew Y. Ng 2002]

Discriminative classifiers differ in their attempt to maximise the quality of the output on training data [Bouchard and Triggs 2004]. [Andrew Y. Ng 2002] demonstrate that discriminative classifiers generally result in a lower asymptotic error than generative classifiers. In contrast, generative classifiers tend to reach their asymptotic error more quickly. Examples of commonly used discriminative classifiers include the Logistic Regression classifier, Perceptron classifier and Support Vector Machines.

Linear discriminant analysis has been selected as an example of how a generative classifier would handle the skin texture features and its rate of learning.

4.5.5 SUPPORT VECTOR MACHINE

Support Vector Machines [Cortes and Vapnik 1995] operate by constructing multi dimensional hyper planes in the feature space to separate data into two or more categories. It is named as such as the closest feature vectors to the dividing hyper-plane are termed supporting vectors, the hyper plane is known to be optimal when the distance (d) between the supporting vectors is its highest value while maintaining class separation.

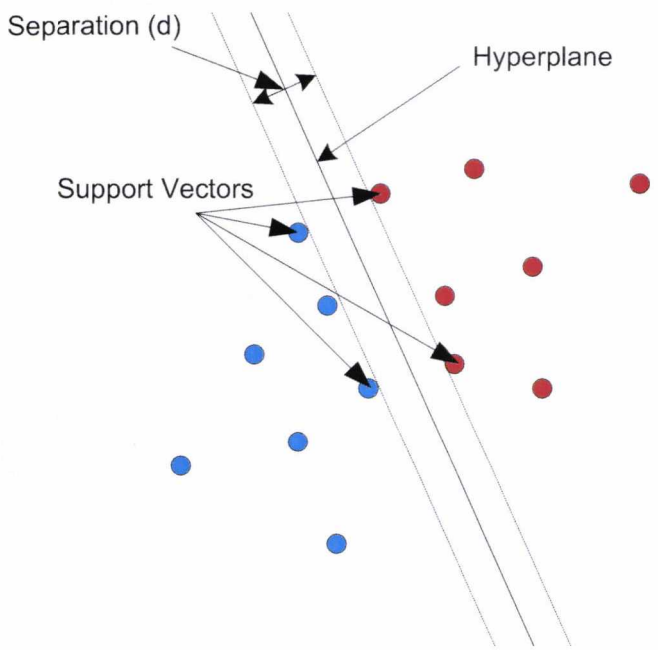


Figure 4.12 - Simplified Support Vector Machine

As shown in Figure 4.12, the support vector machine uses linear hyperplanes to provide feature separation. However, in many circumstances, the feature space is not divisible by straight lines and may seem to require non-linear curves. In these circumstances, support vector machines use "kernels" to map non-linearly separated data to a different feature space where a linear hyperplane may be used. In some instances this involves transforming to a higher dimensional feature space. This process is known as kernel mapping and allows SVM's to work with complicated feature sets whose boundaries may be difficult for other classifiers to determine.

Kernel	**Equation**
Linear	$k(x_i, x_j) = x_i^T x_j$
Polynomial	$k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma \geq 0$
Radial Basis Function (RBF)	$k(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}, \gamma \geq 0$
Sigmoid	$k(x_i, x_j) = \tanh(x_i^T x_j + r)$

Table 4.7 - Support Vector Machine Kernels

A variety of kernels are used in support vector machines, the most popular of which are detailed in Table 4.7. [Khoo, Ong et al. 2008; Rokita, Krzyzak et al. 2008]

The support vector machine used in this study uses the computationally simplest form of kernel - the linear kernel.

Support Vector Machines are one of the newer classifiers that have attracted much interest by scholars and researchers. In 2003 [Meyer, Leisch et al. 2003] a review found them to have "good" overall performance but not "superior" for all datasets.

Support Vector Machines using a linear kernel generally train much faster than a typical neural network but may not always attain the same or greater performance.

4.5.6 CLASSIFIER SUMMARY

Three classifiers have been presented in this section with varying computational requirements:

1. Back Propagation Feed Forward Neural Network (3 Layer, 48 Hidden Nodes). Relative computational cost: Very High
2. Baysean Linear Discriminant Classifier. Relative computational cost: Negligible

3. Support Vector Machine using Linear Kernel. Relative computational cost: Low.

The selection of these classifiers enables a broad assessment of the systems capabilities without bias to a particular classifier and should provide a suitable performance range for an impartial assessment of the extracted features suitability for purpose.

4.6 Evaluation Framework

4.6.1 Introduction

The evaluation framework describes the procedures, approaches and controls that were used during the research to ensure that results are consistent, impartial and presented in a fashion which promotes comparison with other techniques and is recognised and accepted within the field of study.

4.6.2 Systematic Approach

The approaches used to testing and reporting in this thesis are derived from a National Physics Laboratory (NPL) report [Mansfield and Wayman 2002] which propose best practices, terminology and recommended reporting methods.

In addition to best practices, [Mansfield and Wayman 2002] also define several terms in an attempt to minimise the ambiguity that can be present in biometric terminology. In such instances, the terminology of this thesis is aligned with the definitions and procedures described in the aforementioned NPL report.

All methods of recording results in this study are threshold independent. This facilitates a comparison against various other works by allowing the reader to decide whether to bias recognition rates towards false accepts, false rejects or a balance of the two.

4.6.3 Evaluation Type

This evaluation fits the description of a technology evaluation [Mansfield and Wayman 2002]. A technology evaluation aims to compare algorithms using a single dataset or set of databases. Performance against these databases is based on the environment and method of which it is collected and as such a database may be selected that is appropriate for the algorithm.

The terms false match rate (FMR) and false non-match rate (FNMR) are used in this document and are interoperable with the terms false positive and false negative for classifier outputs.

Results are summarised in three principal ways throughout this research:

1. Receiver Operating Characteristic (ROC) Curve and Area Under Curve (AUC).
2. Detection Error Trade-off (DET) Curve.
3. Equal Error Rate (EER) (also known as Crossover Error Rate - CER)

4.6.4 THE ROC CURVE

The receiver operating characteristic (ROC) curve [Metz 1978] is an accepted method for presenting the performance of a pattern matching classification system [Mansfield and Wayman 2002]. ROC curves are advantageous in that they are threshold independent thus allowing an evaluation across all thresholds.

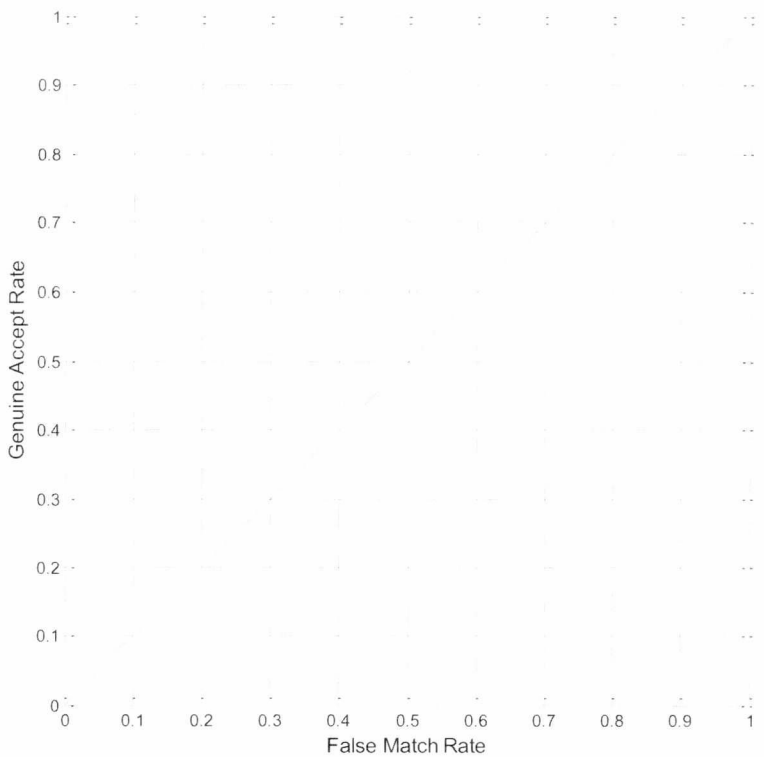


Figure 4.13 - ROC Curve showing Random Class Assignments

In a ROC curve false matches (x-axis) are plotted against genuine accepts (y-axis) as a parametric function of the decision threshold. In a ROC curve, the line defined by y=x (Figure 4.13) represents truly random performance with y=1 representing perfect performance.

(Figure 4.14) shows an example ROC curve taken from the results shown in later chapters. In this plot, the red line shows a higher number of genuine accepts at the same false match rate as the blue line and thus indicates superior performance. As a general rule, the tighter the curve fits to the upper left quadrant (y=1), the better the systems performance.

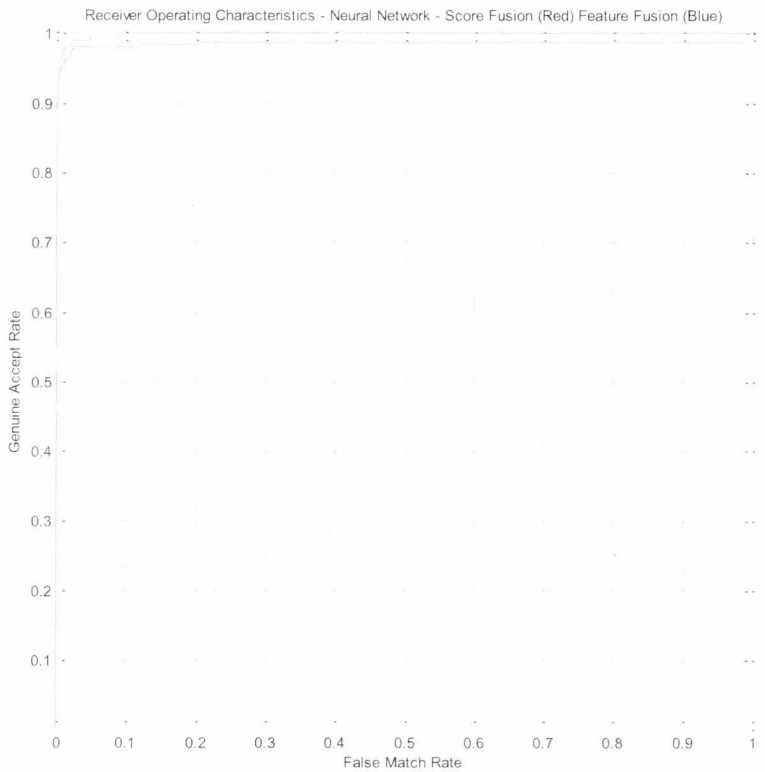


Figure 4.14 - Example ROC Curve

Often it is desirable to have a single quantative numerical representation of a systems performance which can readily be compared to other systems. By assessing the integral of the ROC curve, the aggregate performance of the system is provided in a single value with 1 indicating the system has a 100% chance to correctly assign a class and 0.5 (50% chance of correctly assigning a class) indicating random performance (as provided by the line y=x). This value is called the AUC (Area Under Curve) [Hanley and McNeil 1982], [Hanley and McNeil 1983].

4.6.5 THE DET CURVE

The Detection Error Trade-off (DET) curve was first presented in [Martin, Doddington et al. 1997] to measure speech processing performance and is recommended in [Mansfield and Wayman 2002] report for analysing the performance of biometric systems. It is based on a modified ROC curve which plots error rates on both axis (False Non Match Rate - x axis, False Match Rate - y axis) as the threshold is varied. The inclusion of false non match rate data provides an additional performance statistic for analysis and gives uniform treatment to each type of error. By plotting the results logarithmically, the plot is spread out and can be easier to distinguish between well performing systems.

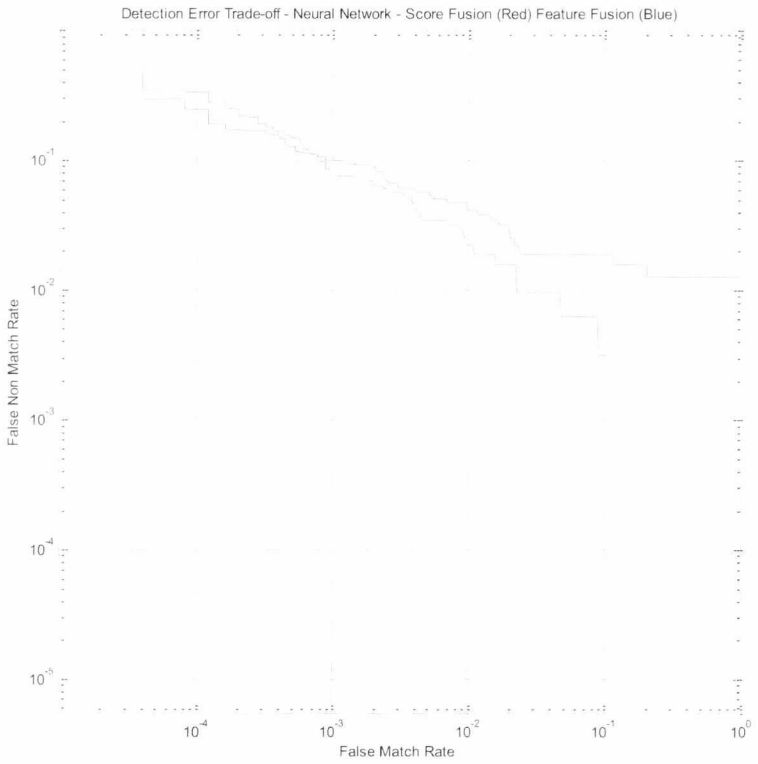


Figure 4.15 - Example DET Curve

The curves in Figure 4.14 and Figure 4.15 are constructed from the same data.

A further observation that can instantly be read from a DET curve is the shape of the curve. A straight line indicates that the underlying class distributions are normal. A curved line indicates an abnormality in these distributions which can signify unreliable or poor classifier performance.

4.6.6 Area Under Curve (AUC) and Equal Error Rate (EER)

The AUC and EER are two values which provide a representation of a classifiers performance. The Area Under Curve (AUC) is based on the integral of the Receiver Operating Characteristic curve as discussed in Section 4.6.4 . The AUC is equal to the probability that classifier output will correctly assign an output to its correct class as opposed to a random erroneous class. [Hanley and McNeil 1982]. For ease of reading, all AUC values supplied in this thesis have been provided in a percentage scale.

The Equal Error Rate (EER) is the error rate at the point at which the false match rate is equal to the false non-match rate. This is found programmatically by differencing the false match and false non-match values, finding the lowest value and correlating the position of this value to the mean of the false match and false non-match rates. Its use is discussed in [Li and Porter 1988].

4.6.7 Fusion Terminology

In the evaluation chapters of this thesis, the following references are used to define permutations of score and feature fusion. These are:

SFSB: Single Feature Type, Single Biometric Region

SFMB: Single Feature Type, Multi Biometric Region

MFSB: Multi Feature Type, Single Biometric Region

MFMB: Multi Feature Type, Multi Biometric Region

While each acronym is explained as it is first used, a summary list here may provide assistance in rapidly locating the evaluation results.

4.7 Chapter Review

This chapter has presented the framework upon which the research is evaluated. The hardware and software platforms allow the accurate reproduction of the results presented in the following sections of this Thesis. The database evaluation assists the reader to understand upon what probes the performance claims have been made and repeat the evaluation for an evolved or competing algorithm.

It has also been stated that the database used in the development of the algorithm is independent from the database used to assess the performance. As such the algorithm has had no specific tuning for the evaluation database and it can be said that the evaluation uses a neutral and fair dataset.

The evaluation framework defines the standards by which the study has been performed and attempts to put in place definitions of terms which may otherwise be ambiguous or open to interpretation as well as describe the presentation format of the results.

With these topics presented it is hoped that the reader will gain a clear understanding of both how the results have been obtained, and a strong basis for comparison.

5 THE TEXTURE OF SKIN

5.1 INTRODUCTION

Following the discussion of texture from Section 3.2 one may make the assumption that the texture is a feature of all visible surfaces and images. If this is true then by inversion one may conclude that all surfaces must contain a property, or series of properties which describe its texture [Zhu, Guo et al. 2005]. These properties are known in the image processing field as image features.

To better visualise this relationship, if the feature space of a surface S is defined by its textural features t then the following is suggested:

$$S = \{t_1 + t_2 + \cdots + t_{n-1} + t_n\} \tag{5.1}$$

Where each t represents a single feature based on the properties of surface S.

This chapter is intended to address two key questions about t. These are:

1. If S represents a region of skin texture from a probe image facial region, does there exist a sufficiently unique dimensionality of features t to permit the autonomous identification of the individual by a classifier?
2. If (1) above is true, how should t be constructed?

In answering these two questions, this chapter will also cover the following topics of interest:

- Preparing an image for feature extraction.
- Locating regions of interest.
- Scale invariance by dynamic frequency selection.

5.2 THE HUMAN FACE IN TWO DIMENSIONS (SCALE)

Human beings have a natural ability to acquire and process images in three dimensions. The co-location of our eyes side by side on the front of our heads allows us to see the same image from two slightly different angles. By alternately closing our eyes, it is easy to observe the slightly different images perceived by each eye. This is known as stereoscopic vision.

Stereoscopic vision gives a natural advantage with depth perception and leads to a natural degree of scale invariance when viewing objects and textures.

This may be simplified by sketching an isosceles triangle of dimensions a, b, c and angles A, B, C (where A=B and a=b) and dividing it in two along the longest axis creating two right angled triangles where d is the distance from the object.

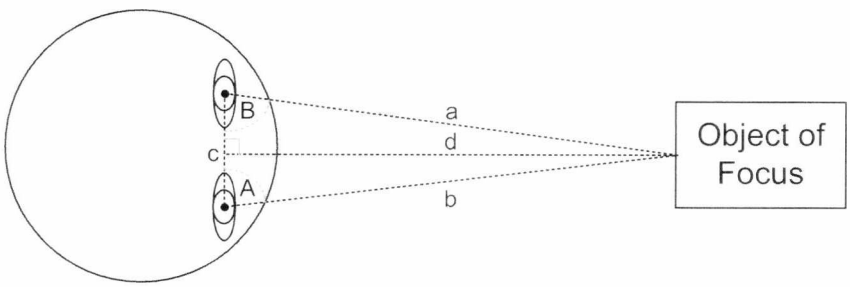


Figure 5.1 - Simplified Human Depth Perception

From the above, we make the following assumptions:

- Distance c is known and constant.
- Angle B = Angle A and is known for the object in focus.

(Note that if the eyes are positioned such that Angle B is not equal to Angle A, the angles are still known as a basis for calculating distance d).

Using an application of the tangent rule, we know that the angle of interest is B and that distance d becomes the opposite side. The adjacent side is formed by halving distance c.

$$d = \frac{c \tan B}{2} \tag{5.2}$$

Thus from (5.2) we conclude that using the parameters known to the human brain, it is possible to use two points of view of the same object of focus to generate a perception of depth, or distance from a point on an image. This not only allows a simple estimate of distance to be performed, but allows the brain to perceive depth and proportional distances between surface objects as a feature.

While replicating this ability is technically possible, the requirements of doing so are computationally intensive. The equipment to obtain this level of information is also not yet commonplace outside of research laboratories. For these reasons, it was decided to pursue the study of skin texture using much more readily available and easy to acquire two dimensional images obtained from a single point of view providing direct frontal face images.

Chapter 0of this document explains that although two dimensional, the image galleries used in this study were obtained in relatively controlled conditions and deliberately include variances in lighting and to a lesser extent scale.

Scaling is an important factor in texture recognition, more so when dealing with two dimensional images which do not contain relative depth information. The first challenge in the study was to ensure that the features extracted were scale invariant this ensuring that the sampling region of interest was not impacted by natural variance in the scale of the probe (test) image in as much as is possible. This is discussed in the following sections.

5.3 TEXTURE AS A FEATURE

As discussed in Chapters 2 and 3 3 it has for some time been commonly accepted that skin patterns such as fingerprints are distinctive to a person. Fingerprints are measured in several ways but generally accepted within the forensic community when a given number of geometric shapes can be matched to an existing record.

In contrast, this study focuses on the texture of skin located on the human face. Specifically, areas below the left and right eye in the upper cheek region as these are generally unobstructed by facial hair or decorations. The use of this area of skin surface for subject identification is an relatively new area of research and its suitability for purpose has yet to be established.

Past research has shown that skin layers may differ in thickness, pigmentation, collagen fibres, protein density and capillary beds. In 1989, a US Patent (4805223) was issued on a method which successfully used a light source to illuminate a small fixed patch of skin with

visible and near infrared light, then measured the light scattered by the skin with a spectroscope [Denyer 1989]. This resulted in a distinct optical pattern for the test subject.

5.4 IMAGE PREPARATION AND PRE-PROCESSING

In order to provide a point of reference for extracting texture regions of interest, a function was written to accept the manual input of eye locations for each image. While there are ways to autonomously acquire the position of a subjects eyes the prime focus of this research is on the performance of texture based identification. For this reason, all other autonomous image processing is performed manually and validated to ensure that any errors can be correctly attributed to the texture based identification process and not to errors introduced during image acquisition or pre-processing.

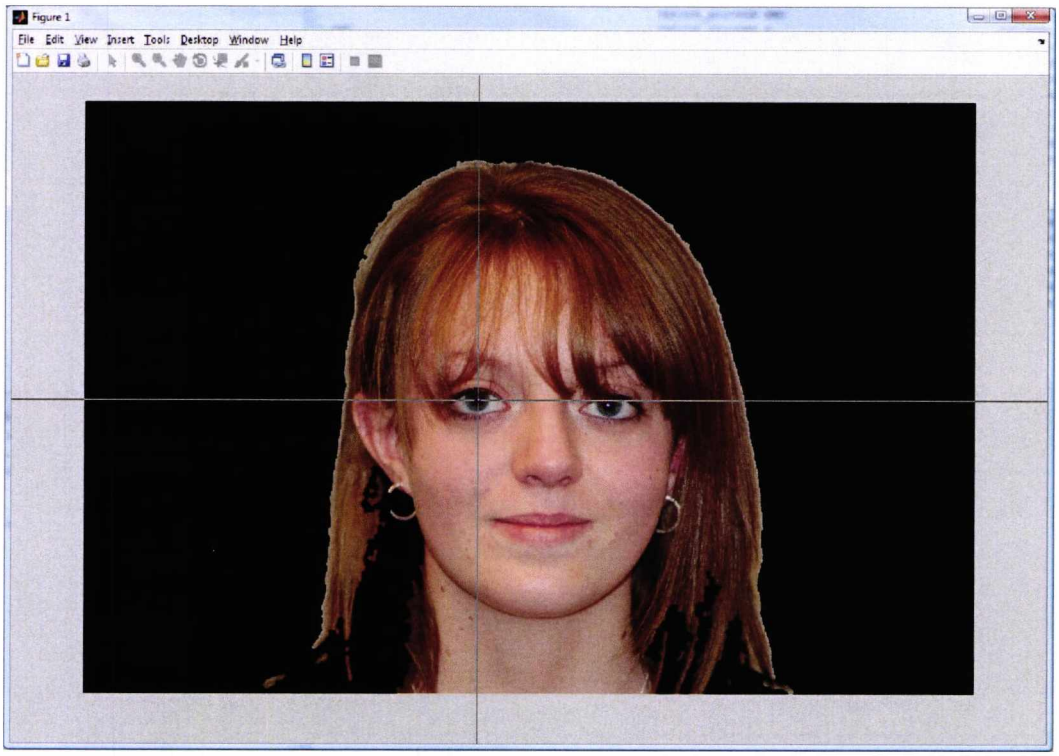


Figure 5.2 - Manual Eye Location

The eye positions are acquired by means of a code which loads each image into memory, presents it to the user and waits for the user to click on each eye in turn. Once acquired, eye locations are stored in a four dimensional array which follows the filename and path constructors, thus allowing the eye locations to be correlated to their source image in subsequent functions. The images were then replayed to the user with the recorded eye locations highlighted to confirm that there were no errors in the marking process.

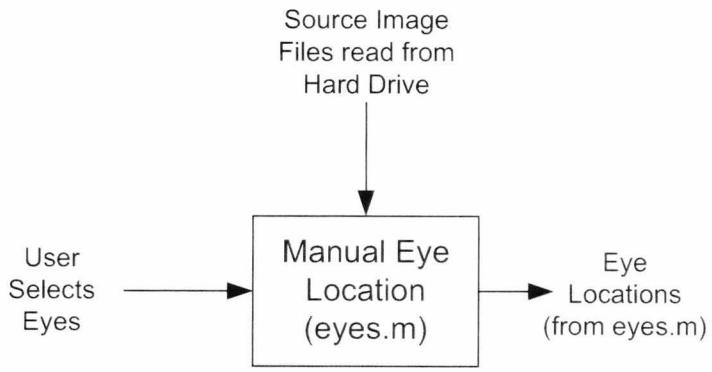


Figure 5.3 - Eye Location Process

Once located, the eye locations and maximum number of classes from which to extract features are passed to the locateROI.m function which re-reads the files from the hard drive and returns a feature vector for each image in a multi dimensional data array.

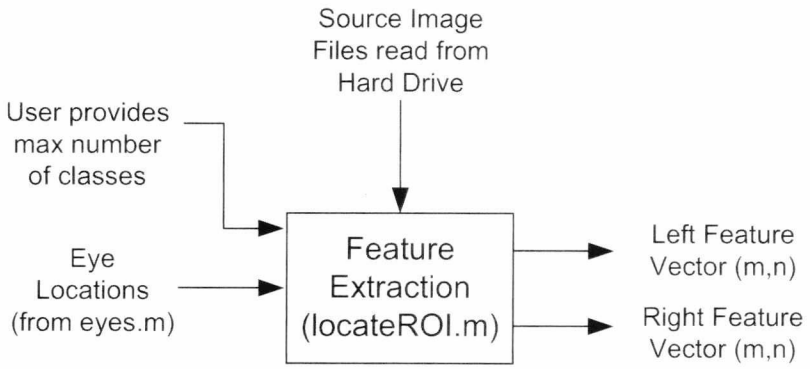


Figure 5.4 - Feature Extraction

The locateROI function has been written in a modular fashion to facilitate the adjustment and fine tuning of the feature extraction mechanisms.

The algorithms presented in this chapter are analysed primarily using the high resolution database initially presented in Chapter 4 during the initial configuration, tuning and setup of the feature extraction algorithms. For comparative purposes, benchmark performance tests will also be taken using BMDB image database (also defined in Chapter 4).

The process for recognition is shown in Figure 5.5.

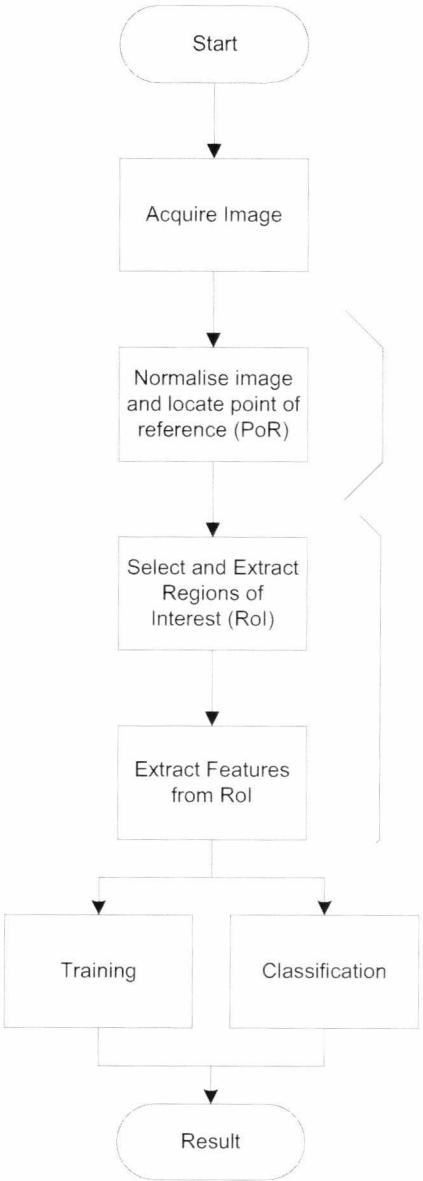


Figure 5.5 – Flowchart of Texture Recognition Process

When compared with generic texture recognition algorithms, skin texture identification presents several challenges. First, one must determine the optimum locations from which to extract the features. Selection of parameters should include considerations such as:

1. Is the area of skin clearly visible?
2. Is the area of skin likely to be obscured by clothing, jewellery, or other objects?
3. Would the area of skin be impacted by shading from other areas of the body?
4. Is the area of skin likely to be artificially changed by makeup or other factors and how will this influence the algorithm?

The human face presents contains areas which may be used in this process and will be the focus of this chapter. Specifically, the focus is on four prime areas which are indicated in Figure 5.6 - Regions of Interest.

These areas were chosen as they present regions of the face that are clearly visible in a frontal portrait subject photograph. Although an individual area may be obscured by hair, facial hair or spectacles, the likelihood of all the areas being obscured to a point of non-recognisability is hopefully minimised.

Throughout this thesis, left and right are used interchangeably with sides "A" and "B" from the perspective of the viewer. Hence when looking at a photograph of a subject image, the left side represents the individual's actual right side. Hence in Figure 5.6, the regions 1A, 2A and 3A are defined as the left side of the face, where regions 1B, 2B and 3B are defined as the right side.

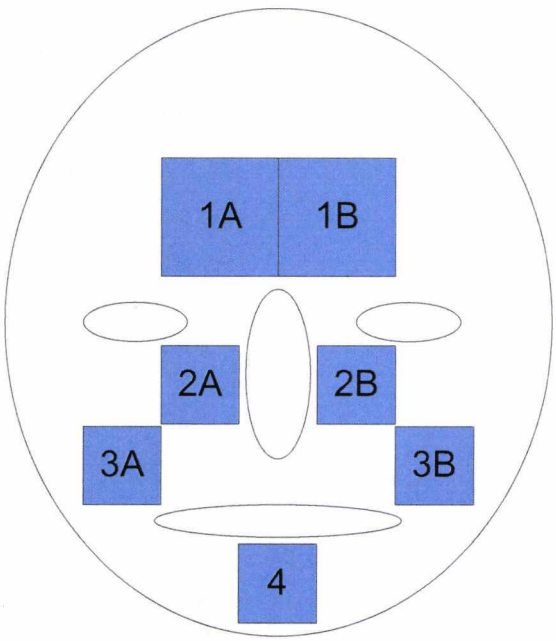


Figure 5.6 - Regions of Interest

Given the large number of images the algorithm is responsible for processing and to expedite the testing process, initial reference points are provided to the algorithm by supplying coordinates for the centre of the eyes.

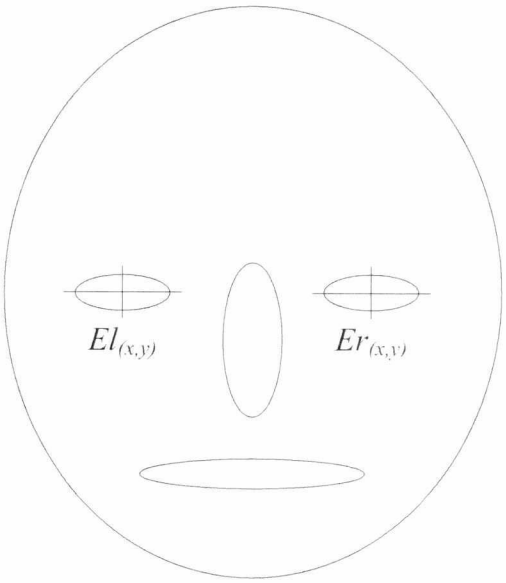


Figure 5.7 - Eye Location

Using $El_{(x,y)}$ and $Er_{(x,y)}$ as reference coordinates for the eyes, determining the scale of the face and eye separation is a relatively simple process as shown below.

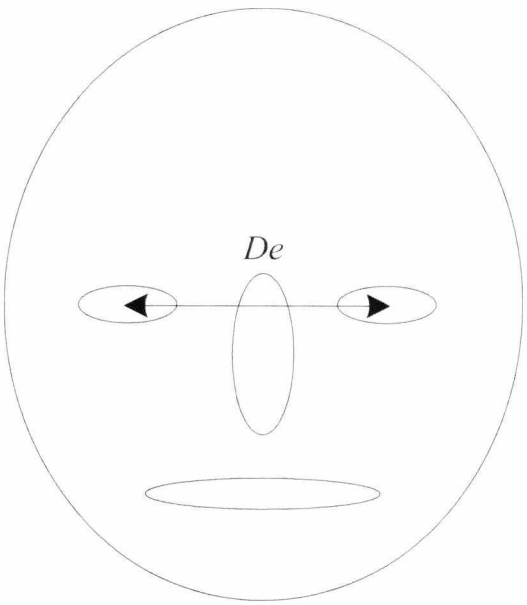


Figure 5.8 - Eye Separation Measurement

5.4.1 Scale Normalisation

Gabor texture filters function by attempting to extract patterns, or repetitions within an image which provide a means for identification. In the following section, frequencies are supplied as input variables to the Gabor feature function. The frequencies are measured in cycles/pixel pattern and as such can be affected by variations in image scale.

$$De = \sqrt{(El_x - Er_x)^2 + (El_y - Er_y)^2} \tag{5.3}$$

By determining the eye separation distance using the Pythagoras Theorem (5.3), it is possible to provide a means to normalise an image based on scale. Although some variance between individuals will exist, one would expect the same person to have a constant eye separation value. Thus this can be utilised in ensuring the Gabor filter supplied frequencies are not impacted by changes in image scaling. The scaling of Gabor filters based on eye separation forms an automatic part of the feature extraction algorithm.

5.4.2 Determination of Feature Extraction Points

Using the eye separation as a baseline measurement and experimental techniques to select region 2b from Figure 5.6 it was found that the feature regions could be isolated by utilising the already known eye locations and eye separation distance along with a scaling factor. The scale values of 0.136 and 0.192 along the x and y axis respectively were appropriate to determine the origin point for the cropped image. An approximation of this is shown in Figure 5.9 and Table 5.1.

Point of reference	X	Y
Left eye	EyeL + (0.136De)	EyeL + (0.192De)
Right eye	EyeR − (0.136De)	EyeR + (0.192De)
Size of cropped area	0.3832 ∗ De	

Table 5.1 - Eye Locations

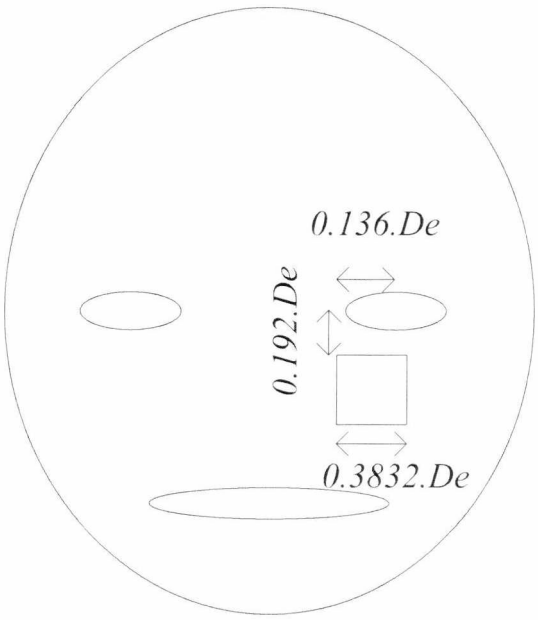


Figure 5.9 - Feature Extraction Point (Not to Scale)

It should be noted that the point of origin for establishing the eye locations is the upper left corner of the image as shown in Figure 5.10.

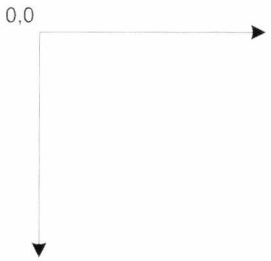


Figure 5.10 - Point of Origin for Feature Extraction Region Calculations

Figure 5.11 and Figure 5.12 show the described skin areas in greyscale from two HRDB subjects. In this example, the textural differences between the images can be seen with the naked eye quite easily.

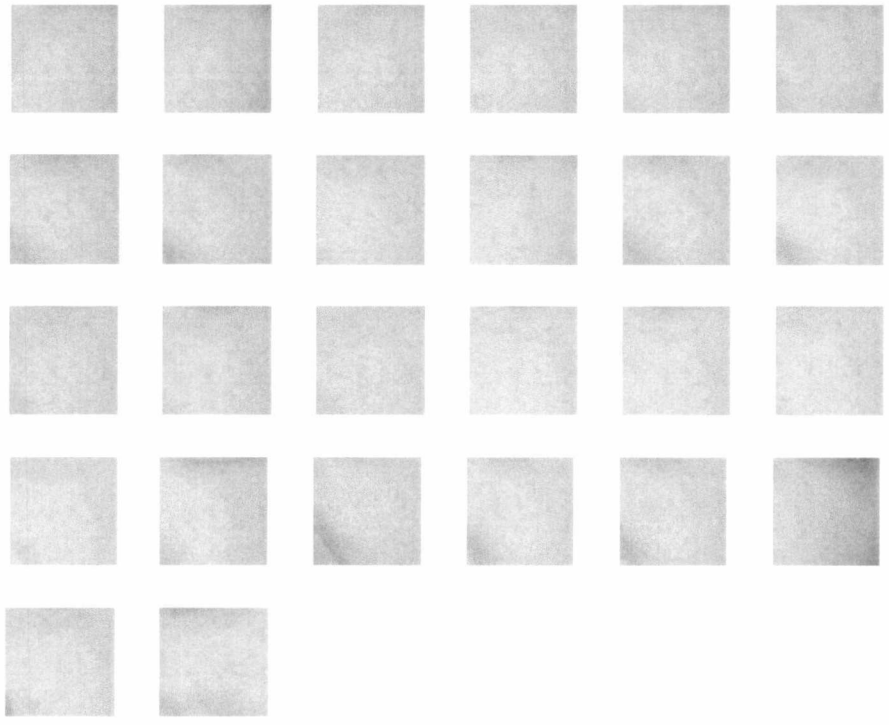


Figure 5.11 - Subject 1

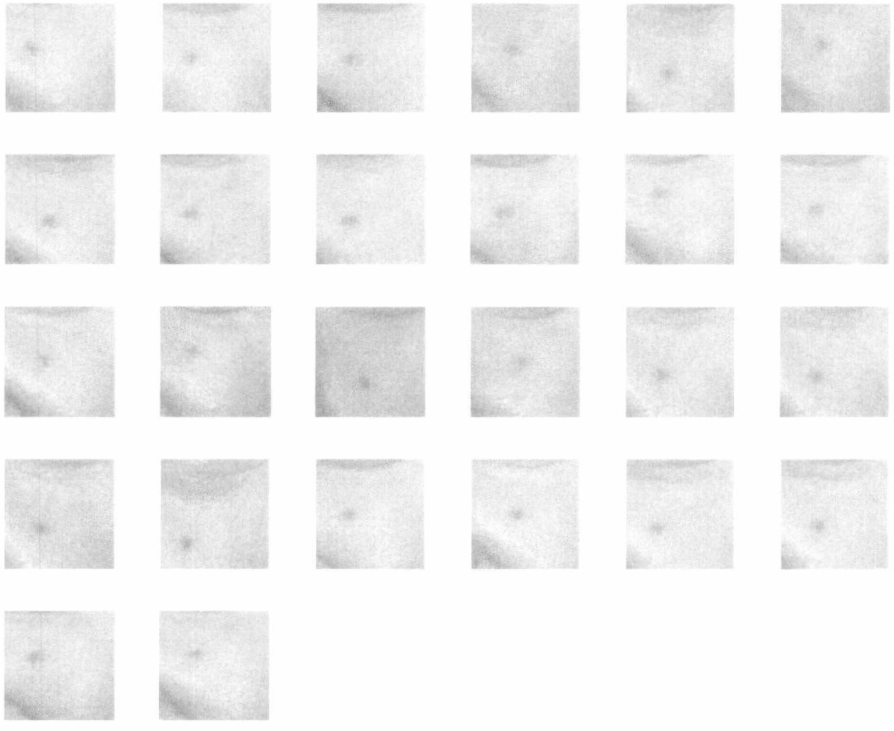


Figure 5.12 - Subject 2

5.5 DESIGNING GABOR WAVELET FILTERS

Gabor filters are complex mathematical functions which result in both a real and imaginary part and are introduced in Chapter 3 .

Taking the Gabor wavelet filter presented in (3.8):

$$f(x,y) = \frac{\alpha\beta}{\pi} e^{(-\alpha^2(x\cos\theta + y\sin\theta)^2 - \beta^2(x\sin\theta + y\cos\theta)^2 + 2j\pi f_0(x\cos\theta + y\sin\theta))} \quad \textbf{(5.4)}$$

Where α represents the frequency as modified by a normalised scalar along the wave, β represents the frequency as modified by a normalised scalar orthogonal to the wave, θ represents the angle and f_0 represents the period as derived from the reciprocal of the pixels/cycle frequency value.

Gabor wavelets are by nature orientated to a set angle depending on the supplied value of θ. The parameters α and β allow a modification of the wavelet along or orthogonal to the waveform respectively. At this stage it was decided to use the natural shape of the Gabor wavelet with no such modification along either axis.

Using a 1:1 scaling ratio:

$$\alpha = f_0, \quad \beta = f_0 \quad \textbf{(5.5)}$$

Selecting initial variables for this equation from Table 5.2 and Table 5.3 and using all available permutations creates a filter bank of 12 Gabor filters each with a unique frequency/orientation combination. ***S*** in this table represents the dimension along the x axis of the source image (which is calculated as being $0.3832 \times De$).

The selected frequencies permit a wide range of frequency patterns to be identified relative to the expected image size. A variety of filtering frequencies were attempted using the HRDB datasets, after these experiments it was found that optimum results were achieved using frequencies of approximately $\frac{20}{s}$, $\frac{10}{s}$ and $\frac{4}{s}$ cycles/pixel. However with the latter of these it was frequently observed that f_0 would rise above 0.25 cycles/pixel at which point studies suggest that this is detrimental to system performance [Shen, Bai et al. 2007].

Experimental results showed that hard coding the upper frequency to 0.2 cycles/pixel avoided this issue.

f_0 (Method)	f_0 (Actual Frequency if s=128)	Typical Value of $\tau = \frac{1}{f_0}$ (for s=128)
0.2	0.2	5
$\frac{10}{s}$	0.078	13
$\frac{4}{s}$	0.031	32

Table 5.2 - Frequency Values for Gabor Feature Extraction

Table 5.2 contains the selected frequencies for which analysis will be performed and includes the period τ in pixels/cycle to allow for better visualisation of the frequency tuning. The selection of these values was based on observations made in [Shen, Bai et al. 2007] that the wavelets perform better in face recognition when f_0 is in the low frequency band as mentioned above. Fixing this value in this manner does not permit for a dynamic scale-dependent frequency to be selected. However it was observed that this method provided superior results against alternatives scale-based values which would permit f_0 to exceed 0.25.

The remaining two frequencies defined Table 5.2 have been selected to provide a non-harmonic pair of filters which under developmental testing against the HRDB dataset showed positive performance.

The selection of how many wavelet filters to create for optimal selection has little theoretical support and researchers can differ significantly in their choice. For this study, the initial objective is a proof of concept that the selected skin regions contain significant information for subject identification. The detailed tuning and selection of candidate filters is not at this stage considered a focal point of the research. To consider examples [Ferrari, Rangayyan et al. 2001] uses a bank of 48 filters in a mammogram asymmetry study whereas [Idrissa and Acheroy 2002] suggests that a 25 wavelet bank is sufficient for texture classification. As this study will cover not only Gabor wavelet features, but also spatial features it is desirable to minimise the initial feature-space to avoid the curse of dimensionality [Bellman and Kalaba 1959].

[Haley and Manjunath 1999] successfully demonstrate a Gabor wavelet based texture classifier that uses a bank of 12 filters. A similar implementation is used in this study of 3

Frequencies and 4 orientations and should allow sufficient flexibility to extend the system at a later date while minimising adverse effects from the curse of dimensionality.

The orientations were selected at 45 degree intervals for the first 180 degrees. (Angles greater than 180 produce duplicate filters).

θ (degrees)
0
45
90
135

Table 5.3 - Theta Values for Gabor Feature Extraction

Table 5.3 shows the values of θ used. Note that it is only required to use a range spanning θ = 0 to θ = 179 degrees as the range 180 - 359 degrees is essentially a reverse direction reproduction and in effect duplicates the preceding range.

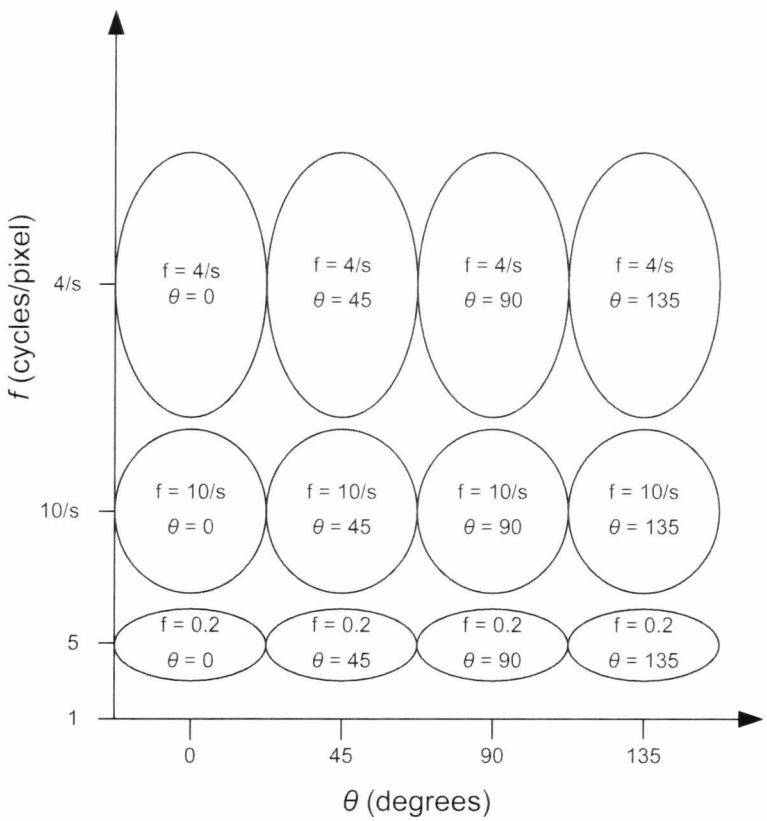


Figure 5.13 - Visual Representation of Filter Bank

Figure 5.14, Figure 5.15 and Figure 5.16 provide a typical output using the above values in a Gabor filter function for a 128x128 filter (only the real part of the filter is shown). As presented in Section 5.4 however, the size of the filter is dynamically created based on the scaling of the source image as calculated from the eye separation distance.

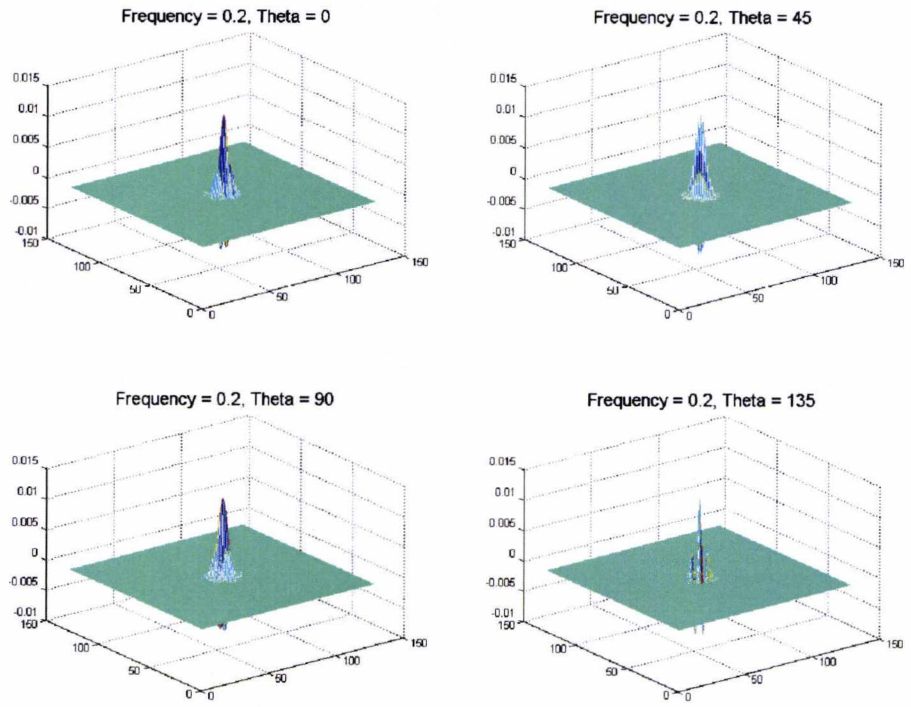


Figure 5.14 - Gabor Filters for $f_0 = 0.2$, $\theta = 0, 45, 90, 135$

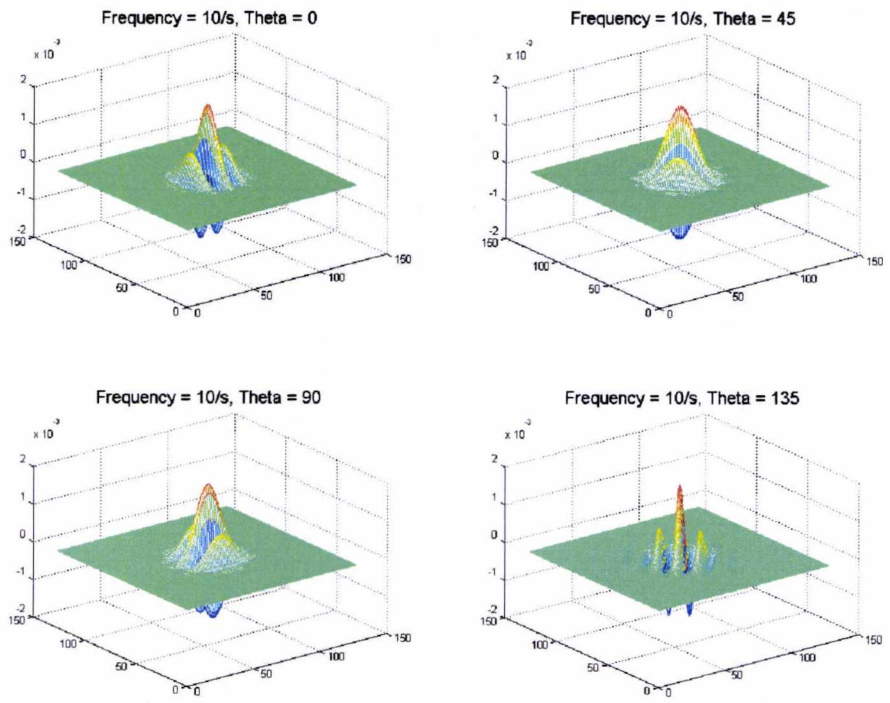


Figure 5.15 - Gabor Filters for $f_0 = \frac{10}{s}(0.078)$, $\theta = 0, 45, 90, 135$

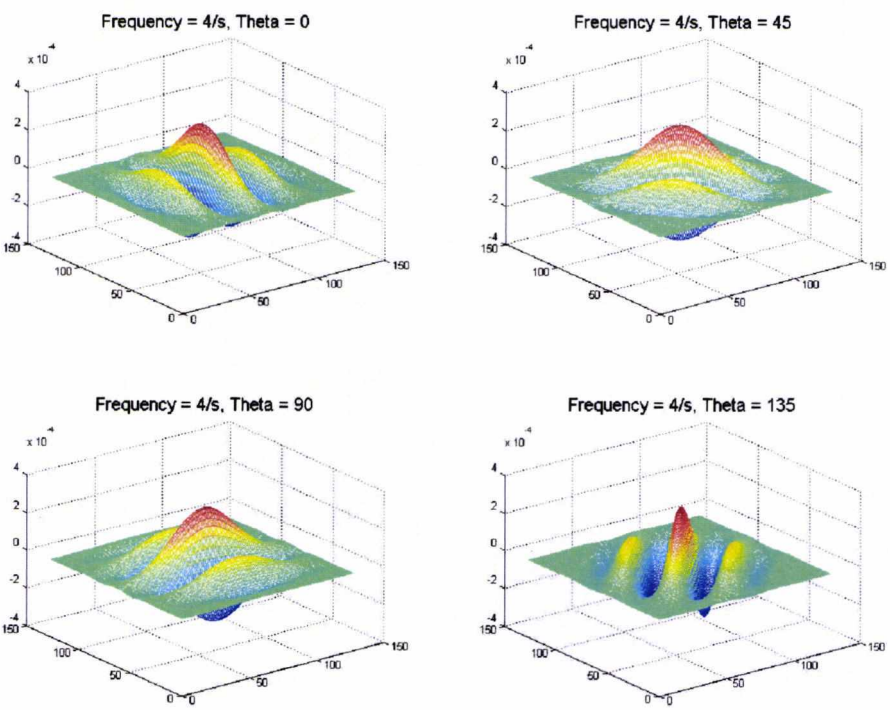


Figure 5.16 - Gabor Filters for $f_0 = \frac{4}{s}(0.031)$, $\theta = 0, 45, 90, 135$

5.6 FEATURE EXTRACTION PROCESS

Having established a mechanism for filtering probe images, the next issue to examine is how to proceed to extract suitable features. The mean energy of the convolved image with its standard deviation provides two simple to acquire distinct features suitable for an initial evaluation of the proposed methods performance.

