FEATURE EXTRACTION AND INFORMATION FUSION IN FACE AND PALMPRINT MULTIMODAL BIOMETRICS

A thesis submitted to the Newcastle University for the degree of Doctor of Philosophy

In the Faculty of Science, Agriculture and Engineering

Muhammad Imran Ahmad

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

June 2013

NEWCASTLE UNIVERSITY

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

I, Muhammad Imran Ahmad, confirm that this thesis and work presented in it are my own achievement.

I have read and understand the penalties associated with plagiarism.

Signed:

Date: 17/06/2013

ABSTRACT

Multimodal biometric systems that integrate the biometric traits from several modalities are able to overcome the limitations of single modal biometrics. Fusing the information at an earlier level by consolidating the features given by different traits can give a better result due to the richness of information at this stage. In this thesis, three novel methods are derived and implemented on face and palmprint modalities, taking advantage of the multimodal biometric fusion at feature level. The benefits of the proposed method are the enhanced capabilities in discriminating information in the fused features and capturing all of the information required to improve the classification performance. Multimodal biometric proposed here consists of several stages such as feature extraction, fusion, recognition and classification.

Feature extraction gathers all important information from the raw images. A new local feature extraction method has been designed to extract information from the face and palmprint images in the form of sub block windows. Multiresolution analysis using Gabor transform and DCT is computed for each sub block window to produce compact local features for the face and palmprint images. Multiresolution Gabor analysis captures important information in the texture of the images while DCT represents the information in different frequency components. Important features with high discrimination power are then preserved by selecting several low frequency coefficients in order to estimate the model parameters.

The local features extracted are fused in a new matrix interleaved method. The new fused feature vector is higher in dimensionality compared to the original feature vectors from both modalities, thus it carries high discriminating power and contains rich statistical information. The fused feature vector also has larger data points in the feature space which is advantageous for the training process using statistical methods. The underlying statistical information in the fused feature vectors is captured using GMM where several numbers of modal parameters are estimated from the distribution of fused feature vector.

Maximum likelihood score is used to measure a degree of certainty to perform recognition while maximum likelihood score normalization is used for classification process. The use of likelihood score normalization is found to be able to suppress an imposter likelihood score when the background model parameters are estimated from a pool of users which include statistical information of an imposter. The present method achieved the highest recognition accuracy 97% and 99.7% when tested using FERET-PolyU dataset and ORL-PolyU dataset respectively.

.

Acknowledgement

Alhamdulillah. Thanks to Allah SWT, whom with His willing gives me the opportunity to complete this thesis. This thesis is a collection of not only hard work, perseverance and continuous efforts in the past four years, but also encouragement, cooperation and support from many people. I would like to take this opportunity to acknowledge these people. First and foremost, I would like to express my utmost gratitude to my supervisor Professor Satnam Dlay and Dr. Wai Lok Woo for giving me the opportunity to pursue a PhD at Newcastle University. My deepest appreciation for my supervisors who nurtures me to become an independent researcher, trained me to critically analyse scientific issue and helped me to understand the concepts of image and signal processing. I am fortunate to have them as my supervisors who are never tired in giving me invaluable support, guidance and encouragement. I would also like to extend my gratefulness to my research colleagues for constructive comments and ideas coming up from our discussions. I am also indebted to my employer and sponsor, Universiti Malaysia Perlis and Ministry of Higher Education Malaysia for their financial support throughout the study. Last but not least, my deepest thankfulness goes to my parent, wife, Naimah and my kids, Ilya and Najwan for their endless love, understanding, sacrifice, care and support. Thank you very much.