A relative high mean energy from a filtered image indicates a high presence of the target frequency at the specified orientation, while lower mean energies indicate a lesser presence. The standard deviation indicates the "spread" or distribution of the energies found in the filtered image and likewise is a descriptor for the texture frequency patterns.

$$\bar{E} = \frac{1}{n} \cdot \sum_{i=1}^{n} E_i \tag{5.6}$$

Where $\bar{E}$ signifies the mean output energy, n signifies the maximum number of pixels and E_i the absolute energy of the filtered output element at position i as calculated using the inbuilt `abs` and `mean` functions of Matlab.

$$\sigma = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (E_i - \bar{E})^2} \tag{5.7}$$

Where σ signifies the standard deviation,

Calculating the mean energy and standard deviation for each of the 12 filters previously presented provides a 24-dimensional feature space vector of the form.

$$\{\bar{E}_1, \bar{E}_2 \ldots \bar{E}_{11}, \bar{E}_{12}, \sigma_1, \sigma_2 \ldots \sigma_{11}, \sigma_{12}\} \tag{5.8}$$

5.7 GABOR WAVELET PERFORMANCE ON THE BMDB DATABASE

5.7.1 TRAINING AND TESTING DATASET CONFIGURATION

All classifiers used in this research require training using a training set of probe images. This involves separating the gallery classes into two sets of data for training and testing as detailed in Section 4.4.5.3 For ease of reading a duplicate of the training/testing assignments is provided in Figure 5.17.

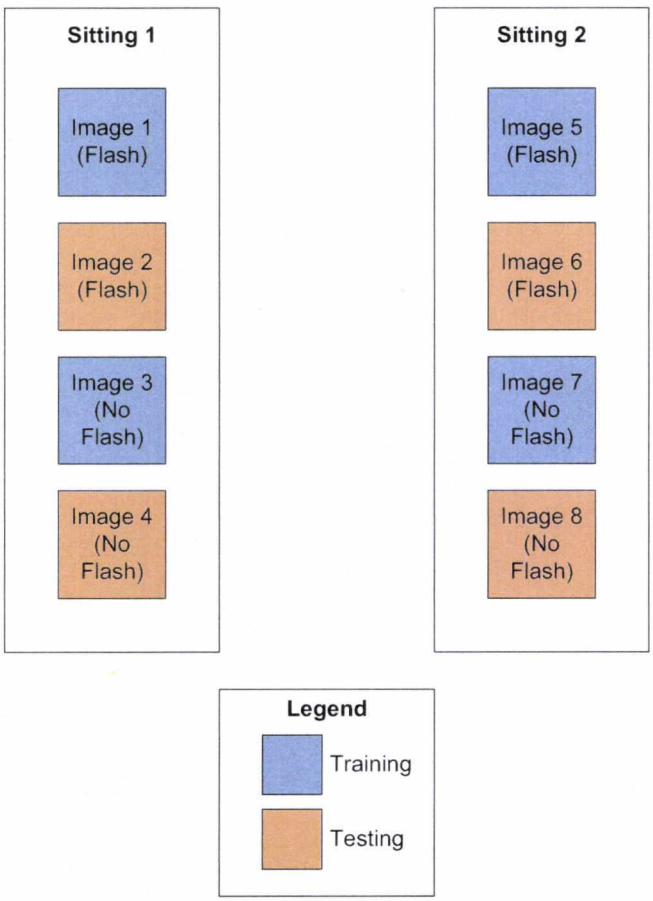


Figure 5.17 - BioSecure DS2 Subset Training and Testing Assignment Process

Using the complete BMDB DS2 Gallery of images which contains 79 unique individual classes and a 50:50 ratio of training to testing images, samples were taken of regions 2A and 2B (left and right) as described earlier in this section.

5.7.2 GABOR FEATURE CLASSIFICATION OF 2A AND 2B

Here a simple comparison between the 2A and 2B facial regions is investigated. Three classifier types are used as described in Section 4.5 to provide an overall performance classifier independent performance metric of the capabilities of the Gabor wavelet bank as designed earlier in this chapter.

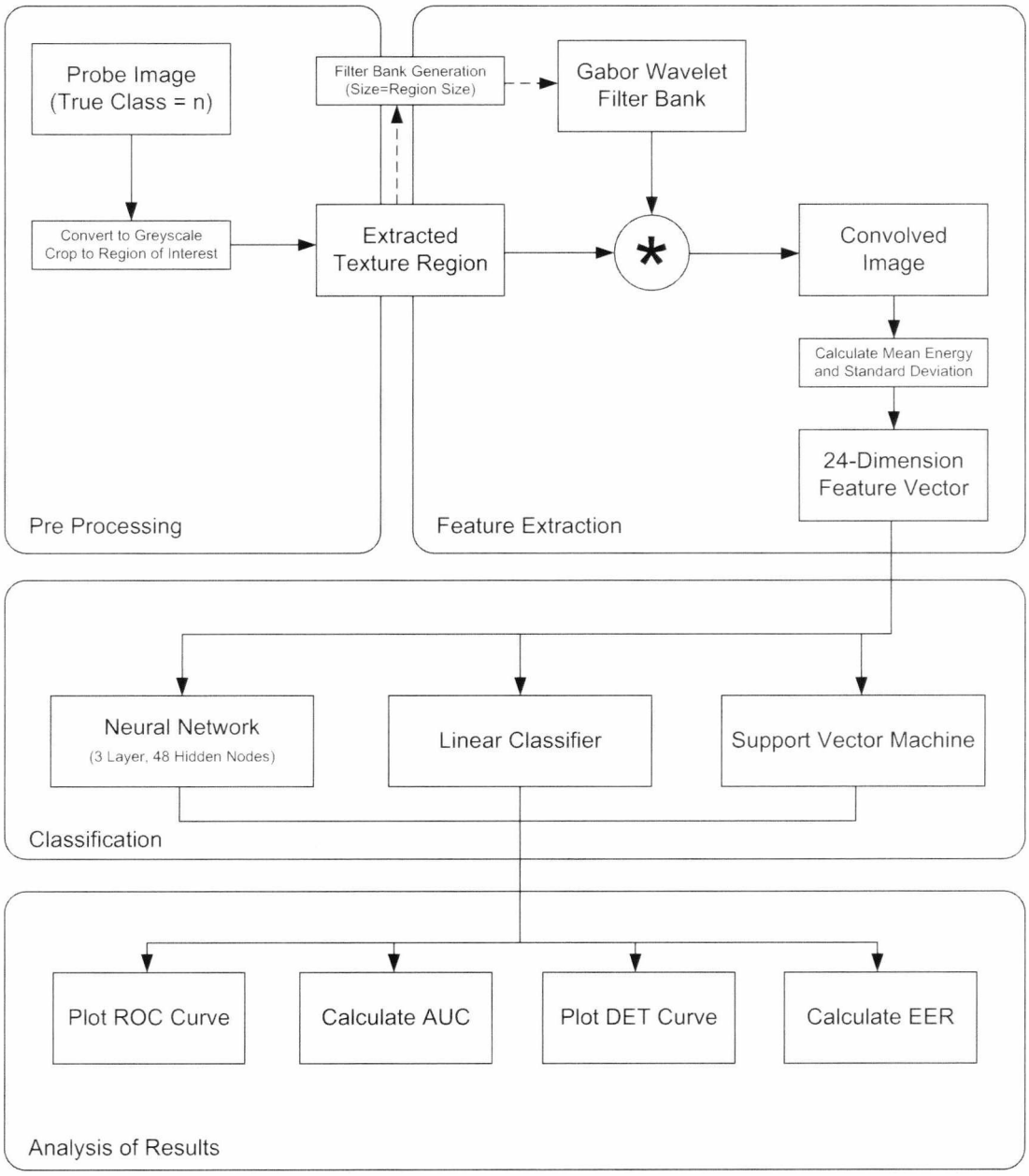


Figure 5.18 - Gabor System Process Diagram

Figure 5.18 shows the processes performed during this evaluation as previous described in the preceding sections of this chapter and chapter 0which discusses the experimental framework.

The subsequent sections of this chapter contain a report on the evaluation results.

5.7.3 EVALUATION RESULTS (SFSB GABOR)

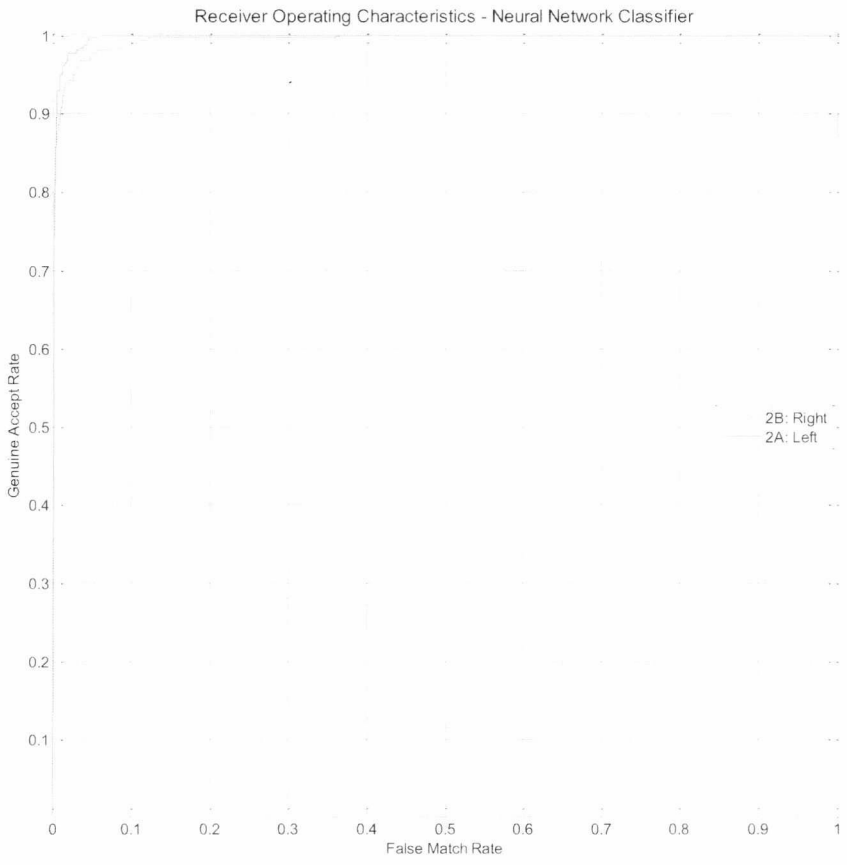


Figure 5.19 - Neural Network 2A and 2B Region Performance Comparison (ROC Curve)

As expected, the two regions perform similarly with the left side marginally outperforming the right. Looking at the DET curve for the same output it is apparent that the resulting outputs are indeed very close.

Both regions demonstrate a surprising initial level of accuracy with ROC curves lying tightly into the upper left region and show that the algorithm performs well against the BMDB Dataset.

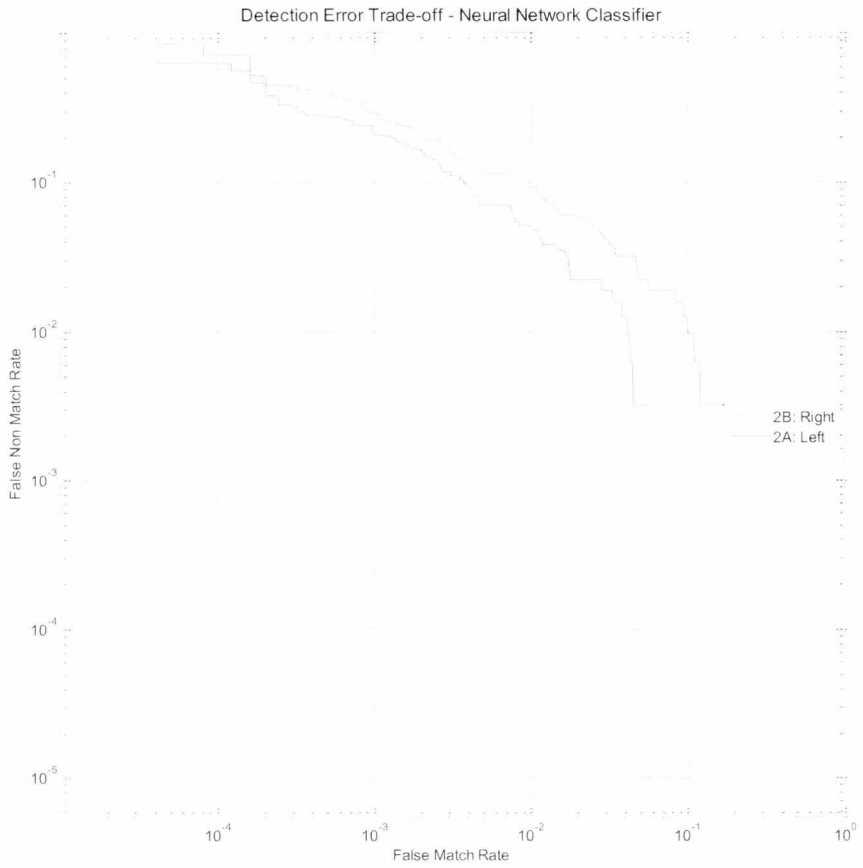


Figure 5.20 - Neural Network 2A and 2B Region Performance Comparison (DET Curve)

Using the same feature data with linear and support vector machines produces less accurate outputs as shown in Figure 5.21 and Figure 5.22. The following DET curves in Figure 5.23 and Figure 5.24 again confirm the performance similarity between independent tests processed on the left and right 2A and 2B regions.

In both regions, the support vector machine performs substantially better than the Baysean linear discriminant classifier. The linear classifier functions by attempting to reduce the feature dimensionality. If this proves difficult, it is possible that the algorithm may be struggling to find strong class separation between the 24 dimensional inputs.

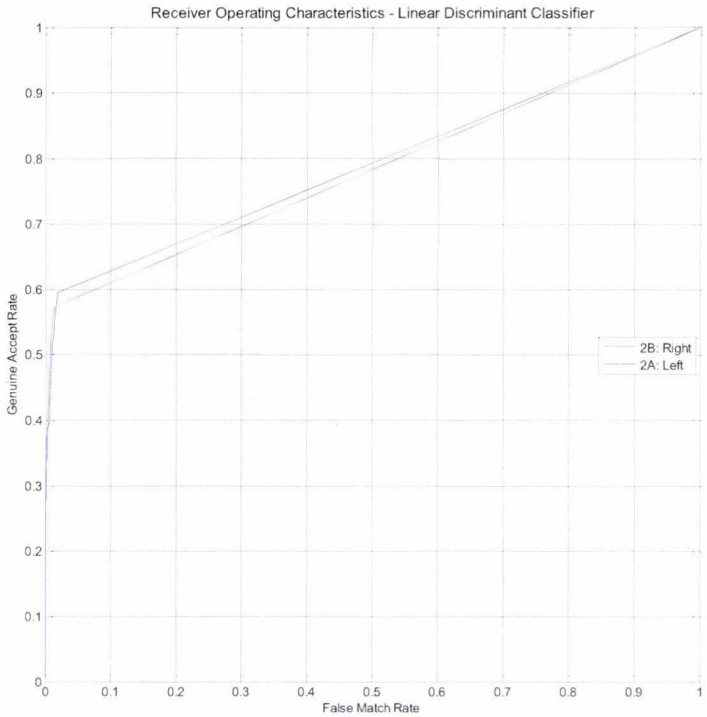


Figure 5.21 - Linear 2A and 2B Region Performance Comparison (ROC Curve)

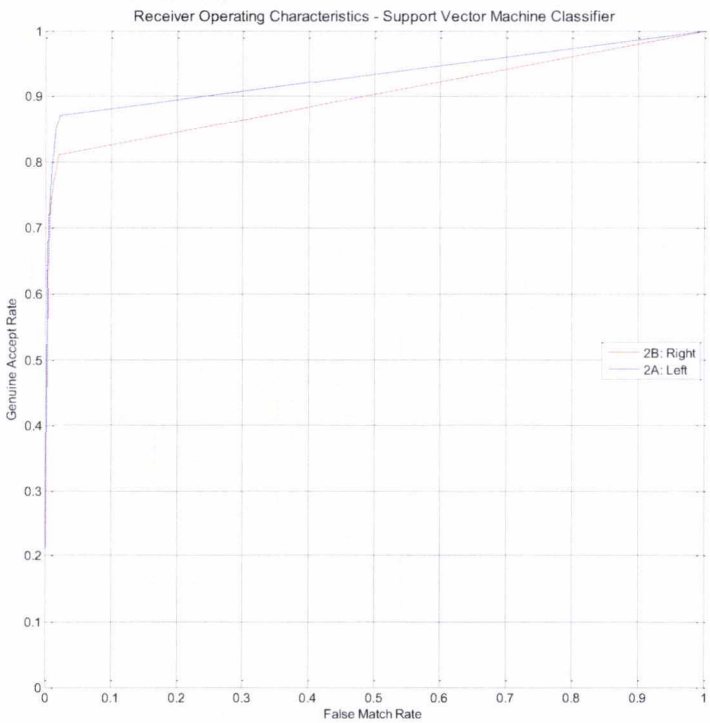


Figure 5.22 - SVM Network 2A and 2B Region Performance Comparison (ROC Curve)

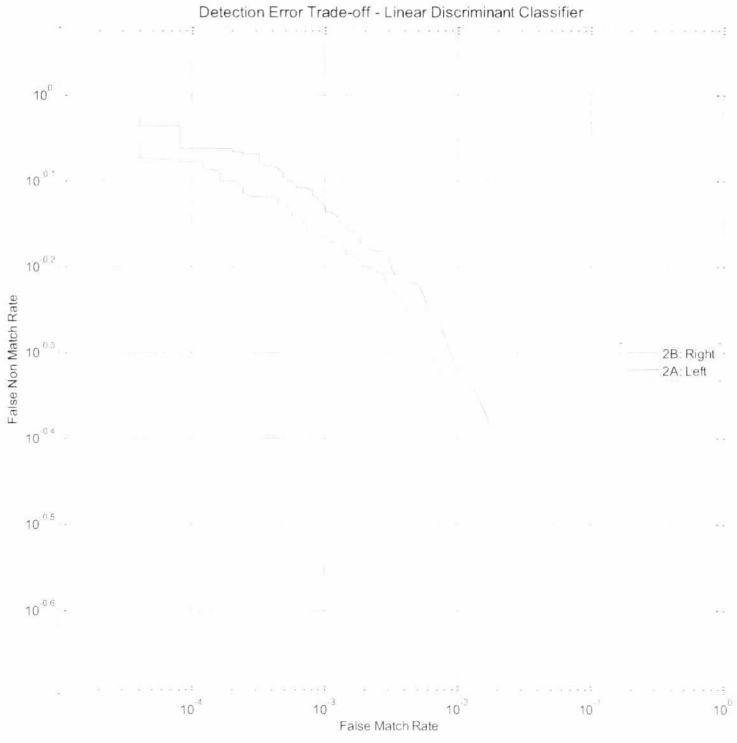


Figure 5.23 - Linear Network 2A and 2B Region Performance Comparison (DET Curve)

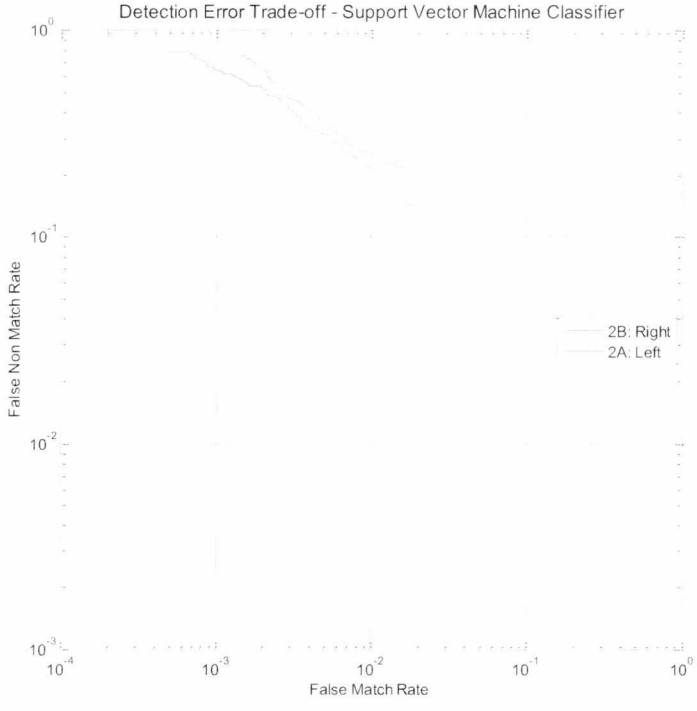


Figure 5.24 - SVM Network 2A and 2B Region Performance Comparison (DET Curve)

In comparing the results between the three classifiers, the neural network shows superior performance as a classifier for the features used. The results indicate that there is a good basis on which to justify further research. This research should be directed to further support the theory that skin regions 2A and 2B of a subjects face are suitable for subject identification.

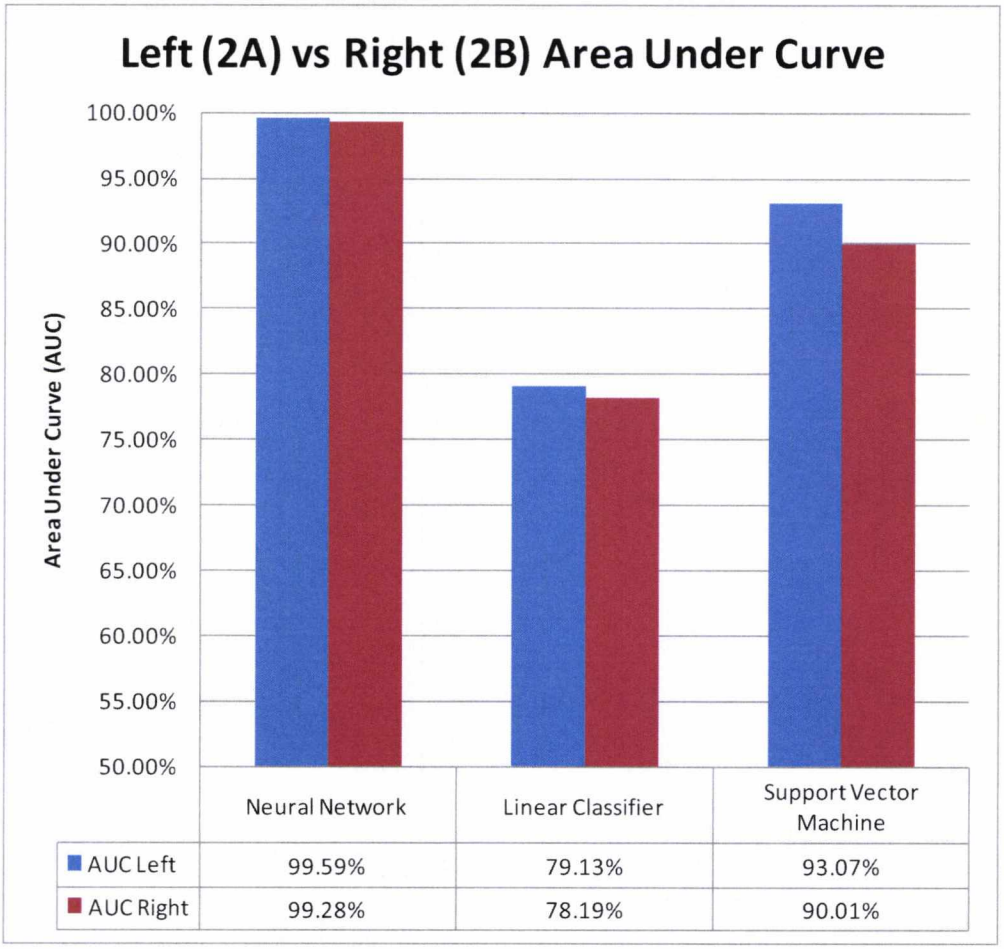


Figure 5.25 - Area under Curve for Left/Right Performance Comparison

All the AUC values in Figure 5.25 have a value greater than 50% indicating which indicates positive classifier performance. In addition, the equal error rate (Figure 5.26) indicates that the system performs particularly well for individual facial regions when classified by the neural network. All training and testing in this and subsequent sections were executed three times for the neural network with the median average being used to record the results. The support vector and linear classifiers produced identical results on repeat attempts.

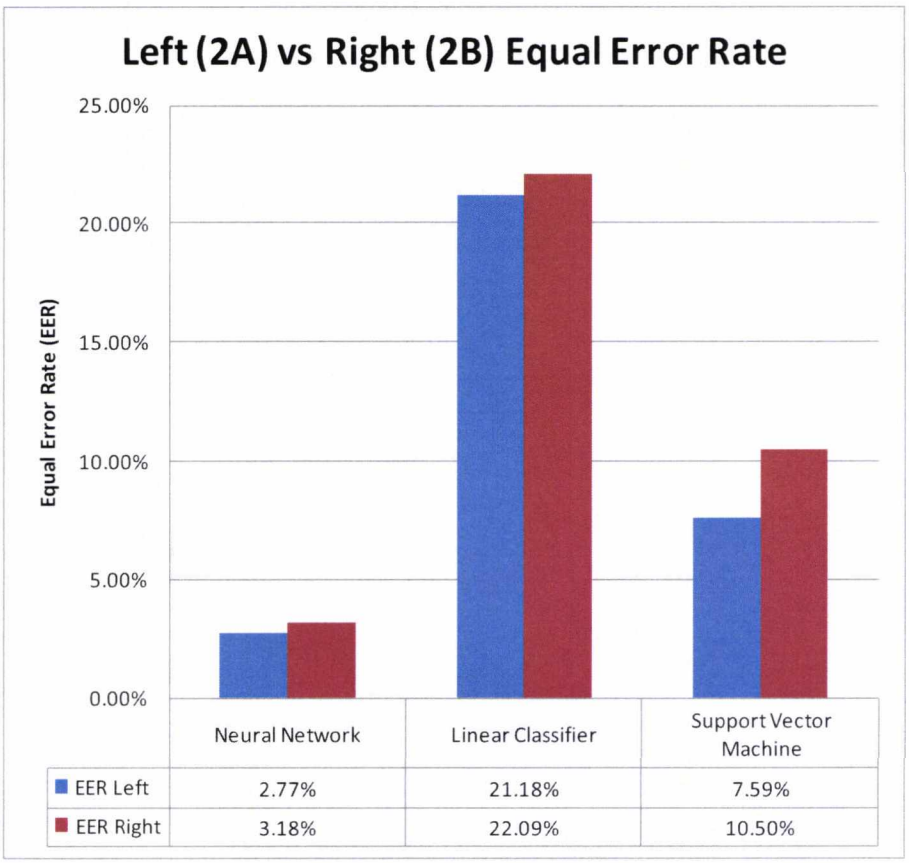


Figure 5.26 - Equal Error Rate for Left/Right Performance Comparison

Figure 5.26 shows the equal error rate in the form of a bar chart for comparison between the various classifiers and regions. A lower equal error rate indicates a more optimum (lower) threshold at which rate of false accepts and rejects are equal. Across all three classifiers, the left side of the face appears to perform better indicating that the features obtained from this side of the face are more suitable for recognition. Although possible, it is unlikely that one side of the face is more suited to recognition than the other. It is more plausible that there exists less than optimal sample data or an obstruction in the underlying dataset which may be impeding recognition.

5.8 CHAPTER REVIEW

In this chapter the following has been presented:

- Introduction to two dimensional texture recognition
- Texture as a feature
- Pre-processing and pre-classification image processing

- Gabor wavelet design and justification for parameter selection
- Conclusive performance report indicating suitability of skin areas 2A and 2B for subject identification

In particular, the evaluation results within this chapter appear to support the initial opening questions and call for research posed in Section 5.1 suggesting the Gabor wavelets provide a suitable means for extracting texture features from skin texture that may be suitable for a face recognition system.

6 FUSION FOR SKIN TEXTURE BASED RECOGNITION

6.1 INTRODUCTION

Chapter 5 presented a supposition of the surface to texture relationship as shown in (6.1) with t representing a single feature based from the feature space of surface S.

$$S = \{t_1 + t_2 + \cdots + t_{n-1} + t_n\} \tag{6.1}$$

It was concluded in Chapter 5 when S was a relatively small greyscale area of skin texture from below a subject's eye defined as region 2A or 2B, that S provided a sufficiently unique set of features to establish autonomous identification with a notable degree of accuracy. Specifically a series of 12 Gabor wavelets were constructed and used to extract an array of 24 texture features t_0 to t_{24} which proved suitable for classification by the three classifiers used in this evaluation.

Chapter 6 is intended to extend upon this evaluation by analysing how the results presented prior to this section may be improved through two primary means:

1. Through the fusion of different regions. In this thesis, this has been termed "single feature-set, multi biometric region" (SFMB).

2. Through the addition and subsequent fusion of spatial frequency feature extraction techniques to the system. This has been termed "multi feature-set, multi biometric region" (MFMB).

For comparison, the evaluation in Chapter 5 falls under the term "Single feature-set, single biometric region" (SFSB).

Both score fusion and feature fusion techniques will be examined as means of combining the features and are discussed in the following section.

At the conclusion of this Chapter the findings will be summarised and a course for future work presented.

6.2 SINGLE FEATURE MULTI REGION FUSION (SFMB GABOR)

6.2.1 EVALUATION CONFIGURATION

In succession from the results presented in Section 5.7 this section extends the system previously analysed by comparing the score and feature fusion of regions 2A and 2B.

In feature fusion, one simply amalgamates the features from each region turning two 24-dimensional feature vectors into a single 48-dimensional feature vector. In score fusion, a classifier is employed for each 24-dimensional feature vector with the resulting classifier outputs being mean averaged.

If the features in region 2A are shown below

$$S_{2A} = \{ta_1 + ta_2 + \cdots + ta_{23} + ta_{24}\} \tag{6.2}$$

and if 2B are similarly shown

$$S_{2B} = \{tb_1 + tb_2 + \cdots + tb_{24} + tb_{24}\} \tag{6.3}$$

Then feature fusion could be depicted as:

$$S_{Feat} = S_{2A} + S_{2B} = \{ta_1 + \cdots + ta_{24} + tb_1 + \cdots + tb_{24}\} \tag{6.4}$$

For score fusion using a mean combining classifier, if the classifier output class probability matrix for region 2A is shown by:

$$Ca = f(\sum_{n=1}^{24} ta_n) \tag{6.5}$$

and for region 2B is shown as:

$$Cb = f(\sum_{n=1}^{24} tb_n) \tag{6.6}$$

Where f represents a classification function on the input features ta and tb then for a mean combining classifier, it may be assumed that:

$$C_{score} = \frac{Ca + Cb}{2} \tag{6.7}$$

While there exist other combiners in score fusion, the intention in this evaluation is to show proof of suitability for further research. Initial experimental development work on the HRDB database showed the mean combining classifier to be suitable for this purpose. The built in PRtools function `meanc` is used in this evaluation.

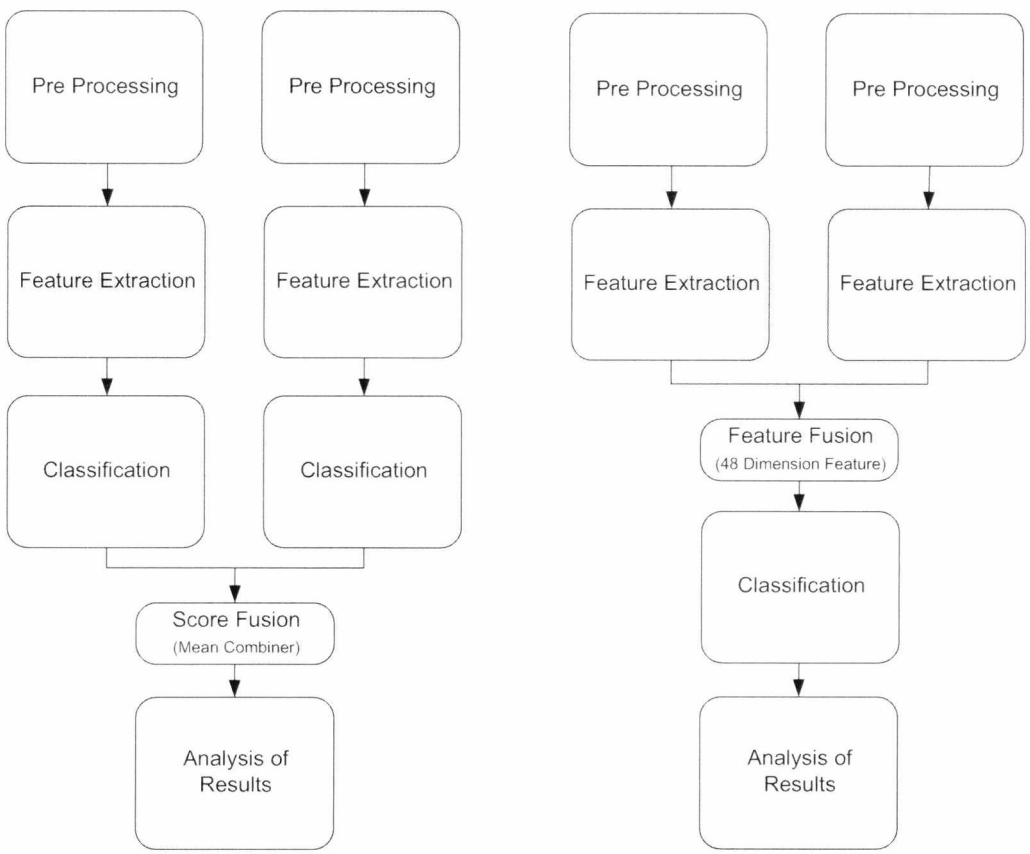


Figure 6.1 - Comparison of Score and Feature Fusion

Figure 6.1 shows a visual side by side comparison of the score and feature fusion methods used while Figure 6.2 and Figure 6.3 show the block design for the score fusion and feature fusion systems. For all score fusion processes, the aforementioned mean combining classifier operating on the parallel outputs of the classifiers is used. Given that each classifier outputs a score between 0 and 1 for each probe image / output class combination, the mean classifier simply calculates the mean score and uses this to determine output class assignment.

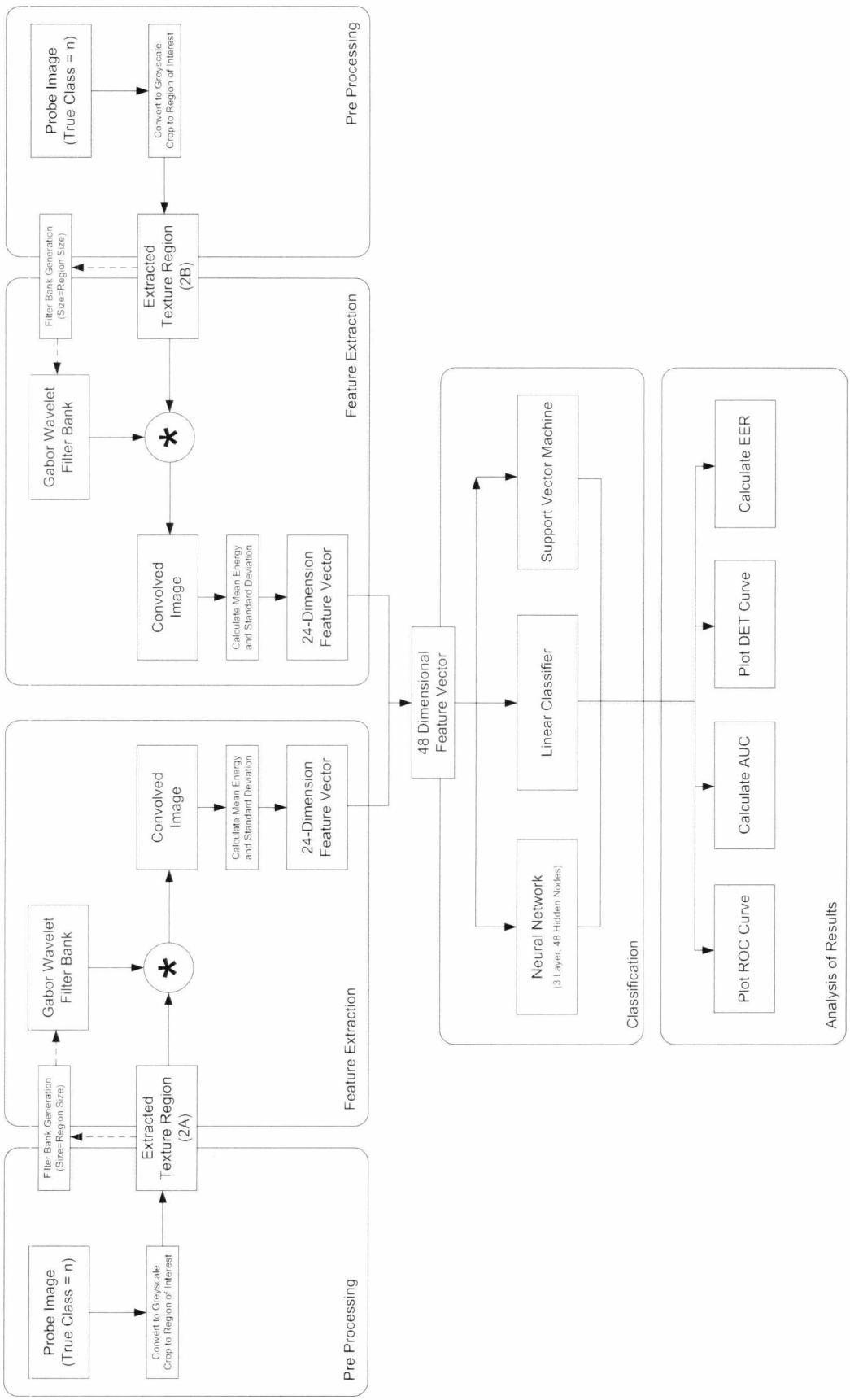


Figure 6.2 - Feature Fusion of Gabor Features

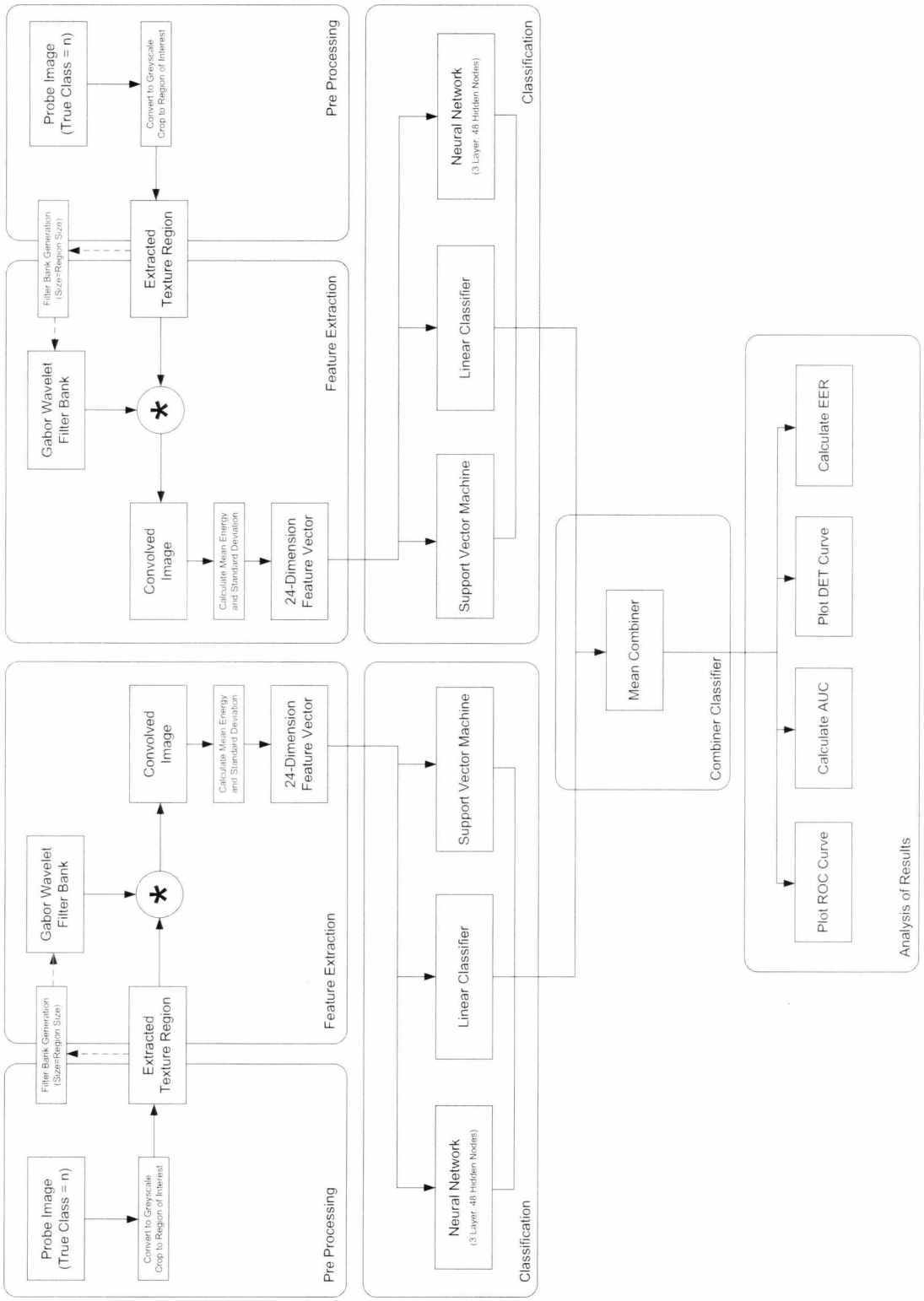


Figure 6.3 - Score Fusion of Gabor Features

In Chapter 5 the left and right skin texture below the eyes were independently examined and found to have similar accuracy and error rates when passed through separately trained classifiers of the same design. This section will extend this by evaluating if the results obtained are in fact almost identical, or if sufficient crossover exists between the two texture

feature vectors to allow an improvement in the overall recognition accuracy of the system by fusing either the feature vectors prior to classification or the respective scores of individual classifiers. In summary, this section should answer two specific questions:

1. Can the accuracy of the 2A and 2B classifiers presented in Chapter 5 be improved by fusion techniques?

2. If (1) is true, how do score and feature fusion techniques compare in this objective?

6.2.2 Evaluation Results (SFMB Gabor)

The neural network ROC curve is shown in Figure 6.4 and enlarged in Figure 6.5.

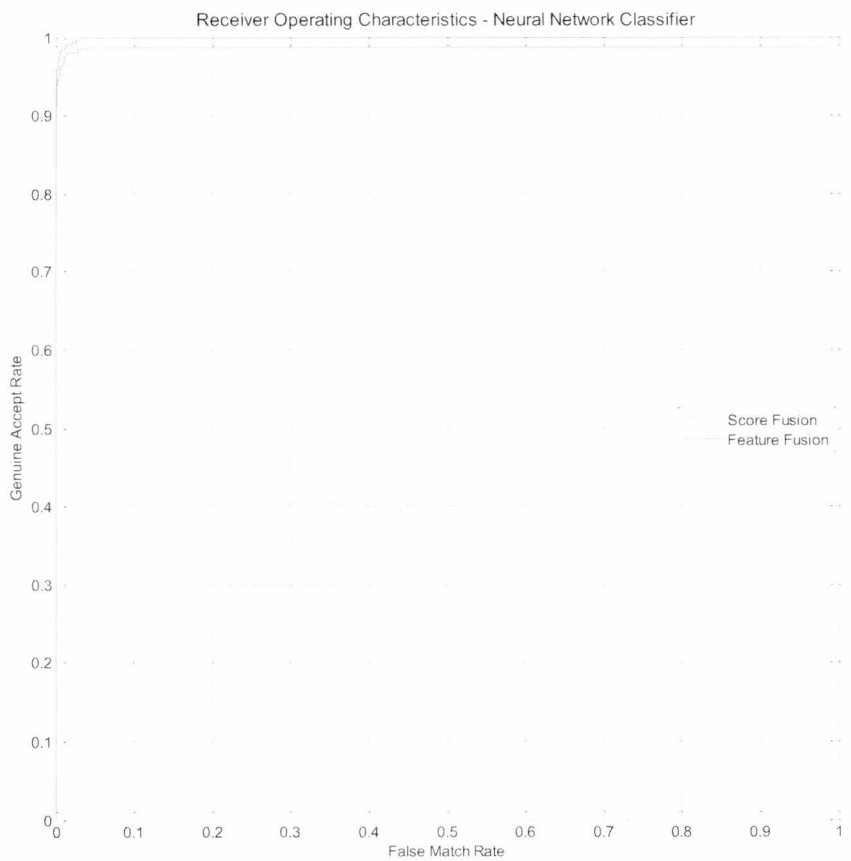


Figure 6.4 - Neural Network Score vs. Feature Fusion Performance (ROC Curve)

It can be observed that there is a performance difference between the two techniques with score fusion showing a larger area under the curve and thus being favoured by the neural network classifier.

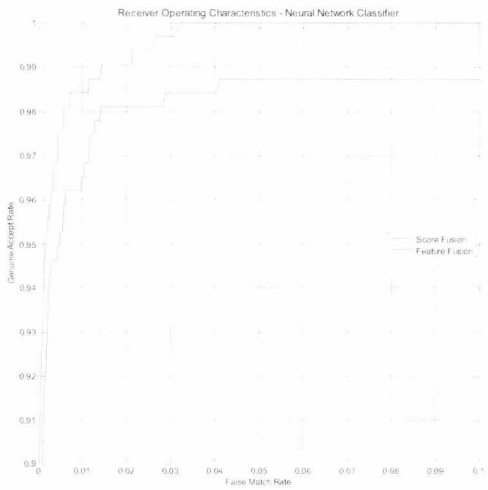


Figure 6.5 - NN Score and Feature Fusion (ROC Curve Enlarged)

The DET Curve for the neural network classifier (Figure 6.6) again highlights the disparity between score and feature fusion in this scenario with score fusion showing better overall performance.

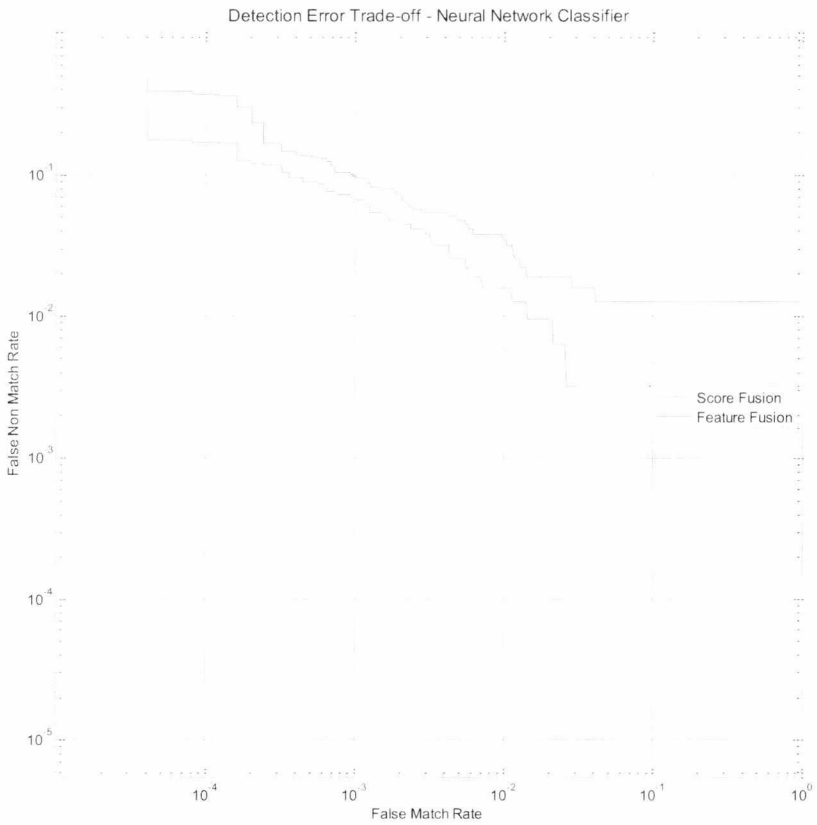


Figure 6.6 - Neural Network Score vs. Feature Fusion Performance (DET Curve)

Repeating the comparison with the linear classifier and support vector machine favours score fusion heavily over feature fusion which appears to be suffering from the curse of dimensionality [Bellman and Kalaba 1959] when the classifiers are supplied with a 48 dimension feature vector. In both instances however, the score fused classifier output shows a marked improvement over the previous separate regional analysis as performed previously with the support vector machine also showing a lesser but still noticeable improvement with score fusion when compared with individual face areas.

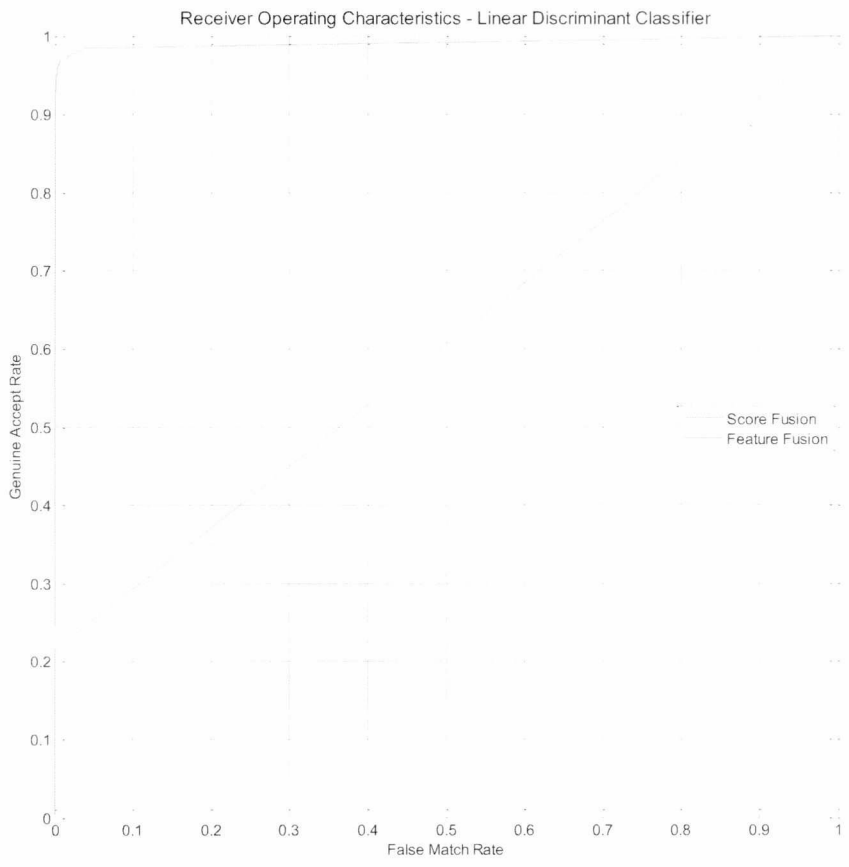


Figure 6.7 - Linear Classifier Score vs. Feature Fusion Performance (ROC Curve)

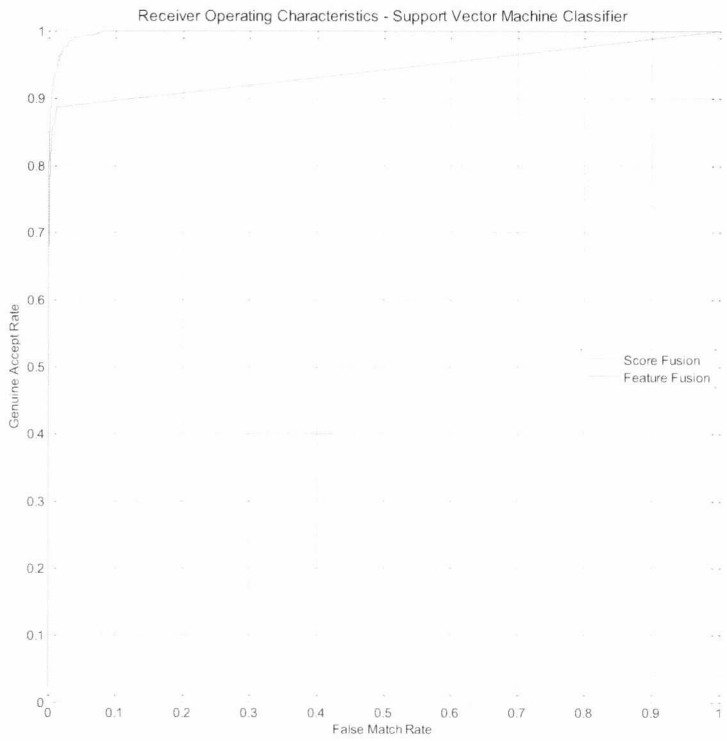


Figure 6.8 - Support Vector Machine Score vs. Feature Fusion Performance (ROC Curve)

Figure 6.9 - Linear Classifier Score vs. Feature Fusion Performance (DET Curve)

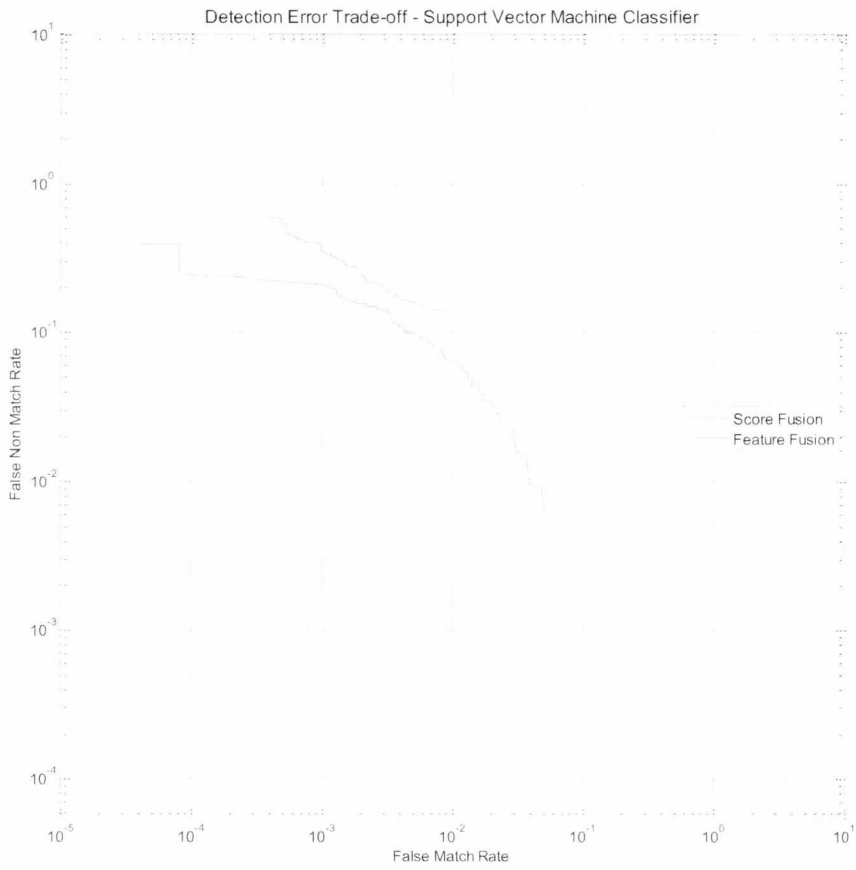


Figure 6.10 - Support Vector Machine Score vs. Feature Fusion Performance (DET Curve)

Inspecting the DET curves for the three classifiers indicates a general improvement using score based fusion against feature fusion, this is not to say that score fusion is in all instances an ideal candidate. The DET curve for the support vector classifier shows a noticeable curve indicating a potential abnormality in the output class distribution. Interestingly the linear classifier fails to register for feature fusion on the DET curve (see Figure 6.9) as the performance is so far below that of the score classifier.

Summarising the AUC and EER values from these experimental results below the performance of each classifier may be compared.

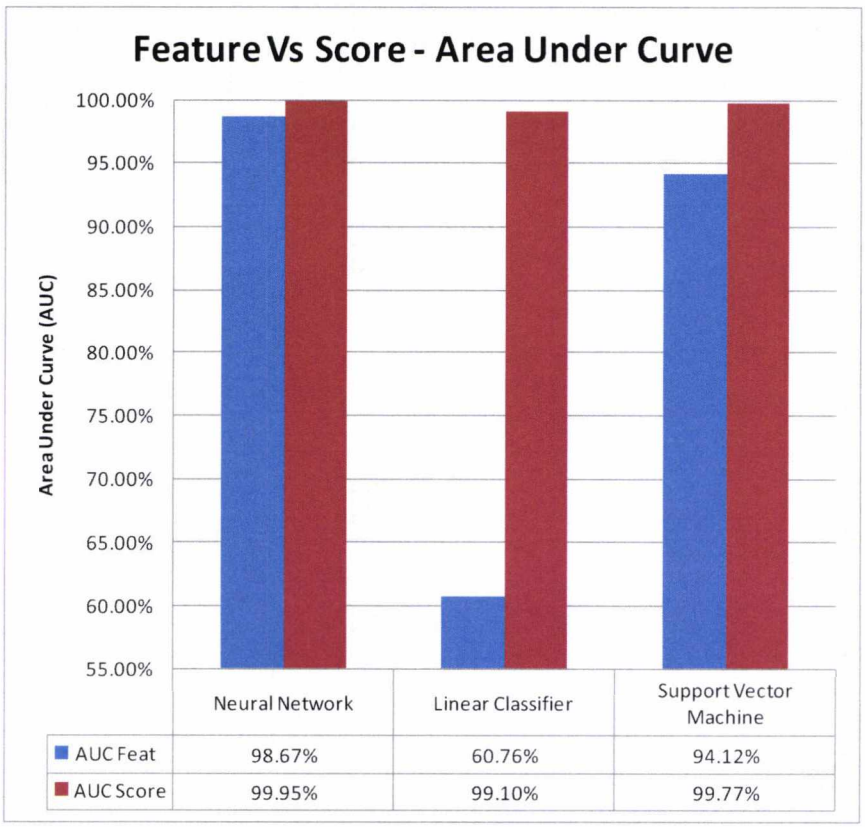


Figure 6.11 - AUC Comparison of Feature and Score Fusion (Gabor Features)

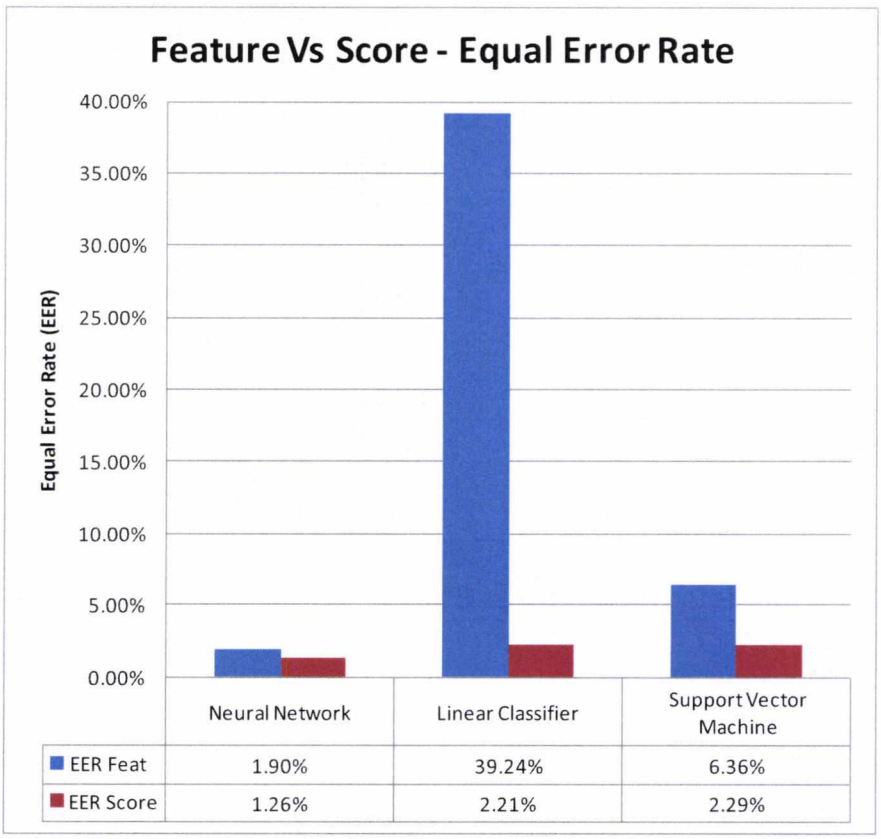
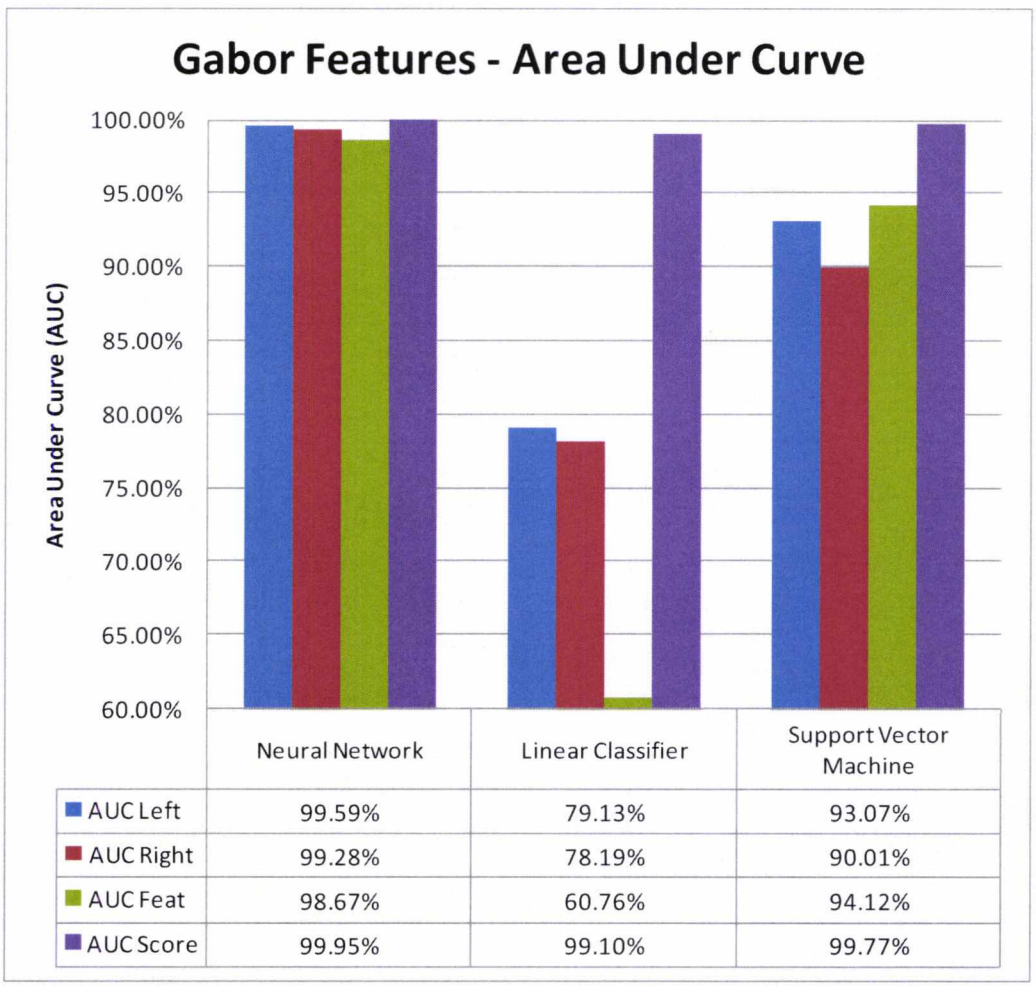


Figure 6.12 - EER Comparison of Feature and Score Fusion (Gabor Features)

6.3 PERFORMANCE OF INDIVIDUAL REGIONS VS. FUSED REGIONS

When comparing the various fused results alongside those of individual region of interest classification as performed in Section 5.7 the neural network classifier shows a marked performance lead over its rival classifiers for all input patterns. It should be highlighted however that the training of a neural network is several times more computationally intensive than either of the other two classifiers (this is discussed further in Chapter 6).

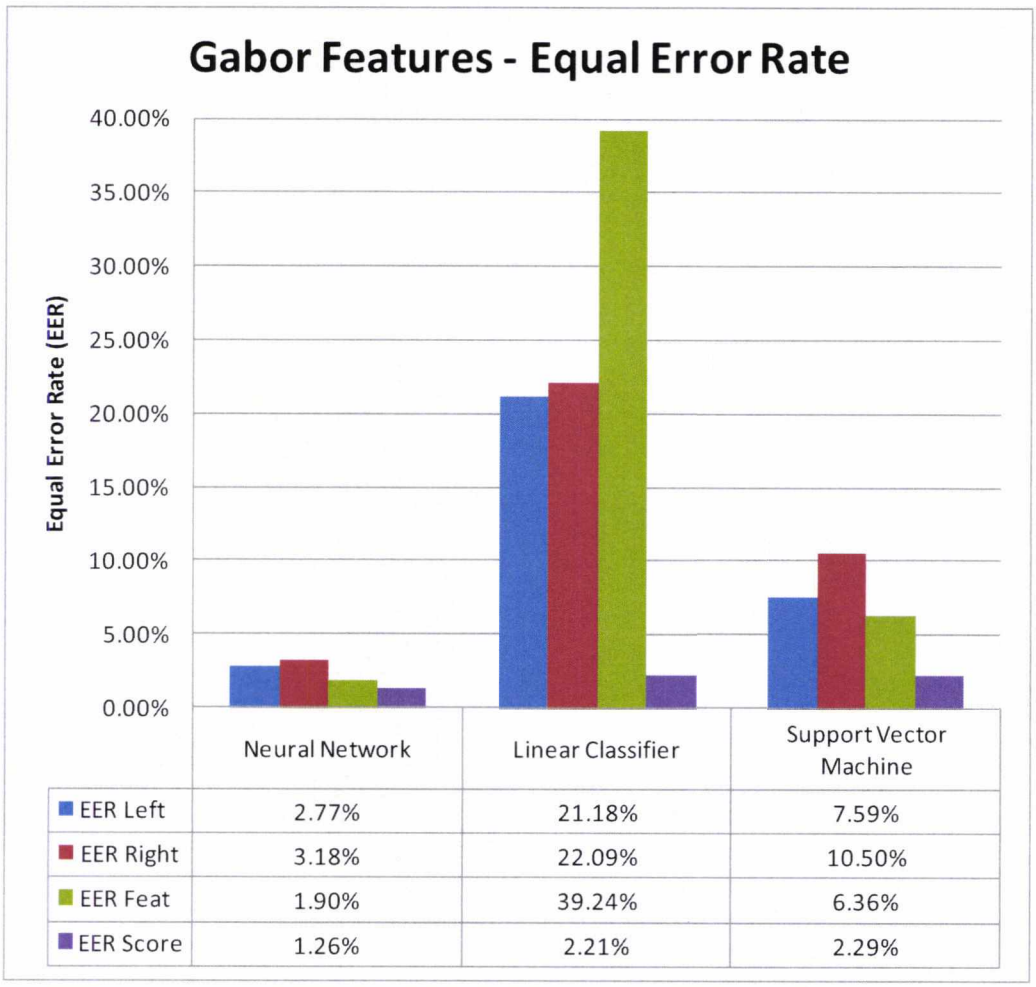
Worst performing of the classifiers is the linear classifier. Although able to achieve results of good accuracy when using score fusion classification this classifier suffers greatly when using feature fusion. Although much less pronounced than its Baysean linear counterpart, the neural network also demonstrates lower performance AUC value when provided with a feature fused input vector. Overall the neural network seems best equipped to handle both types of fusion with both types giving an AUC of over 99% and an EER of under 2%. The best results are obtained when using score fusion with a neural network classifier.



	Neural Network	Linear Classifier	Support Vector Machine
AUC Left	99.59%	79.13%	93.07%
AUC Right	99.28%	78.19%	90.01%
AUC Feat	98.67%	60.76%	94.12%
AUC Score	99.95%	99.10%	99.77%

Figure 6.13 - AUC Comparison of Score, Feature and Individual Gabor Feature Classification

Overall only the Support Vector Machine classifier shows an AUC based performance increase when using feature fusion. The most marked observation from Figure 6.13 is the factor of the performance enhancement brought by score fusion to all three classifier types.



	Neural Network	Linear Classifier	Support Vector Machine
EER Left	2.77%	21.18%	7.59%
EER Right	3.18%	22.09%	10.50%
EER Feat	1.90%	39.24%	6.36%
EER Score	1.26%	2.21%	2.29%

Figure 6.14 - EER Comparison of Score, Feature and Individual Gabor Feature Classification

Figure 6.15 shows the equal error rate box plot when taking averages from each classifier to group by region, or fusion method. From this plot it can be observed that score fusion presents a significantly lower overall error rate as well as a small variance spread. This indicates the stability of score fusion for grouping of Gabor features when compared with feature fusion.

It should be noted that a potential failing of feature fusion in this instance could be attributed to the curse of dimensionality and insufficient training data. An avenue for further research would be to investigate how feature fusion responds to varying input training sizes.

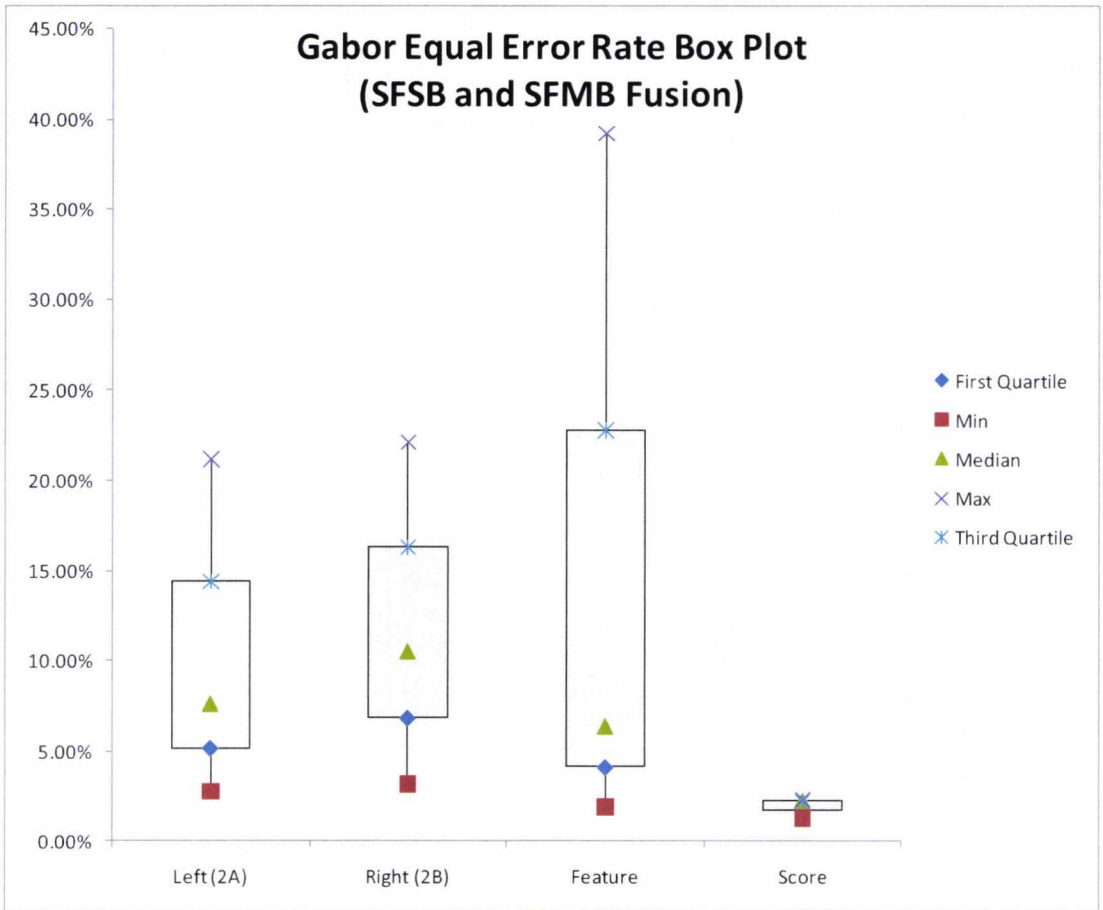


Figure 6.15 - Gabor Equal Error Rate Box Plot

In conclusion it may be seen that the use of fusion techniques can provide performance benefits to the system. This is particularly true for score fusion which provides stable outputs across all three classifiers. The choice of classifier used in feature fusion results appears to be a much more important consideration than with score.

6.4 Enhancing the System with Spatial Features

6.4.1 Limitations of the Frequency Domain

Wavelet based frequency transforms such as Gabor wavelet filters are noted for their ability to identify medium and low frequency patterns. Nyquist sampling theory states that accurate signal reconstruction is not possible when the frequencies are higher than half of the sampling frequency. In a 2 dimensional image, the maximum frequency is thus 0.5 cycles/pixel. However, [Shen, Bai et al. 2007] extended this to suggest that 0.25 cycles/pixel is in fact the maximum frequency for information that is usable in face recognition. This study has limited the Gabor frequency to a maximum f_{max} of 0.20 cycles/pixel.

Such limitations in frequency domain based filters mean that there is minimal effective capacity to extract any useful high frequency information from a source image. This may be desirable for some applications where high frequency data may not be required or may introduce errors (such as highly compressed images). In skin texture, there may be significant information hidden which could assist a system in providing an accurate identification of a subject.

Cox et al [G.S.Cox 1992] applied Co-occurrence matrix features [R.M. Haralick November 1973] to assist in the identification and diagnosis of lung cancer. The study showed that extracting pixel intensity relationships between nearby pixels allowed features to be generated that were representative of high frequency changes within an image and thus suitable to detect anomalous textures within an image.

Previous work on multi-biometrics from a single facial region have shown an improvement of overall system performance. [Hiremath and Danti 2006] [Yu, Teng et al. 2006].

6.4.2 Grey Level Co-Occurrence Probability Matrix (GLCP)

The Grey Level Co-Occurrence technique was first proposed by [R.M. Haralick 1973] as a means of extracting useful information from aerial imaging and satellite imaging for differentiating terrain types and rock photographs for differentiating between sandstone types. The operation of the GLCP system is straightforward and used in many practical applications ([Clausi 2001], [Clausi 2002], [Soh and Tsatsoulis 1999], [Latif-Amet, Ertuzun et al. 2000]).

In a typical greyscale image of known dimensions each pixel contains a value (generally an 8-bit value representing the grey-level of said pixel). The image is quantised if desired to reduce the number of grey levels (G). (See Figure 6.16 for an example).

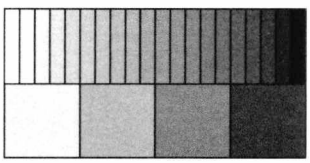


Figure 6.16 - Example Quantisation from 20 to 4 Gray Levels

Using a 2-Dimensional co-occurrence matrix with dimensions i and j equal to the number of grey levels (G), the original image is examined by looking at pixel (a) and its neighbour pixel (b) at a predefined pixel spacing and direction (distance δ and angle θ). These values construct a fixed-size window 'N'. The co-occurrence matrix is then completed by looking

up the grey level value of (a) on the x-axis, the grey level value of (b) on the y-axis and incrementing the matrix value (P) held at these coordinates.

$$P(i,j) = \Pr(i,j|\delta,\theta,G,N) \tag{6.8}$$

Once established, features are then extracted from the co-occurrence matrix that describe the frequencies identified by the matrix. [R.M. Haralick 1973] defines 14 GLCP features which may be used to describe the two dimensional probability function $P(i,j)$. [Kisner, Essoh et al. 2007] narrows this selection to six features. Although popular with some authors, correlation is cautioned against by [Jobanputra and Clausi 2006] with particular regard to high contrast images. Although facial texture is not generally considered a high contrast texture the use of this feature has been avoided as quite simply, there are features that do not suffer in this way. This minimises any potential impact from spectacles, birthmarks or other skin markings that may introduce a high contrast element.

[Barber and Ledrew] suggest that maximum texture discrimination for GLCPs occurs with three features. Hence three of the features from the list recommended by [Kisner, Essoh et al. 2007] have been selected for this study. Following a paper by [Baraldi and Parmiggiani 1995] the entropy and contrast (inertia) have been selected. The third feature to be used is homogeneity or angular second-moment (ASM) which indicates whether the GLCP contains a large number of small entries or a small number of large entries.

$$ASM = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{ij}^2 \tag{6.9}$$

$$Ent = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P_{ij} \log P_{ij} \tag{6.10}$$

$$Ine = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j)\cdot(i-j)^2 \tag{6.11}$$

where G indicates the number of grey levels and P the probability matrix at position (i, j).

6.4.3 OPTIMISING GREY LEVEL QUANTISATION FOR SKIN TEXTURE

In designing a GLCP matrix, there are three key values that must be decided. G (the number of grey-levels) in the image may be optimised to the most efficient number of levels required to identify the image [Randen and Husoy 1999] reports that for an image of 32x32 pixels, G is optimised at 8 levels.

Using the BMDB database in this study, typical image sizes vary from 120x120 pixels to 180x180 pixels. The optimum value of G will be ascertained by evaluating a series of greyscale quantisation levels on the original image and passing these through the three classifiers used elsewhere in this study for regions 2A and 2B.

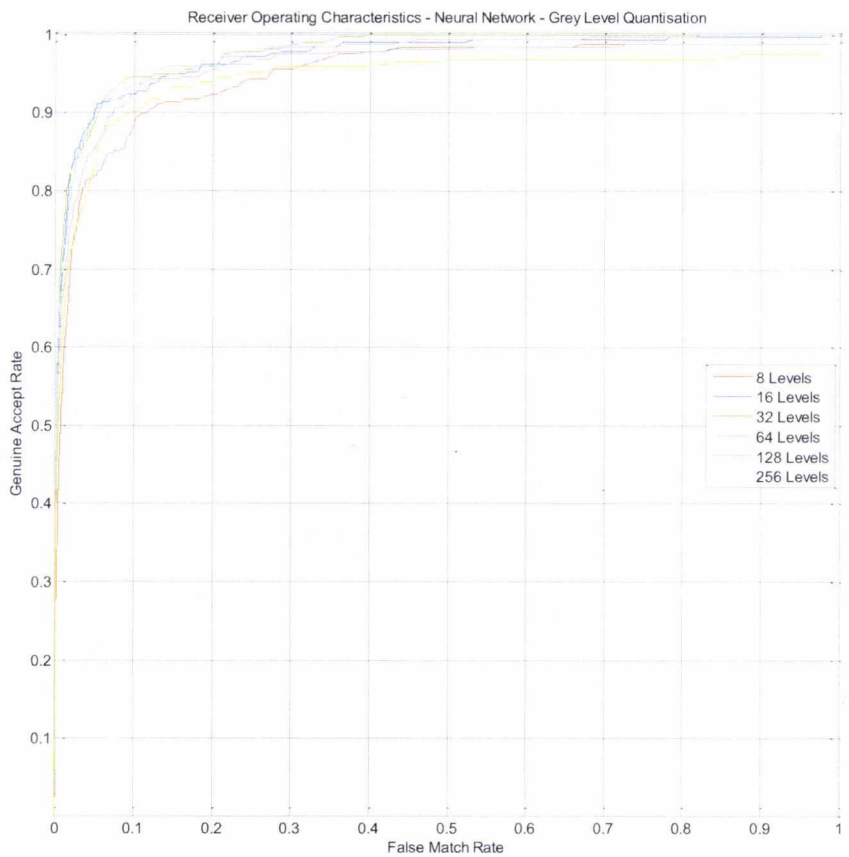


Figure 6.17 - ROC Curve for Neural Net (Region 2A - Left)

It is evident from Figure 6.17 and Figure 6.18 that the neural network shows a marginal performance increase when G = 32. Region 2B indicates an abnormal result at G = 16. However repeating the test could not eliminate this outlier and it is marked as a statistical anomaly, the DET curves further accentuate this anomaly.

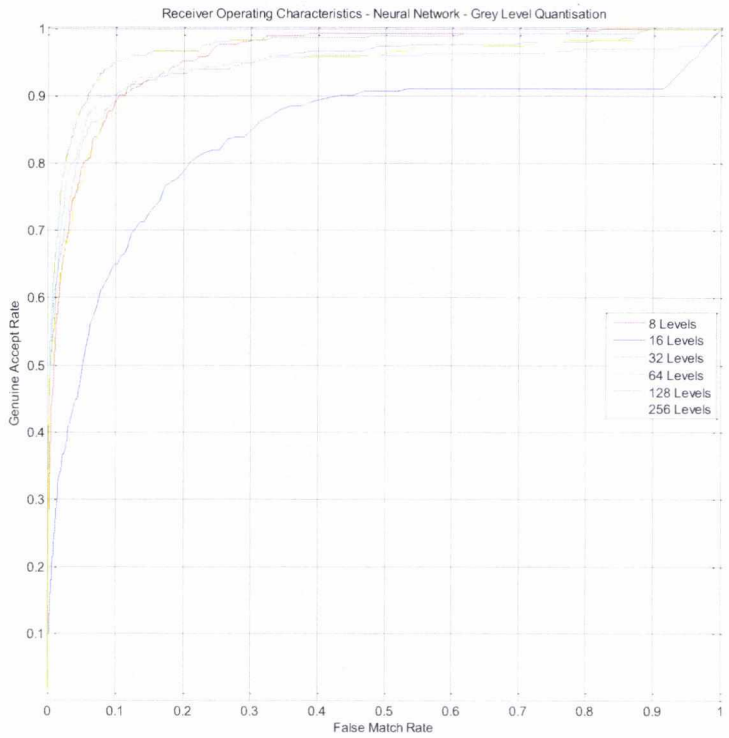


Figure 6.18 - ROC Curve for Neural Net (Region 2B - Right)

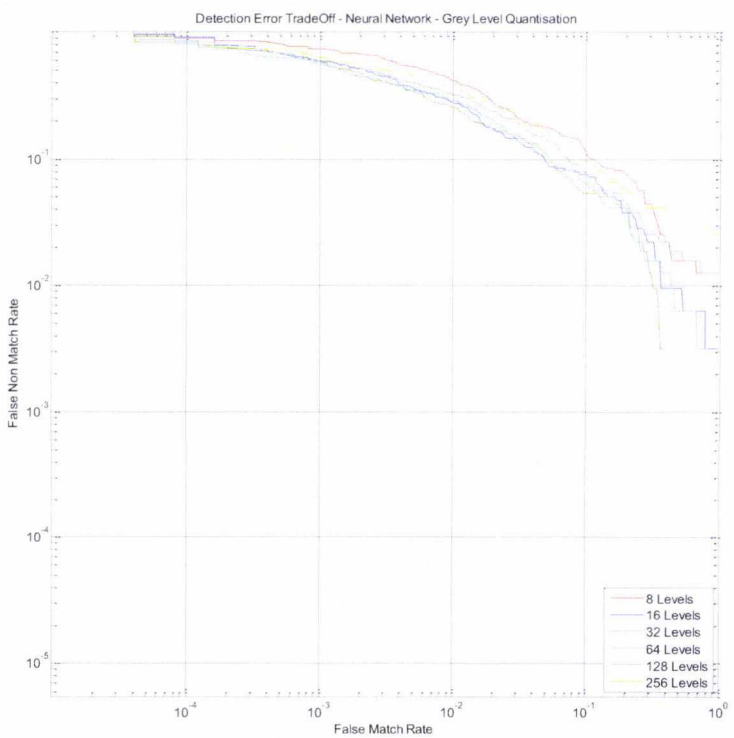


Figure 6.19 - DET Curve for Neural Net (Region 2A - Left)

Figure 6.20 - DET Curve for Neural Net (Region 2B - Right)

The sequencing of grey quantisation performance is similar when used with a Support Vector Machine classifier showing marginally improved performance with G = 16 and G = 32. The Linear Discriminate classifier appears to heavily favour low grey levels and shows optimal results when G = 8. The ROC and DET curves for the linear and SVM outputs are contained within Appendix A.

Amalgamating the area under the ROC curve (AUC) for the three classifiers at each grey level shows that the most confident results are encountered when G = 32. This is indicated by the tallest bar in Figure 6.21 and further sustained by Figure 6.22 showing the lowest overall equal error rate across all classifiers.

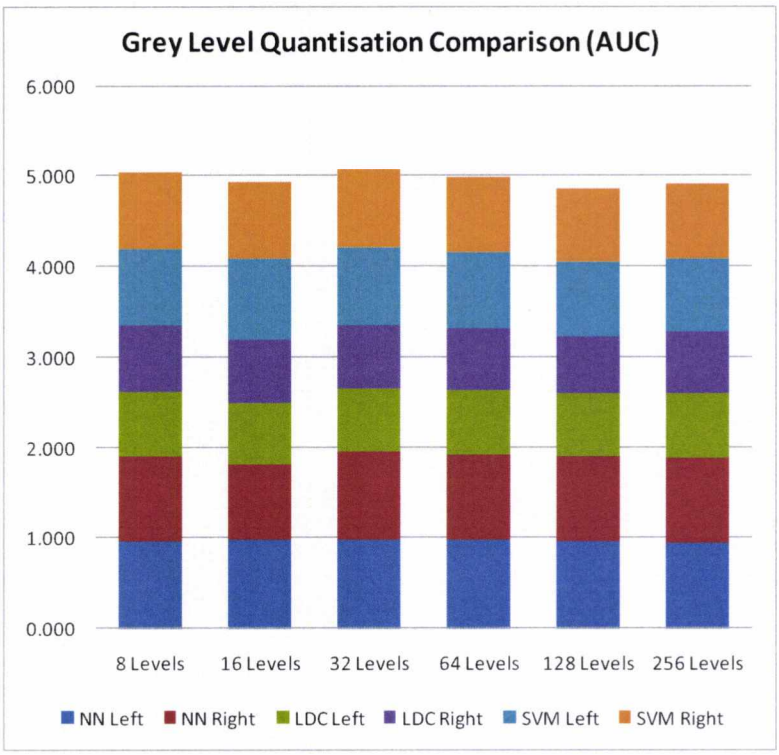


Figure 6.21 - Area under ROC Curve by grey level (GLCP)

The equal error rates shown in Figure 6.22 also support optimal results being provided with 32 level quantisation applied to the source image.

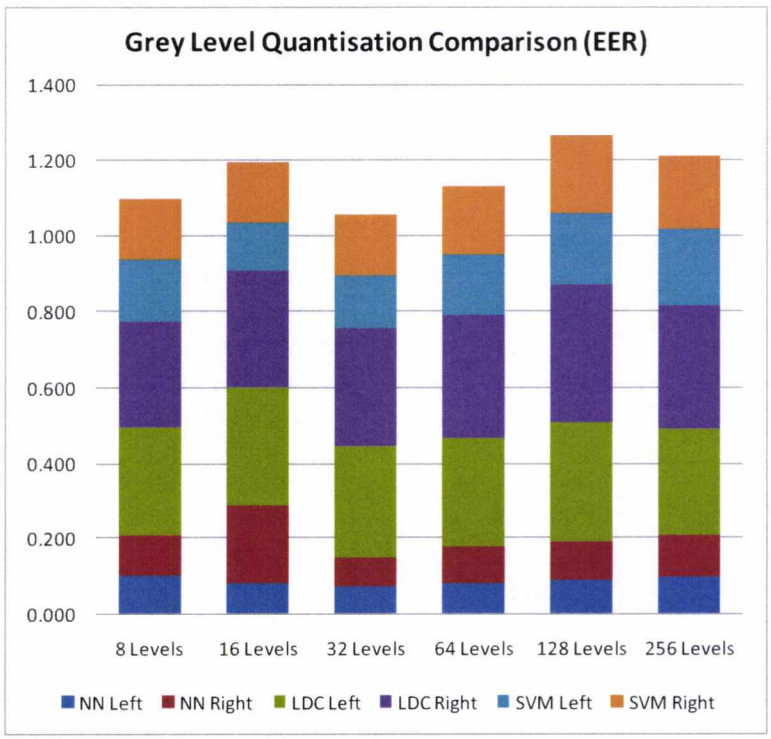


Figure 6.22 - Equal Error Rate (EER) by grey level (GLCP)

This marks an interesting point that may present an opportunity for investigative research. [Randen and Husoy 1999] propose that for a 32 x 32 pixel image, the optimum quantisation level for "G" is 8. The above observations suggest that for the size of texture under analysis, the optimum quantisation level for "G" is 32. It is hence plausible that the quantisation level is directly proportional to the size of the source image and there will hence be a falloff point where an image resolution may simply be too high to enable classifiers to accurately discriminate image features even when provided with 8 bit grey level quantisation (256 grey levels).

6.4.4 GLCP Distance and Angle Parameters

As the objective in this study of the GLCP algorithm is to identify high frequency patterns outside the range of detection of the Gabor Wavelets, the following parameters have been assigned:

Distance (δ)	Angle (θ)
1	0°
1	45°
1	90°
1	135°
2	0°
2	45°
2	90°
2	135°

Table 6.1 - GLCP Distance and Direction Parameters

Figure 6.23 and Figure 6.24 show the practical implementation of Table 6.1 in creating the GLCP matrix. Due to the movement of window '*N*' across the image, the information that would be retrieved by creating $\boldsymbol{\theta}$ angles of 180° - 360° would be duplicated. For this reason these angles are not implemented.

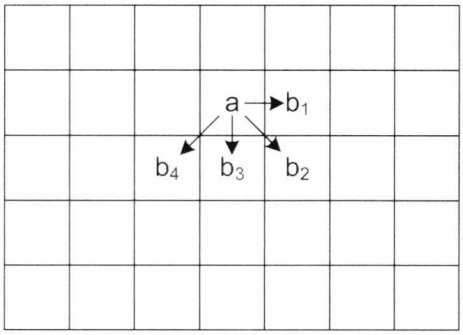


Figure 6.23 - GLCP at $\delta = 1$

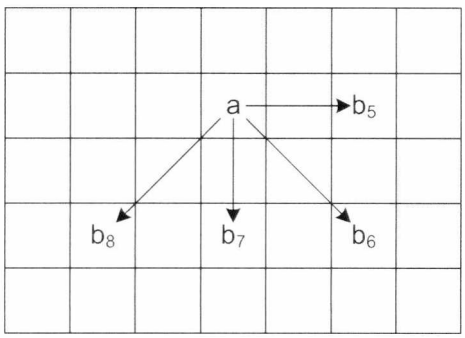


Figure 6.24 - GLCP at $\delta = 2$

6.5 GLCP Feature Classification of 2A and 2B

6.5.1 Evaluation Configuration

Using the defined parameters in GLCP feature extraction results in a 24 dimensional feature vector which is in turn fed to each of the three classifiers (neural network, linear discriminant and support vector machine).

All GLCP feature extractions are preceded by quantising the image to 32 grey levels as justified in Section 6.4.3 The overall process for the GLCP system is shown in Figure 6.25. After quantisation, the GLCP matrices are generated for each permutation of δ and θ. (6.9), (6.10) and (6.11) (page 106) are then used to find the angular second-moment, entropy and inertia which are fused into a 24 dimension feature vector prior to classification.

This process is performed individually for region 2A and 2B and subsequently with fusion of score and feature fusion to assess the complete capabilities of the GLCP features in this scenario and compare this against the Gabor feature classification provided earlier in Chapter 5 and Section 6.2 . To avoid duplication, equations (6.2) - (6.7) may be referenced to the score and feature fusion methods used in this section.

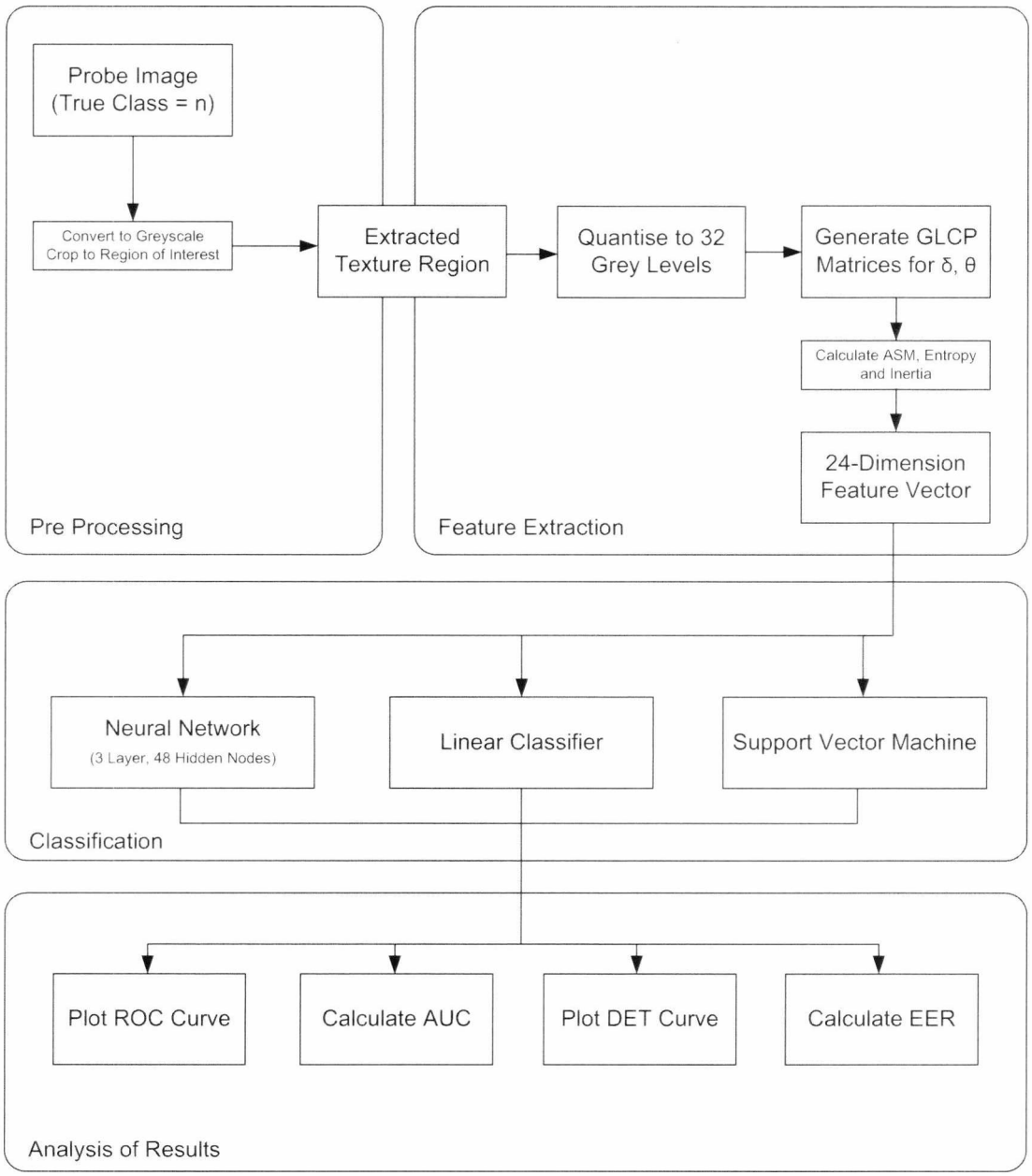


Figure 6.25 - GLCP Single Region Process

Figure 6.26 and Figure 6.27 detail the processes of pre-processing, feature extraction, classification and analysis stages for both score and feature fusion when applied to the GLCP system.

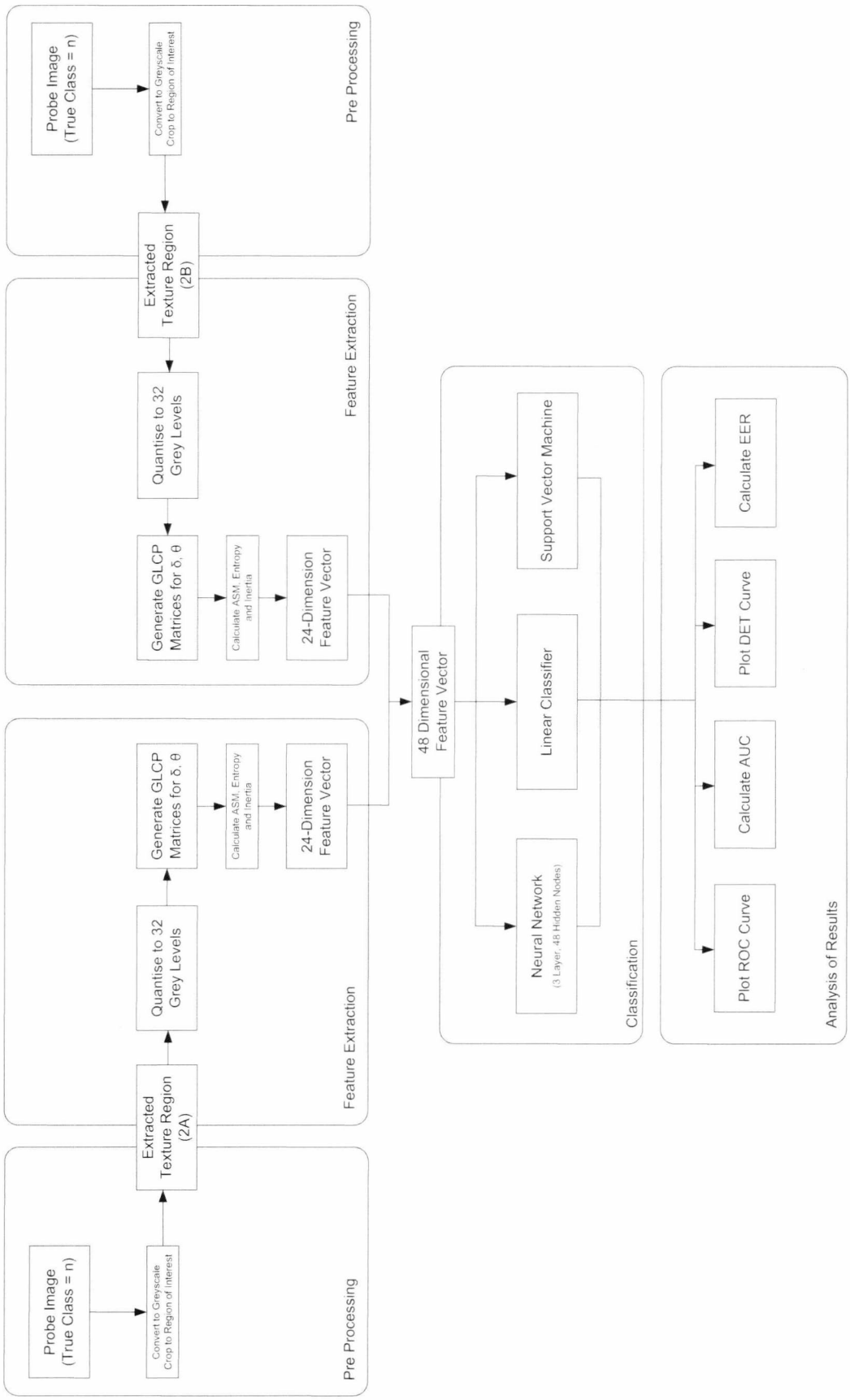


Figure 6.26 - GLCP Feature Fusion Process

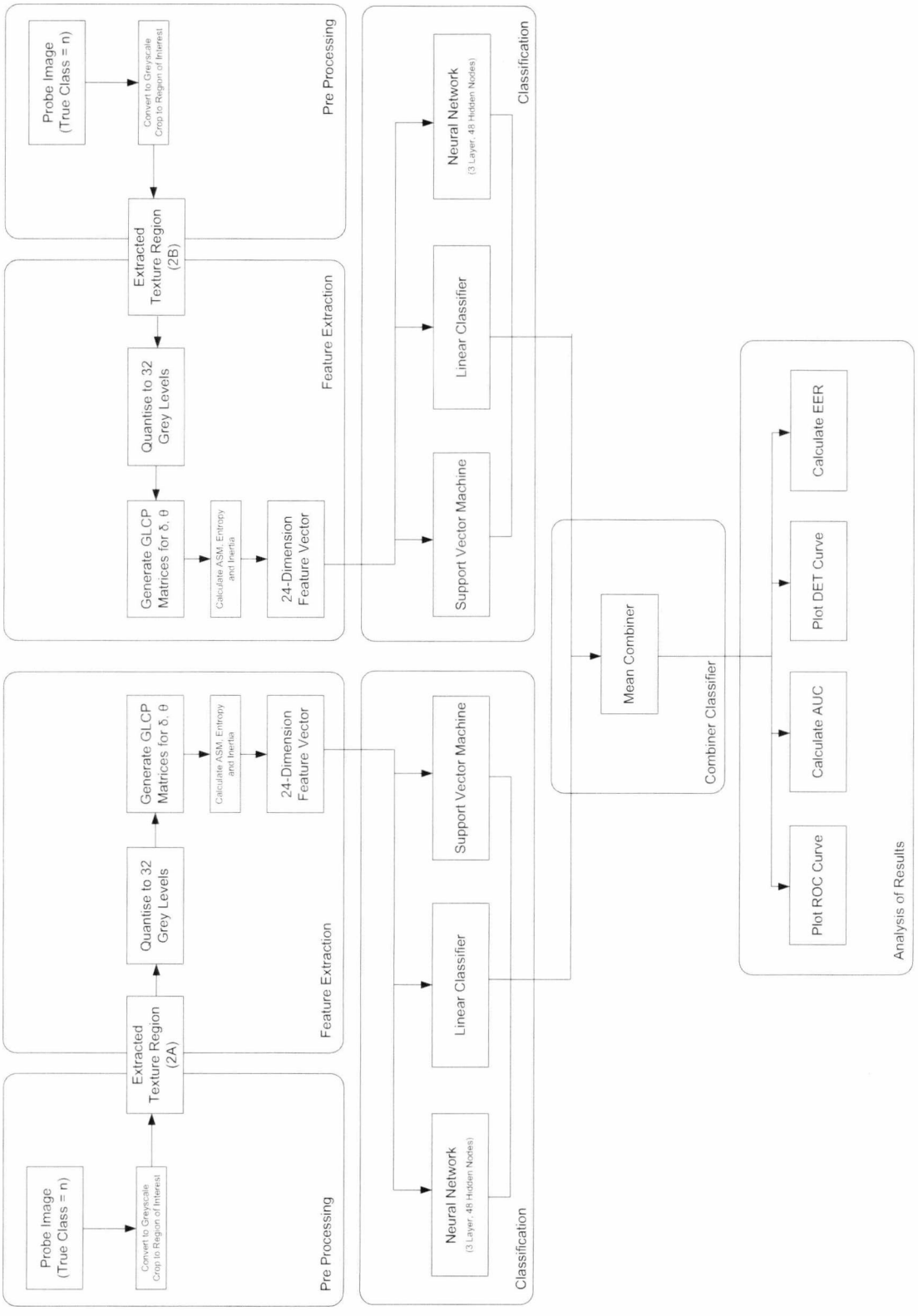


Figure 6.27 - GLCP Score Fusion Process

6.5.2 EVALUATION RESULTS (SFSB VS. SFMB GLCP)

The individual 2A and 2B results were presented as part of the GLCP Grey Level analysis earlier in this chapter. To avoid duplication by presenting these results independently, they have been included in the appropriate ROC and DET curves with the score and feature fusion results below.

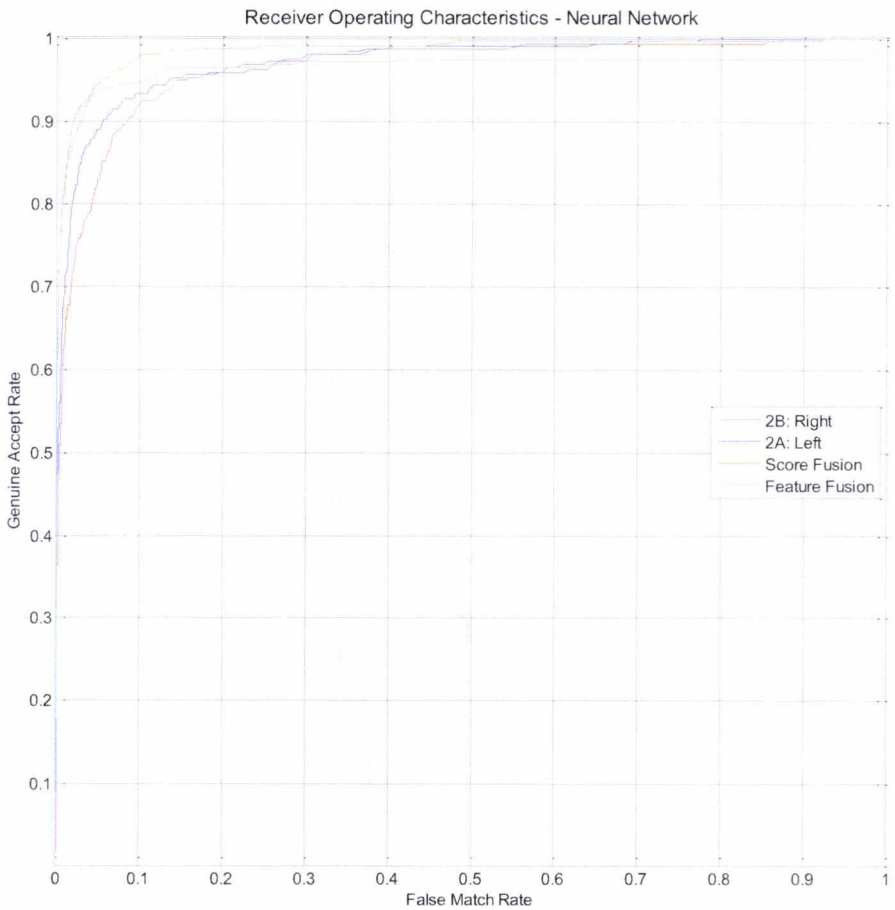


Figure 6.28 - Neural Network Receiver Operating Characteristics (GLCP)

Following the expectations set by the Gabor wavelet results, the score fusion presents a much closer curve to the ideal (y = 1). Examining the DET curve in Figure 6.29 confirms these findings showing soft curves for both fusion techniques indicating a well distributed classification output. Even when considered alone, the individual classification for 2A and 2B shows an impressive level of accuracy for the relatively limited frequency features extracted by the GLCP.

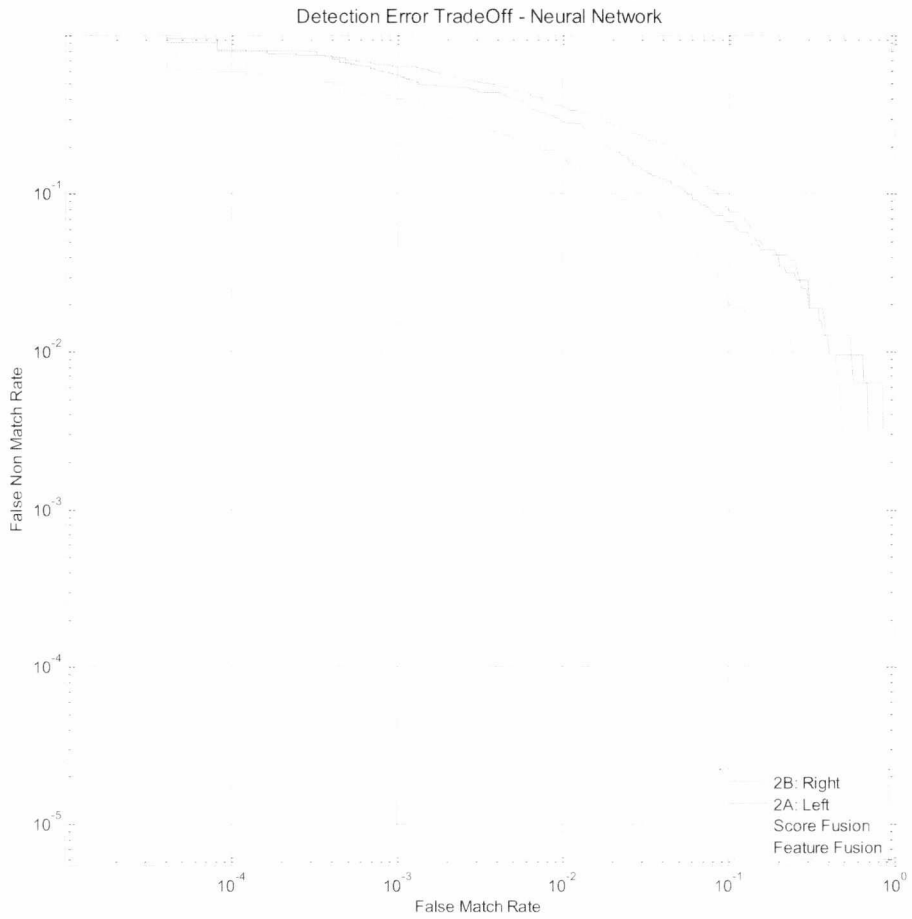


Figure 6.29 - Neural Network Detection Error Trade-off (GLCP)

The linear classifier also shows excellent potential for GLCP classification with score fusion applied but performs very poorly when confronted with feature fusion. When provided with a feature converged 48 dimensional co-occurrence vector the system fails to provide satisfactory results. In the DET curve (Figure 6.31) the feature fusion curve falls out of the range displayed as observed with the Gabor features. This could again be due to the curse of dimensionality or insufficient training data as observed during a similar configuration using Gabor filters in Section 6.2 .

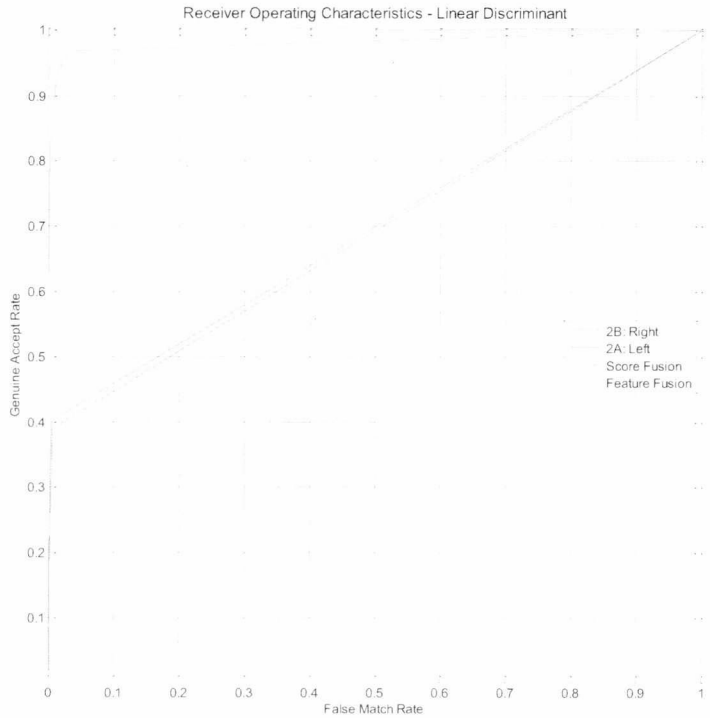


Figure 6.30 - Linear Classifier Receiver Operating Characteristics (GLCP)

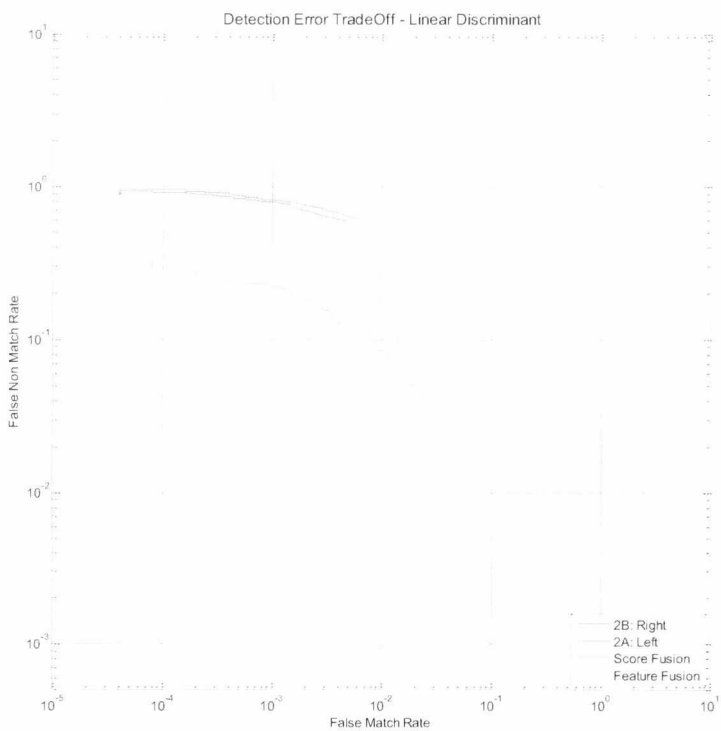


Figure 6.31 - Linear Classifier Detection Error Trade-off (GLCP)

A similar pattern is observed for score fusion in the support vector machine ROC curve, however both feature fusion and separate regional classification fails to give as satisfactory results (Figure 6.32). Although all outputs are well above the random classification line (thus indicating a measure of successful identification) there is a disparity between the results obtainable with score fusion and the rival techniques. Evidently score fusion again gives a substantial advantage.

The DET curve (Figure 6.33) confirms these findings showing a decreased overall error rate for score classification.

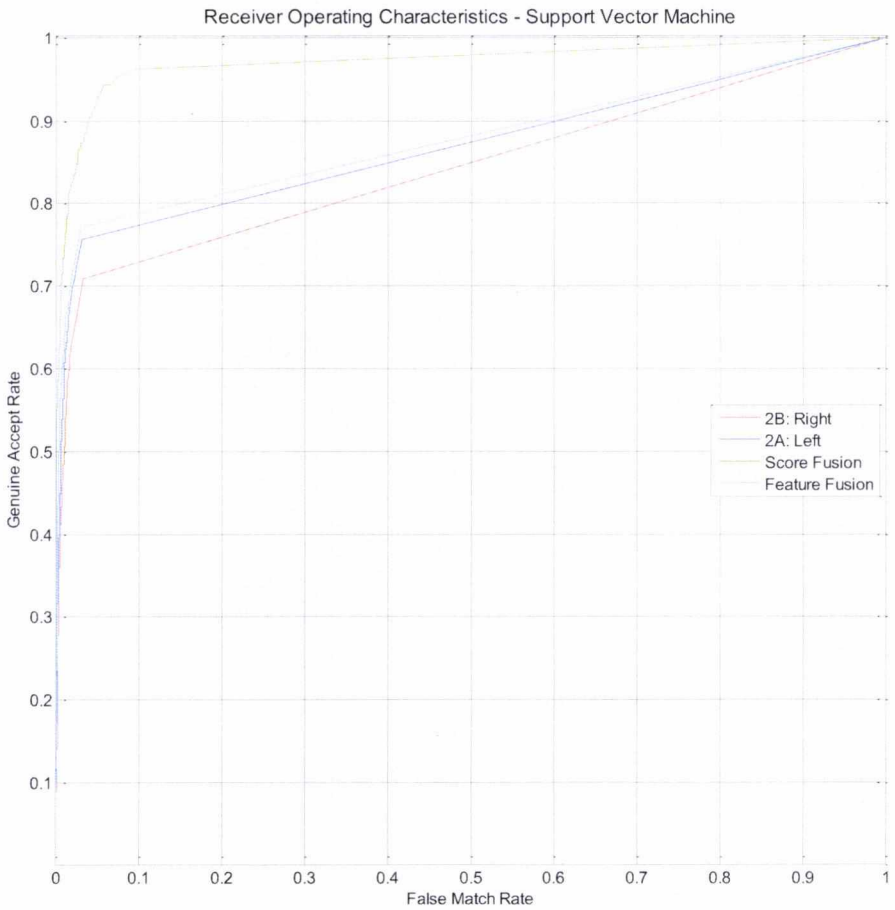


Figure 6.32 - Support Vector Machine Receiver Operating Characteristics (GLCP)

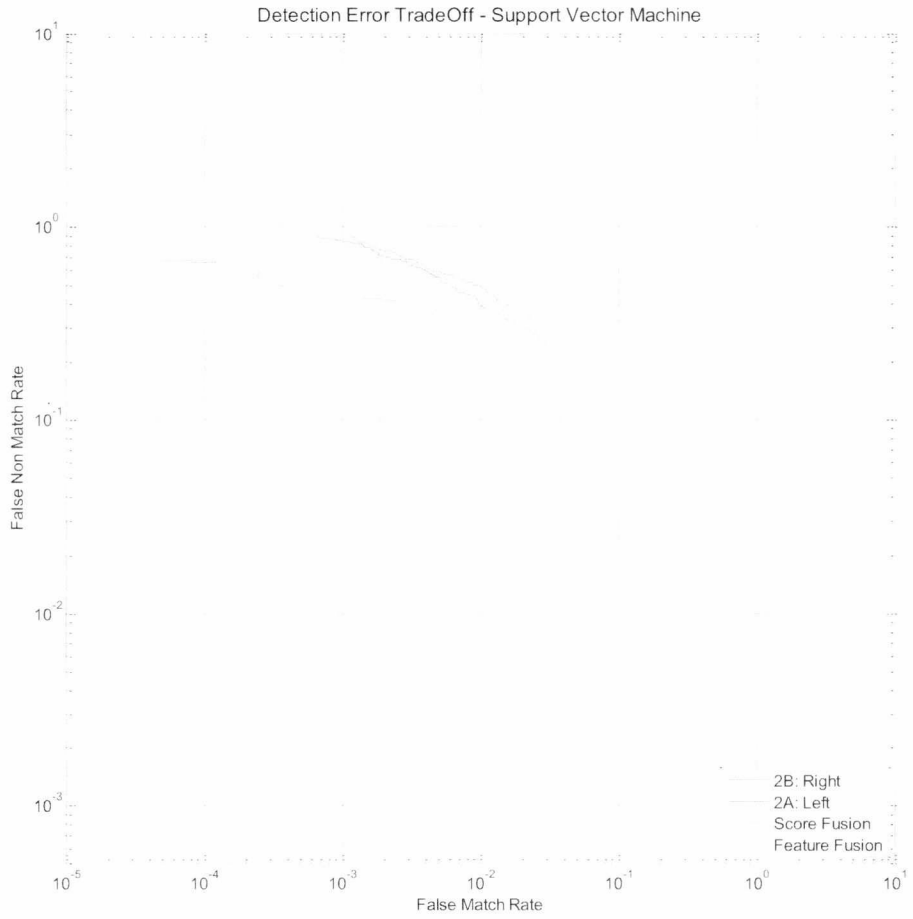
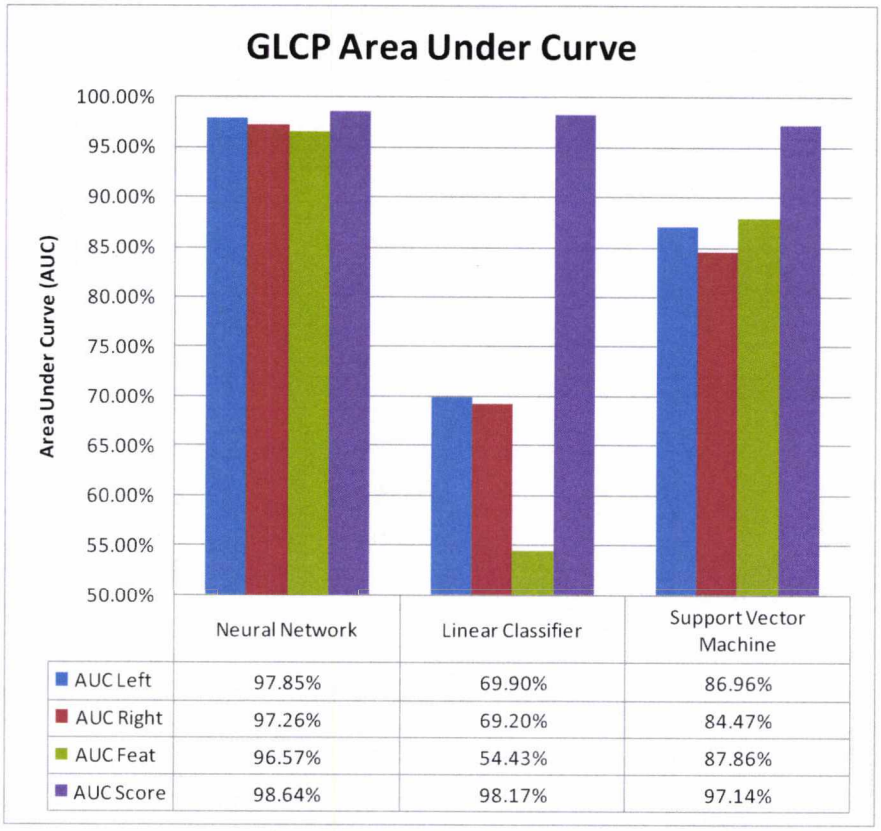


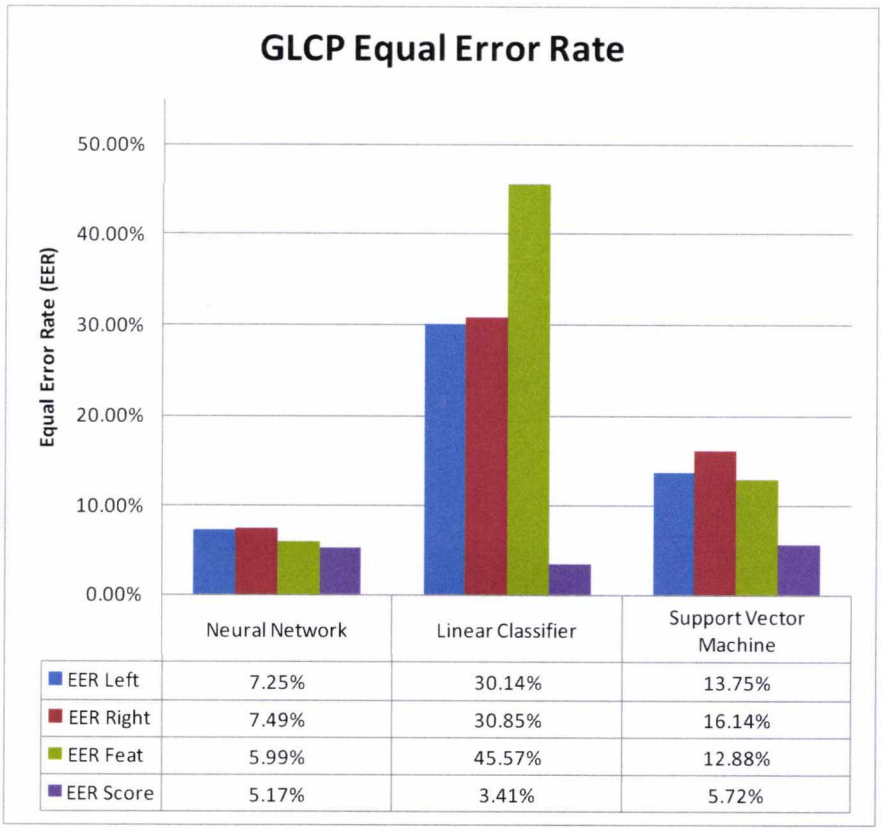
Figure 6.33 - Support Vector Machine Detection Error Trade-off (GLCP)

Comparing the area under each receiver operating characteristic curve allows a simple side by side comparison of the performance of each organisational permutation and classifier. Figure 6.34 clearly shows the lead of the neural network over the linear discriminant classifier and support vector machines. It also highlights the shortcomings of the linear classifier when dealing with large vectors produced with GLCP features. In contrast to this, the support vector machine struggles to maintain equivalent performance to the neural network, but does manage to provide consistent results.



	Neural Network	Linear Classifier	Support Vector Machine
AUC Left	97.85%	69.90%	86.96%
AUC Right	97.26%	69.20%	84.47%
AUC Feat	96.57%	54.43%	87.86%
AUC Score	98.64%	98.17%	97.14%

Figure 6.34 - GLCP AUC Classification Results for Regions 2A and 2B



	Neural Network	Linear Classifier	Support Vector Machine
EER Left	7.25%	30.14%	13.75%
EER Right	7.49%	30.85%	16.14%
EER Feat	5.99%	45.57%	12.88%
EER Score	5.17%	3.41%	5.72%

Figure 6.35 - GLCP EER Classification Results for Regions 2A and 2B

By combining the equal error rate statistical averages, quartiles and extremes for each output permutation and constructing a box plot, the advantages of score classification for GLCP features becomes apparent.

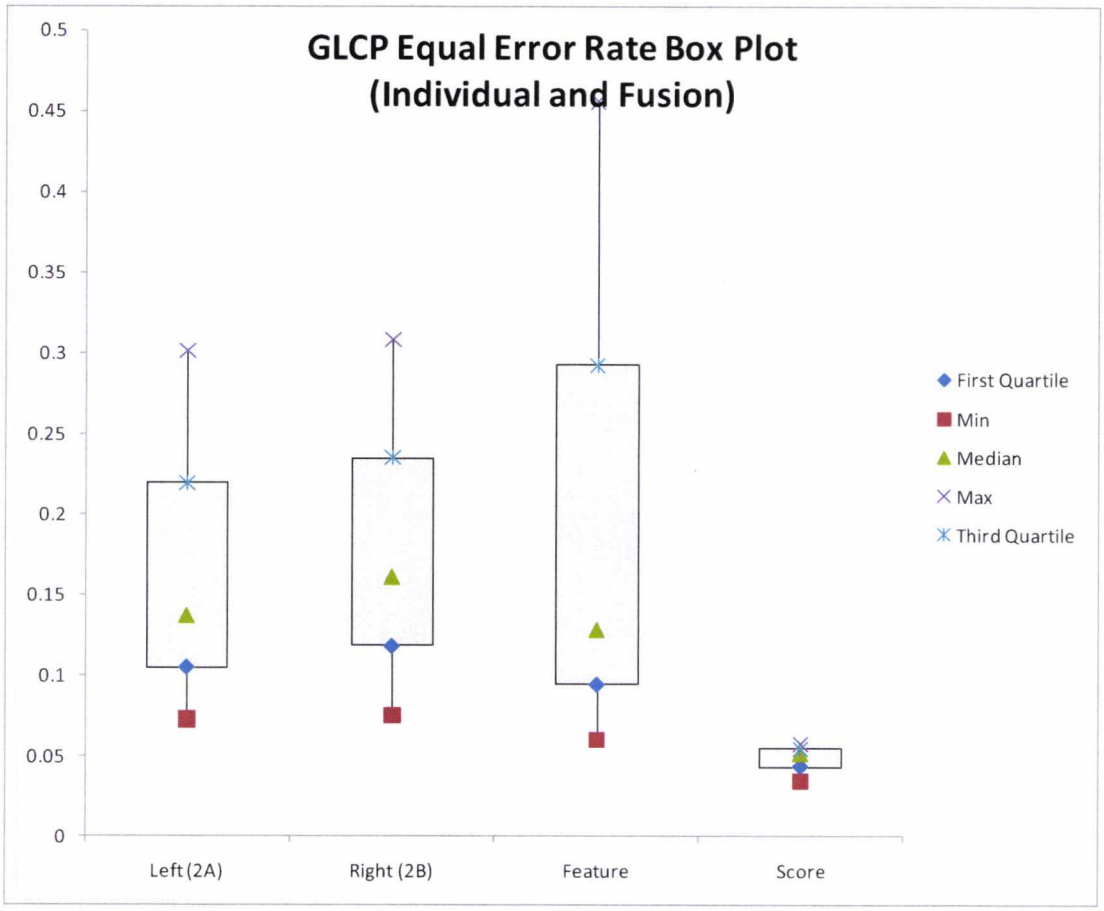


Figure 6.36 - GLCP EER Box Plot

The box plot indicates the level of variance between classifiers per feature set. The centre box represents the middle 50% of the data sample with the upper and lower whiskers showing the first and third quartiles. The green triangle indicates the median equal error rate. Of particular interest is the kurtosis of the distribution which shows the size of the peak, or how close the classifier error rates are. A good kurtosis will be shown by a small box while a spread distribution is shown by a larger box such as the feature fusion example in Figure 6.36.

6.5.3 CONCLUSIONS: GLCP FEATURE IDENTIFICATION PERFORMANCE

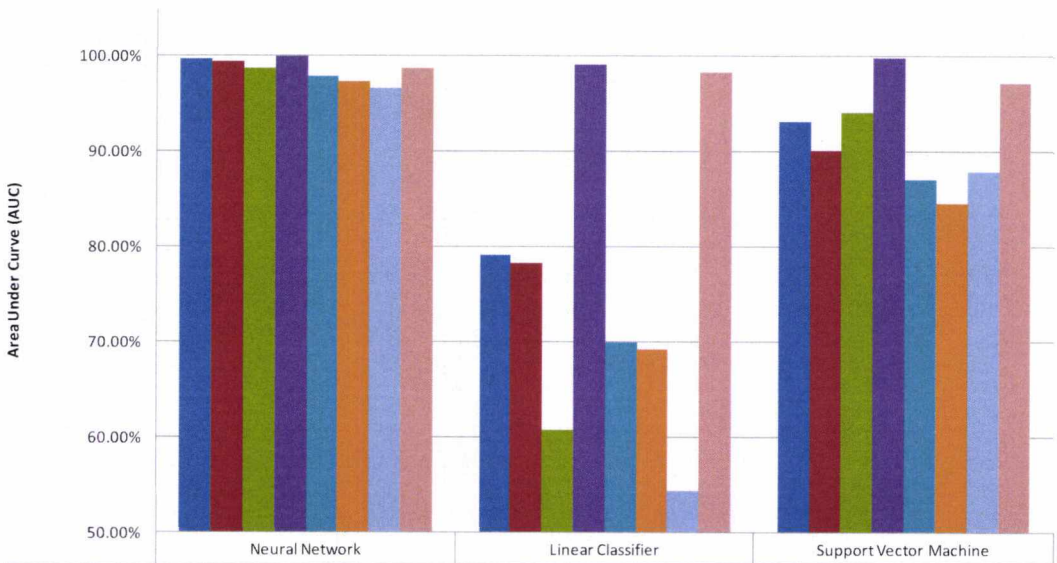
The grey level co-occurrence probability matrix has shown a strong capability for skin texture recognition by identifying high frequency pixel patterns within an image. A quantisation level for the GLCP system has been identified and this has been used to produce

optimal features for classifier discrimination. A theory has been proposed connecting the optimum quantisation level to the input image size.

It has been identified that score fusion presents the best way of combining the multi-biometric information from regions 2A and 2B.

6.6 FREQUENCY VERSUS SPATIAL FEATURES

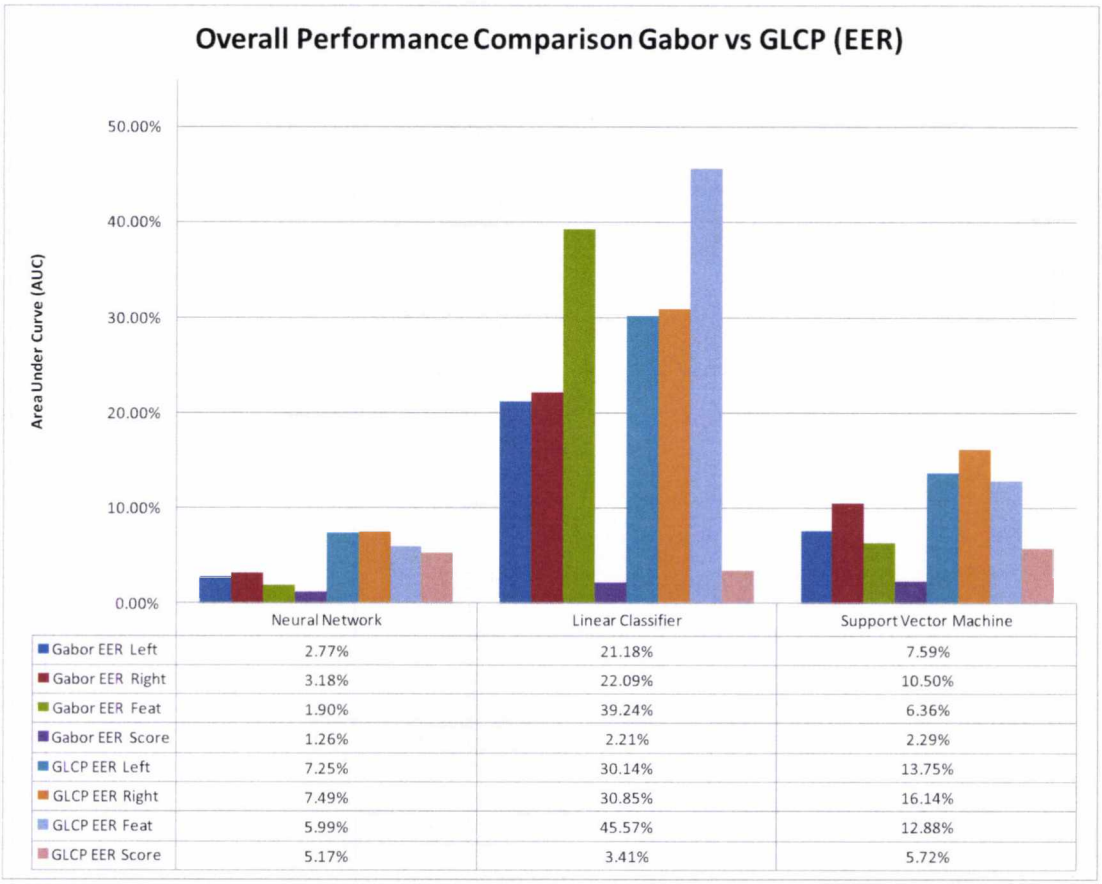
The GLCP was initially considered to be a secondary supporting feature set to enhance the medium to low frequency patterns identified by the Gabor wavelets in Chapter 5 . Recent sections have suggested that in fact the GLCP method proposed is capable of independent feature extraction and classification to produce impressive results in its own right.



	Neural Network	Linear Classifier	Support Vector Machine
Gabor AUC Left	99.59%	79.13%	93.07%
Gabor AUC Right	99.28%	78.19%	90.01%
Gabor AUC Feat	98.67%	60.76%	94.12%
Gabor AUC Score	99.95%	99.10%	99.77%
GLCP AUC Left	97.85%	69.90%	86.96%
GLCP AUC Right	97.26%	69.20%	84.47%
GLCP AUC Feat	96.57%	54.43%	87.86%
GLCP AUC Score	98.64%	98.17%	97.14%

Figure 6.37 - Gabor vs. GLCP Comparison (AUC)

Comparing Gabor Wavelet features with the GLCP matrix features side by side reveals that the two systems are closely matched. While there can be little doubt that Gabor based features offer superior performance, both sets of results are merit worthy in their own right.



	Neural Network	Linear Classifier	Support Vector Machine
Gabor EER Left	2.77%	21.18%	7.59%
Gabor EER Right	3.18%	22.09%	10.50%
Gabor EER Feat	1.90%	39.24%	6.36%
Gabor EER Score	1.26%	2.21%	2.29%
GLCP EER Left	7.25%	30.14%	13.75%
GLCP EER Right	7.49%	30.85%	16.14%
GLCP EER Feat	5.99%	45.57%	12.88%
GLCP EER Score	5.17%	3.41%	5.72%

Figure 6.38 - Gabor vs. GLCP Comparison (EER)

Figure 6.39 shows a side by side comparison of the Gabor and GLCP features across all three classifiers and provides an indicator to the stability of the features across a base of classifiers. Here, the advantage of the score fusion techniques for both feature spaces is clear with optimum performance achieved using the Gabor feature set. It all configurations, the GLCP performance tails slightly behind the Gabor performance.

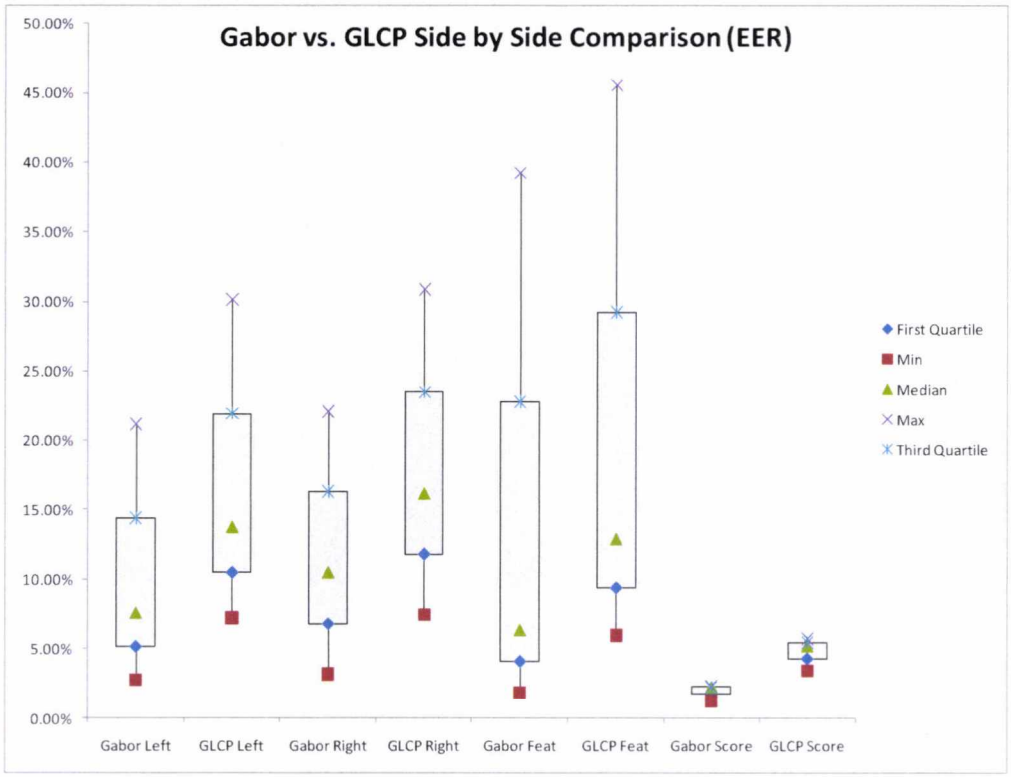


Figure 6.39 - Gabor vs. GLCP Side by Side Box Plot Comparison (EER)

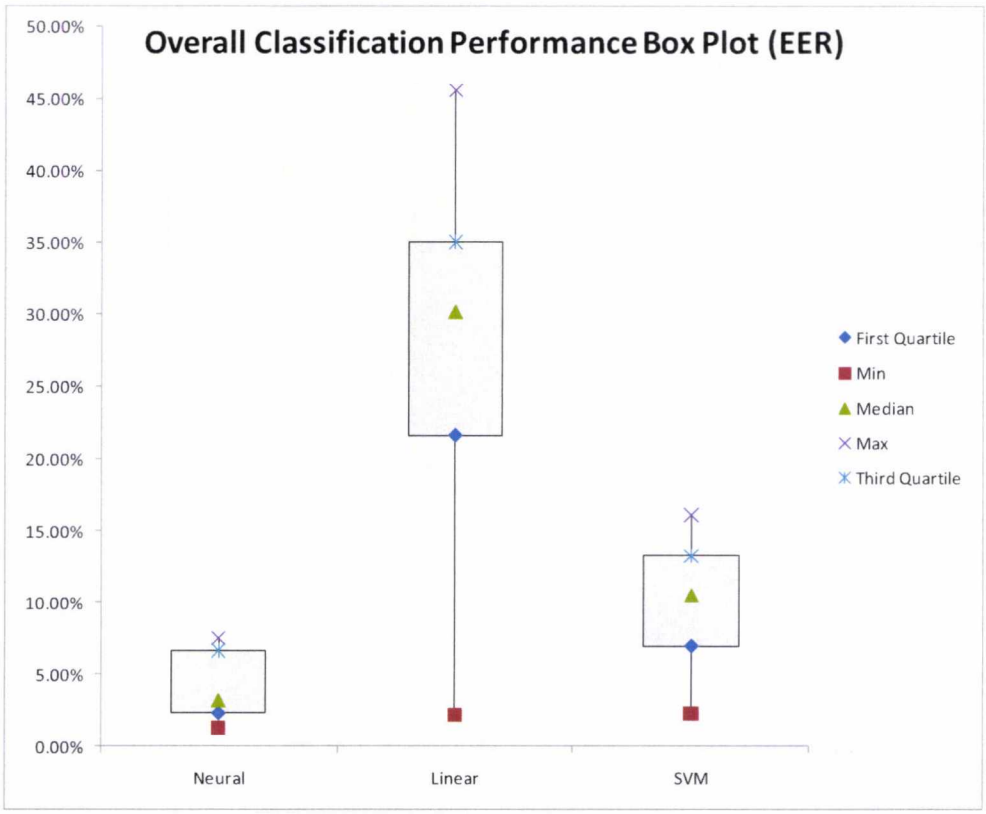


Figure 6.40 - Equal Error Rate Overall Classifier Performance Box Plot

Evident from the preceding figures is the performance lead maintained by the neural network in accurately identifying subjects from the source probe images when provided with either feature set. By representing the equal error rate averages for each classifier type across all input patterns, the performance differences across the classifier baseline become even more apparent (Figure 6.40).

It is observed from this plot that the linear discriminant classifier suffers with large dimension feature vectors.

6.7 Fusion of Frequency and Spatial Features

Although it has been demonstrated that Gabor and GLCP feature based classification exhibits similar performance trends when passed through the three classifiers, the feature set acquired by each method is inherently unique due the distinct frequency coverage (see Figure 6.41). This should provide an excellent foundation for more complete texture feature coverage and should provide a classifier with more useful (as opposed to repetitious) information with which to work.

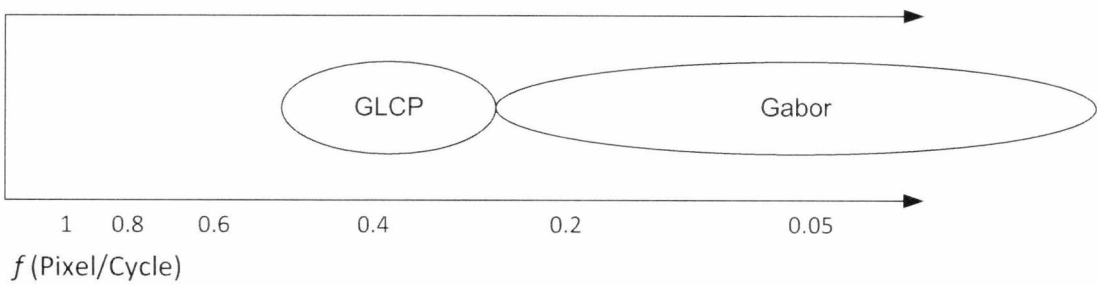


Figure 6.41 - Frequency Coverage of GLCP and Gabor Filters

It should however be noted that GLCP features may show a greater dependence on the scale of the supplied images. This will be examined further in Chapter 7.

In order to provide a complete review of the systems performance and potential capabilities, all permutations of classifier inputs shall be considered (see Figure 6.42). In total there are 16 possible configurations of independent classification, score and feature fusion.

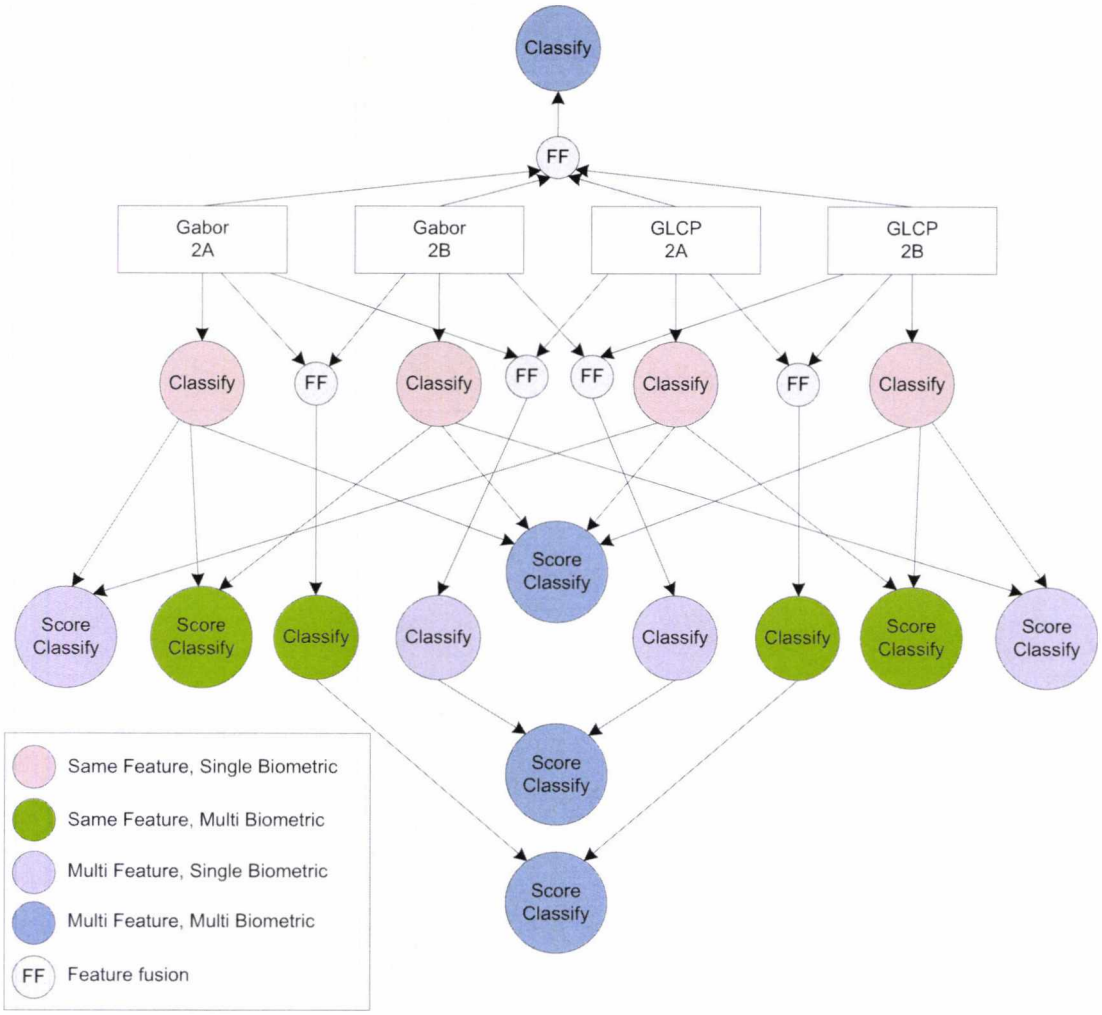


Figure 6.42 - Classification Permutations

For the purposes of clarity within this section the term "same-feature set" is used to describe systems using a single type of features (ie. Gabor or GLCP). Inversely, "multi-feature set" is used to describe systems containing both Gabor and GLCP features.

Likewise "single biometric" is used to describe systems based on a single skin region. Multi-biometric is used to describe systems based on two or more skin regions.

Using these definitions, it can be stated that all possible same-feature set permutations have been previously described in this study (in Figure 6.42 single feature systems are coloured light red and green). Subsequent sections of this document focus on multi-feature set systems.

6.8 MULTI FEATURE TYPE SINGLE REGION (MFSB)

6.8.1 EVALUATION CONFIGURATION

This section considers combining Gabor and GLCP features when handling a single biometric region (ie. a single region 2A or 2B). In Figure 6.42 these instances are highlighted in light purple. The process diagrams show the system differences between feature and score fusion.

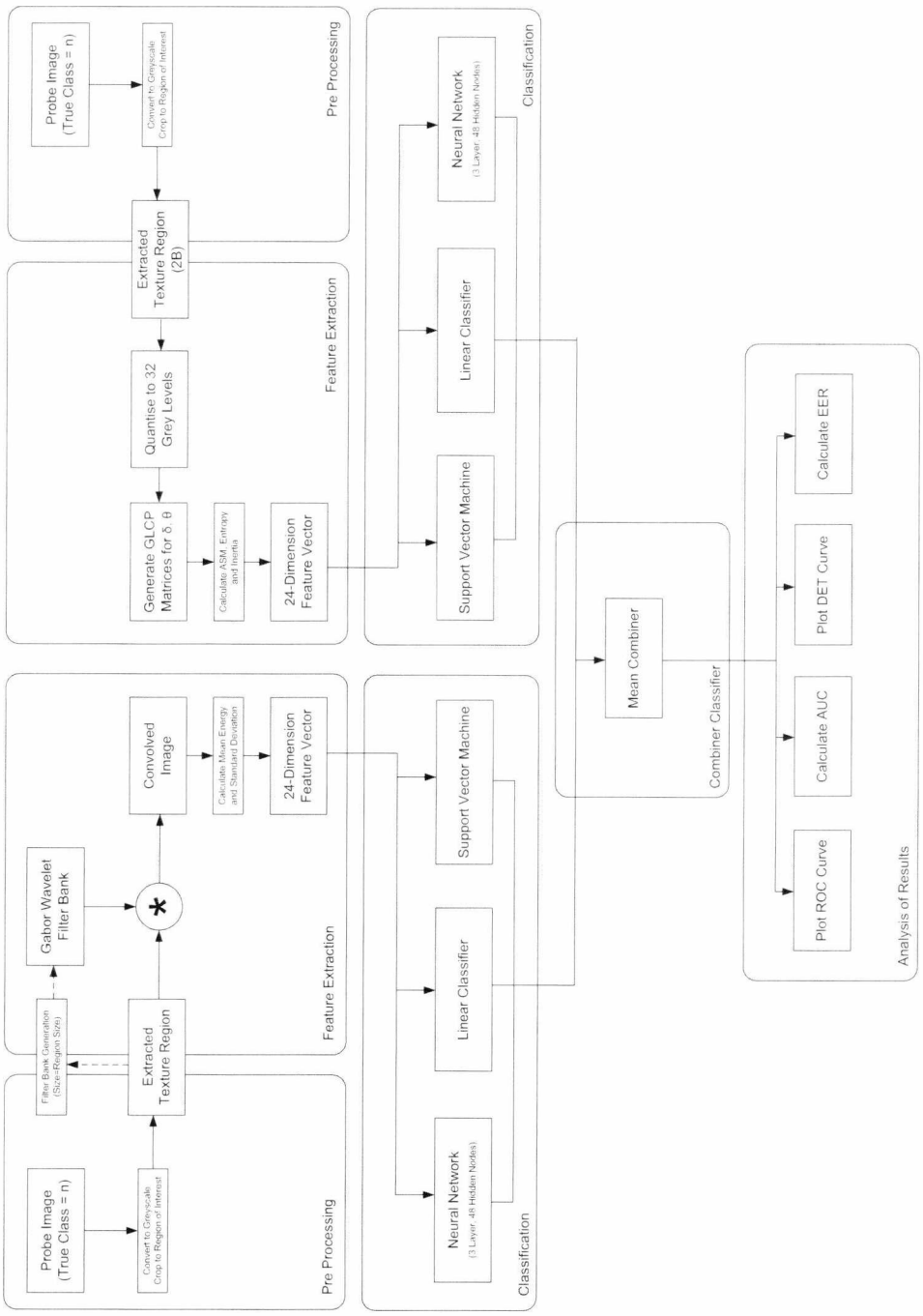


Figure 6.43 - Multi Feature Type, Single Biometric Region Score Fusion Process

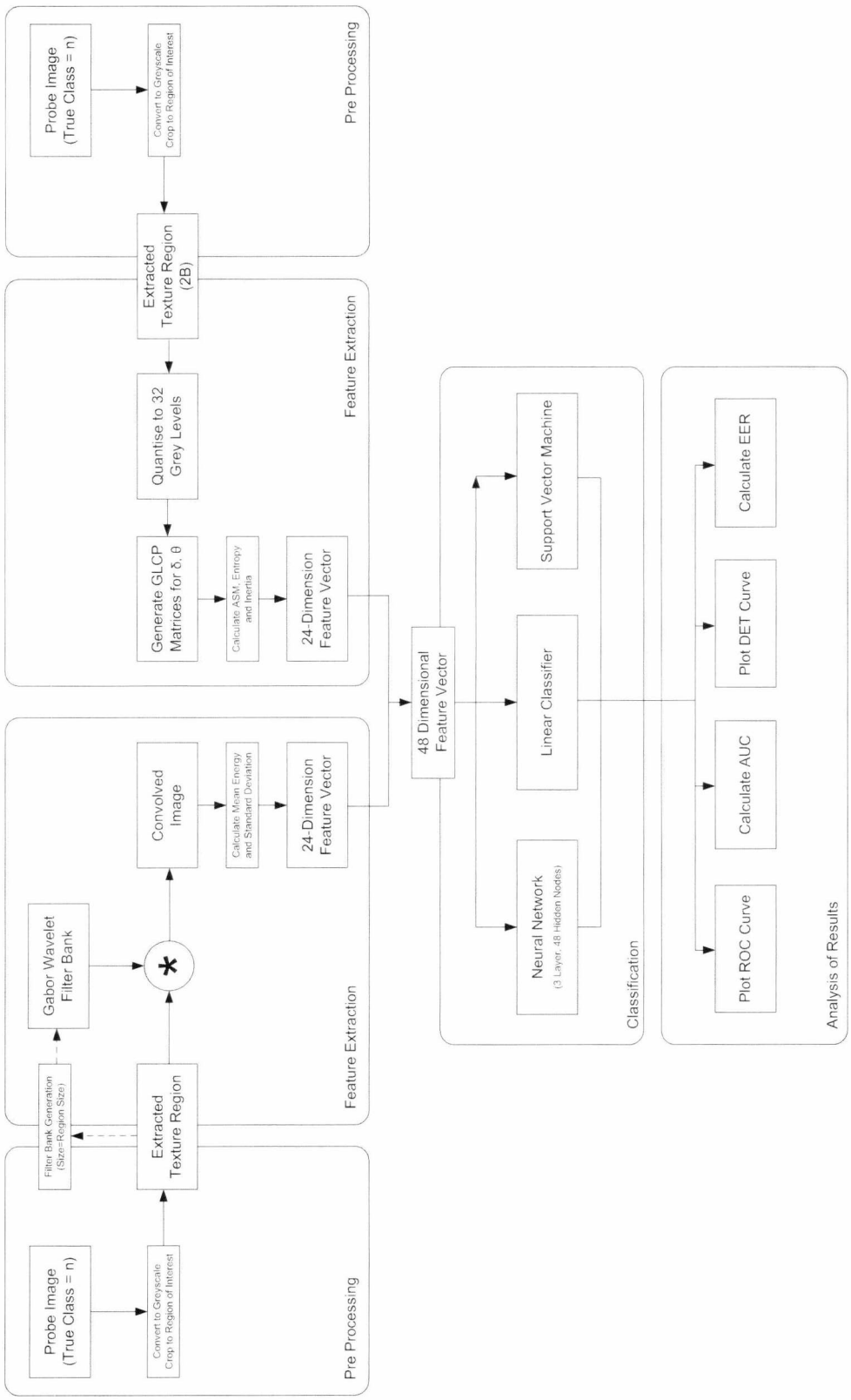


Figure 6.44 - Multi Feature Type, Single Biometric Region Feature Fusion Process

Each of these processes are repeated for each individual region (2A and 2B).

6.8.2 EVALUATION RESULTS (MFSB)

The evaluation results for the multi feature type, single biometric region are shown below:

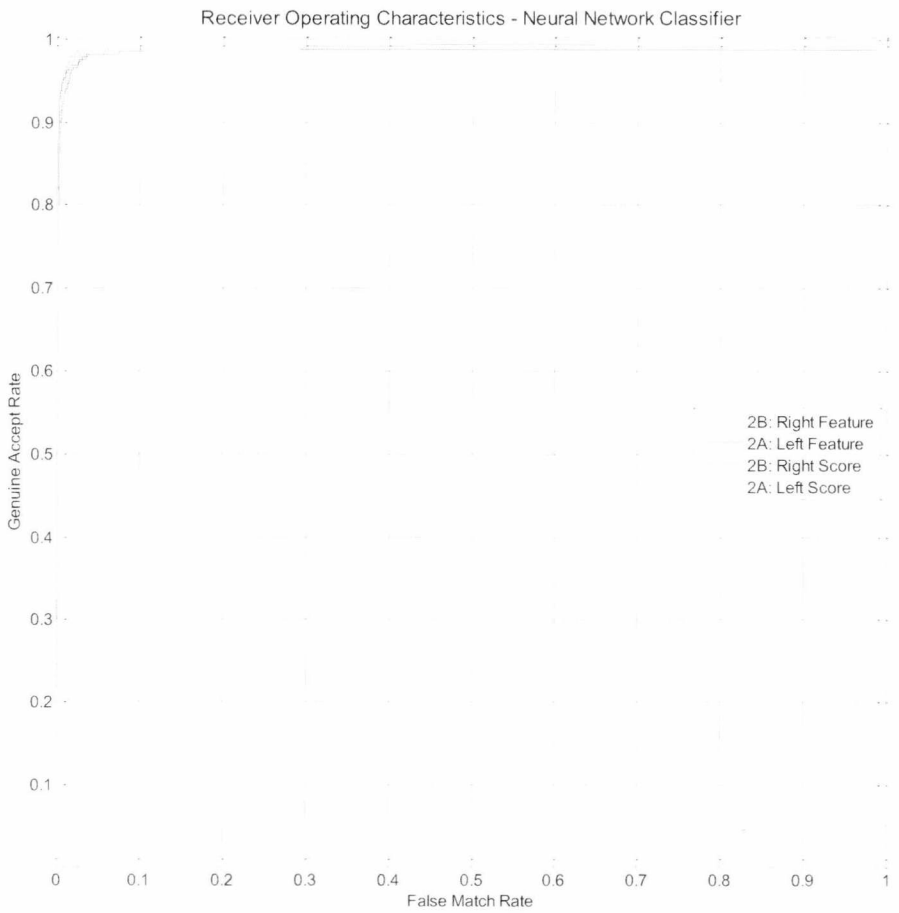


Figure 6.45 - MFSB Neural Network ROC Curve

The neural network output shows good performance on the ROC curve from all permutations. The right region (2B) performs slightly below (2A) when score fusion is used. Although this could be attributed to a weak neural network convergence, similar trends have been seen when analysing results of the left and right regions. It is suggested in Chapter 5 that this may be attributable to an underlying anomaly in the dataset. The ROC curve is shown with the upper left region enlarged in Figure 6.46. The performance of score fusion remains above that of feature fusion as previously observed indicating that the system performs better multiple classifiers are assigned a smaller feature space .

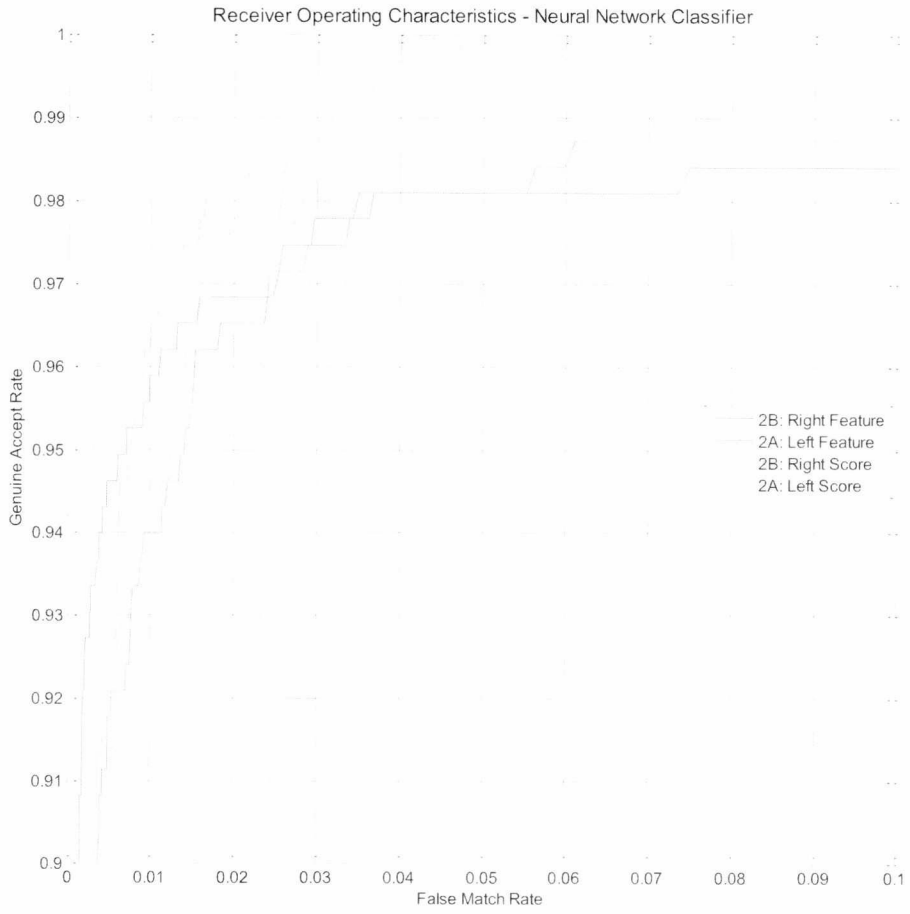


Figure 6.46 - MFSB Neural Network ROC Curve (Enlarged)

The detection error trade off curve shows a similar pattern with all four systems showing similar classification performance (Figure 6.47). A distinct lower error rate is visible in the systems employing score fusion in this graph.

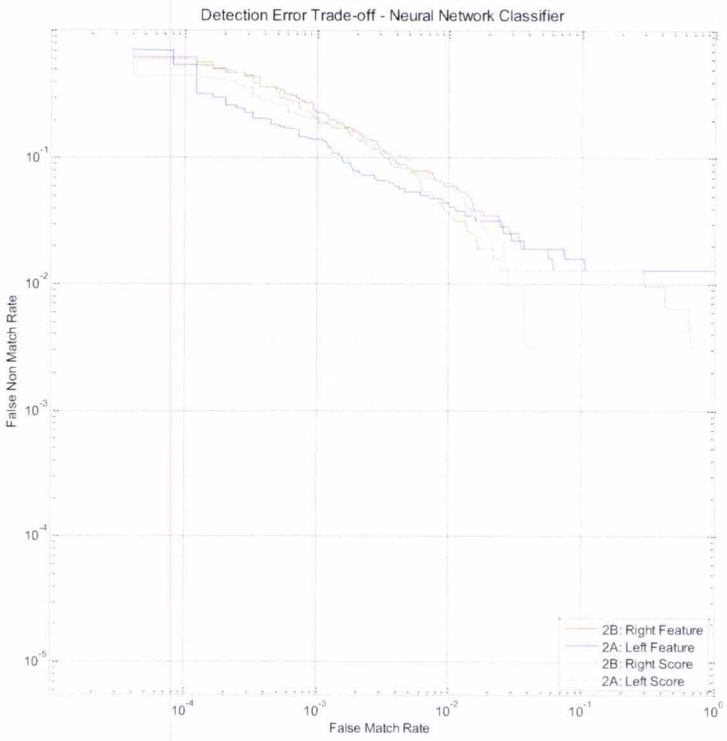


Figure 6.47 - MFSB Neural Network DET Curve

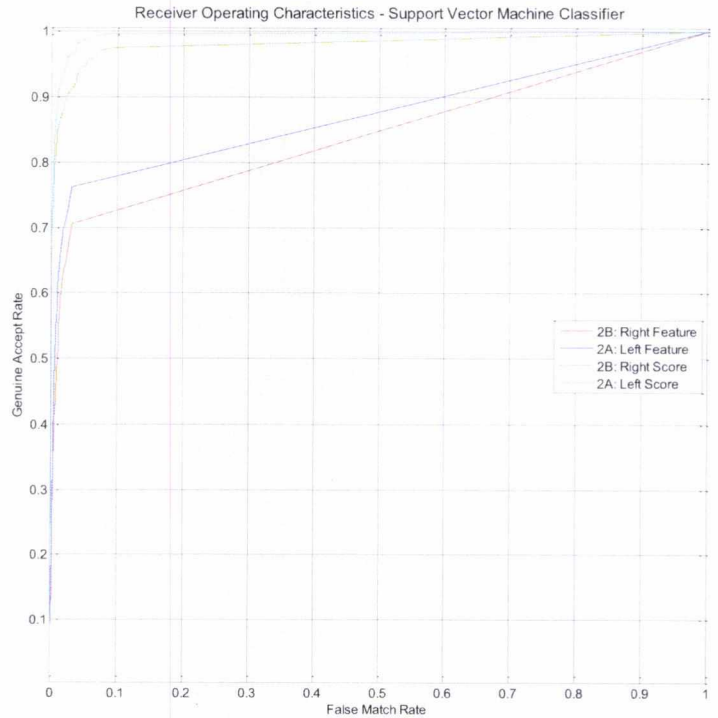


Figure 6.48 - MFSB Support Vector Machine ROC Curve

With the support vector machines, the weaknesses of feature fusion become more obvious and the system has more difficulty in establishing support vector separation of the higher dimensional feature vectors. This disparity between the two is evident in both the ROC and DET curves (See Figure 6.48 and Figure 6.49).

Figure 6.49 - MFSB Support Vector Machine DET Curve

The disparity between score and feature fusion presented by the Support Vector Machine is accentuated with the Linear classifier where the ROC curves for each type of fusion show greater divergence (Figure 6.50). The feature fusion outputs barely register on the DET curve and exhibit a worse response which is reflected in the equal error rate.

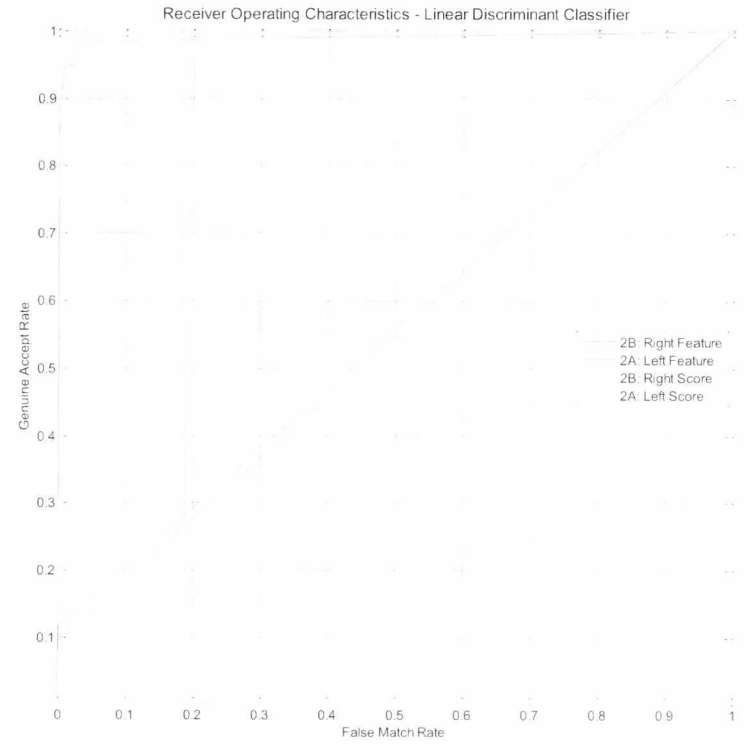
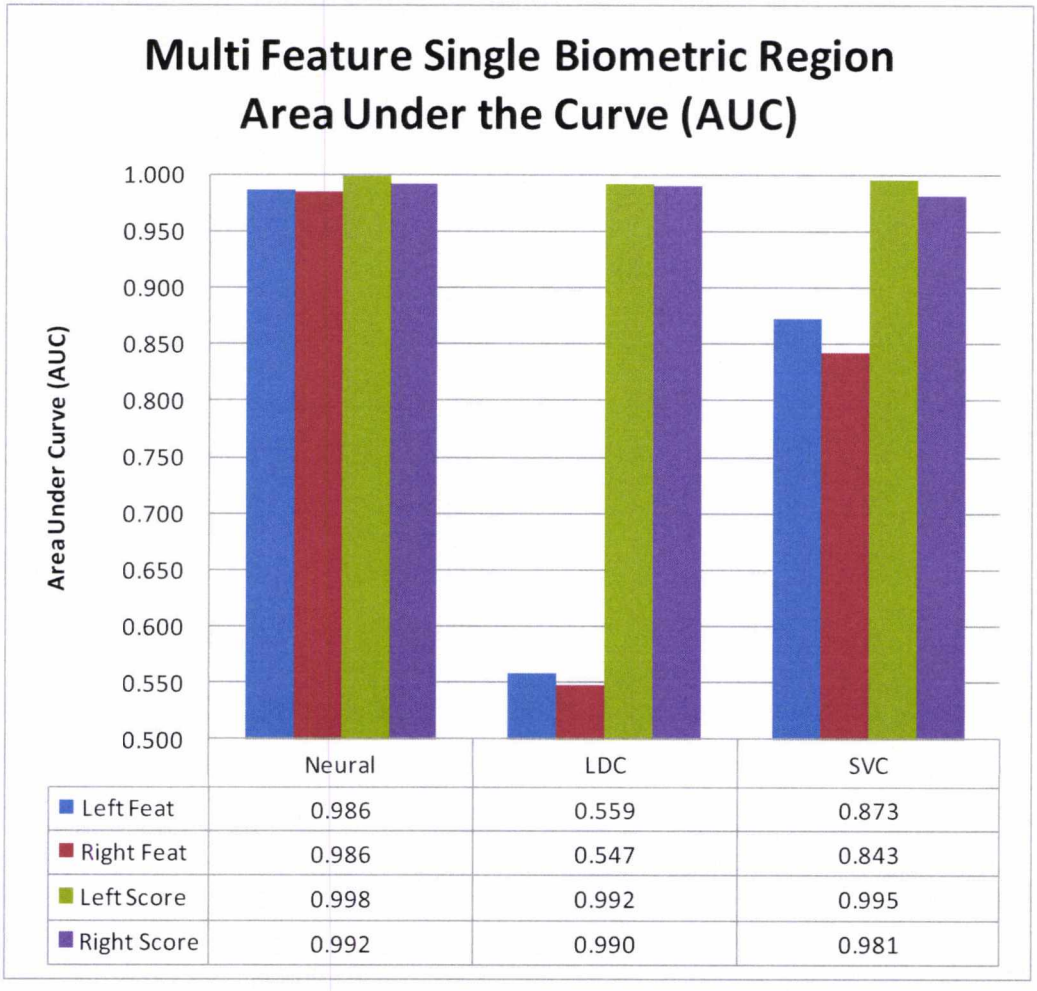


Figure 6.50 - MFSB Linear Classifier Machine ROC Curve

Figure 6.51 - MFSB Linear Classifier Vector Machine DET Curve

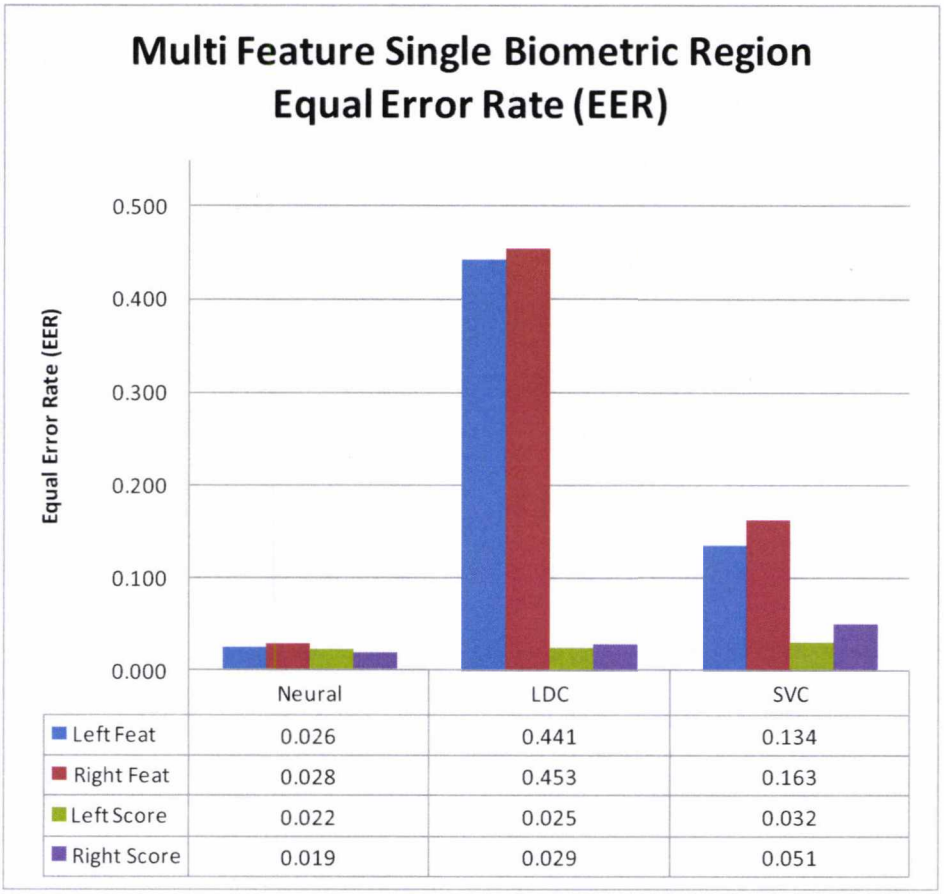
In a side by side comparison of the AUC and EER values for the single biometric comparison, the neural network tolerates both 24 and 48 dimensional feature vectors with little difference between the outputs for either. In these instances, the ability of neural networks to learn feature patterns without relying on feature-space boundaries appears to give it a strong advantage particularly with the 48-dimensional feature vector.



	Neural	LDC	SVC
Left Feat	0.986	0.559	0.873
Right Feat	0.986	0.547	0.843
Left Score	0.998	0.992	0.995
Right Score	0.992	0.990	0.981

Figure 6.52 - MFSB AUC Comparison (all classifiers)

Output for score fusion in general across all three classifiers is good with the box plot (Figure 6.54) showing relatively small variance for this method. The box plot has been formed of cumulative averages between the Neural Network, Support Vector Machine and Linear Discriminant classifiers. It also highlights the extent of the equal error rate variation when feature fusion is used suggesting that this method is less than ideal when compared with score fusion. It can also be observed that features from the left region (2A) classify marginally better than those from the right region.



	Neural	LDC	SVC
Left Feat	0.026	0.441	0.134
Right Feat	0.028	0.453	0.163
Left Score	0.022	0.025	0.032
Right Score	0.019	0.029	0.051

Figure 6.53 - MFSB EER Comparison (all classifiers)

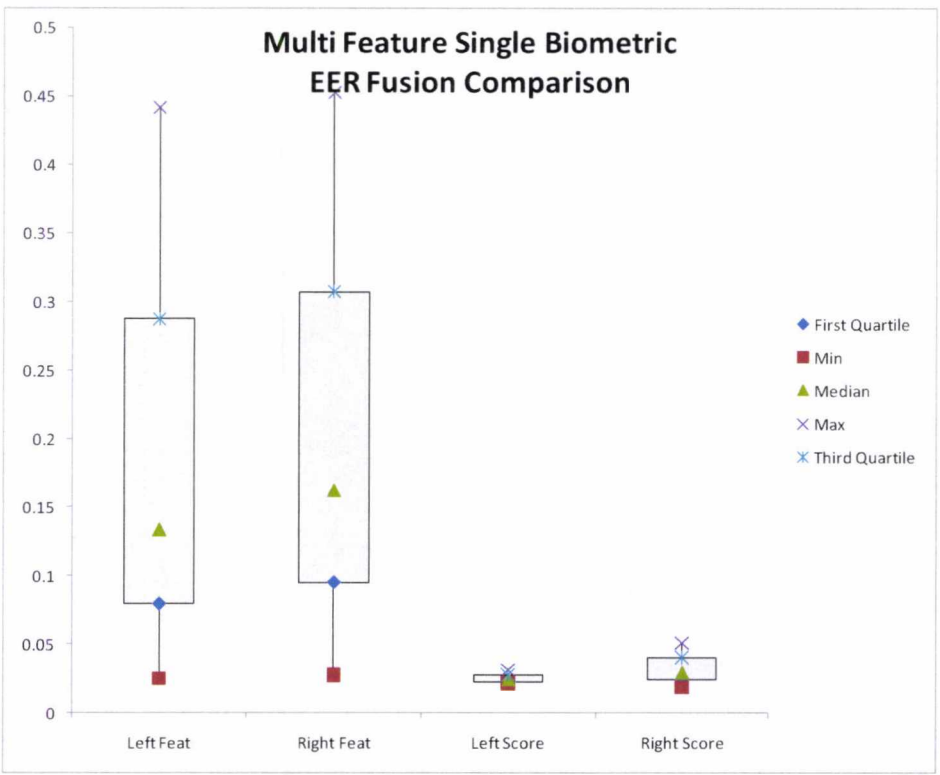


Figure 6.54 - MFSB EER Box Plot (All Classifiers)

6.9 MULTI FEATURE MULTI REGION COMPARISON (MFMB)

6.9.1 EVALUATION CONFIGURATION

So far the use of small greyscale skin texture extracts from high resolution photographs of subjects faces has been suggested to offer a sufficiently diverse feature set to enable subject identification by a variety of classifiers. The evaluation has shown that by fusing the scores from both the Gabor and GLCP feature classifications the overall system performance is improved (Section 6.8). The evaluation has also shown that when score fusion is used to combine single feature multi-biometric information (2A and 2B) the systems performance is also improved (Section 6.2). These improvements have been observed across all classifiers.

It has also been observed that the neural network classifier shows much greater performance to feature fusion techniques than either the support vector machine or the linear discriminant classifier although this remains marginally under the neural networks score fusion performance.

Knowing this information it is anticipated that in a multi feature multi biometric environment where both Gabor and GLCP features are combined for both skin regions, that an additional overall performance enhancement will be observed for score fusion. Conversely with feature fusion now reaching a potential 96-dimensional input vector the curse of dimensionality may now begin to have a more severe impact on these results.

To assist with identifying the four permutation patterns used in this section, Table 6.2 displays a matrix of each permutation with an accompanying reference to the process diagram. Note that in order to keep the diagrams legible they have been simplified from past diagrams into simple block diagrams, the inner workings of each diagram is based on previous process diagrams.

Identifier	Gabor and GLCP Features	Region 2A and 2B Features	Process Diagram Reference
1	Fused	Fused	Figure 6.55
2	Score	Fused	Figure 6.56
3	Fused	Score	Figure 6.56
4	Score	Score	Figure 6.57

Table 6.2 - MFMB Permutations

System 1 uses a simple feature fusion system to generate a 96 dimensional feature vector where *A* identifies facial region 2A and *B* identifies facial region 2B.

$$S_1 = \{A_{GAB} + B_{GAB} + A_{GLCP} + B_{GLCP}\} \quad \textbf{(6.12)}$$

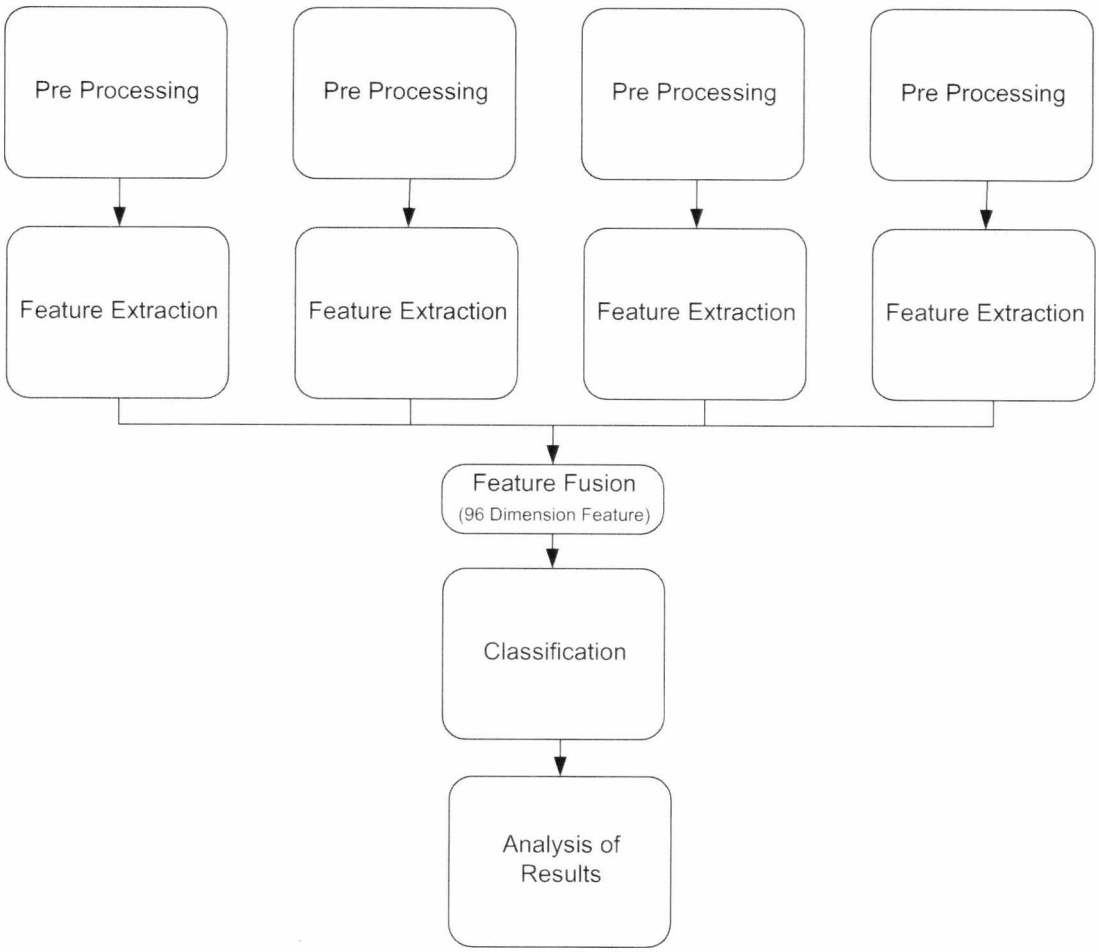


Figure 6.55 - System 1 - Full Feature Fusion

System 2 performs feature fusion of each region independently and performing score fusion of the two facial regions giving an output vector of:

$$S_2 = \{[A_{GAB} + A_{GLCP}] \mid [B_{GLCP} + B_{GAB}]\} \quad \textbf{(6.13)}$$

System 3 inverses this by performing feature fusion of similar feature space sets and score fuses the two feature set outputs.

$$S_3 = \{[A_{GAB} + B_{GAB}] \mid [A_{GLCP} + B_{GLCP}]\} \quad \textbf{(6.14)}$$

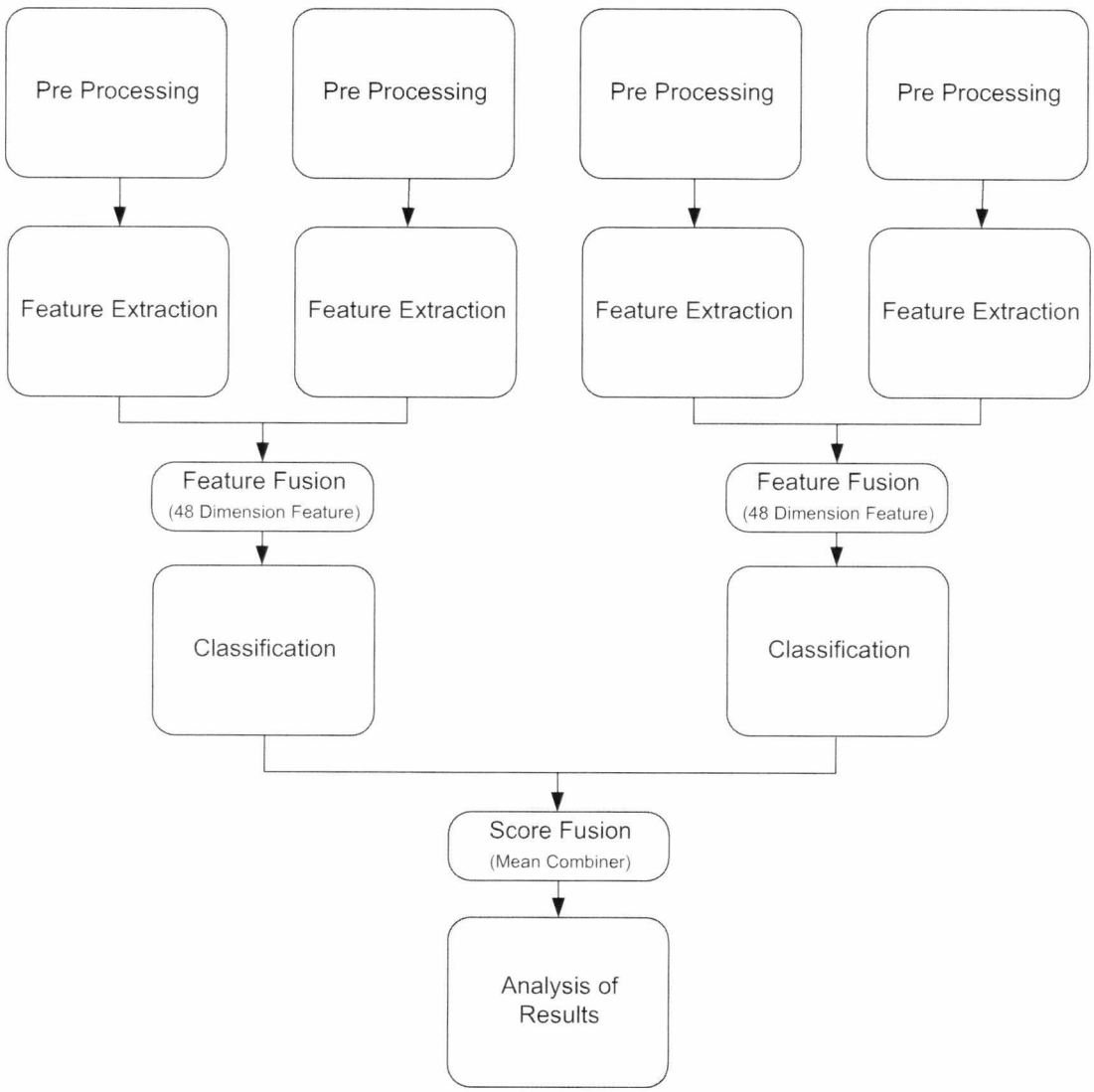


Figure 6.56 - Systems 2 and 3 - Hybrid Feature and Score Fusion

System 4 does not use any form of feature fusion and instead simply classifies each facial region within its own feature space (Gabor or GLCP) relying entirely on the score fusion method to identify class outputs.

$$S_4 = \{A_{GAB} \mid B_{GAB} \mid A_{GLCP} \mid B_{GLCP}\} \quad \textbf{(6.15)}$$

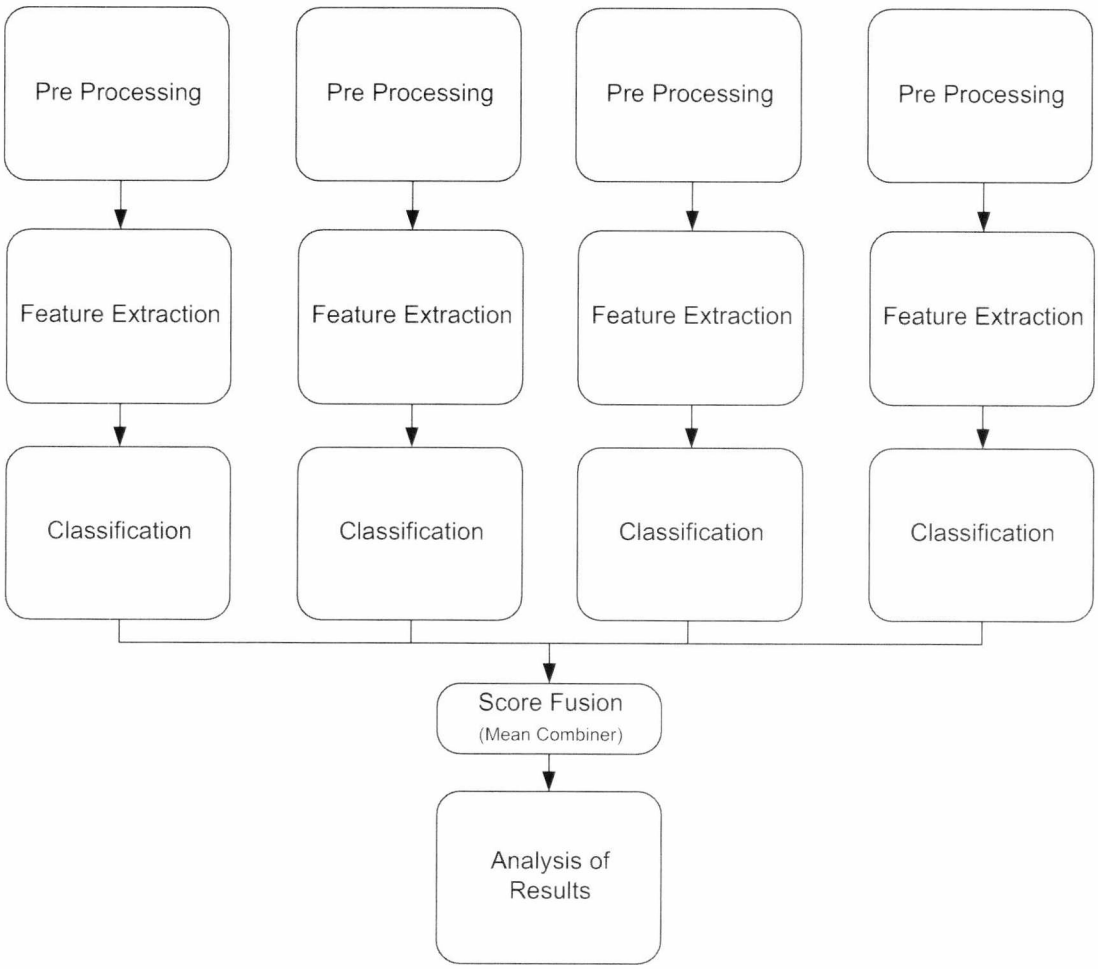


Figure 6.57 - System 4 - Full Score Fusion

6.9.2 EVALUATION RESULTS (MFMB)

As shown by previous models, the neural network continues to perform well with all permutations as shown in Figure 6.58. Expanding the upper left region of the ROC curve as shown in Figure 6.59 shows the performance differences between the systems. There is a clear performance impact with System 1 using a 96-dimension single feature fused vector as can be seen in the ROC and DET curves. From the ROC curves it is difficult to discern between the remaining systems and the DET become more necessary to differentiate the various systems performance.

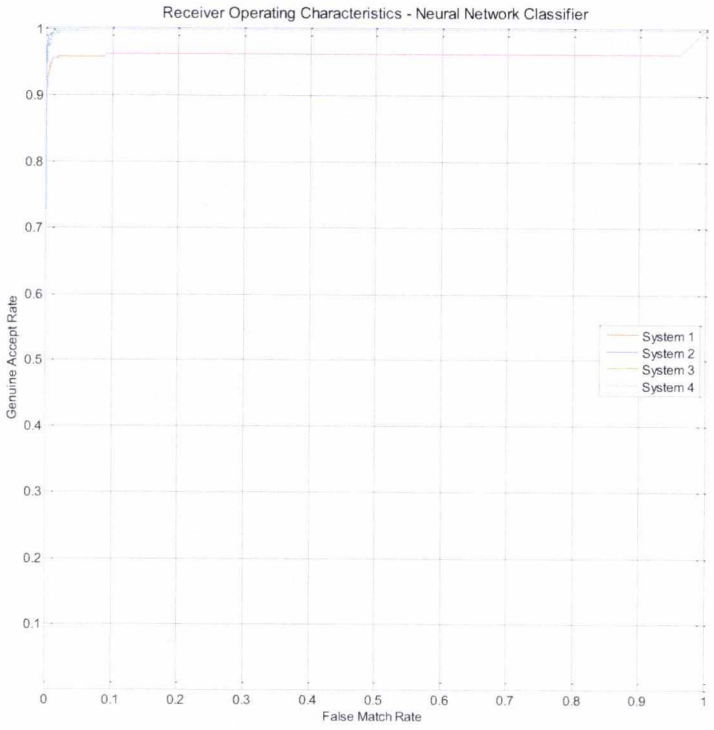


Figure 6.58 - MFMB Neural Network ROC Curve

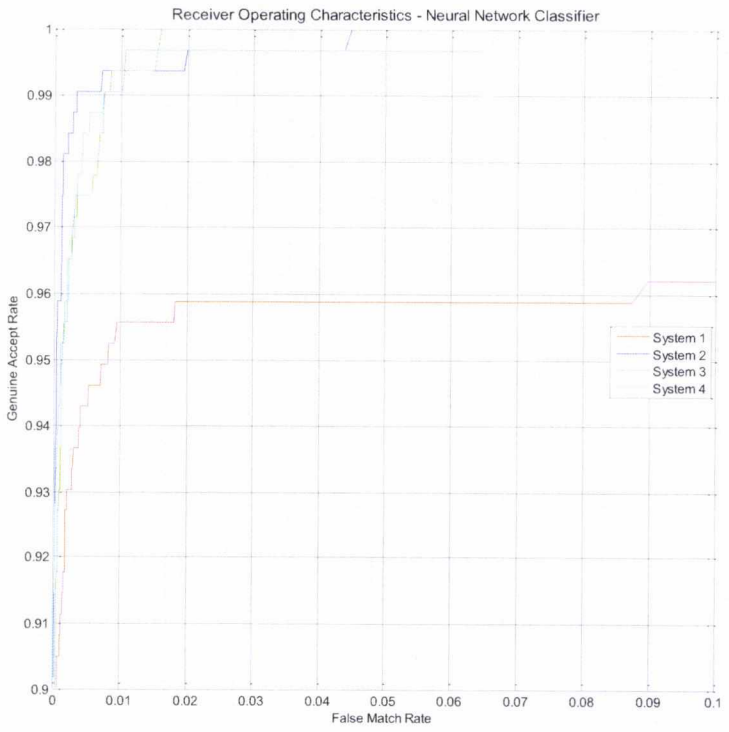


Figure 6.59 - MFMB Neural Network ROC Curve (Enlarged)

The Neural Network DET curve displays the difference between the four systems in greater clarity. It can be seen as the false non match rate declines the false match rate is higher for System 4. With only 316 training images and 79 classes, a 96 dimensional feature vector is more than likely within the territory of the curse of dimensionality. If true the effect is likely to be further accentuated in the SVM and Linear classifiers. System 2 appears to perform marginally better with the neural network.

It should be noted at this stage that the neural network evaluation results for Systems 2-4 are extremely close and may hence be subject to random convergence within the neural network. As discussed in Chapter 1.1a)i)6there may be insufficient data in the evaluation databases to enable an accurate comparison between these systems using a neural network.

Figure 6.60 - MFMB Neural Network DET Curve

An initial viewing of the support vector machine outputs for the same systems shows an even more pronounced difference between the class assignment outputs. In this instance the falloff in performance of System 1 can clearly be seen as predicted.

Now also apparent is the difficulty of System 2 which relies on feature fusion of two disparate feature spaces when compared with system 3 (which fuses common Gabor features from both regions, and then fuses the scores with the feature fused GLCP outputs).

This lower performance of System 2 compared with System 3 is likely attributed to the different feature ranges produced by GLCP and Gabor filters and the impact of mixing these two vector types in the same feature space.

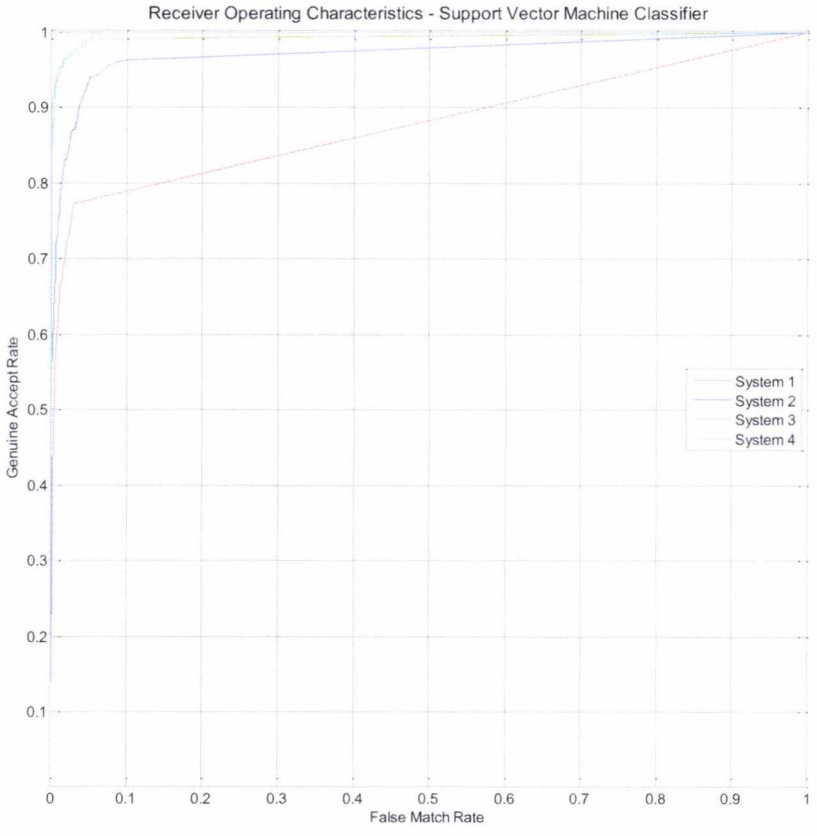


Figure 6.61 - MFMB Support Vector Machine ROC Curve

System 4 performs best in the SVM evaluation closely followed by System 3. Both of these systems avoid fusing different feature types in the same feature space and their success may be attributed to this. Enhancing the upper left region Figure 6.62 and inspecting the DET curve Figure 6.63 similar output classification performance as shown by the neural network is present. Systems 1 and 2 suggest a more serious performance impact as previously noted.

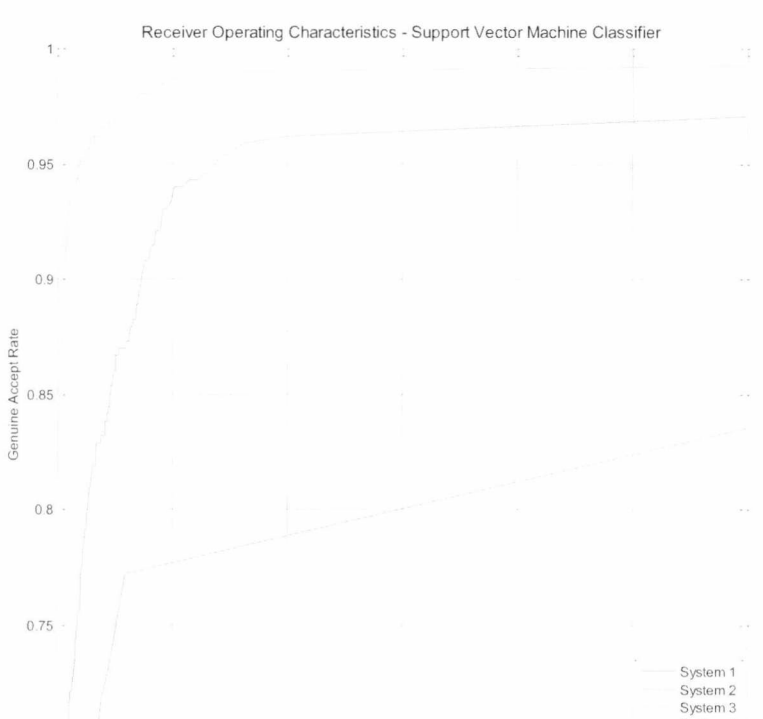


Figure 6.62 - MFMB Support Vector Machine ROC Curve (Enlarged)

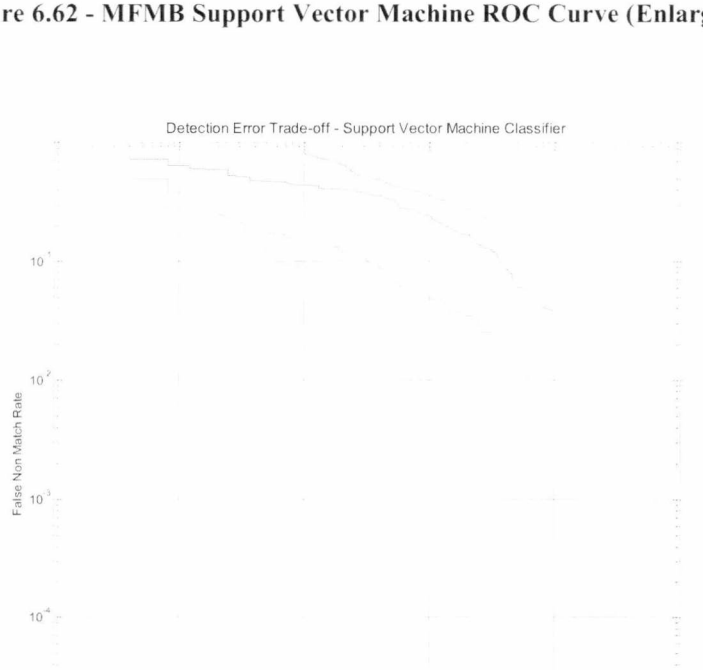


Figure 6.63 - MFMB Support Vector Machine DET Curve

The linear classifier now shows clearly the limitations of the system when dealing with the feature fused vectors in System 1 with an apparent failure to classify. In the ROC curve, System 1 indicates completely random classifier output by following the y=x straight line path. It is probable that the combination of using a 96 dimensional vector with the simple classification methods of the linear discriminant classifier have resulted this inability to classify.

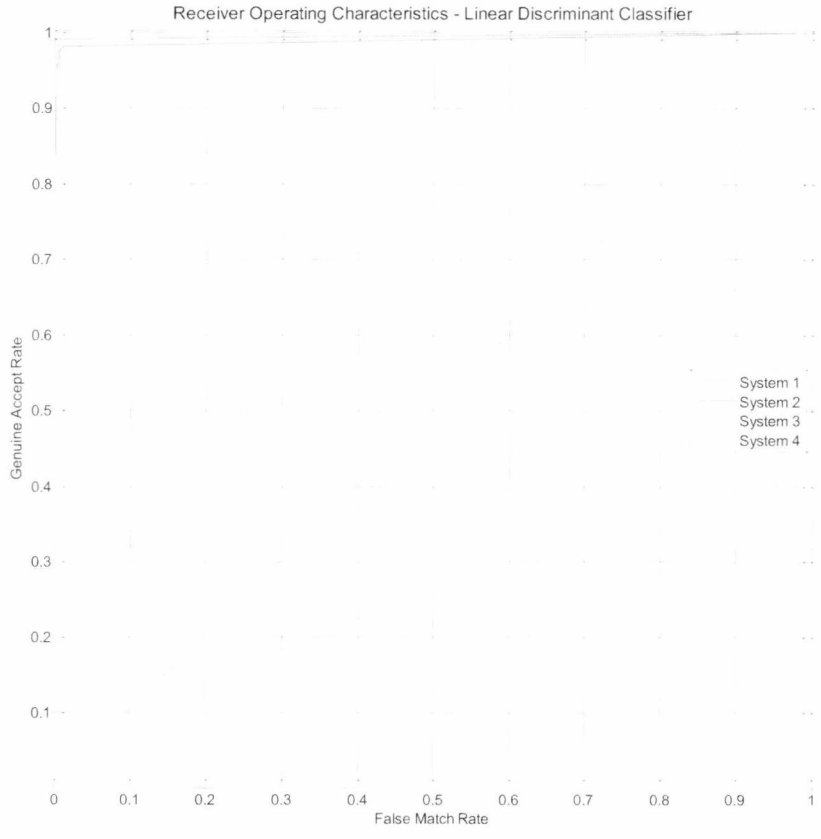


Figure 6.64 - MFMB Linear Classifier ROC Curve

Systems 2, 3 and 4 show good performance using this classifier with similar trends as perceived in the support vector machine. While System 2 trails slightly behind 3 and 4, the difference between 3 and 4 is not clear from the curves and the EER and AUC values will be used to differentiate between these two systems.

Due to its exceptionally poor performance, System 1 fails to register on the DET curve.

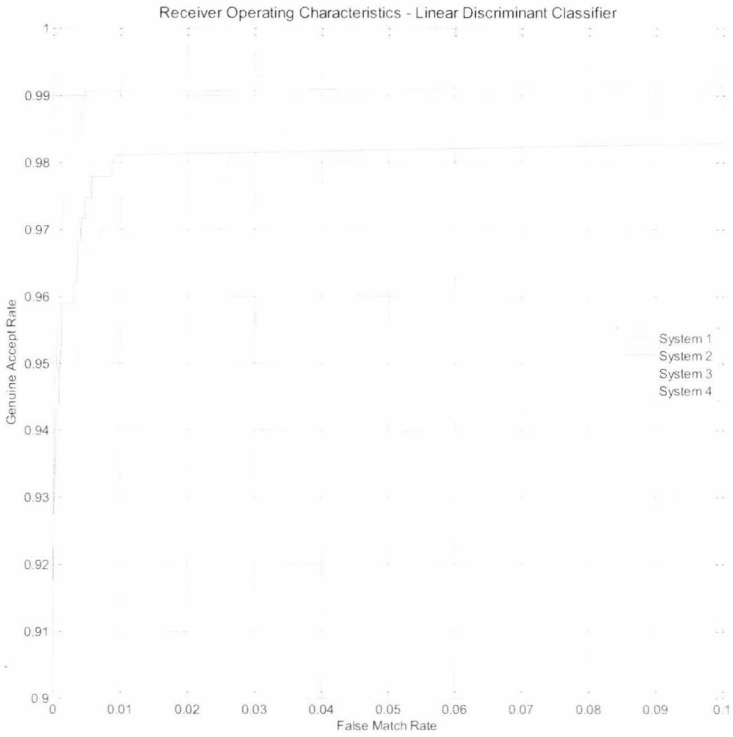


Figure 6.65 - MFMB Linear Classifier ROC Curve (Enlarged)

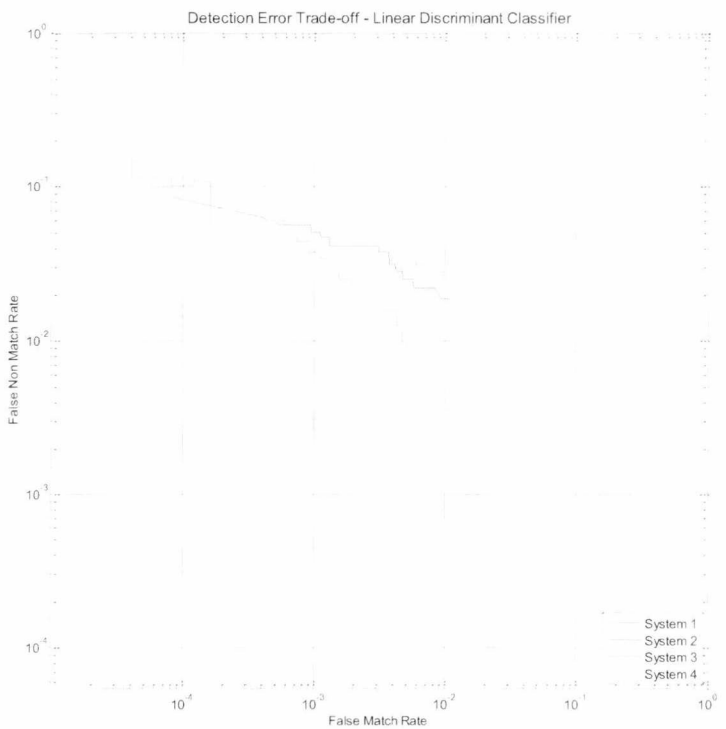
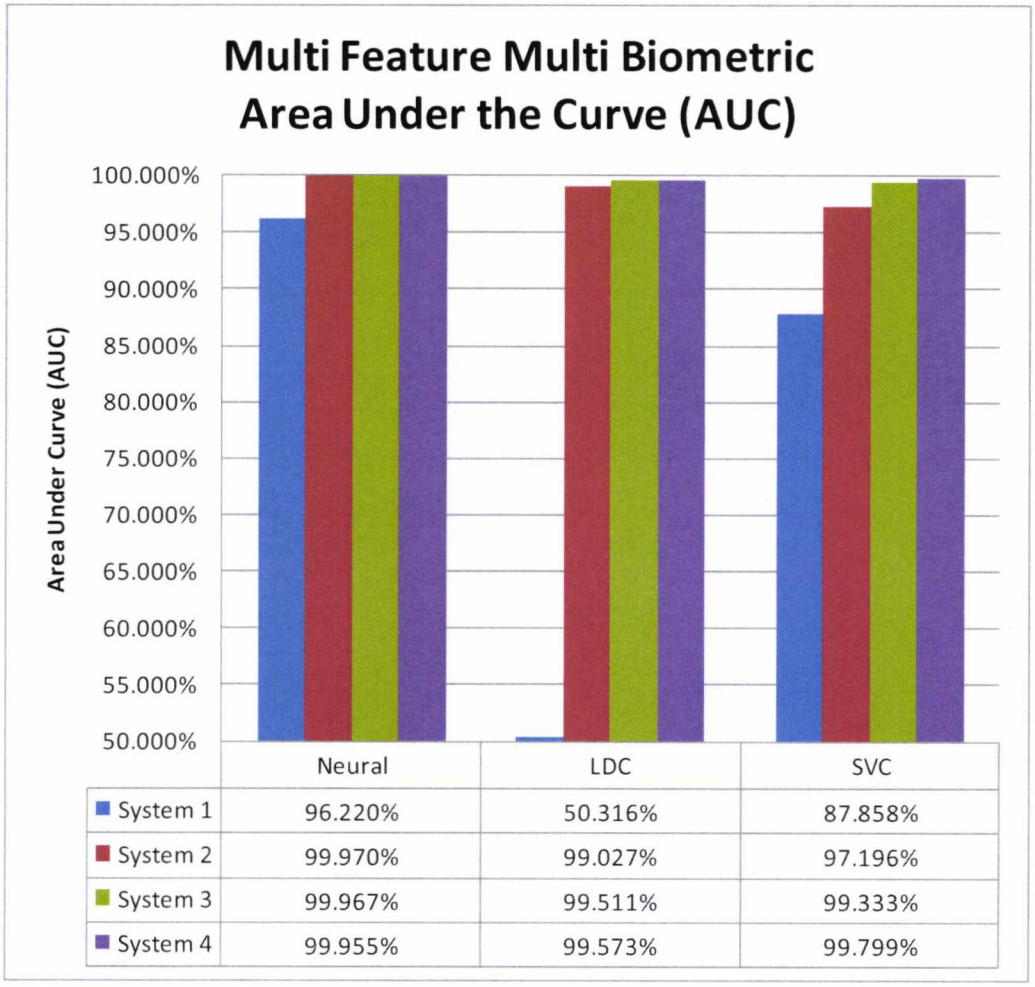


Figure 6.66 - MFMB Linear Classifier DET Curve



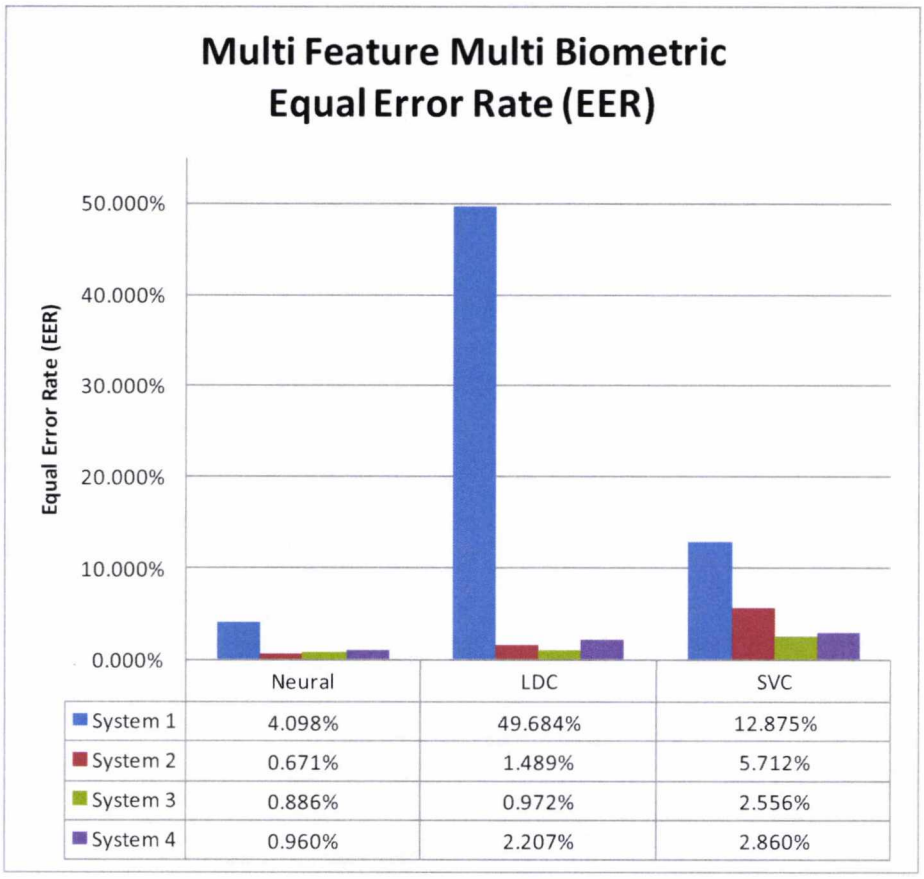
	Neural	LDC	SVC
System 1	96.220%	50.316%	87.858%
System 2	99.970%	99.027%	97.196%
System 3	99.967%	99.511%	99.333%
System 4	99.955%	99.573%	99.799%

Figure 6.67 - MFMB AUC Comparison - System Performance

Figure 6.67 and Figure 6.68 show the overall AUC and EER comparison between the classifiers and fusion permutations. Clearly the neural network shows a substantial performance lead over its rival classifiers achieving the highest area under the curve for all four Systems and over 99.9% for Systems 2, 3 and 4. The equal error rates show that with the neural network, Systems 2, 3 and 4 all achieved an EER of less than 1% which equates to 3 erroneous classifications out of the 316 probe image test set.

Figure 6.69 shows the performance advantage obtained when keeping feature types within the same input feature vector either by feature fusion of similar vectors or independent classification of all vectors.

The box plot (Figure 6.71) shows the relative invariance between outputs of the neural network.



	Neural	LDC	SVC
System 1	4.098%	49.684%	12.875%
System 2	0.671%	1.489%	5.712%
System 3	0.886%	0.972%	2.556%
System 4	0.960%	2.207%	2.860%

Figure 6.68 - MFMB EER Comparison - System Performance

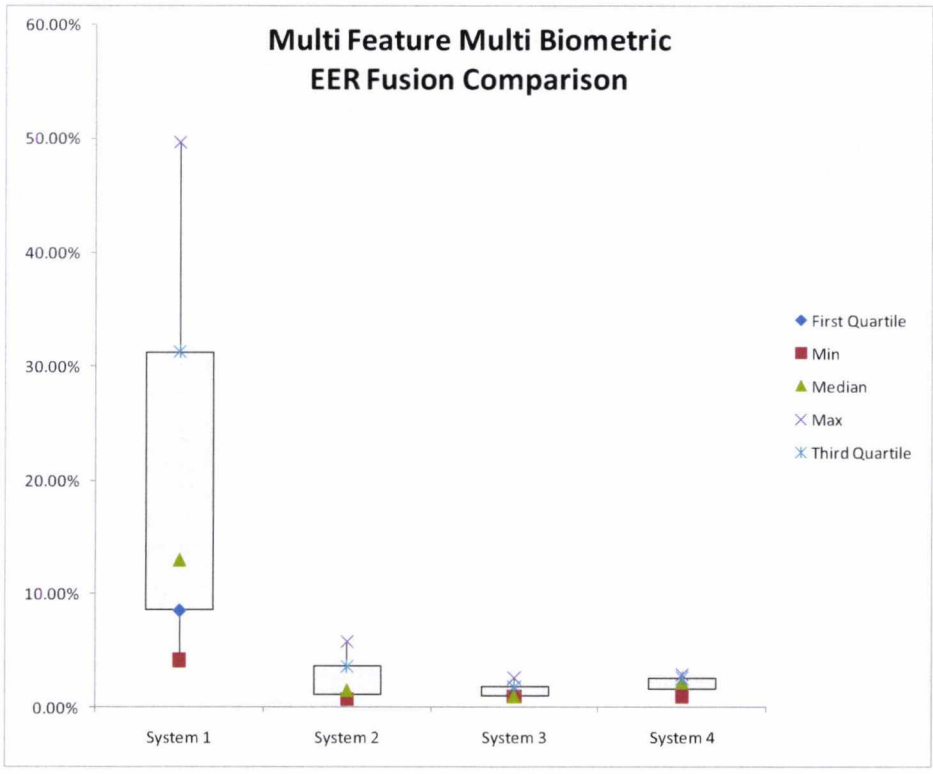


Figure 6.69 - MFMB EER Box Plot - System Performance

The performance between System 3 and System 4 is extremely close on all output graphs. Figure 6.69 suggests that System 3 may provide lower variance in the equal error rates whereas the probability of a correct classification improves with System 4 as shown in Figure 6.70.

With accuracy results now in excess of 99% it is suggested that the tests be repeated with a much larger biometric face database in order to provide a definitive comparison between Systems 3 and 4.

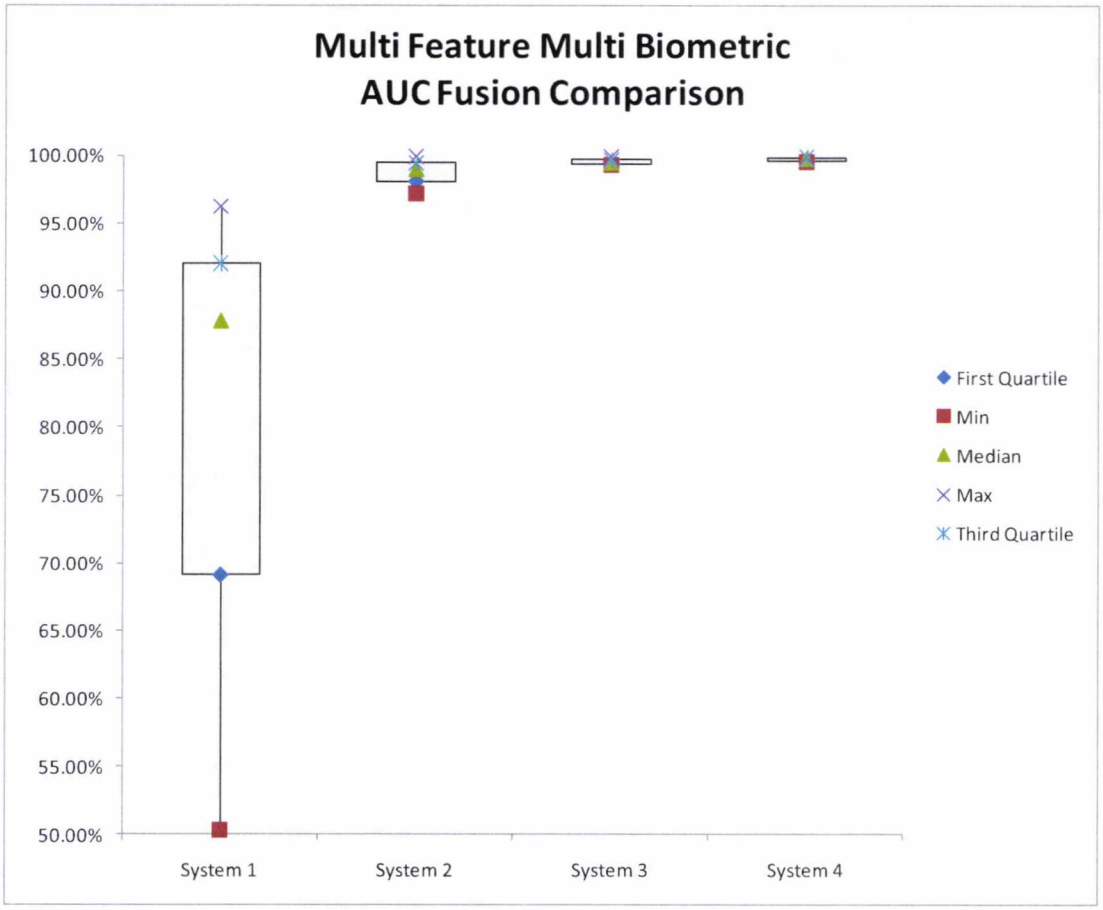


Figure 6.70 - MFMB AUC Box Plot - System Performance

Figure 6.71 shows the relative performance of each classifier across all four systems and again supports previous evidence that the neural network provides the highest level of adaptability even when influenced by the curse of dimensionality. As the feature size increases, the linear discriminant classifiers performance worsens significantly and is the only one of the three classifiers that was completely unable to perform positive classification of System 1.

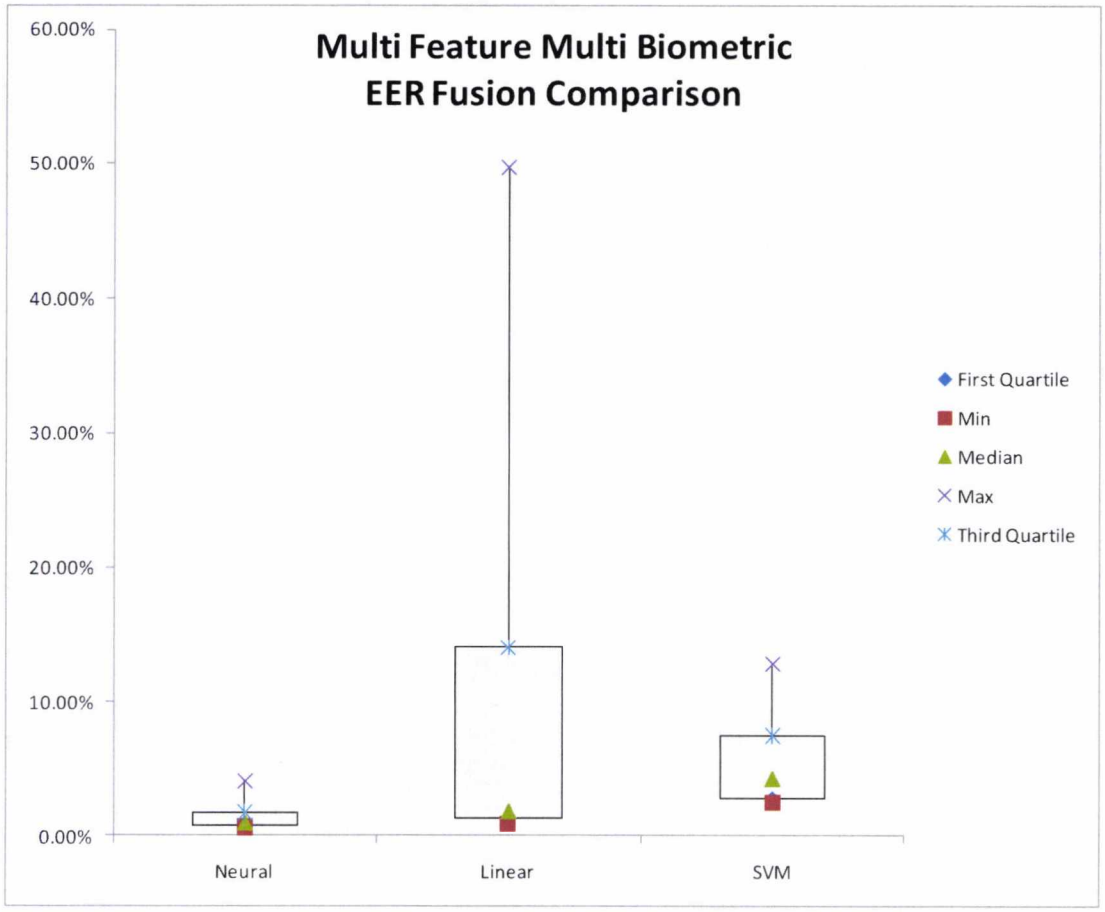


Figure 6.71 - MFMB EER Box Plot - Classifier Performance

6.10 RESULTS AND CONCLUSIONS

The results have indicate that facial skin texture may contain suitable biometric features for a face recognition system. The results also lean towards the selection of a full score fusion configuration as an optimum system configuration. By combining multiple skin texture feature vectors (GLCP, Gabor) and multiple regions the probability of a correctly chosen class is as high as 99.97% with an equal error rate of under 0.7% when a neural network classifier is applied (System 2).

It is also suggested that in the multi-feature type multi biometric region evaluation, System 3 and System 4 offer excellent results. It is highlighted however, that an observation of the results at this level of accuracy may be flawed (see Section 7.5 for the rationale behind this statement). It is recommended that further testing against a larger database is performed in order to strengthen these conclusions.

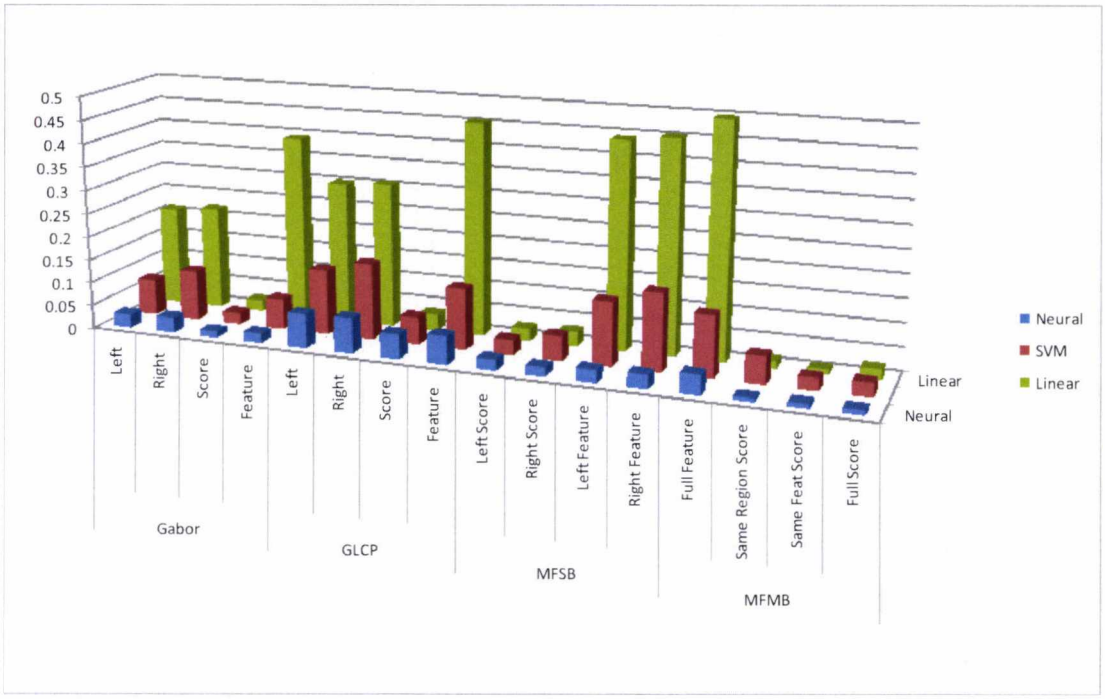


Figure 6.72 - Equal Error Rate for all Permutations

Figure 6.72 shows a summary of the equal error rates from the complete evaluation by both classifier and feature type. In the rear row, the linear classifier shows an equal error rate of under 5% in all score fusion scenarios and shows a clear benefit from using multiple feature sets as opposed to a single region and feature type. As the feature space dimensionality is increased, the error rate increases in some instances exponentially and indicates that the classifier struggles to cope with these. As previously mentioned, this may be attributable to the curse of dimensionality and generative classifiers.

It should be reiterated that the discussion in Chapter 4 of experimental framework draws attention to the fact that the classifiers have not been "tuned" for the data presented. Indeed, it is believed that their performance may improve with a more detailed analysis of their configuration. In this light, the results should not been seen as an optimal output, but more as an indicative measure of the systems capabilities.

Chapter 7 discusses the statistical confidence that may be placed in a system taking into account the size of the image database used. This suggests it is difficult to compare between results with an equal error rate less than 1% which is found particularly in the neural network outputs.

Taking this into account, the neural network consistently shows a superior ability to correctly classify outputs when compared with both the support vector machine and linear classifiers. The neural network however, was subjected to partial tuning as discussed in Chapter 4 where

experimental results found that a 48 neuron hidden layer produced optimal results when compared with 24 and 96 neuron systems. On this basis it is unfair to state that the neural network is optimised for the features discussed as the classifier has been subjected to partial optimisation.

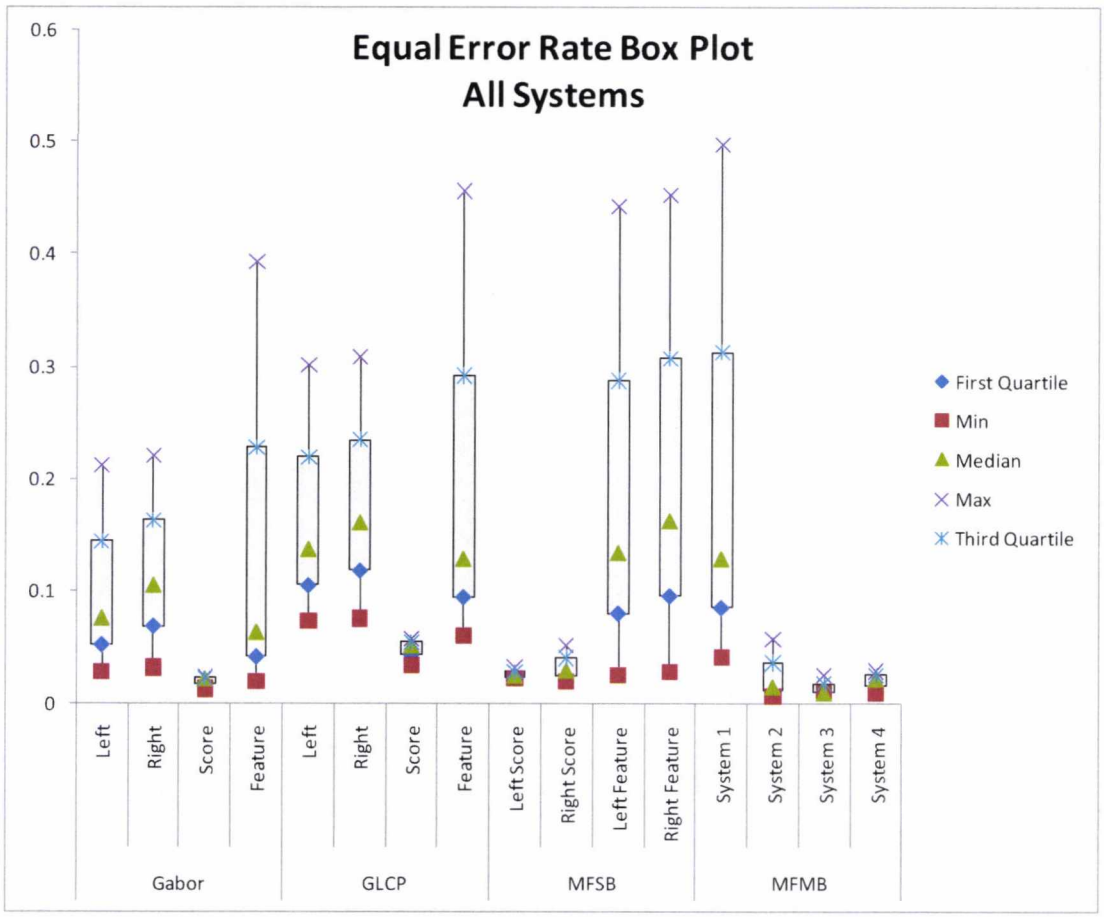


Figure 6.73 - Equal Error Rate Comparison - All Systems

In an attempt to drawn an average balance across the classifier types and move the focus to the features, Figure 6.73 shows the relative performance of the systems in relation to the feature type and presentation (score/feature fusion). From this analysis, the small boxes for all score fusion types indicate that this method of fusing features offers a good reduction in the equal error rates obtained.

6.11 CHAPTER REVIEW

In this chapter, the following conclusions have been drawn in response to the initial objectives laid out in the introduction:

1. Fusing the classification outputs (score fusion) using mean averaging can lower the equal error rate by around 50% (neural network). Even greater percentage gains are seen in the linear and support vector machine classifiers..

2. By using GLCP spatially collected texture features, face recognition is possible and shows performance slightly behind that of frequency domain based Gabor wavelet filters.

3. Performance can be improved when GLCP and Gabor wavelet features are fused.

4. The results indicate that score fusion of small feature vectors provide superior performance to feature fusion generating large feature vectors. [Bellman and Kalaba 1959] provide a plausible explanation for this behaviour with their presentation of the curse of dimensionality.

5. The optimal configurations for the least output variance across classifiers is demonstrated when fusing features of different types is avoided. When following this guidance, it is possible to obtain an EER score of under 1% (Neural Network).

6. The neural network shows the greatest flexibility to adapt to various forms of fused feature input configurations. Support Vector Machines also show a good capacity to maintain low output variance.

7 DISCUSSION OF SYSTEM PERFORMANCE

7.1 INTRODUCTION

This chapter examines the performance of the system as a whole and examines the practical aspect of such a systems implementation in terms of hardware requirements. This chapter is divided into two sections the first of which examines the performance of the system in terms of computational requirements.

The second section of this chapter investigates the impact of image resolution on the feature extraction and accuracy process. As the feature vectors created remain the same size changing these factors may not impact the classification performance, but using lower resolution images could potentially have a large impact on the initial feature extraction process.

7.2 SYSTEM PERFORMANCE

Measurement of the systems performance allows a user of the technology to evaluate the hardware processing requirements to ensure a satisfactory system is presented. The exact design and selection of classifiers is very often dependent on both the performance and functional requirements of the end system.

In all systems, the final objective will be to provide some form of (timing) performance versus accuracy for an overall system which will allow the outputs to be tuned to the users requirements.

There exist a variety of methods to convert an image to greyscale and crop relevant sections. For this reason these timings are not included in the assessment.

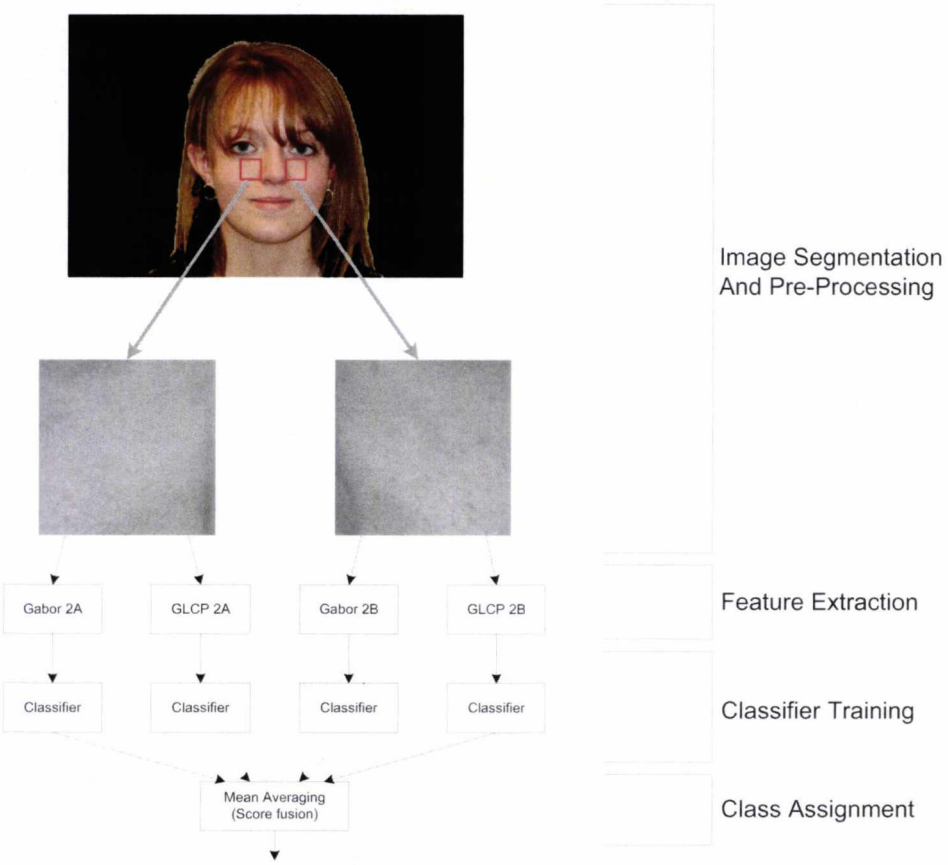


Figure 7.1 - Stages of Face Identification

Figure 7.1 highlights the four stages in the identification process. Class assignment using a fully trained classifier is an extremely rapid process using all three of the proposed classifiers. This section will provide information on the two most intensive parts of the above process, namely feature extraction and classifier training.

All timings were recorded three times with the mean value recorded. In all instances, a variance of under 1% was seen from the mean in the timings.

7.2.1 FEATURE EXTRACTION

Although the feature extraction performance has not impeded this research, [Choi, Tse et al. 2008] recently described a method of simplifying the Gabor wavelet filtering process that claims to reduce the computational time by a factor of over 4 times. An investigation into the performance impact of this technique for facial texture recognition could prove interesting in a practical application of the technology.

The MatLab functions "tic" and "toc" were used to accumulate the amount of time spent extracting the Gabor and GLCP features from the BMDB DS2 dataset from both 2A and 2B skin regions at the original size as supplied in Chapter 5 (total 1264 images).

These figures exclude Disk IO of loading the file into memory and the manual process of eye location and automatic process of cropping to the region of interest.

For the GLCP matrices, these figures include creation of the Co-occurrence Matrices and computation of the features. For the Gabor Wavelet Filters, these figures include the FFT of the source image, creation of the wavelet and all intermediate processes prior to the provision of a 24 dimensional feature vector.

	Gabor	**GLCP**
Total Time for 1264 Images	941 Seconds	123 Seconds
Time per image	0.74 Seconds	0.10 Seconds

Figure 7.2 - Feature Extraction Performance (Avg. over 3 Passes)

7.2.2 CLASSIFICATION

In this study, three classifiers have been used (see Section 4.5 page 52). Each of these classifiers has exhibited both performance and accuracy variances under differing conditions.

In order to measure performance, the test described in Section 6.9 (page 137) was repeated and the time to train the classifier measured using the `tic` and `toc` functions built into MatLab. The system was put under test control by ensuring that it was not used for any other purposes during the training period. The tests were executed overnight by a script. Each test was executed twice and the mean training period recorded.

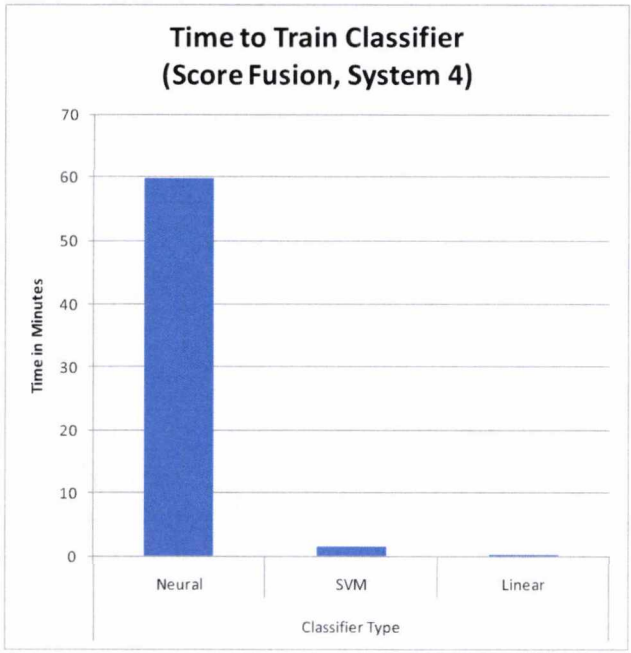


Figure 7.3 - Time to Train Classifier

The neural network has performed consistently across all tests. Neural networks are highly adaptive at the expense of processing power required to train them. The training process for a neural network is a supervised iterative convergence where the network is repeatedly trained until optimal perceptron weights are achieved. As can be seen from Figure 7.3, this puts it at a significant disadvantage in terms of training time required when compared with both the support vector machine and linear discriminant classifiers. Although this chart represents a worse case performance scenario (four independent classifiers using score fusion), the linear classifier manages to achieve full training in under 3 seconds.

For full details on the hardware used to perform these tests, please review Section 4.3 (page 41).

7.3 Performance at Varying Image Resolutions

This section discusses the accuracy to resolution trade-off in two ways. Firstly, the performance of the system when the training and probe image resolution are reduced together. For each resolution, the system is re-trained and re-tested using the training and probe datasets described in Chapter 0

Secondly, the performance of the system when only the resolution of the probe data is reduced is analysed. This gives an indicator of how the system might perform when training

images are provided under controlled conditions, but future probe images are generated in a different environment at a lower resolution.

Finally this section will discuss the confidence in the proposed system and performance metrics supplied in terms of statistical accuracy and reliability.

All image resizing described in this section is performed using bicubic interpolation resizing as this has shown to be a reliable method that maintains image quality when downsizing images. [Keys 1981].

7.3.1 Matched Training and Testing Resolutions

By reducing the training and probe image resolutions together, the minimum required resolution for successful classifier training can be measured. Table 7.1 provides a graph legend for the ROC and DET curves along with the cropped texture dimensions and eye separation distance which is further illustrated in Figure 7.4. It should be noted that due to the manual segmentation process described in Section 5.4 there is a difference of 47.3% between the smallest eye separation and largest eye separation. (For a scale factor of 1.0 the eye separation will lie between 330 pixels and 486 pixels).

Scale Factor	Min Eye Separation	Min Dimensions	Graph Colour
1.0	330	128	Red (Dotted)
0.9	297	115	Red
0.8	264	102	Blue
0.7	231	89	Green
0.6	198	77	Cyan
0.5	165	64	Black
0.4	132	51	Yellow
0.3	99	39	Magenta
0.2	66	26	Magenta (Dotted)
0.1	33	13	Blue (Dotted)

Table 7.1 - Resolution Scaling

The initial neural network results show an impressive series of results as the training and probe image resolution are reduced and interestingly show an increase in accuracy which appears to be inversely proportional to the image size.

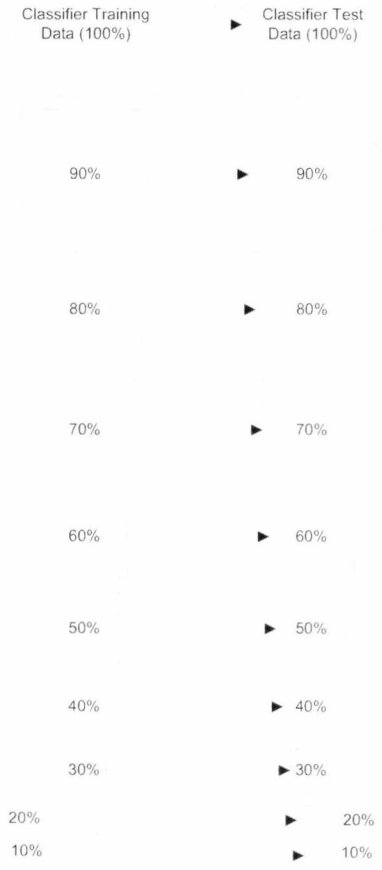


Figure 7.4 - Reduction of Training and Probe Image Size

One reason for this behaviour may be attributable to the effective reduction in relevant features. In effect as image size is reduced the dynamically sized Gabor filters converge to less useful frequencies whereas the GLCP become more capable of reading informational from the medium to low frequency range that was previously undetectable to them.

This in effect reduces the feature space as features are effectively removed. If this theory is true, one would expect to see a saturation point reached with low resolution images at the point which the image size is too small to enable sufficient variation between features. Due to the relatively small size of the BMDB database, this behaviour may not be visible until much larger databases are used.

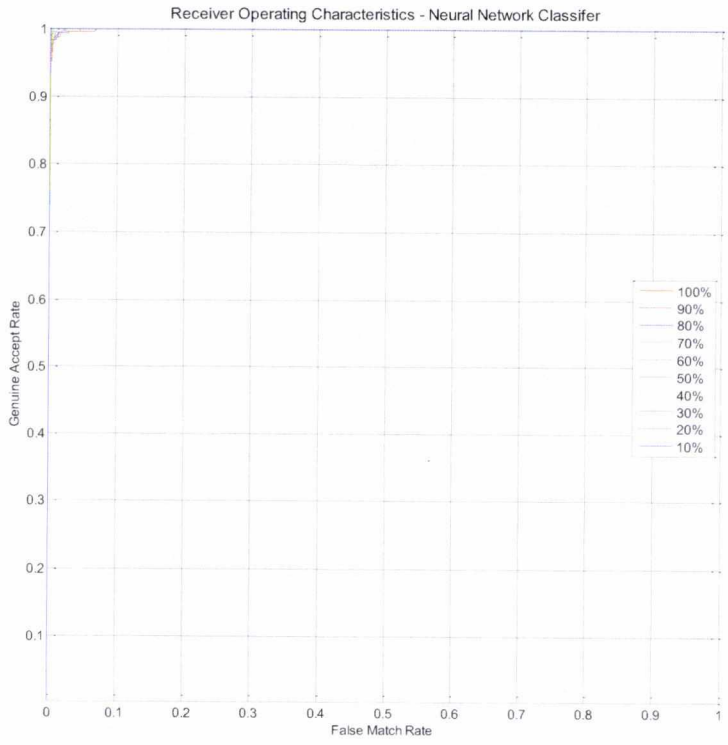


Figure 7.5 - ROC Curve of Resolution Scaling using a Neural Network

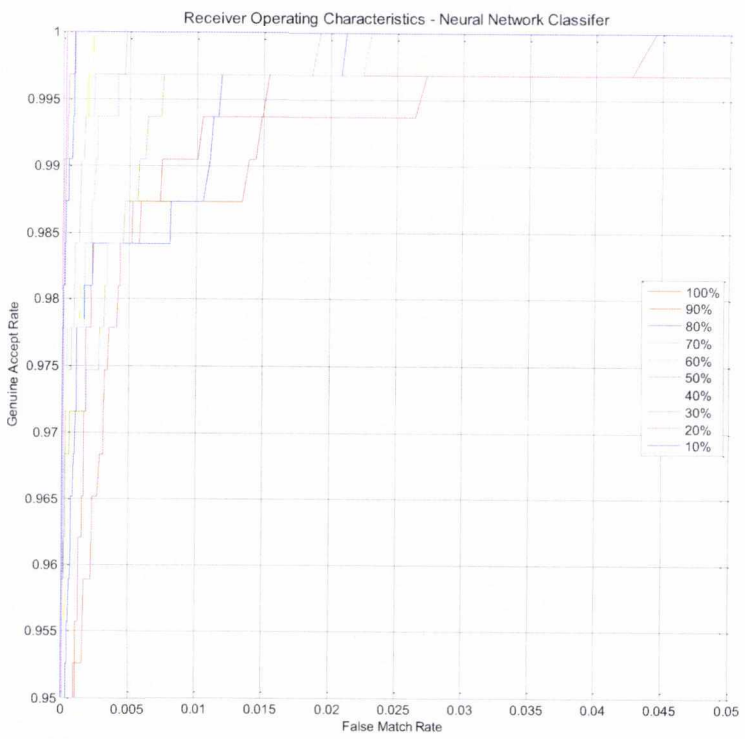


Figure 7.6 - Enlarged ROC Curve of Resolution Scaling using a Neural Network

Figure 7.7 - DET Curve of Resolution Scaling using a Neural Network

A closer examination of Figure 7.7 reveals that in fact, the source image resolution is outperformed by several other permutations. Analysis of the AUC and EER scores indicate that optimum recognition when a scaling of 30% is applied to the source image. This in all likelihood represents the ideal resolution for optimum feature extraction. The AUC at this resolution indicates a 99.9998% probability of correct image classification with an equal error rate (EER) of 0.01%. While this makes for an interesting observation, these tests would need to be repeated on a much larger database to substantiate this level of accuracy.

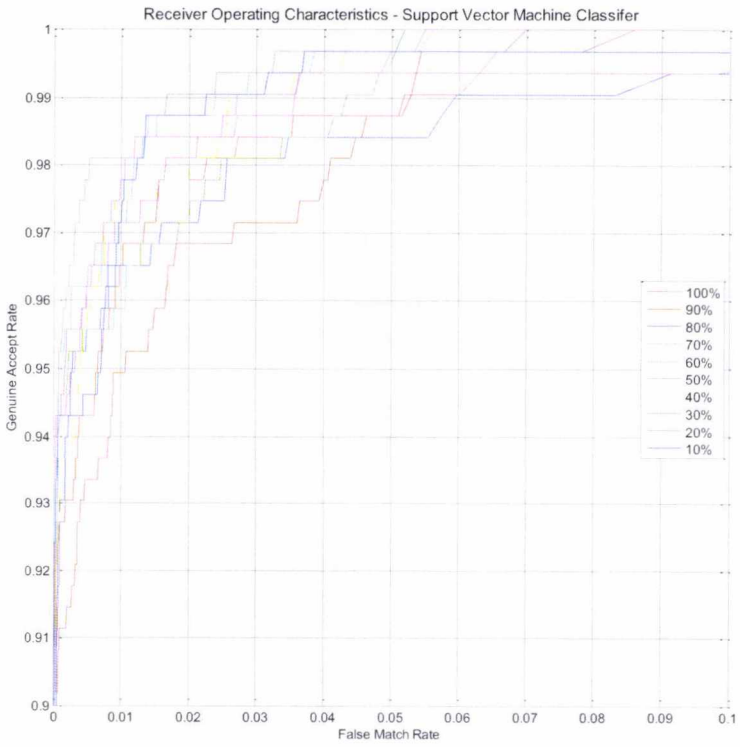


Figure 7.8 - Enlarged ROC Curve of Resolution Scaling using a Support Vector Machine

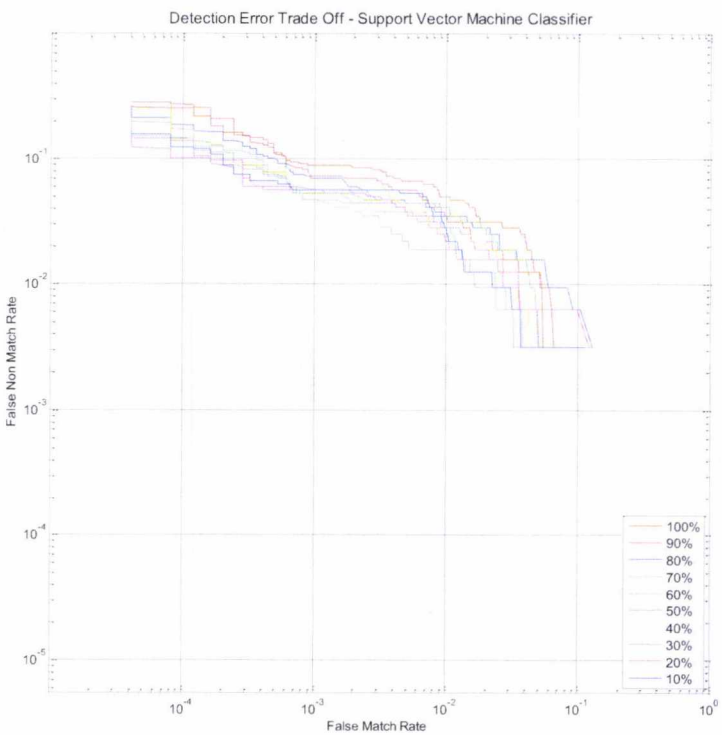


Figure 7.9 - DET Curve of Resolution Scaling using a Support Vector Machine

The support vector machine classification (Figure 7.8 and Figure 7.9) performance above indicates similar performance to the neural network. With an optimum scaling of 50% the AUC based chance of correct classification is 99.88% and EER of 1.33%.

Note that for the support vector machine and linear discriminant classifiers only the enlarged version of the ROC curves are only shown.

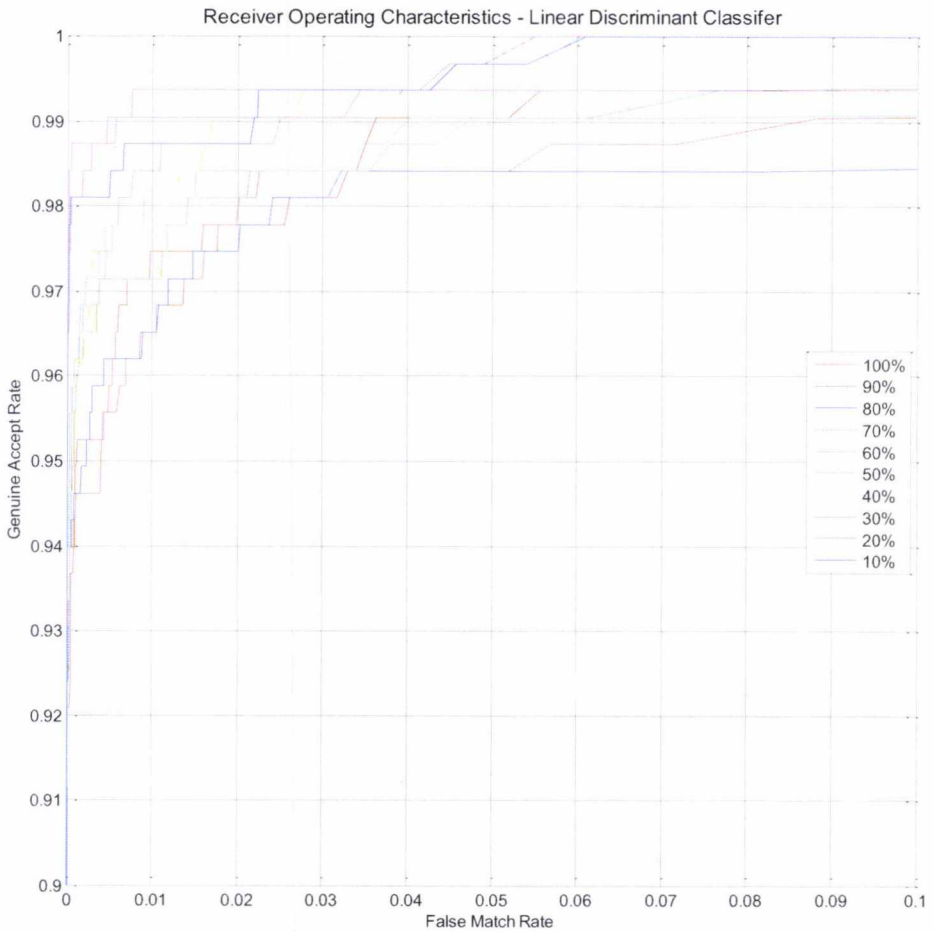


Figure 7.10 - Enlarged ROC Curve of Resolution Scaling using a Linear Classifier

Similar results are also achieved using the linear classifier with optimum recognition taking place at a scaling of 20% providing a probability of correct classification of 99.96% with an equal error rate of 0.7%.

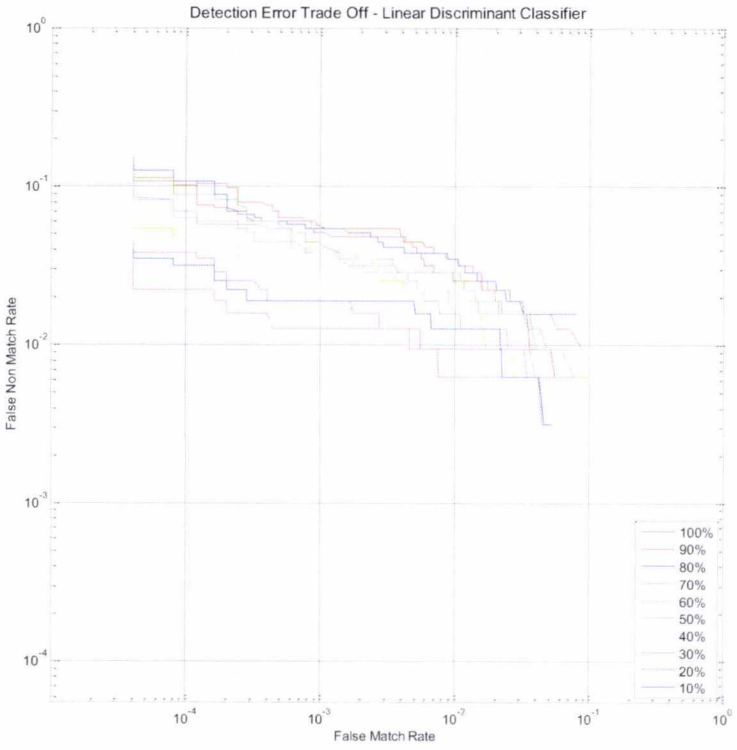


Figure 7.11 - DET Curve of Resolution Scaling using a Linear Classifier

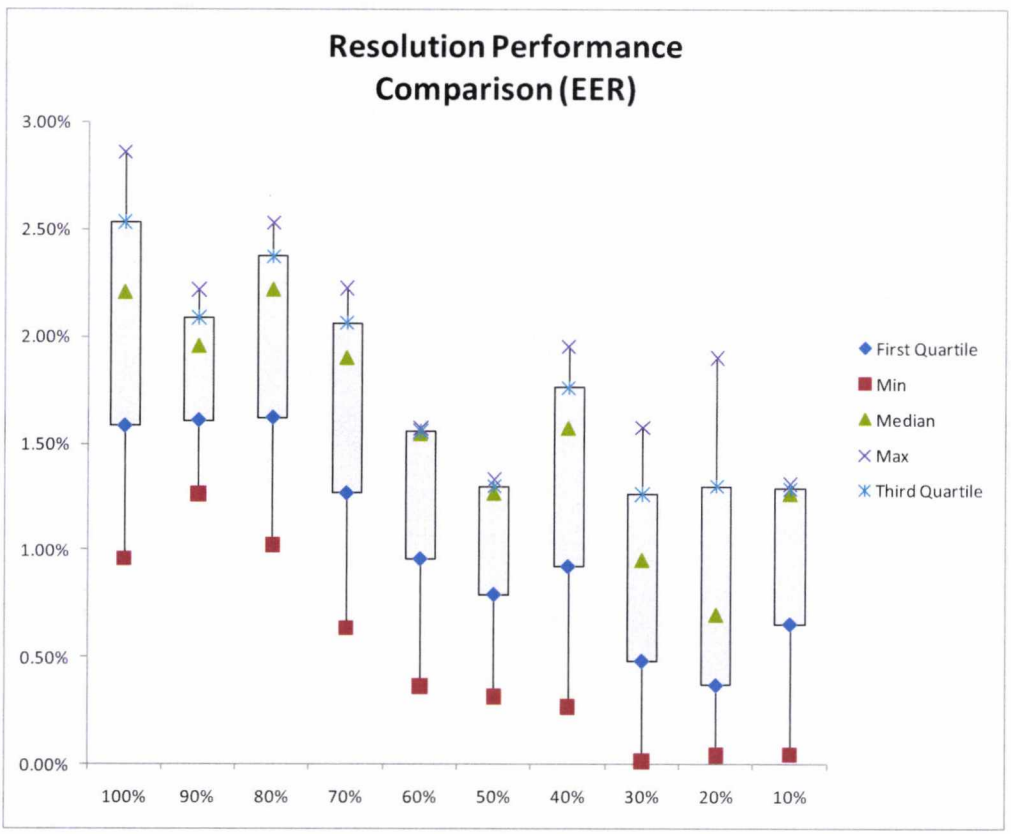
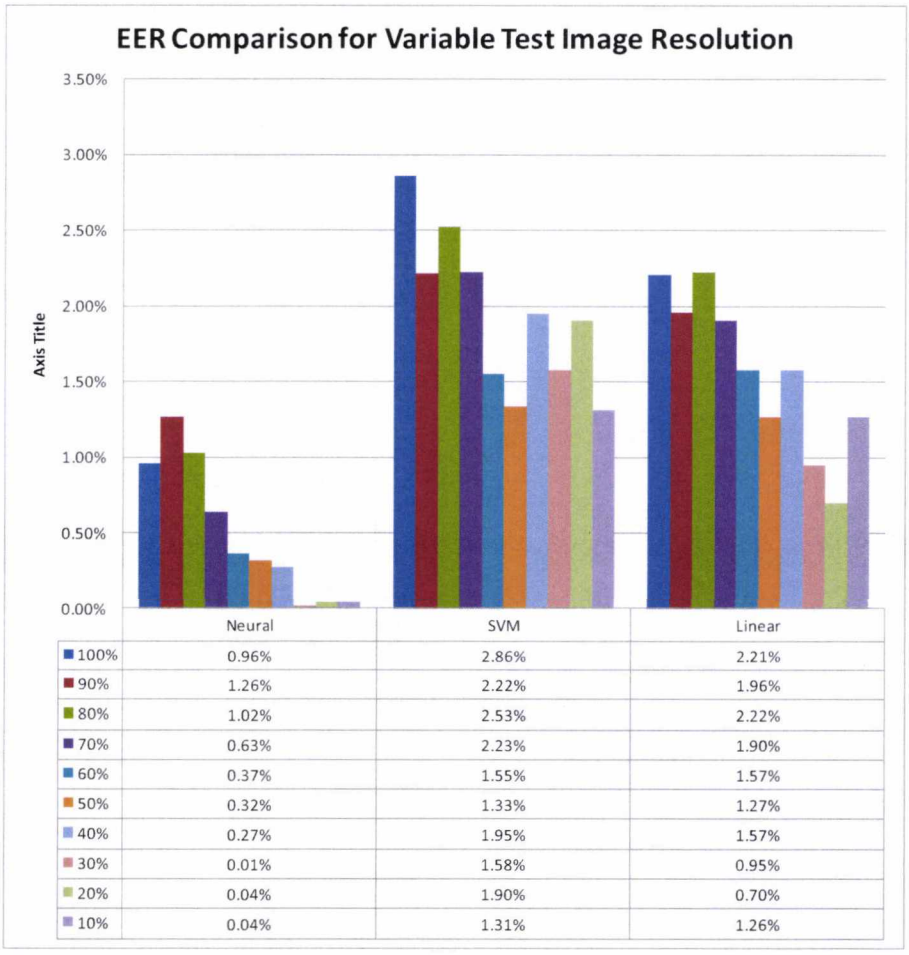


Figure 7.12 - Box plot showing the effect of Image Resolution on Classification Performance

The box plot in Figure 7.12 indicates that for the most consistent results across all classifiers, a resolution of 90% or 50% (representing a minimum eye separation of 165 pixels at 50%) may be preferred for the filter frequencies defined. It is plausible that an adaptation of the Gabor filter frequencies applied may hence improve results without the source image resolution being changed.

The above ROC Curves show similar performance for all resolution inputs with very little falloff as the resolution of the source image is decreased. Initially it may seem surprising that such accuracy can be maintained even with very small image resolutions. However as seen by earlier tests, the higher information metric is contained within the medium-low frequency patterns found within the Gabor filtered image.

These findings support the theory that as the image scale is reduced, these low and medium frequency patterns become less visible to the Gabor Filters and more visible to the high frequency GLCP features. Thus through the fusion of the Gabor and GLCP feature scores the system maintains the ability to identify facial skin texture even at extremely low resolutions with a good degree of accuracy.



	Neural	SVM	Linear
100%	0.96%	2.86%	2.21%
90%	1.26%	2.22%	1.96%
80%	1.02%	2.53%	2.22%
70%	0.63%	2.23%	1.90%
60%	0.37%	1.55%	1.57%
50%	0.32%	1.33%	1.27%
40%	0.27%	1.95%	1.57%
30%	0.01%	1.58%	0.95%
20%	0.04%	1.90%	0.70%
10%	0.04%	1.31%	1.26%

Figure 7.13 - Comparison of Equal Error Rate as Image Resolution is Decreased

At this stage it is unwise to draw the conclusion that lower resolution images are best suited for texture based recognition. It is more plausible that the GLCP part of the algorithm increases in its performance as described above. With larger frequencies, the Gabor filters may handle larger images with increased efficiency. Further analysis is required to make either claim.

7.3.2 TESTING AT AN UNTRAINED RESOLUTION

In the scenario where probe images are provided at a lower resolution than that which the system has been trained on, it is of interest to understand how the system will perform. Table 7.2 indicates the scaling parameters and graph colouring used within this section.

Scale Factor	Min Eye Separation	Min Dimensions	Graph Colour
1.0	330	128	Red (Dotted)
0.9	297	115	Red
0.8	264	102	Blue
0.7	231	89	Green
0.6	198	77	Cyan
0.5	165	64	Black
0.4	132	51	Yellow
0.3	99	39	Magenta
0.2	66	26	Magenta (Dotted)
0.1	33	13	Blue (Dotted)

Table 7.2 - Resolution Scaling

Figure 7.14 shows the configuration of this evaluation with each classifier being trained on the original texture features only. The same probe image set as used in Section 7.3.1 earlier is then used to assess the performance of each classifier. The expectation is that the performance will diminish as the probe images are downsized. Unlike the previous tests where the training image size was also reduced, this battery of tests are intended to evaluate how the system will respond to images that may have been acquired at a lower resolution once training has already taken place.

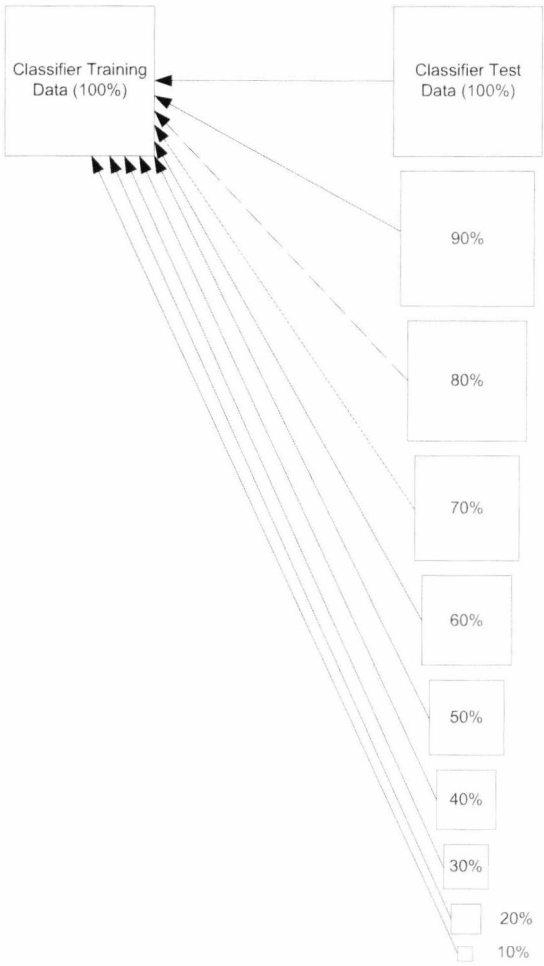


Figure 7.14 - Reduction of Probe Image Size

The performance difference between changing just the testing probe image size from the previous section where both training and testing image sizes were changed is immediately apparent. There is an exponentially increasing falloff as input image resolution is decreased which becomes dramatically pronounced once the source image is decreased below 50% of the size of the training image.

Figure 7.15 - Wavelength versus Image Size

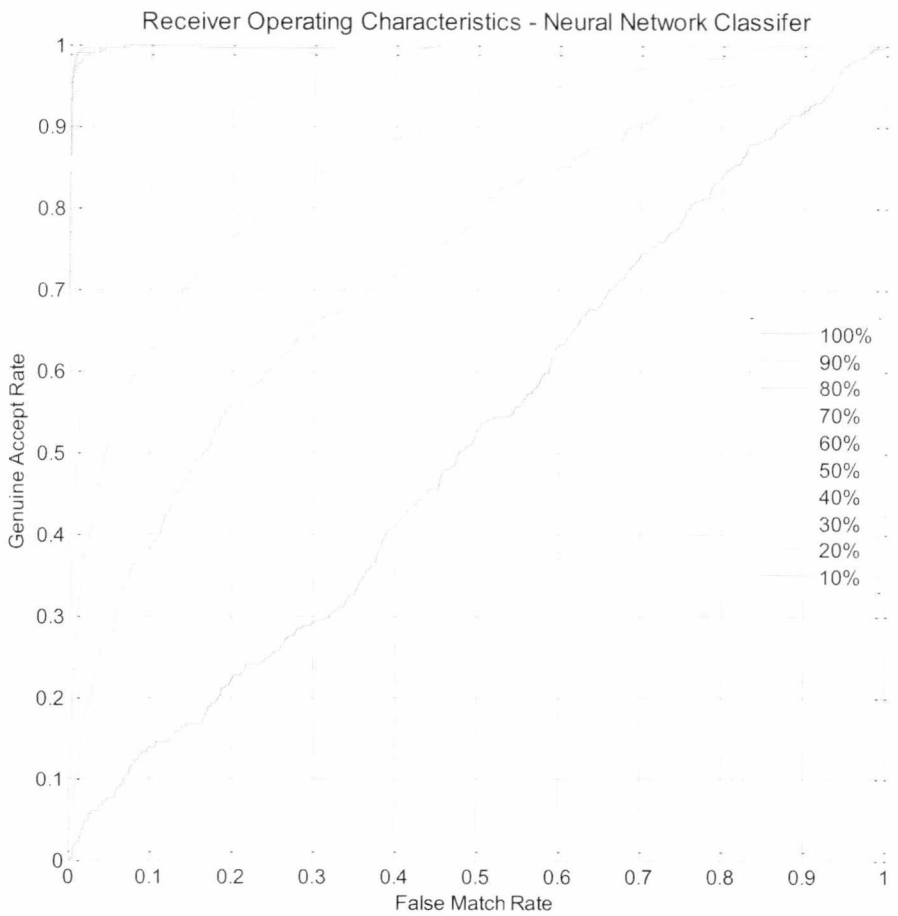


Figure 7.16 - ROC Curve of Image Resolution Scaling using a Neural Network

Taking the initial Gabor parameters supplied from Section 5.5 and plotting the filter wavelength versus the image dimensions in pixels it can be observed from Figure 7.15 that as the image resolution crosses 50 pixels (using Table 7.2 this equates to 40% of the input image size) the $f = \frac{4}{x}$ filter becomes redundant which correlates with the classification response showing a marked drop-off after 40%. At this point, the classifiers are now relying almost totally on the $f = \frac{10}{x}$ and $f = 0.2$ Gabor filters (as it is plausible that the GLCP filters are providing little useful information for class separation due to their reliance on relative high frequency spatial locations which are likely lost in the shrinking process).

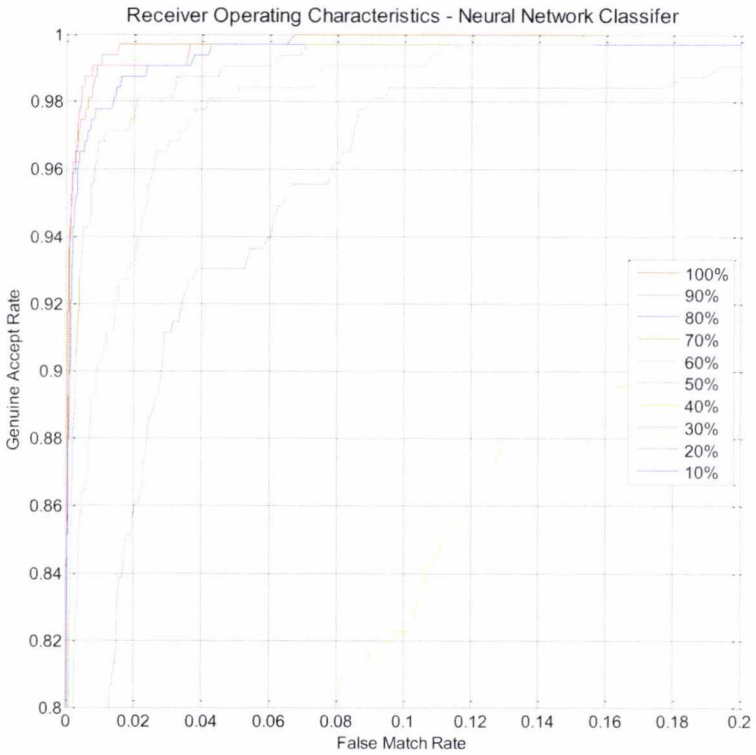


Figure 7.17 - ROC Curve of Image Resolution Scaling using a Neural Network (Enlarged)

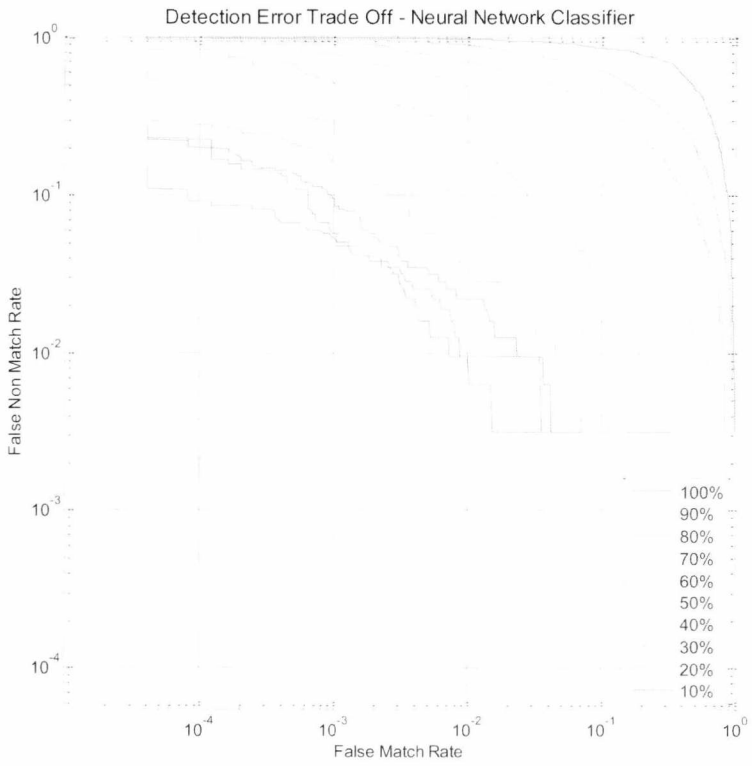


Figure 7.18 - DET Curve of Image Resolution Scaling using a Neural Network

Similar trends are also noticeable in both the support vector machine and linear classifiers as shown in the below ROC and DET curves. In the case of the support vector machine, the performance follows a slightly balanced trend with performance loss remaining fairly well linearly spaced between 60% and 20% of the source image.

The GLCP filters used in the system is particularly susceptible to variations in image size as no normalisation is performed for this. However the results show that the classifiers may be capable of using the feature information from the Gabor filters to offset this failing. Only when the Gabor filters themselves become negatively influenced by the reduced input image scale do the results begin to exponentially deteriorate.

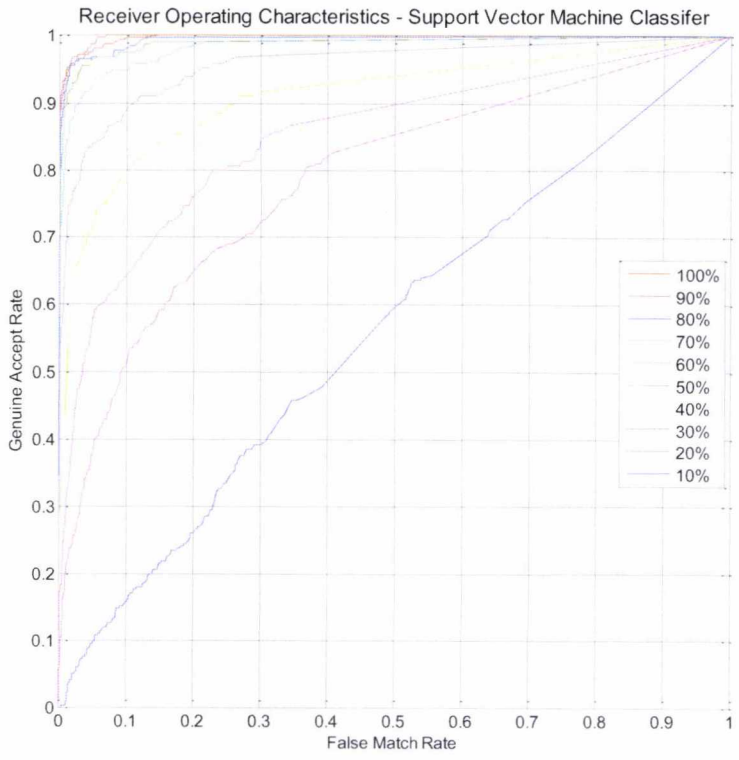


Figure 7.19 - ROC Curve of Image Resolution Scaling using a Support Vector Machine

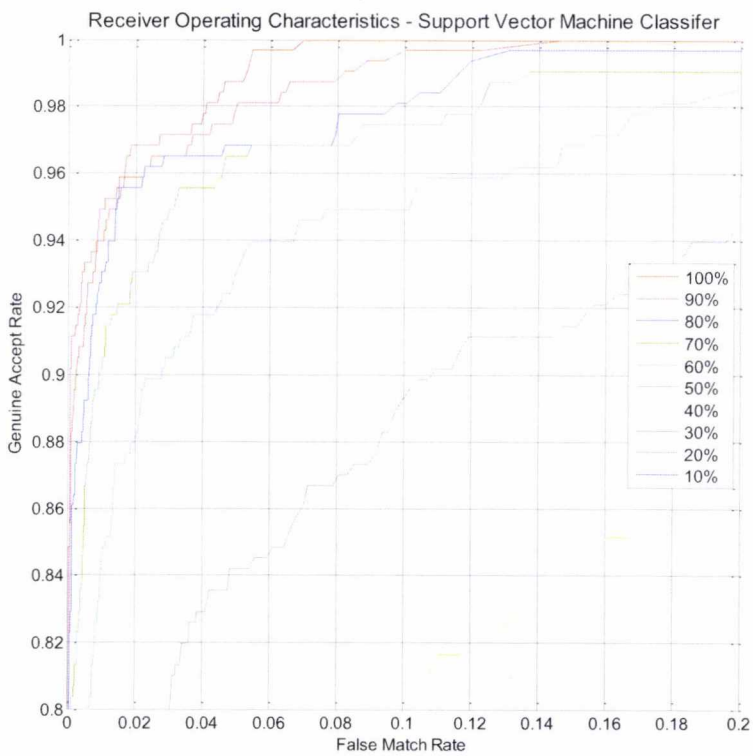


Figure 7.20 - ROC Curve of Resolution Scaling using a Support Vector Machine (Enlarged)

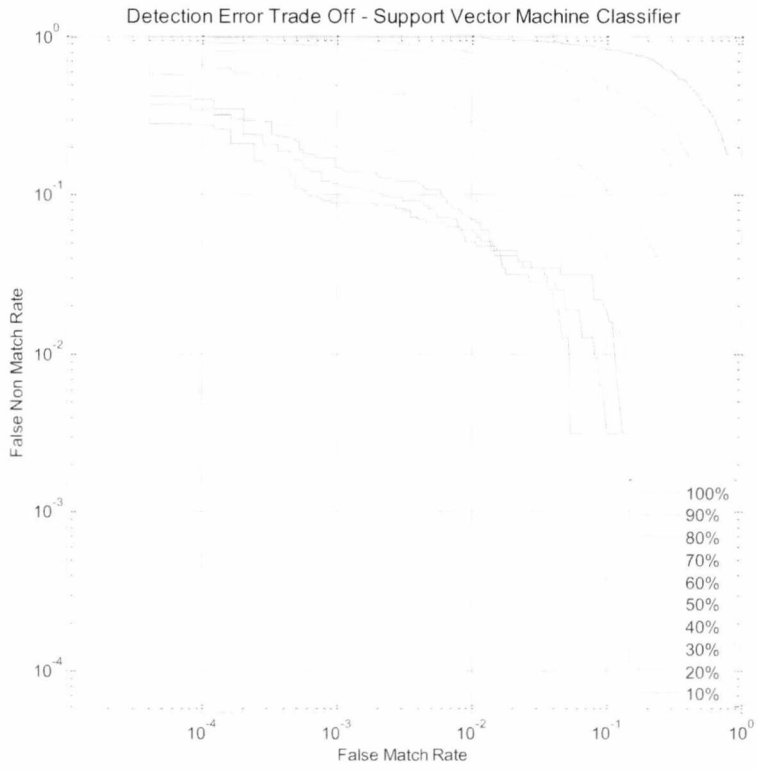


Figure 7.21 - DET Curve of Probe Image Resolution Scaling using a Support Vector Machine

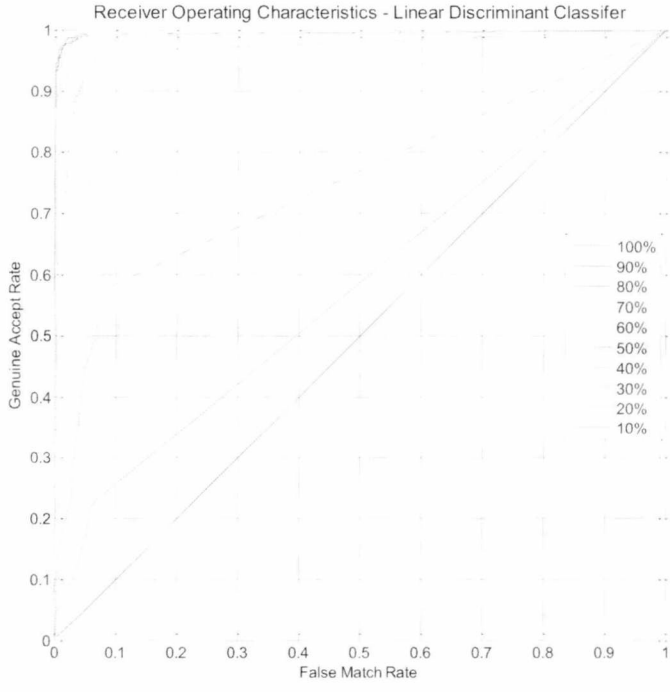


Figure 7.22 - ROC Curve of Image Resolution Scaling using a Linear Classifier

It is also possible that this is merely an anomaly which would not be sustained given a larger dataset.

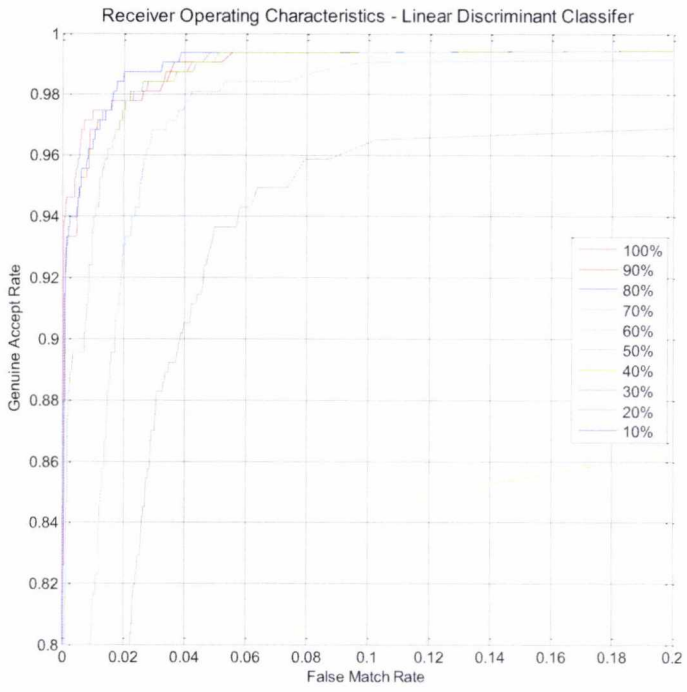


Figure 7.23 - ROC Curve of Image Resolution Scaling using a Linear Classifier (Enlarged)

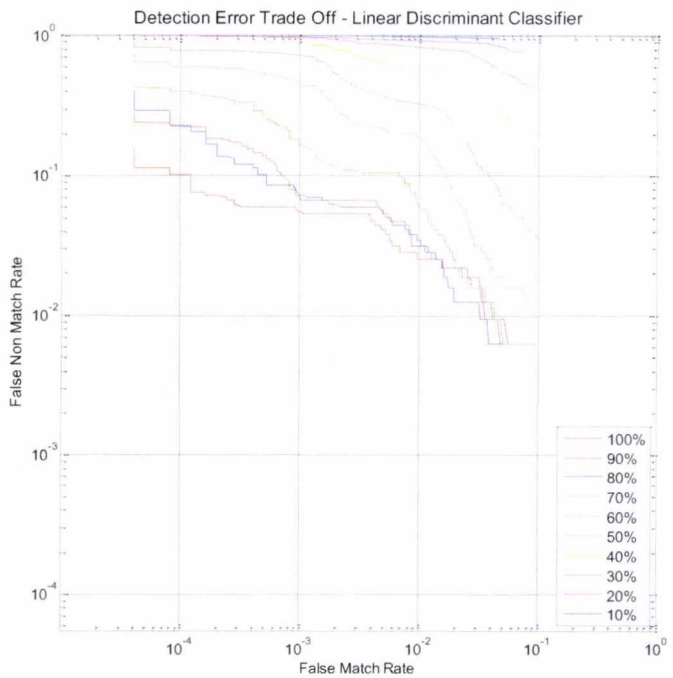
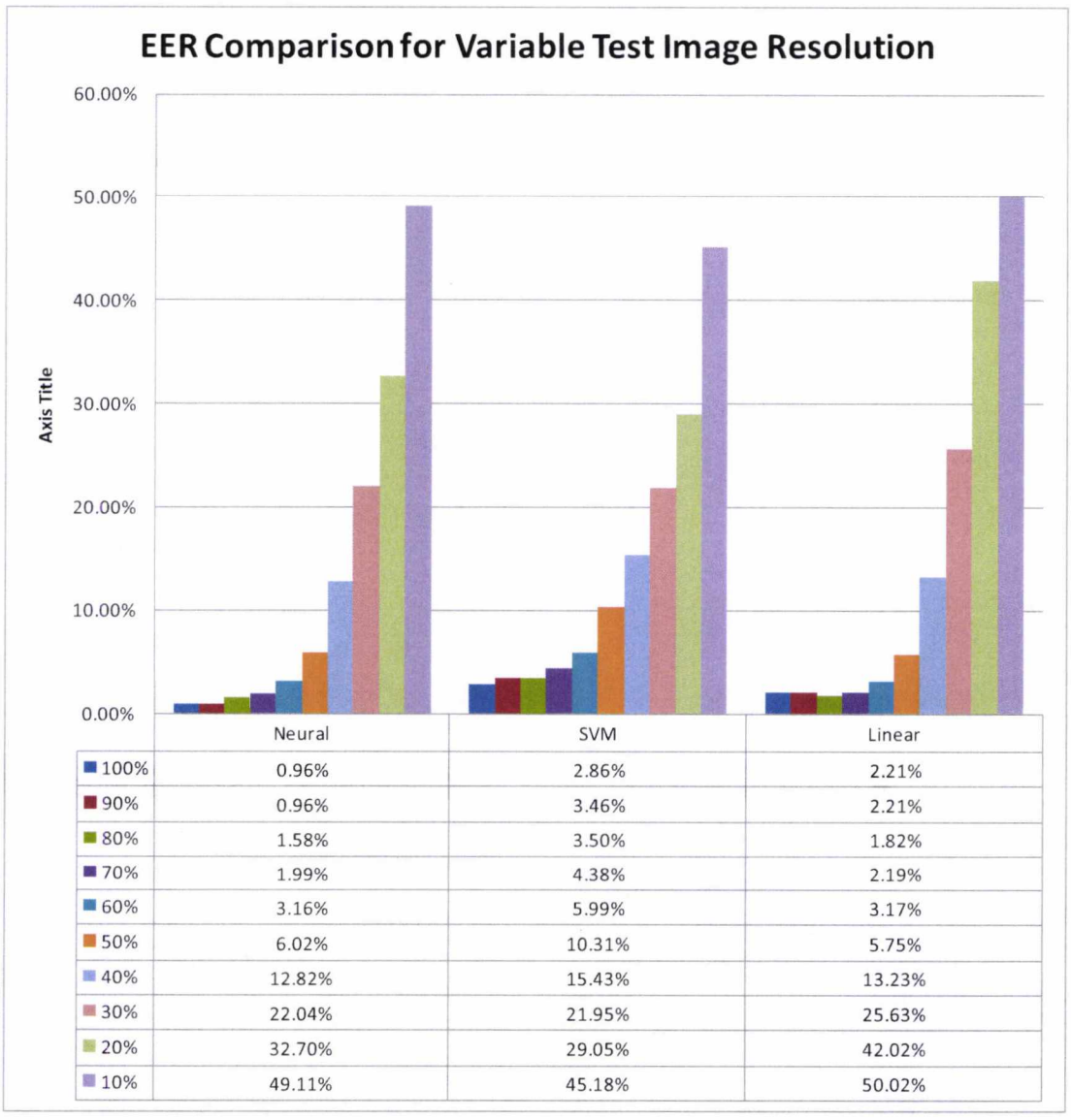


Figure 7.24 - DET Curve of Probe Image Resolution Scaling using a Linear Classifier

The linear classifier assumes the same performance trends as the neural network showing a marked falloff in performance after 50% of the image resolution (representing the point where the $f = \frac{4}{x}$ filter becomes redundant).

On a close examination of the ROC curve for the linear classifier (Figure 7.23) it is interesting to observe that performance shows a slight increase over the source image at a resolution scaling of 80%. The reasons for this behaviour are not obvious. Linear classifiers operate by attempting to select the most descriptive features using a feature dimensionality reduction algorithm. This given, it is possible that in this instance, the features selected provide a better description of the textures at this resolution.



	Neural	SVM	Linear
100%	0.96%	2.86%	2.21%
90%	0.96%	3.46%	2.21%
80%	1.58%	3.50%	1.82%
70%	1.99%	4.38%	2.19%
60%	3.16%	5.99%	3.17%
50%	6.02%	10.31%	5.75%
40%	12.82%	15.43%	13.23%
30%	22.04%	21.95%	25.63%
20%	32.70%	29.05%	42.02%
10%	49.11%	45.18%	50.02%

Figure 7.25 - Comparison of Equal Error Rate as Probe Image Resolution is Decreased

The equal error rates reinforce previous findings that the features extracted from low resolution images bear less resemblance to training set data. The manner in which the decline accelerates after 50% and again after 20% corresponds with the dynamically sized Gabor filters crossing the 0.2 cycles/pixel frequency threshold.

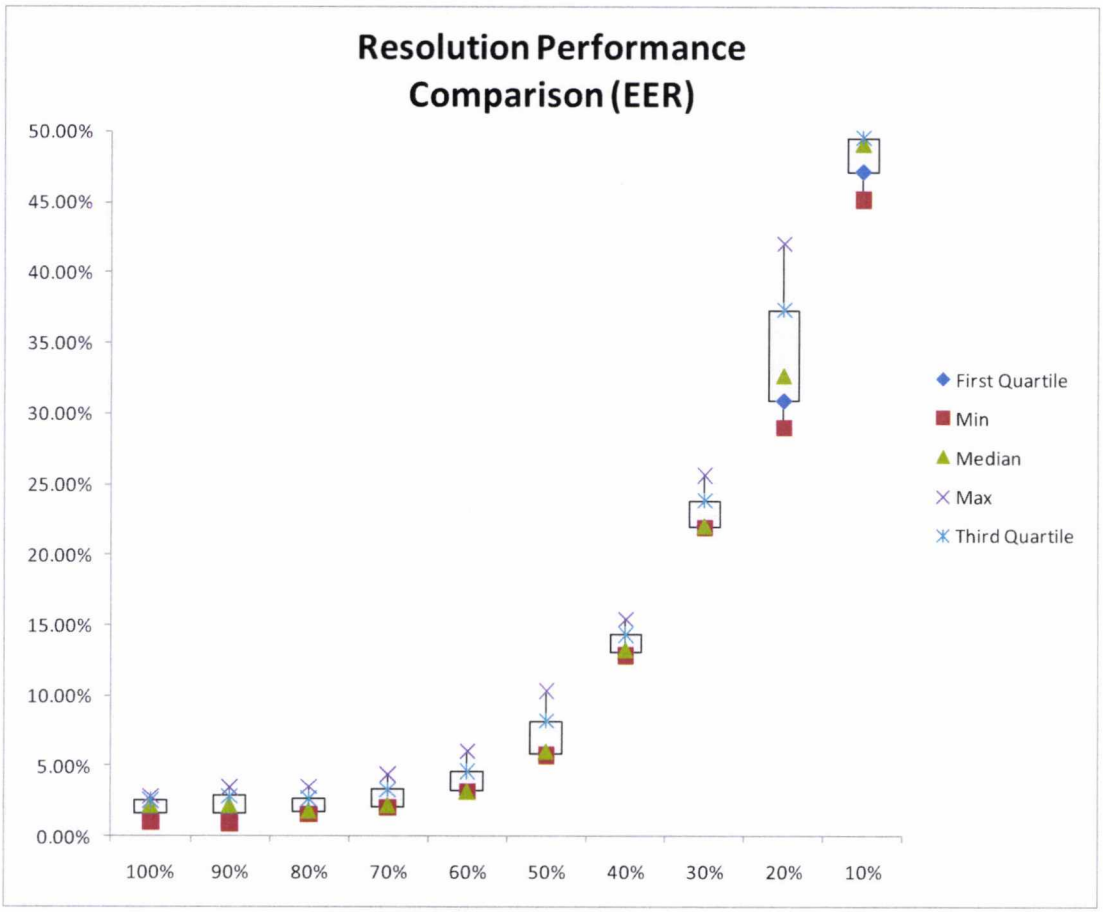


Figure 7.26 - Plot showing the effect of probe image resolution on classification performance

A box style plot of the output classification scores clearly shows the falloff that occurs after the 50% scaling factor although it should be noted that 40% scaling in some cases still maintains an equal error rate of under 15%. This clearly indicates the robustness of the system against variations in scale with the Neural Network achieving an EER of 6.0% at the 50% scaling threshold and 12.8% at 40% scaling.

7.4 CROSSOVER OF MULTI-BIOMETRIC REGIONS

Thus far consideration has only been given to providing test data of the same skin region to which the classifier has received training for. For example, a classifier trained on skin patches for region 2A has only been tested on skin probe images of the same region.

This section shows the performance of the system when classifiers trained for region 2A are tested using skin patches for 2B and vice versa and hence shows a measure of uniqueness of a person's facial skin in general.

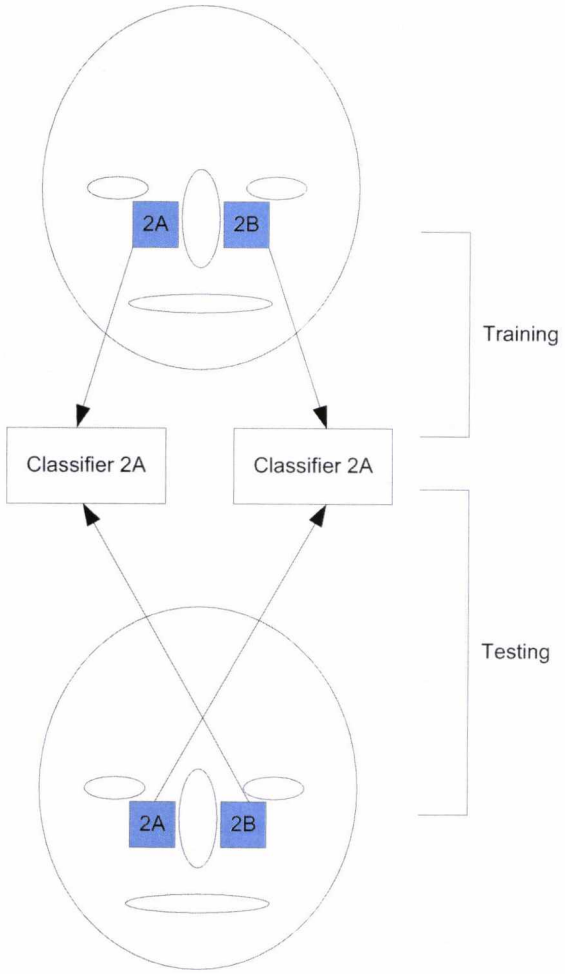


Figure 7.27 - Cross Region Classification Illustration

By investigating this question, it can be shown whether the skin texture is unique to the region from which it is extracted, or to the person in general.

Classifier	Region (train)	Region (test)	Colour
Neural Network	2A	2B	Red
Neural Network	2B	2A	Blue
Support Vector Machine	2A	2B	Green
Support Vector Machine	2B	2A	Blue
Linear	2A	2B	Magenta
Linear	2B	2A	Black

Table 7.3 - Cross Region Classification Parameters

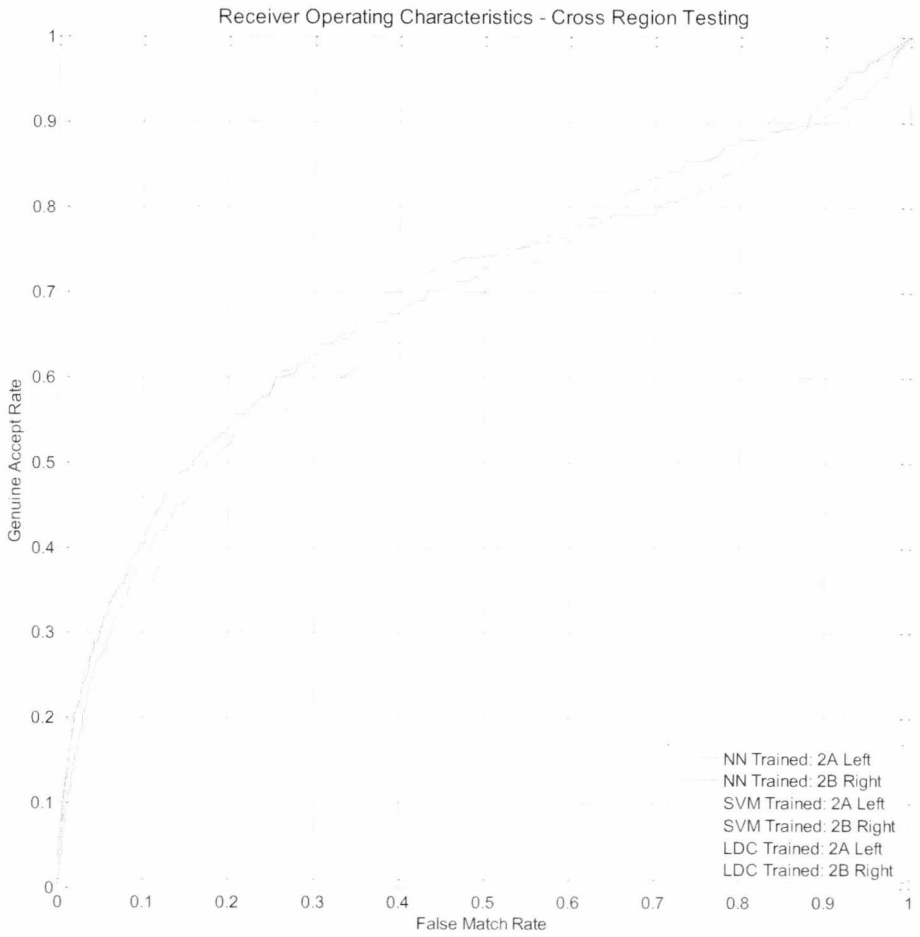


Figure 7.28 - Cross Region Testing - ROC Curve

It is seen clearly from the ROC and DET curves that the performance is substantially decreased when cross-region matching is attempted although the results remain positive which shown a potential for matching using a rank based performance measurement system which would score results based on the probability of them being within a defined number of outputs.

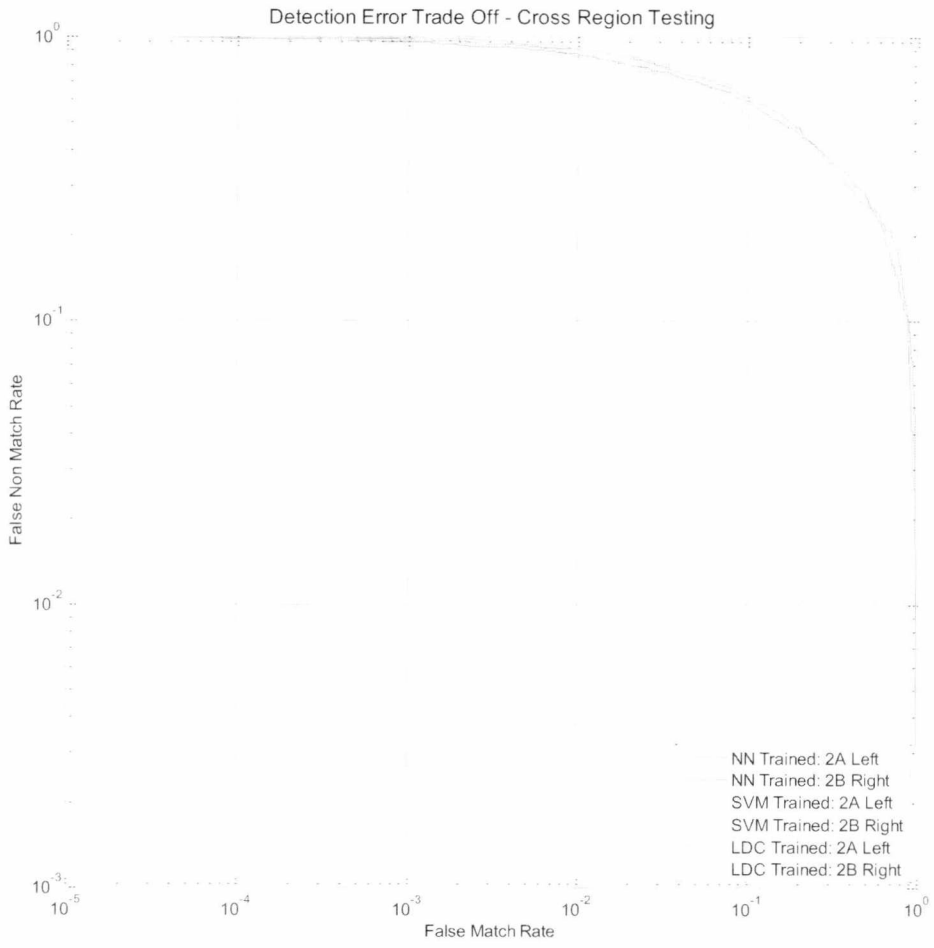


Figure 7.29 - Cross Region Testing - DET Curve

Comparing the AUC and EER levels again confirms that positive classification does occur, but the accuracy levels fall well below that of using the same region for both classification training and testing.

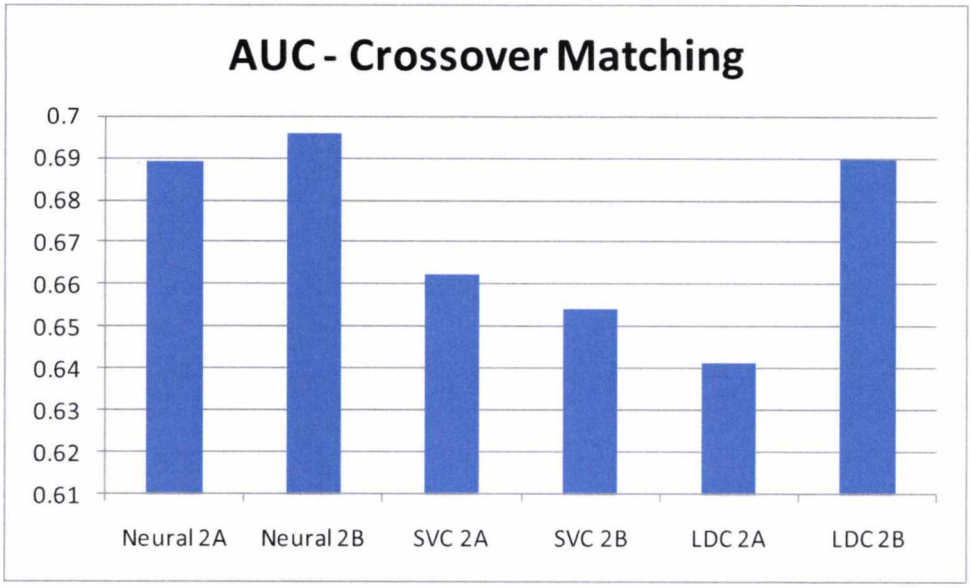


Figure 7.30 - Cross Region Testing - Area Under the Curve

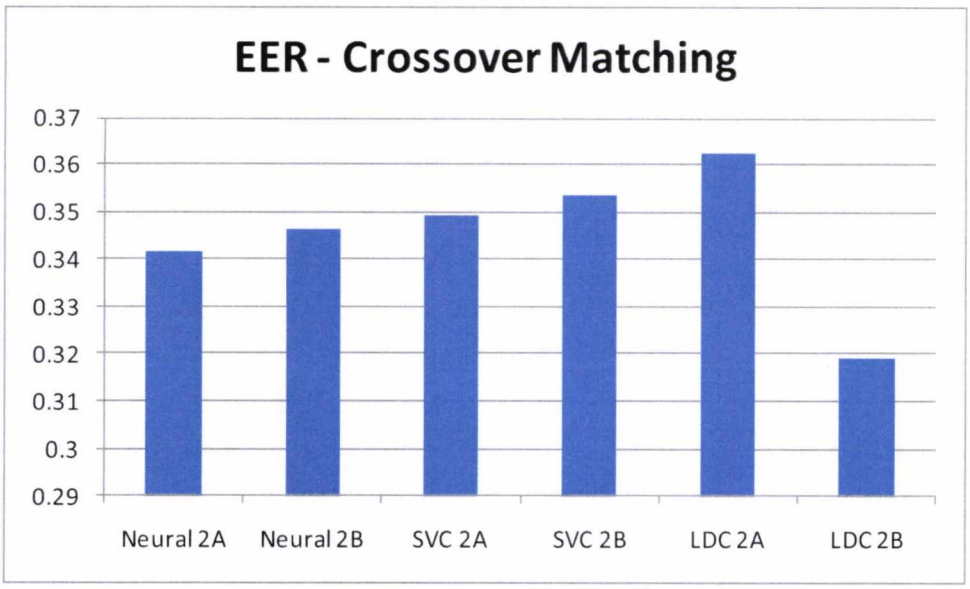


Figure 7.31 - Cross Region Testing - Equal Error Rate

All classifiers show similar outputs within 5% EER and 6% AUC which indicates a potential to perform filtering based on skin texture samples from other areas. It is hypothesised that the categorisation performed by the classifiers at this stage is recognising "skin types" rather than the person's identity. This is to say that it may be possible to classify an individual based on his or her similarity to a core thinned set of skin textures known by the classifier rather than his or her uniqueness as an individual. An application similar to the Eigenfaces method may be an alternate line of research to improve these results.

7.5 System Confidence

Several claims have been made pertaining to the accuracy of the algorithms and classification systems presented in this Thesis. Although substantiated by the evidence presented, it is useful to know how the system would perform against a larger dataset and to what level of confidence these findings may be employed.

The "Rule of 3" [Mansfield and Wayman 2002] states that the lowest error rate that can be statistically established with a given number N of comparisons is equal to:

$$p \approx \frac{3}{N} \text{ for a 95\% confidence level.} \tag{7.1}$$

Using the BMDB database, a total of 316 test comparisons have been made which indicates that the lowest error rate that may be statistically established as being 95% confident is 0.95%.

It is apparent that the DS2 BMDB dataset employed in this research is on the limits of the accuracy threshold for the performance claims. While in Chapter 0there are several suggestions to enhance the accuracy of the system, attempting any of these would require a larger dataset to ensure that any improvements to the accuracy of the system would give a sufficient level of confidence.

Neural networks in particular exhibit slightly different classification outputs depending to some extent on random seed values used during training and network convergence. It has been observed during this evaluation that a variation of up to 0.1% in the AUC and EER values may be seen between repeated neural network performance outputs using the same training probe images. For this reason it is suggested that the relative performance of differing systems be weighted towards the support vector machine and linear discriminant classifiers with the neural network viewed as a potential performance output.

7.6 Chapter Review

In this Chapter, a critical review of the systems performance, accuracy and confidence is discussed that has covered the following:

- Time taken to extract features (excluding file I/O)

- Time taken to train a classifier
- Accuracy of the system in relation to input image resolution variance from 100% to 10% of original image when both training and probe images are resized together.
- Accuracy of the system in relation to input image resolution variance from 100% to 10% of original image when only the probe images are resized.
- Statistical confidence of the reported system accuracy.
- Comparison with recent evaluations.

It should be highlighted that although the results for Section 7.3.1 appear impressive with almost 100% reported accuracy and tiny EER values, this should be read with relative caution. As highlighted in the latter sections of this chapter it is considered highly likely that the algorithm will be unable to maintain such high accuracy rates on large datasets. In contrast however, it is plausible that the feature vectors may benefit from some form of reduction such as principal component analysis (PCA). This would be an interesting future avenue of research.

Section 7.3.2 shows the robustness of the system when the resolution of input images and indicates that even at half the original image resolution an equal error rate of 6% may be achievable.

8 CONCLUSIONS

8.1 INTRODUCTION

This chapter concludes the study with a critical appraisal of the research. A statement is made concerning the nature of the work presented previously and its scope. This statement is followed by a summary of the findings and a performance comparison against similar techniques in the last five years. The preliminary nature of the research is outlined with a call for future research. Possible areas of future study are presented with a recommended approach to further the field. In closing, a brief overview of the framework and protocols are given with final chapter conclusions.

8.2 SUMMARY OF WORK

This thesis began with a hypothesis suggesting that facial skin texture from two specific regions may be a biometric feature suitable for face recognition. A review of feature extraction methods capable of describing texture was presented and Gabor filters were selected to perform the evaluation. A mathematical basis for a filters construction has been presented with filter parameters selected based on referenced literature and experimental results conducted as part of this study on a training dataset (HRDB). The HRDB dataset was collected as part of this research for developmental and training purposes but is not publicly accessible for comparison.

Following initial experimentation on the HRDB dataset, a testing dataset (BMDB) was selected for analysis. The BMDB database is a recently collected database which includes 8 high resolution images of each subject taken at two different sittings with two different lighting compositions. The dataset was divided into a training and probe set of 4 images per subject for each purpose. An evaluation of the proposed algorithm was then executed. This indicated that features from the left skin region gave an area under the receiver operating characteristics curve (AUC) of 99.6% and the right region of 99.3% when using a neural network with respective equal error rates (EER) of 2.8% and 3.2% for the same classifier. Although promising, the support vector machine and linear classifiers struggles to attain a similar level of accuracy with the SVM reaching a peak AUC of 93.1% (10.5% EER) and the linear classifier having an AUC of 79.1% (21.2% EER).

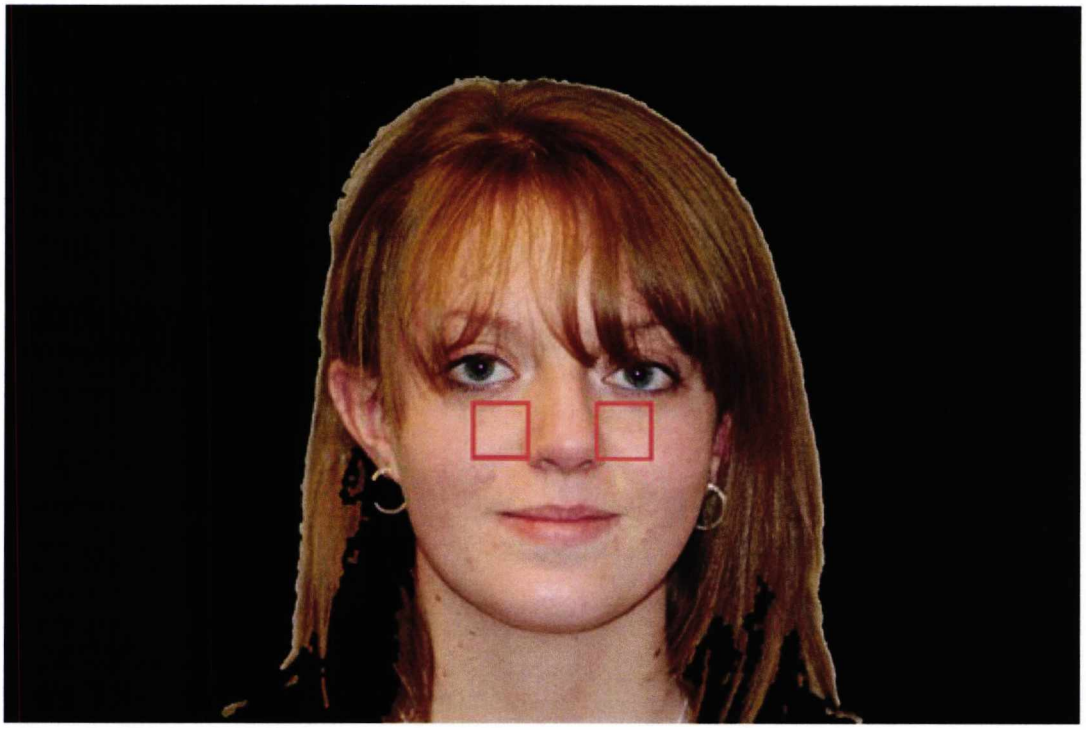


Figure 8.1 - Skin Texture Regions used in this Evaluation

The algorithm was then developed further to evaluate the performance benefits of fusing the two proposed skin texture regions which showed an improvement on the base performance metrics. The system was then extended once more to provide the ability of extracting features from high resolution images using the Grey Level Co-occurrence Probability Matrix (GLCP) method. A total of 16 possible permutation configurations were illustrated and each configuration evaluated including the separate testing of the GLCP system in isolation from the Gabor features. The GLCP system showed capable performance in its own right attaining an average AUC across all classifiers and regions of 84.3% with an EER of 17.6%.

It was found that a full score fusion configuration offered the best overall AUC and EER values across all classifiers with minimum variance in the outputs (AUC 99.8%, EER 2.0%). This was then used to measure performance against variations in scale. A baseline was established to show the systems reliance on resolution. The system showed minimal variations when the training and probe image resolution were reduced together and the system retrained on the lower resolution training images. A theory for this behaviour was noted in Section 7.3.1 .

The system was then evaluated to assess performance against lower resolution of probe images and was found to have a good level of robustness against variations when probe image size was reduced independently of the training images maintaining an EER of under 6% (neural network and linear classifiers) when probe image resolution was 50% of training

image resolution. After this point the EER rapidly rose reaching 22% at 30% probe image resolution with an unusable level of classification at 10% resolution.

The final evaluation consisted of attempting to classify regions of skin from sides of the face to classifiers trained on the opposite region. This showed performance substantially under that which was observed when the same classifiers classified similar skin regions with AUC values averaging 67.2% and an average EER of 34.5% across all classifiers and cross-regions. Interestingly this does still indicate that some limited recognition is possible between skin texture regions from different areas of the face.

A summary of the non-scaled results taking the mean AUC and EER values across all classifiers is shown in Figure 8.2:

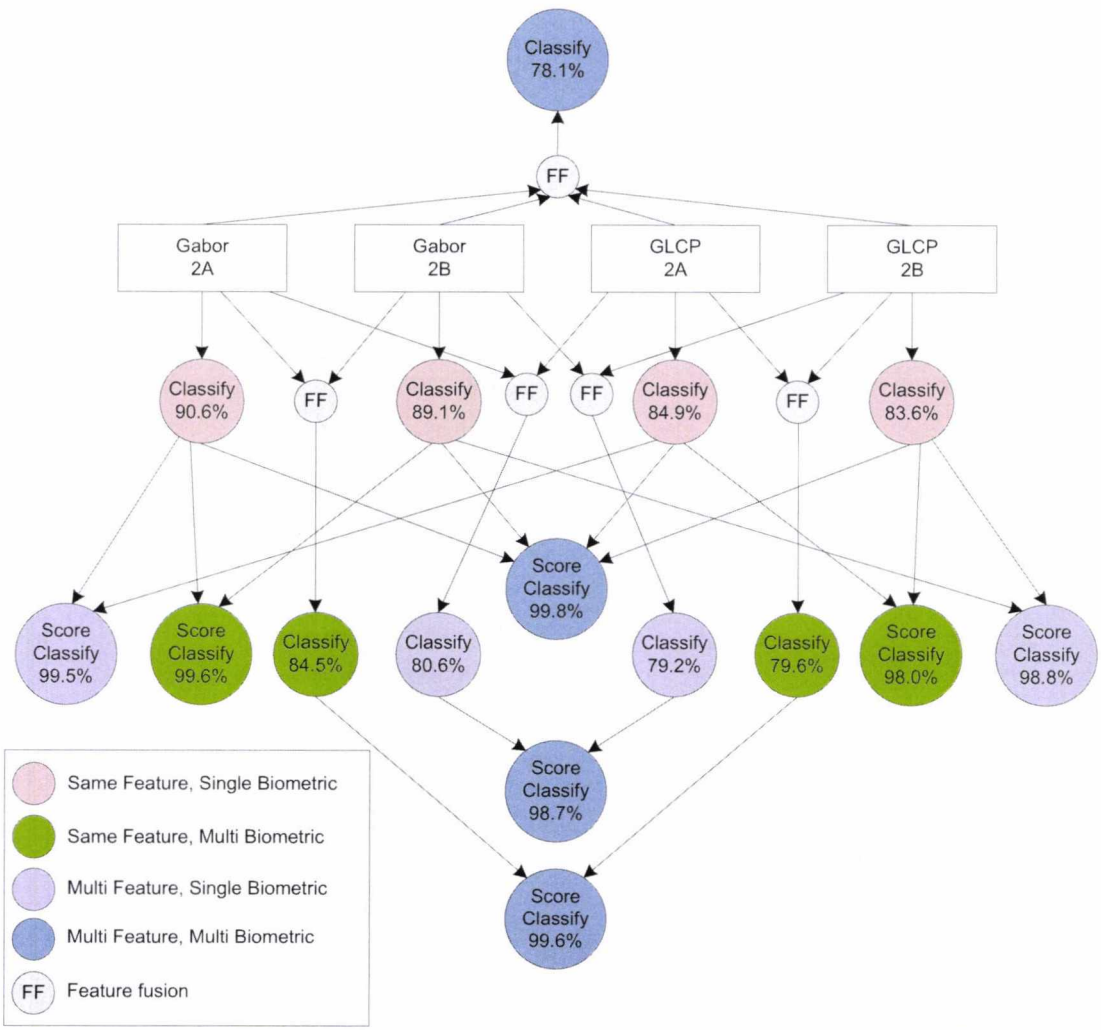


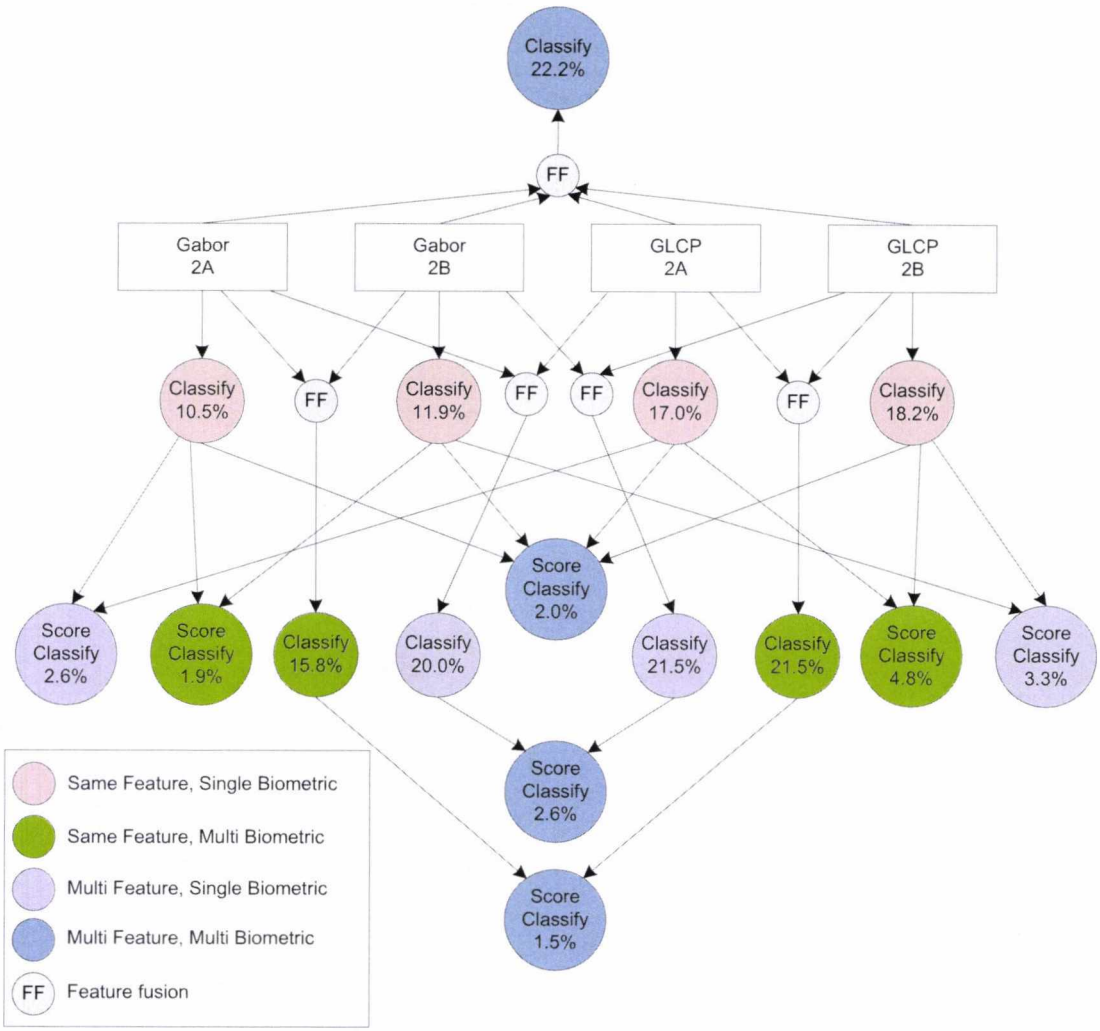
Figure 8.2 - Area Under Curve Comparison (Average Classifier Outputs)

Ranking each configuration in terms of probability of a successful classification (as indicated by the AUC value) sequences the systems as shown in Table 8.1. Note that these results are independent of classifier and have been calculated using the average values shown in the above figure. It can be seen that score fusion performs towards the top end of the table with full feature fusion occupying bottom place in the rankings. For ease of comparison, the colours used in Figure 8.2 are replicated in the score table.

Rank	System	Description
1	MFMB System 4	Full score fusion of all features
2	SFMB Score Gabor	Gabor score fusion (Both Regions)
3	MFMB System 2	Score fusion by feature type (feature fusion of different regions)
4	MFSB 2A Score	Score fusion both feature types for left eye (Region 2A)
5	MFSB 2B Score	Score fusion both feature types for right eye (Region 2B)
6	MFMB System 3	Score fusion by region (feature fusion of different features)
7	SFMB Score GLCP	GLCP score fusion (Both Regions)
8	SFSB 2A Gabor	Gabor features for left eye (Region 2A)
9	SFSB 2B Gabor	Gabor features for right eye (Region 2B)
10	SFSB 2A GLCP	GLCP features for left eye (Region 2A)
11	SFMB Feat Gabor	Gabor feature fusion (Both Regions)
12	SFSB 2B GLCP	GLCP features for right eye (Region 2B)
13	MFSB 2A Feature	Feature fusion of both feature types for left eye (Region 2A)
14	SFMB Feat GLCP	GLCP feature fusion (Both Regions)
15	MFSB 2B Feature	Feature fusion of both feature types for right eye (Region 2A)
16	MFMB System 1	Full feature fusion of all inputs

Table 8.1 - Ranked Configurations based on Average AUC

Performing the same analysis on the equal error rates shows a similar but not identical order of performance. The maximum change of ranked position between the AUC and EER performance tables is 2. An observation that can be made at this stage is how well Gabor features perform when used without GLCP. Repeating the evaluation on a much larger database (see Section 8.3 and 8.5.1 for a discussion of possible databases) would be beneficial to see if these trends persist across large databases.



Rank	System	Description
1	MFMB System 2	Score fusion by feature type (feature fusion of different regions)
2	SFMB Score Gabor	Gabor score fusion (Both Regions)
3	MFMB System 4	Full score fusion of all features
4	MFMB System 3	Score fusion by region (feature fusion of different features)
5	MFSB 2A Score	Score fusion both feature types for left eye (Region 2A)
6	MFSB 2B Score	Score fusion both feature types for right eye (Region 2B)
7	SFMB Score GLCP	GLCP score fusion (Both Regions)
8	SFSB 2A Gabor	Gabor features for left eye (Region 2A)
9	SFSB 2B Gabor	Gabor features for right eye (Region 2B)
10	SFMB Feat Gabor	Gabor feature fusion (Both Regions)
11	SFSB 2A GLCP	GLCP features for left eye (Region 2A)
12	SFSB 2B GLCP	GLCP features for right eye (Region 2B)

Rank	System	Description
13	MFSB 2A Feature	Feature fusion of both feature types for left eye (Region 2A)
14	MFSB 2B Feature	Feature fusion of both feature types for right eye (Region 2A)
15	SFMB Feat GLCP	GLCP feature fusion (Both Regions)
16	MFMB System 1	Full feature fusion of all inputs

Table 8.2 - Ranked Configurations based on Average EER

8.3 COMPARISON OF RESULTS

Due to the recent release of the BMDB database, it is not currently possible to provide a direct comparison of competing methods of face recognition against the same dataset. The system appears to perform favourably when compared with other recent approaches to face recognition.

[Yu, Teng et al. 2006] demonstrate a combining system that leverages Gabor Jets, Eigenface, Spectroface and independent component analysis recognition systems and show a peak accuracy of 92.50% at Rank 1 using 40 individuals from the Feret face database.

[Choi, Tse et al. 2008] construct a simplified Gabor wavelet which shows dramatic computational performance advancements over traditional fast Fourier transform based systems. Against the Yale B face database which consists of 10 individuals with 64 images of each subject an optimum performance score of 95.47% is achieved using a bank of 16 filters.

[Singh, Vatsa et al. 2008] use Gabor features extracted from visible and long-wave infrared face images with a support vector machine classifier and achieve results of 95.85% and 94.98% on the Notre Dame face database (159 individuals) and Equinox face database (95 individuals, respectively. In both cases, 3 images were used for training with all remaining images used as probe images.

[Yu, He et al. 2010] combine both Gabor phase and magnitude information to show that both are in fact complimentary to superior performance. They achieve a recognition rate of 95% against the Feret database using 4 training samples and 2 test samples. This demonstrated a 0.42% recognition rate increase against a purely phase based system, and 5.42% against a system which used features solely from the filtered magnitude.

Recently, [Zhang and Qiao 2010] present a method of using gradient based Gabor filters with a support vector machine to perform face recognition on a large database from The Face

Recognition Grand Challenge (FRGC) [Phillips, Flynn et al. 2005]. The database consists of 24,042 images of 466 subjects. The system achieved a recognition rate of 93.1% when using native resolution images (of between 1600x1200 pixels and 2272x1704 pixels). Normalising the images to a resolution of 128 x 144 pixels resulted a recognition rate of 78.4% with a false acceptance rate of 0.1%.

Bar	Team	Type	Database Size Images (Subjects)	Recognition Rate
A	[Yu, Teng et al. 2006]	Hybrid	240 (40)	92.50%
B	[Yu, Teng et al. 2006]	Gabor	180 (30)	80.67%
C	[Singh, Vatsa et al. 2008]	Gab-Multi Spectral	? (95)	94.98%[1]
D	[Choi, Tse et al. 2008]	Simplified Gabor Wavelet	640 (10)	95.47%
E	[Yu, He et al. 2010]	GMPTR+NDLA	720 (120)	95.00%
F	[Zhang and Qiao 2010]	Gradient Gabor	24042 (466)	93.10%
G	This study (FMR=0.001)	Gabor-GLCP	632 (79)	95.25%
H	This study (at EER)	Gabor-GLCP	632 (79)	99.04%[2]

Table 8.3 - Comparison of Other Recognition Methods utilising Gabor Filters

In comparison with these results, skin texture appears to have basis as a biometric suitable for face recognition. Further work is necessary to optimise performance with particular respect to classifier configuration along with an evaluation on a larger database (for example, FRGC).

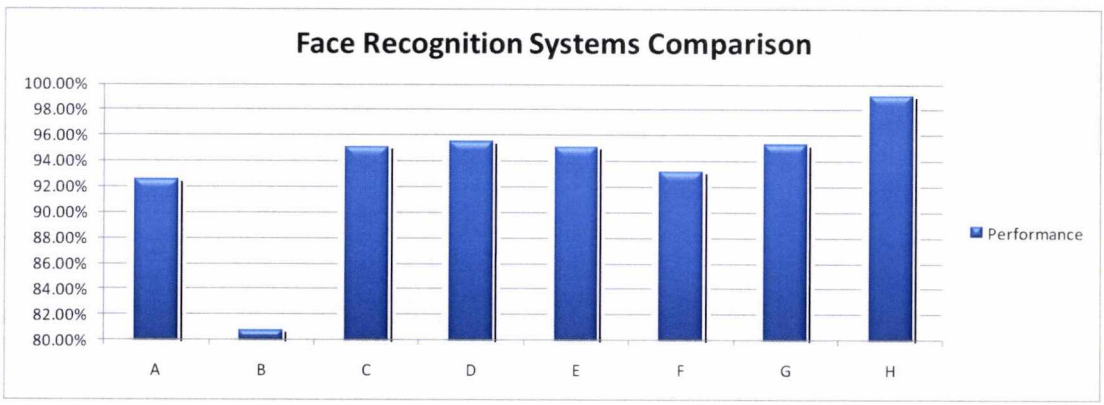


Figure 8.3 - Relative Performance of Presented vs. Other Systems

Figure 8.3 provides a visual representation of the relative performance between the discussed systems following the key in Table 8.3. From this analysis it appears that facial skin texture may be usable for subject identification.

[1] Experimental basis was verification, not identification.
[2] Recognition rate calculated as Genuine Accept Rate at a False Match Rate of 0.0096.

8.4 COMPARISON OF RESULTS AGAINST COMMERCIAL OFFERINGS

The National Institute of Standards and Technology (NIST) conducted what has become the major commercial benchmark biometrics evaluation of the last 5 years. [Phillips, Scruggs et al. 2007]. Known as the Face Recognition Vendor Test (FRVT and previously FERET), evaluations from 1994, 1995, 1996 (FERET), 2000 and 2002 (FRVT) were followed up with a large scale evaluation in 2006 which was extended to include Iris recognition systems.

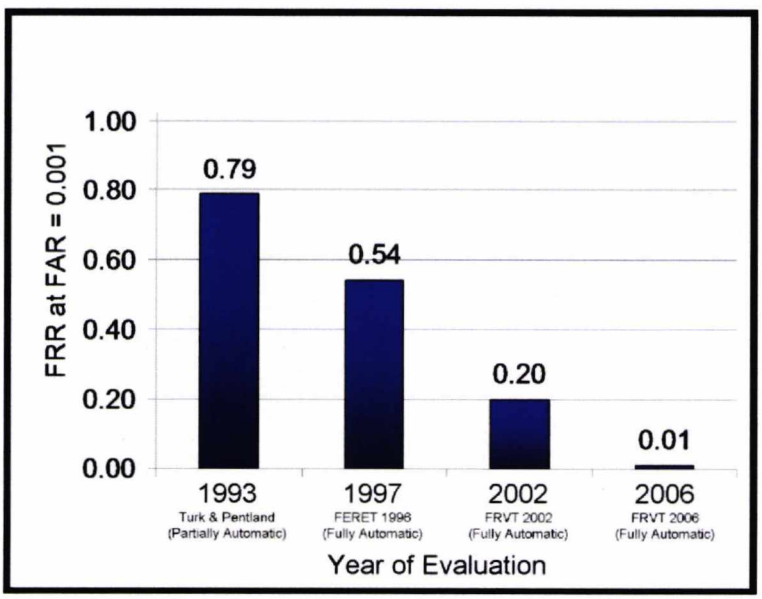


Figure 8.4 - Reduction in Error Rate for State of the Art Biometric Systems (source FRVT 2006)

Figure 8.4 shows the trend in commercial state of the art biometrics system's error rates over the past 17 years. The best commercial offerings described by [Phillips, Scruggs et al. 2007] could reach a False Non-Match Rate/False Reject Rate (FNMR/FRR) under 0.01 at a False Match Rate/False Accept Rate (FMR/FAR) of 0.001 when using very high resolution still images (6 Megapixels).

The evaluation in the present work lends itself to comparison against these results by its threshold independent method of presenting data (see Section 4.6 for an explanation of this). Evaluating the MFMB System 4 described in Section 6.9.2 at an FNMR of 0.009 yields a FMR of 0.003. With the FMR at a value of 0.001, the FNMR is 0.05. These results place its performance the best that was obtained from the 2002 to the 2006 evaluations (Figure 8.5).

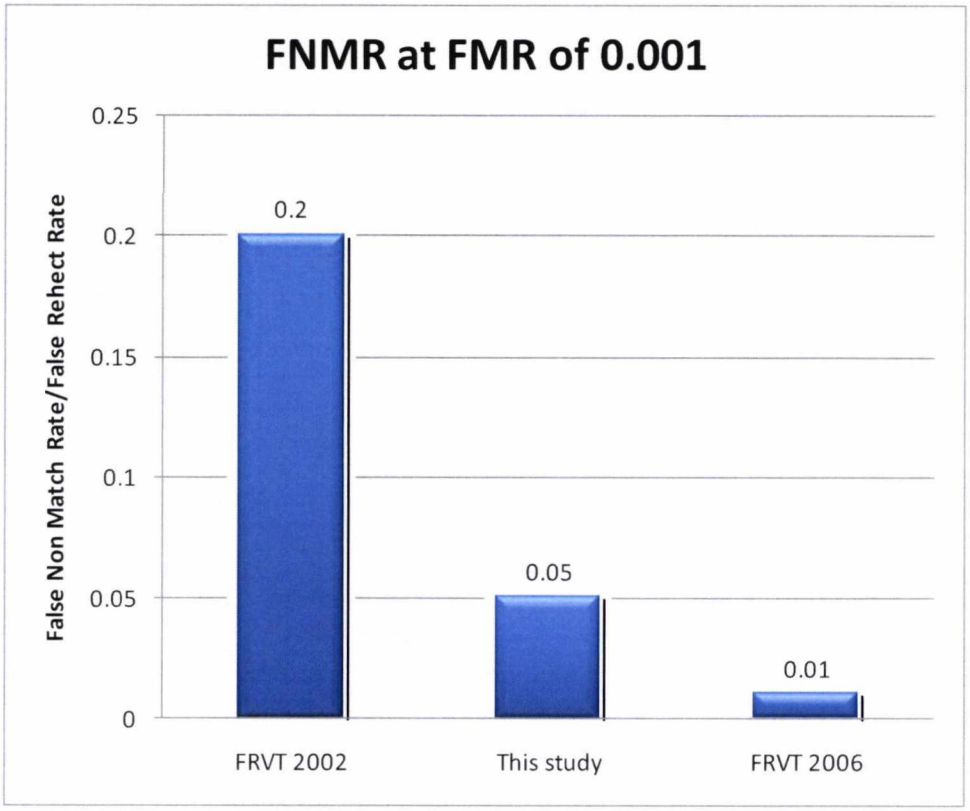


Figure 8.5 - This study (Neural Network) against FRVT 2006

Evaluating against the 30% scaling of the training and probe images described in Section 7.3.1 at an FNMR of 0.009 gives an FMR of 0.0001. Basing the results on an FMR of 0.001 gives an FNMR of 0. At this scaled resolution both of these are excellent results and appear to initially indicate potential performance on par with the FRVT results. (The performance increase at lower resolutions may be attributable to increased texture visibility to GLCP matrices and suboptimal Gabor filter sizes as commented previously).

When viewing results in this manner however, three key considerations must be taken into account. These are:

1. The FRVT evaluation was performed against the FRGC database. This is a much larger database than the BMDB database. (Section 7.5 states that the lowest error rate that may be statistically established with 95% confidence is 0.95% for the BMDB database. This indicates that an FMR value of 0.001 is not statistically significant).

2. The FRVT evaluation was a "fully automatic" evaluation. The study in this paper was supported by manual eye location prior to feature extraction. This process should be fully automated for a fair comparison.

3. The FRVT evaluation attempts to compensate for lighting variations between images. This evaluation makes no such attempt at this stage.

Given these differences, it is important to take a critical view of the above comparisons until the results presented in this work can be repeated using a fully automated process on a larger test database.

8.5 Reflections and Proposals

The material covered in this thesis has been both enlightening and fascinating. From the initial hypothesis that skin texture regions outside of fingerprints, palm prints and iris patterns would be suitable candidates for a biometric recognition system to the investigation of the optimal way to extract such features indicates that there may still remain many novel implementations of biometric recognition systems yet to be discovered.

With the limitations imposed by number of subjects and images in the BMDB database it is hoped that this research will be an indicator to the feasibility of facial skin texture as a biometric in its own right as well as a supporting biometric in multi-biometric systems that can be used to improve overall recognition accuracy.

During the course of this research several other avenues of research have been discussed which involve using skin texture as a biometric for face recognition. While some of these avenues have been subject to some preliminary investigations in this thesis, there are several potential areas of research for which future research may be beneficial. This section explores some of these areas.

8.5.1 Large Scale Evaluation

Sections 7.5 and 8.3 discuss the confidence and competitive performance in the evaluation presented in this thesis. Due to the size and recent release of the BMDB database, it may be difficult to express a high level of confidence that the performance will remain consistent when evaluated against other larger databases. It would be an interesting avenue of research to repeat the evaluation with a much larger dataset where the resultant performance data could be placed head to head with other face recognition techniques using the same databases.

The evidence indicates that image resolution is not as critical to performance for skin texture as initially thought. An ideal candidate for a re-assessment of data would be the FRGC v2 database as presented by [Phillips, Flynn et al. 2005].

8.5.2 GABOR AND GLCP FILTER TUNING

The results in Section 7.3.1 indicate that performance may be increased by adjusting the scale of the source images. This could suggest that the Gabor filter frequencies may not be optimal for the source image size. It would be interesting to evaluate the comparison of increasing base filter sizes for the Gabor and GLCP features against decreasing the source image size. It may also be (given that the BMDB database is stored in JPEG format) that by reducing the image size, the effect of compression artefacts in the source image is lessened.

8.5.3 CLASSIFIER OPTIMISATION

Section 4.5.2 discusses the optimisation of classifiers that took place as part of this research. Recent evidence by [Zhang and Qiao 2010], [Cho and Wong 2008] and other authors indicate that classifier optimisation has the potential to offer significant performance improvements. Indeed, outside of the three classifiers presented in the course of this study it may be that other genres of classifiers provide superior performance when dealing with skin texture regions.

It is the recommendation of the author that the attempts to optimise classification techniques take place with or following a larger scale evaluation in order that the accuracy of the results is statically confident.

8.5.4 CROSS REGION SKIN TEXTURE RECOGNITION

As an extension to the work presented in Section 7.4 it is suggested that a technique similar to Eigenfaces may be used to generalise face textures into a subset of definable patterns. This may provide a more generic method of face recognition that is less dependent on a specific region of skin.

8.5.5 MULTI SPECTRAL PROCESSING

The study performed in this thesis focussed on using greyscale image conversions to facilitate simpler algorithms for the purpose of proof of concept of the initial hypothesis. Several techniques for multi-spectral processing do exist. [Manian and Vasquez 2002] indicates that the HSV colour spectrum is particularly suitable for autonomous processing of skin texture surfaces using Gabor filters. Furthermore [Singh, Vatsa et al. 2008] present a novel means for fusing the image to create a multi-spectral source image and retrieve

composite feature vectors. [Singh, Vatsa et al. 2008] also compare feature and score fusion techniques for face recognition using log-Gabor wavelets and support vector machine based classifiers.

[Wang, Zhou et al. 2006] demonstrate how colour information may be preserved as a feature using Quaternion-Gabor filters for colour segmentation of images with promising results compared to existing algorithms such as JSEG [Deng and Manjunath 2001].

Also of interest is [Jones and Abbott 2004] paper on optimised conversion of colour images to greyscale to preserve the maximum amount of useful information from the source image. This shows potential improvements of 4% to 14% in recognition rates.

Outside of the visual spectrum, evidence from [Chen and Jing 2006] suggests that infrared may provide a promising alternative to visible light in Gabor feature based facial recognition and [Shi and Healey 2003] discuss the use of Gabor filters to extract textural information from hyper-spectral images. However, the use of such systems is limited to specifically designed biometric systems that use infra-red adapted image acquisition equipment.

8.5.6 Multi Biometric Processing

Chapter 6 discussed in depth the ability to both feature and score fuse biometrics from different facial regions as well as biometrics of different types to enhance the overall recognition accuracy of the system. Similar techniques have been employed in [Yao, Jing et al. 2007] and [Rokita, Krzyzak et al. 2008] to use palm print and facial biometrics in a multi-modal identification system.

In addition to the other facial areas mentioned in Chapter 5, the BMDB database includes biometrics from other regions such as finger and palm prints which could lead to an investigation in the applicability of the algorithms in this document to other regions of skin.

8.5.7 Aging

It is commonly held that a person's fingerprints change very little with age. The hypothesis that this may hold true for facial skin texture patterns is an area worthy of further research. The difficulty in substantiating this claim is the lack of databases which provide suitable candidate information and period biometric samples across a prolonged period.

A recent study indicates that as a person ages, the epidermis thickens resulting in more pronounced pore patterns on the face [Sugiyama-Nakagiri, Sugata et al. 2008]. It is

unknown what impact this would have on the features used in this study. It is recommended that an analysis of this be undertaken.

8.6 Note on Matlab Code

As part of the research activities, several Matlab functions have been written that fit within the framework presented. In particular, a function was written to extract genuine accept rates, false match rates and false non match rates directly from a PRTools classifier output dataset. This information is an essential precursor to establishing DET and ROC performance curves that are presented throughout this thesis. As no publicly available function which performs this task is currently included with PRTools, this has been included in Appendix B. All other code used in this evaluation may be made available on request for research purposes.

8.7 Closing Statement

The study was introduced with a discussion about facial regions of skin texture as a biometric. Section 1.6 summarised 5 specific questions that have been addressed in detail in this thesis.

In addition, a framework for presenting and analysing performance data has been presented which facilitates a relative comparison with existing technologies. By following the framework established by [Mansfield and Wayman 2002] the results are displayed in a standards based manner which is also independent of classification acceptance thresholds. This enables a comparison against other techniques that may be evaluated against FAR or FNMR error biases as well as a visual interpretation of how the system performs across a range of threshold values.

It was found that the system performs well when compared with other recently published methods and has a degree of robustness against variations in the probe image data. It was also found that the initial hypothesis of performance being directly related to image resolution was false. This will enable performance testing to be repeated against larger lower resolution datasets that are more widely established and recognised in the community.

It is hoped that the introduction of specific regions of skin texture as a biometric will provoke further analysis on its practical advantages and large scale performance. And that this in turn may increase the efficiency, capability and confidence of biometric systems.

BIBLIOGRAPHY

Abate, A. F., M. Nappi, et al. (2006). Multi-Modal Face Recognition by Means of Augmented Normal Map and Pca. IEEE International Conference on Image Processing (ICIP 2006), Atlanta, GA, 649-652.

Allan M. Haggar, J. P. W., David A. Reimann, David O. Hearshen, Jerry W. Froelich, (1989). "Eigenimage Filtering in Mr Imaging: An Application in the Abnormal Chest Wall." Magnetic Resonance in Medicine **11**(1): 85-97.

Amin, M. A., N. V. Afzulpurkar, et al. (2005). Fuzzy-C-Means-Determines the Principle Component Pairs to Estimate the Degree of Emotion from Facial Expressions. Fuzzy Systems and Knowledge Discovery, Pt 1, Proceedings. **3613:** 484-493.

Andrew Laine, J. F. (1993). An Adaptive Approach for Texture Segmentation by Multi-Channel Wavelet Frames, Center for Computer Vision and Visualization.

Andrew Y. Ng, M. I. J. (2002). "On Discriminative Vs. Generative Classifiers: A Comparison of Logistic Regression and Naive Bayes." NIPS **14**.

Aujol, J. F., G. Aubert, et al. (2003). "Wavelet-Based Level Set Evolution for Classification of Textured Images." IEEE Transactions on Image Processing **12**(12): 1634-1641.

Baraldi, A. and F. Parmiggiani (1995). "An Investigation of the Textural Characteristics Associated with Gray Level Cooccurrence Matrix Statistical Parameters." Geoscience and Remote Sensing, IEEE Transactions on **33**(2): 293-304.

Barber, D. G. and E. F. Ledrew (1991). "Sar Sea Ice Discrimination Using Texture Statistics - a Multivariate Approach." Photogrammetric Engineering and Remote Sensing **57**(April): 385-395.

Bellman, R. and R. Kalaba (1959). "On Adaptive Control Processes." IRE Transactions on Automatic Control **4**(2): 1-9.

Bernoulli, D. (1753). "Sur Le Mélange De Plusieurs Espéces De Vibrations Simples Isochrones, Qui Peuvent Coexister Dans Un Mêmesystème De Corps." Mem. Berlin.

Bertillon, A. (1896). Signaletic Instructions Including the Theory and Practice of Anthropometrical Identification. Chicago, National Library of Medicine.

Bishop, C. (1995). Neural Networks for Pattern Recognition. Oxford, New York, Oxford University Press.

Bodnarova, A., M. Bennamoun, et al. (2002). "Optimal Gabor Filters for Textile Flaw Detection." Pattern Recognition **35**(12): 2973-2991.

Bouchard, G. and B. Triggs (2004). The Trade-Off between Generative and Discriminative Classifiers. Compstat 2004 Symposium. Prague.

Bourgeat, P., F. Meriaudeau, et al. (2006). "Gabor Filtering for Feature Extraction on Complex Images: Application to Defect Detection on Semiconductors." Imaging Science Journal **54**(4): 200-210.

Brodatz, P. (1966). Textures: A Photographic Album for Artists and Designers. New York, Dover Publications.

Cao, W., T. Q. Peng, et al. (2004). A Remote Sensing Image Classification Method Using Color and Texture Feature. Advances in Neural Networks - Isnn 2004, Pt 1. **3173:** 965-970.

Chen, C. M., H. H. S. Lu, et al. (2001). "A Textural Approach Based on Gabor Functions for Texture Edge Detection in Ultrasound Images." Ultrasound in Medicine and Biology **27**(4): 515-534.

Chen, H. Y., C. L. Huang, et al. (2008). "Hybrid-Boost Learning for Multi-Pose Face Detection and Facial Expression Recognition." Pattern Recognition **41**(3): 1173-1185.

Chen, J. D., J. H. Lai, et al. (2004). Gabor-Based Kernel Fisher Discriminant Analysis for Pose Discrimination. Advances in Biometric Person Authentication, Proceedings. **3338:** 153-161.

Chen, X. R. and Z. L. Jing (2006). "Infrared Face Recognition Based on Log-Gabor Wavelets." International Journal of Pattern Recognition and Artificial Intelligence **20**(3): 351-360.

Cho, S. Y. and J. J. Wong (2008). "Human Face Recognition by Adaptive Processing of Tree Structures Representation." Neural Computing & Applications **17**(3): 201-215.

Choi, W. P., S. H. Tse, et al. (2008). "Simplified Gabor Wavelets for Human Face Recognition." Pattern Recognition **41**(3): 1186-1199.

Chow, T. W. S. and M. K. M. Rahman (2006). "Face Matching in Large Database by Self-Organizing Maps." Neural Processing Letters **23**(3): 305-323.

Clausi, D. A. (2001). "Comparison and Fusion of Co-Occurrence, Gabor, and Mrf Texture Features for Classification of Sar Sea Ice Imagery." Atmosphere & Oceans **39**(4): 183-194.

Clausi, D. A. (2002). "An Analysis of Co-Occurrence Texture Statistics as a Function of Grey Level Quantization." Canadian Journal of Remote Sensing **28**(1): 45.

Clausi, D. A. and B. Yue (2004). "Comparing Co-Occurrence Probabilities and Markov Random Fields for Texture Analysis of Sar Sea Ice Imagery." IEEE Transactions on Geoscience and Remote Sensing **42**(1): 215-228.

Cortes, C. and V. Vapnik (1995). "Support-Vector Networks." Machine Learning **20**(3): 273-297.

Deng, Y. and B. S. Manjunath (2001). "Unsupervised Segmentation of Color-Texture Regions in Images and Video." Pattern Analysis and Machine Intelligence, IEEE Transactions on **23**(8): 800-810.

Denyer, P. B. (1989). Patent: Skin-Pattern Recognition Method and Device. United States, The Quantum Fund Limited (Edinburgh).

Donato, G., M. S. Bartlett, et al. (1999). "Classifying Facial Actions." Ieee Transactions on Pattern Analysis and Machine Intelligence **21**(10): 974-989.

Duc, B., S. Fischer, et al. (1999). "Face Authentication with Gabor Information on Deformable Graphs." IEEE Transactions on Image Processing **8**(4): 504-516.

Duda, R. O., P. E. Hart, et al. (2001). Pattern Classification. New York, John Wiley and Sons.

Eraldo Ribeiro, E. R. H. (1998). 3-D Planar Orientation from Texture: Estimating Vanishing Point from Local Spectral Analysis. York, Department of Computer Science, University of York.

Ersi, E. F. and J. S. Zelek (2005). Rotation-Invariant Facial Feature Detection Using Gabor Wavelet and Entropy. Image Analysis and Recognition. **3656:** 1040-1047.

Escofet, J., R. Navarro, et al. (1998). "Detection of Local Defects in Textile Webs Using Gabor Filters." Optical Engineering **37**(8): 2297-2307.

Escofet, J., R. Navarro, et al. (1998). "Detection of Local Defects in Textile Webs Using Gabor Filters." Optical Engineering **37**(8): 2297-2307.

Euler, L. (1749). "On Vibrations of Excited Chords." Nova acta eruditorum **512-527**.

Euler, L. (1750). "On the Propagation of Pulses in an Elastic Medium." Novi Commentarii academiae scientiarum Petropolitanae **1**: 67-105.

Ferrari, R. J., R. M. Rangayyan, et al. (2001). "Analysis of Asymmetry in Mammograms Via Directional Filtering with Gabor Wavelets." Ieee Transactions on Medical Imaging **20**(9): 953-964.

Fourier, J. (1955). The Analytical Theory of Heat. New York, Dover.

G.S.Cox, F. J. H., G. de Jagger (1992). Experiments in Lung Cancer Nodule Detection Using Texture Analysis and Neural Network Classifiers. South Africa, Dept of Electrical Engineering, University of Capetown.

Gabor, D. (1946). "Theory of Communication." **November 93**(26 pp 429-457).

Gabor, D. (1948). "Holography." Nature **161**: 777-778.

Garcia, C., G. Zikos, et al. (1999). Face Detection in Color Images Using Wavelet Packet Analysis. 6th International Conference on Multimedia Computing and Systems (IEEE ICMCS 99), Florence, Italy, 703-708.

Gneushev, A. N. (2007). "Construction and Optimization of a Texture-Geometric Model of a Face Image in the Space of Basic Gabor Functions." Journal of Computer and Systems Sciences International **46**(3): 418-428.

Gomez, G. (2002). "On Selecting Colour Components for Skin Detection." 16th International Conference on Pattern Recognition **2**: 20961.

Gonzalez-Jimenez, D. and J. L. Alba-Castro (2007). "Shape-Driven Gabor Jets for Face Description and Authentication." Ieee Transactions on Information Forensics and Security **2**(4): 769-780.

Haley, G. M. and B. S. Manjunath (1999). "Rotation-Invariant Texture Classification Using a Complete Space-Frequency Model." IEEE Transactions on Image Processing **8**(2): 255-269.

Hanazawa, A. (2004). "Coding of Texture and Shading in Monkey Areav4." International Congress Series **1269**: 89-92.

Hanley, J. A. and B. J. McNeil (1982). "The Meaning and Use of the Area under a Receiver Operating Characteristic (Roc) Curve." Radiology **143**(1): 29-36.

Hanley, J. A. and B. J. McNeil (1983). "A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases." Radiology **148**(3): 839-843.

Hawker, S. and C. Soanes (2005). The Compact Oxford English Dictionary, 3rd Edition. Oxford, Oxford Press, ISBN: 0-19-861022-X.

Hellstrom, B. J. and L. N. Kanal (1989). The Introduction of Hidden Units to Optimizing Neural Networks. TENCON '89. Fourth IEEE Region 10 International Conference, 127-131.

Hiremath, P. S. and A. Danti (2006). Combining Geometric and Gabor Features for Face Recognition. Computer Vision - Accv 2006, Pt I. **3851:** 140-149.

Idrissa, M. and M. Acheroy (2002). "Texture Classification Using Gabor Filters." Pattern Recognition Letters **23**(9): 1095-1102.

Imaoka, H. and K. Okajima (2004). "Algorithm for the Detection of Faces on the Basis of Gabor Features and Information Maximization." Neural Computation **16**(6): 1163-1191.

Jeon, I., K. Kwon, et al. (2004). Optimal Gabor Encoding Scheme for Face Recognition Using Genetic Algorithm. Knowledge-Based Intelligent Information and Engineering Systems, Pt 2, Proceedings. **3214:** 227-236.

Jobanputra, R. and D. A. Clausi (2006). "Preserving Boundaries for Image Texture Segmentation Using Grey Level Co-Occurring Probabilities." Pattern Recogn. **39**(2): 234-245.

Jones, C. F. and A. L. Abbott (2004). "Optimization of Color Conversion for Face Recognition." Eurasip Journal on Applied Signal Processing **2004**(4): 522-529.

Kaplan, L. M. (1999). "Extended Fractal Analysis for Texture Classification and Segmentation." IEEE Transactions on Image Processing **8**(11): 1572-1585.

Keys, R. (1981). "Cubic Convolution Interpolation for Digital Image Processing." IEEE Transactions on Acoustics, Speech and Signal Processing **29**(6): 1153-1160.

Khoo, H.-K., H.-C. Ong, et al. (2008). Image Texture Classification Using Combined Grey Level Co-Occurrence Probabilities and Support Vector Machines. Fifth International Conference on Computer Graphics, Imaging and Visualisation**:** 180-184.

Kim, D. S., I. Jeon, et al. (2006). "Embedded Face Recognition Based on Fast Genetic Algorithm for Intelligent Digital Photography." Ieee Transactions on Consumer Electronics **52**(3): 726-734.

Kisner, T., A. Essoh, et al. (2007). Visualisation of Network Traffic Using Dynamic Co-Occurrence Matrices. Second International Conference on Internet Monitoring and Protection.

Kotropoulos, C., A. Tefas, et al. (2000). "Frontal Face Authentication Using Morphological Elastic Graph Matching." IEEE Transactions on Image Processing **9**(4): 555-560.

Krishna, S., J. Black, et al. (2006). Using Genetic Algorithms to Find Person-Specific Gabor Feature Detectors for Face Indexing and Recognition. Advances in Biometrics, Proceedings. **3832:** 182-191.

Kyrki, V., J. K. Kamarainen, et al. (2004). "Simple Gabor Feature Space for Invariant Object Recognition." Pattern Recognition Letters **25**(3): 311-318.

Lai, J. H. and P. C. Yuen (2006). "Face and Eye Detection from Head and Shoulder Image on Mobile Devices." International Journal of Pattern Recognition and Artificial Intelligence **20**(7): 1053-1075.

Lam, K. W. L. H. N. (1995). "Optimal Sizing of Feedforward Neural Networks: Case Studies." Artificial Neural Networks and Expert Systems, New Zealand Conference **0**.

Latif-Amet, A., A. Ertuzun, et al. (2000). "An Efficient Method for Texture Defect Detection: Sub-Band Domain Co-Occurrence Matrices." Image and Vision Computing **18**(6-7): 543-553.

Lee, J. S., Y. M. Kuo, et al. (2006). The Adult Image Identification Based on Online Sampling. IEEE International Joint Conference on Neural Network, Vancouver, CANADA, 2566-2571.

Li, K. P. and J. E. Porter (1988). Normalizations and Selection of Speech Segments for Speaker Recognition Scoring. Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on, 595-598 vol.591.

Li, Y. F., Z. Y. Ou, et al. (2005). Face Recognition Using Gabor Features and Support Vector Machines. Advances in Natural Computation, Pt 2, Proceedings. **3611:** 119-122.

Lian, H. C. and B. L. Lu (2007). "Multi-View Gender Classification Using Multi-Resolution Local Binary Patterns and Support Vector Machines." International Journal of Neural Systems **17**(6): 479-487.

Liu, C. J. (2006). "Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance." Ieee Transactions on Pattern Analysis and Machine Intelligence **28**(5): 725-737.

Liu, D. H., K. M. Lam, et al. (2004). "Optimal Sampling of Gabor Features for Face Recognition." Pattern Recognition Letters **25**(2): 267-276.

Maio, D. and L. Nanni (2005). "An Efficient Fingerprint Verification System Using Integrated Gabor Filters and Parzen Window Classifier." Neurocomputing **68**: 208-216.

Mak, K. L. and P. Peng (2008). "An Automated Inspection System for Textile Fabrics Based on Gabor Filters." Robotics and Computer-Integrated Manufacturing **24**(3): 359-369.

Manian, V. and R. Vasquez (2002). "Approaches to Color- and Texture-Based Image Classification." Optical Engineering **41**(7): 1480-1490.

Mansfield, A. J. and J. L. Wayman (2002). Best Practices in Testing and Reporting Performance of Biometric Devices**:** 32.

Martin, A., G. Doddington, et al. (1997). "The Det Curve in Assessment of Detection Task Performance." In Proceedings EuroSpeech **4**: 1895--1898.

Massot, C. and J. Herault (2008). "Model of Frequency Analysis in the Visual Cortex and the Shape from Texture Problem." International Journal of Computer Vision **76**(2): 165-182.

McKean, H. D. a. H. P. (1972). Fourier Series and Integrals. New York, Academic Press.

Meade, M., S. C. Sivakumar, et al. (2005). "Comparative Performance of Principal Component Analysis, Gabor Wavelets and Discrete Wavelet Transforms for Face Recognition." Canadian Journal of Electrical and Computer Engineering-Revue Canadienne De Genie Electrique Et Informatique **30**(2): 93-102.

Metz, C. (1978). "Basic Principles of Roc Analysis." Seminars on Nuclear Medicine **8**(4): 283-298.

Meyer, D., F. Leisch, et al. (2003). "The Support Vector Machine under Test." Neurocomputing **55**(1-2): 169-186.

Minksy, M. (1961). "Steps Towards Artificial Intelligence." Proceedings of the Institute of Radio Engineers (IRE) **49**(1): 8-30.

Movellan, J. R. (2005). Tutorial on Gabor Filters. San Diego, University of California.

Munir, H. M. (2008). Security Thru Biometrics & Other Cutting Edge Technologies. International Symposium on Biometrics & Security Technologies (ISBAST), Islamabad, IEEE.

Nanni, L. and D. Maio (2007). "Weighted Sub-Gabor for Face Recognition." Pattern Recognition Letters **28**(4): 487-492.

Nixon, K. A. and R. K. Rowe (2005). Multispectral Fingerprint Imaging for Spoof Detection. Conference on Biometric Technology for Human Identification II, Orlando, FL, 214-225.

Ortega-Garcia, J., J. Fierrez, et al. (2009). "The Multi-Scenario Multi-Environment Biosecure Multimodal Database (Bmdb)." IEEE Transactions on Pattern Analysis and Machine Intelligence.

Pan, S., G. Shavit, et al. (2006). "Automated Classification of Protein Crystallization Images Using Support Vector Machines with Scale-Invariant Texture and Gabor Features." Acta Crystallographica Section D-Biological Crystallography **62**: 271-279.

Phillips, P. J., P. J. Flynn, et al. (2005). Overview of the Face Recognition Grand Challenge. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, IEEE Computer Society: 947-954.

Phillips, P. J., W. T. Scruggs, et al. (2007). Frvt 2006 and Ice 2006 Large-Scale Results. Gaithersburg, MD, National Institute of Standards and Technology (NIST).

Polyak, B. T. (1964). "Some Methods of Speeding up the Convergence of Iterative Methods." USSR Computational Mathematics and Mathematical Physics **4**: 1-17.

R.M. Haralick, K. S., I. Dinstein (1973). "Textural Features for Image Classification." **SMC-3**.

R.M. Haralick, K. S., I. Dinstein (November 1973). "Textural Features for Image Classification." **SMC-3**.

R.R. Coifman, M. V. W. (1992). "Entropy-Based Algorithms for Best Basis Selection." **38**: 713-718.

Randen, T. and J. H. Husoy (1999). "Filtering for Texture Classification: A Comparative Study." Ieee Transactions on Pattern Analysis and Machine Intelligence **21**(4): 291-310.

Rangayyan, R. M. and F. J. Ayres (2006). "Gabor Filters and Phase Portraits for the Detection of Architectural Distortion in Mammograms." Medical & Biological Engineering & Computing **44**(10): 883-894.

Rokita, J., A. Krzyzak, et al. (2008). "Multimodal Biometrics by Face and Hand Images Taken by a Cell Phone Camera." International Journal of Pattern Recognition and Artificial Intelligence **22**(3): 411-429.

Rumelhart, D. E., G. E. Hinton, et al. (1986). "Learning Internal Representations by Error Propagation." Parallel Distributed Processing: Explorations in the Microstructure of Cognition **1**: 318-362.

Santoro, M., R. Prevete, et al. (2006). Mass Detection in Mammograms Using Gabor Filters and Fuzzy Clustering. Fuzzy Logic and Applications. **3849:** 334-343.

Shen, L., L. Bai, et al. (2007). "Gabor Wavelets and General Discriminant Analysis for Face Identification and Verification." Image and Vision Computing **25**(5): 553-563.

Shi, M. H. and G. Healey (2003). "Hyperspectral Texture Recognition Using a Multiscale Opponent Representation." Ieee Transactions on Geoscience and Remote Sensing **41**(5): 1090-1095.

Shih, F. Y. and C. F. Chuang (2004). "Automatic Extraction of Head and Face Boundaries and Facial Features." Information Sciences **158**: 117-130.

Singh, R., M. Vatsa, et al. (2008). "Hierarchical Fusion of Multi-Spectral Face Images for Improved Recognition Performance." Information Fusion **9**(2): 200-210.

Singh, R., M. Vatsa, et al. (2008). "Integrated Multilevel Image Fusion and Match Score Fusion of Visible and Infrared Face Images for Robust Face Recognition." Pattern Recognition **41**(3): 880-893.

Soh, L.-K. and C. Tsatsoulis (1999). "Texture Analysis of Sar Sea Ice Imagery Using Gray Level Co-Occurrence Matrices." IEEE Transactions on Geoscience and Remote Sensing **37**(2): 780-795.

Sugiyama-Nakagiri, Y., K. Sugata, et al. (2008). "Age-Related Changes in the Epidermal Architecture around Facial Pores." Journal of Dermatological Science **50**(2): 151-154.

Tang, L., L. Tian, et al. (2003). "Classification of Broadleaf and Grass Weeds Using Gabor Wavelets and an Artificial Neural Network." Transactions of the Asae **46**(4): 1247-1254.

Tarini, M., H. Yamauchi, et al. (2002). Texturing Faces. Graphics Interface 2002 Conference, Calgary, Canada, 89-98.

Turk, M. and A. Pentland (1991). "Eigenfaces for Recognition." J. Cognitive Neuroscience **3**(1): 71-86.

Unser, M. (1993). Texture Discrimination Using Wavelets. Bethesda, Biomedical Engineering and Instrumentation Program.

US-DHS (2009) "United States Is Collecting 10 Fingerprints from International Travelers at Major U.S. Ports of Entry." US Visit **March 2009**.

Wang, L., Y. P. Li, et al. (2006). A Novel 2d Gabor Wavelets Window Method for Face Recognition. Multimedia Content Representation, Classification and Security. **4105:** 497-504.

Wang, X. H., Y. Zhou, et al. (2006). Color Texture Segmentation Based on Quaternion-Gabor Features. Progress in Pattern Recognition, Image Analysis and Applications, Proceedings. **4225:** 345-353.

Wang, Y. J., C. S. Chua, et al. (2001). Face Recognition from 2d and 3d Images. Audio- and Video-Based Biometric Person Authentication, Proceedings. **2091:** 26-31.

Wang, Y. J., C. S. Chua, et al. (2002). "Facial Feature Detection and Face Recognition from 2d and 3d Images." Pattern Recognition Letters **23**(10): 1191-1202.

Wen L. Hwang, C.-S. L., Pau-Choo Chung (1996). Shape from Texture: Direct Estimation of Planar Surface Orientations Using Continuous Wavelet Transform. Taipei, Taiwan, Institute of Informational Science, Academia Sinicia.

Wen L. Hwang, C.-S. L., Pau-Choo Chung (Circa 1996). Shape from Texture: Direct Estimation of Planar Surface Orientations Using Continuous Wavelet Transform. Taipei, Taiwan, Institute of Informational Science, Academia Sinicia.

Wickerhauser, M. V. (1992). Acuoustic Signal Compression with Wavelet Packets. Wavelets: A Tutorial in Theory and Applications, Accademic Press.

Xie, J., Y. F. Jiang, et al. (2005). "Segmentation of Kidney from Ultrasound Images Based on Texture and Shape Priors." Ieee Transactions on Medical Imaging **24**(1): 45-57.

Yao, Y. F., X. Y. Jing, et al. (2007). "Face and Palmprint Feature Level Fusion for Single Sample Biometrics Recognition." Neurocomputing **70**(7-9): 1582-1586.

Yu, L., Z. He, et al. (2010). "Gabor Texture Representation Method for Face Recognition Using the Gamma and Generalized Gaussian Models." Image and Vision Computing **28**(1): 177-187.

Yu, W. W., X. L. Teng, et al. (2006). "Face Recognition Fusing Global and Local Features." Journal of Electronic Imaging **15**(1).

Zhang, B. C. and Y. Qiao (2010). "Face Recognition Based on Gradient Gabor Feature and Efficient Kernel Fisher Analysis." Neural Computing & Applications **19**(4): 617-623.

Zhang, B. H., S. G. Shan, et al. (2007). "Histogram of Gabor Phase Patterns (Hgpp): A Novel Object Representation Approach for Face Recognition." IEEE Transactions on Image Processing **16**(1): 57-68.

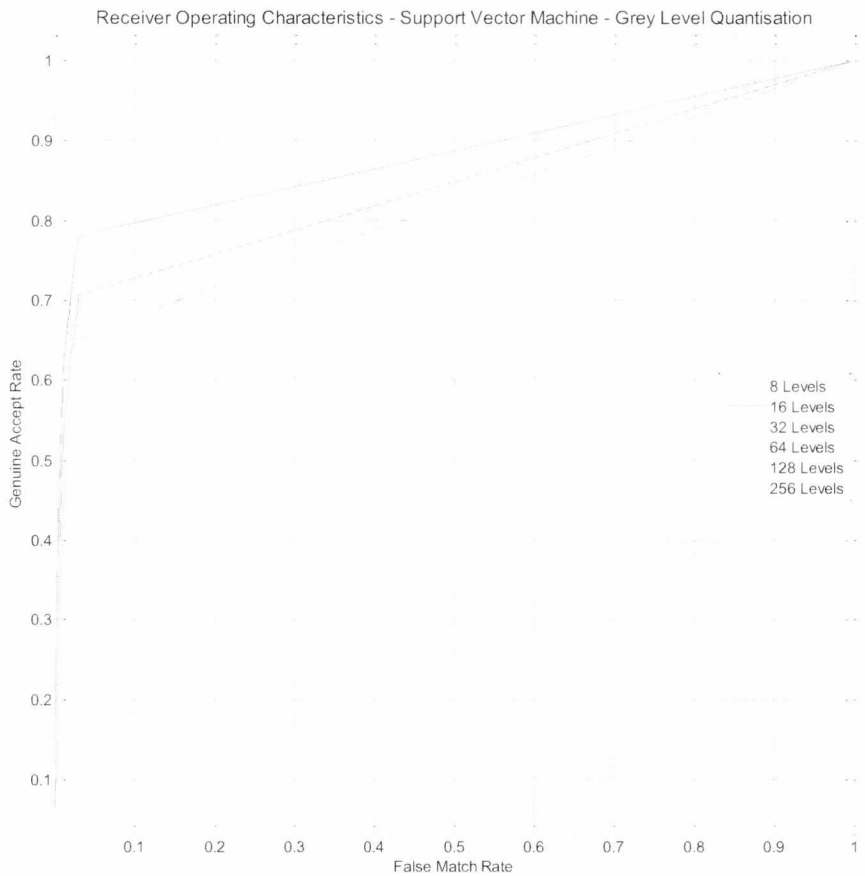
Zhu, C. and X. Yang (1998). "Study of Remote Sensing Image Texture Analysis and Classification Using Wavelet." International Journal of Remote Sensing **19**(16): 3197-3203.

Zhu, S. C., C. E. Guo, et al. (2005). "What Are Textons?" International Journal of Computer Vision **62**(1-2): 121-143.

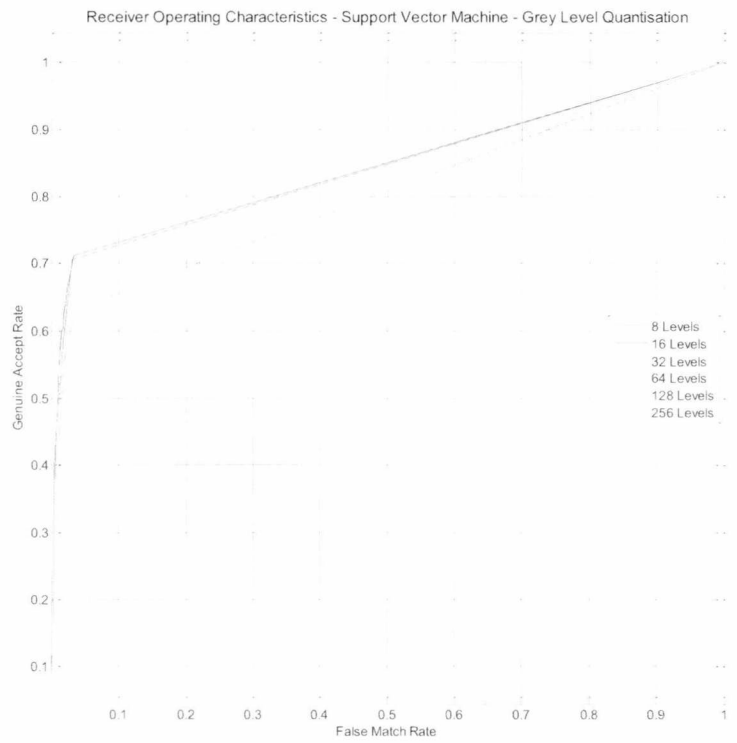
APPENDIX A

GLCP Grey Level Optimisation ROC Curves

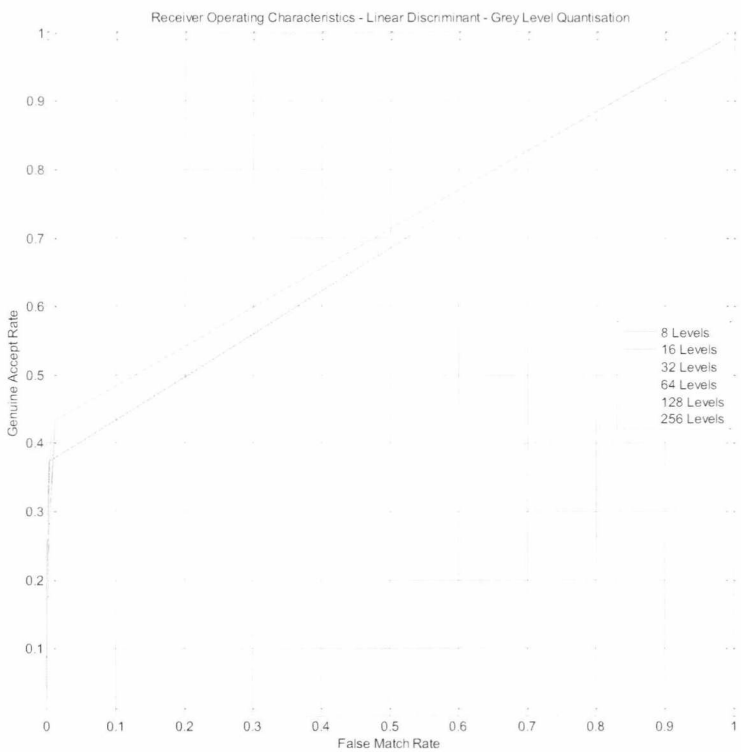
The following figures are in support of Section 6.4.3 page 107 and include ROC curves for the support vector machine and linear discriminant classifiers obtained during the GLCP quantisation process.



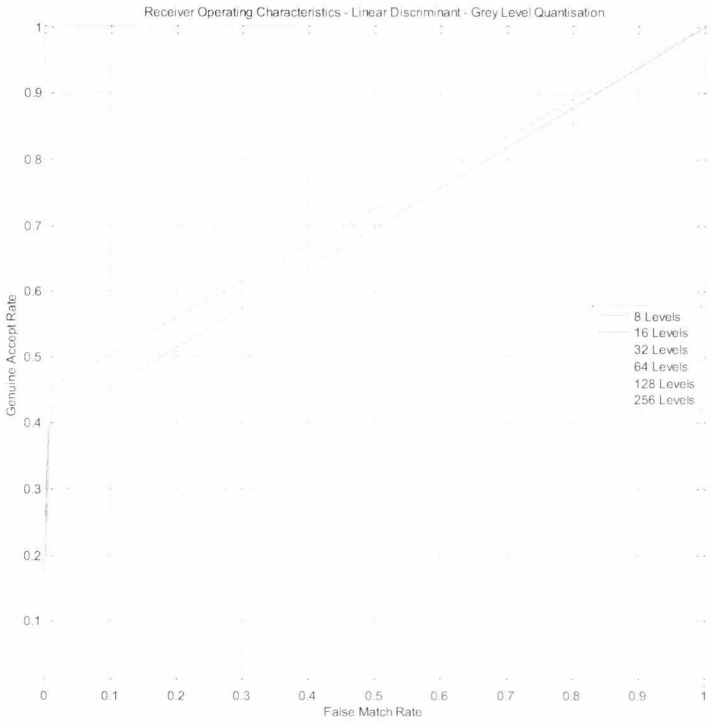
ROC Curve for SVM (Region 2A - Left) at Varying GLCP Quantisation Levels



ROC Curve for SVM (Region 2B - Right) at Varying GLCP Quantisation Levels



ROC Curve for Linear Classifier (Region 2A - Left) at Varying GLCP Quantisation Levels



ROC Curve for Linear Classifier (Region 2B - Right) at Varying GLCP Quantisation Levels

APPENDIX B

MATLAB Code for GAR/FAR/FNMR Generation from PRTools DataSet Output

After test classification, PRTools outputs a dataset type object with a .data field corresponding to a classification score matrix. This code takes the PRTools dataset and outputs Genuine Accept, False Accept and False Non Match Rates. These can then be used in the plotting of Receiver Operating Characteristic and Detection Error Trade-off curves. This code is free to use and integrate into other works of research as long as the author is referenced as the source.

```
function [GAR FNMR FMR]=prepareROC(classifier_output, number_of_samples,
samples_per_class)

%Function to extract and plot Genuine Accept Rate, False Nonmatch and False
%Match rates.  (C) 2009 Dale Rowe.

%Extract classifier output matrix from PRTools dataset (remove .data suffix
%if not required).
classifier_data = classifier_output;

%Normalise data.
Normalised_Value = max(max(classifier_data));

%Calculate increment value for threshold, set initial threshold to 0 and
%initialise and allocate variables prior to loop.

threshold_increment = Normalised_Value/number_of_samples;
threshold = 0;
FMR = zeros(number_of_samples,1);
FNMR = zeros(number_of_samples,1);
GAR = zeros(number_of_samples,1);
index = 1;

truepos = 0;
falsepos = 0;
falseneg = 0;
i=0;

while (i<number_of_samples)
    i=i+1;
    truepos = 0;        %Reset variables
    falsepos = 0;
    falseneg = 0;
    for n=1:size(classifier_data,2)      %For each CLASS
        for m=1:size(classifier_data,1)      %For each OBJECT
            if (n==floor((m-1)/samples_per_class)+1) %if should be true positive
                if (classifier_data(m,n) >= threshold)
                    truepos = truepos +1;    %Increment True Positive found
                else
                    falseneg = falseneg + 1;  %Else it SHOULD be positive and is
hence false negative
                end
            else
                if (classifier_data(m,n) >= threshold)  %If not in correct class,
mark
                    falsepos = falsepos + 1;     %as false positive.
                end
            end
        end
    end
```

```
    %Calculate FMR, FNMR and GAR

    FMR(index) = falsepos/(size(classifier_data,1)*(size(classifier_data,2)-1));
    FNMR(index) = falseneg/(size(classifier_data,1));
    GAR(index) = 1-FNMR(index);
    index=index+1;
    threshold = threshold + threshold_increment;
end
```