Abbreviations

ATM	Automatic Telling Machines
CBM	Cohort Background Model
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DLDA	Direct Linear Discrimination Analysis
EER	Equal Error Rate
EM	Expectation Maximization
FAR	False Acceptance Rate
FFT	Fast Fourier Transform
FLD	Fisher Linear Discrimination
FRR	False Rejection Rate
GAR	Genuine Acceptance Rate
GMM	Gaussian Mixture Model
HMM	Hidden Marcov Model
ICA	Independent Component Analysis
IFFT	Inverse Fast Fourier Transform
KDCV	Kernel discrimination common vector
KDDA	Kernel Direct Discriminant Analysis
KL	Karhunen Loeve
LBP	Local Binary Pattern
LDA	Linear Discrimination Analysis
ML	Maximum Likelihood
PCA	Principle Components Analysis
PDF	Probability Density Function
PSO	Particle Swarm Optimization
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SVM	Support Vector Machine
UBM	Universal Background Model

Table of Contents

3. FEATURE EXTRACTION AND COMPACT REPRESENTATION

4. FEATURE FUSION AND PARAMETER ESTIMATION

5. LIKELIHOOD SCORE NORMALIZATION

List of Publications

M. I. Ahmad, W. L. Woo, S. S. Dlay, “Multimodal biometric fusion at feature level: Face and palmprint,” *Proc. of 7th IEEE International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP) , pp.* 801-804, 2009.

M. I. Ahmad, S. S. Dlay, W. L. Woo, “Feature extraction in face and palmprint multimodal biometrics,” *Submitted to IEEE Transaction on systems, man, and cybernetics.*

M. I. Ahmad, S.S. Dlay, W.L. Woo, “Feature fusion by preserving non-stationary information in face and palmprint multimodal biometrics,” *Submitted to IEEE Transaction on image processing.*

List of Tables

List of Figures

Chapter 1

1.1. Introduction

The verification of the identities of individuals is becoming an increasingly important requirement in a variety of applications, especially involving automatic access control. Examples of such applications are telebanking, the control of physical access, and automatic telling machines (ATMs). Traditional approaches make use of passwords, personal cards, PIN numbers and keys to achieve verification. However, security can easily be breached in these systems when a card or key is lost or stolen or when a password is compromised. Furthermore, difficult passwords may be hard for a legitimate user to remember and simple passwords are easier for an imposter to guess. The use of biometrics offers an alternative means of identification which helps avoid the problems associated with conventional methods. A biometric identification system is defined as the recognition of an individual by using information about certain physical characteristics or personal traits held in a database. Recognition could be achieved by the measurement of features in any of three categories of intrinsic; extrinsic; and hybrid biometric. Intrinsic biometric identify the individual's generic make-up for example from fingerprints or iris patterns. Extrinsic biometrics involves the individual's learned behaviour, such as signatures and keystrokes. Finally, hybrid biometric is based on a combination of the individual's physical characteristics and personal traits such as characteristic of the voice. This leads to the question of what biological measurements can be interpreted as biometrics. In fact any human physiological or behavioural characteristic can be used as a biometric characteristic as long as it satisfies the following requirements [1, 2]:

- Universality – every person will have their own characteristics.
- Distinctiveness – any two persons should be sufficiently different in terms of the selected biometric identifier.
- Permanence – the characteristic should be sufficiently invariant over a period of time.
- Collectability – the characteristic can be measured quantitatively.

However, in a practical biometrics system which employs biometrics for personal recognition, there are a number of other issues that should be considered [1]:

- Performance – the accuracy and speed of achievable recognition in terms of the resources required, as well as the operational and environmental factors which affect accuracy and speed.
- Acceptability – the degree to which people are willing to accept the use of a particular form of biometric identification in daily life.
- Circumvention – how easily the system can be fooled using fraudulent methods.

A practical biometrics system should have specified levels of recognition accuracy, speed, and resource requirements, be harmless to all users, be accepted by the intended population, and be sufficiently robust to resist various fraudulent methods and attacks on the system. The success of a biometric system relies on how the relevant information is captured, the learning strategy used, and the extent to which it is robust to input data variation.

1.2. Biometric Systems

A biometrics system is basically a pattern classification process that operates by obtaining biometric data from an individual, extracting a feature set from the data acquired and comparing this feature set against the template set in the database. This system consists of four main modules, as shown in Figure 1.1:

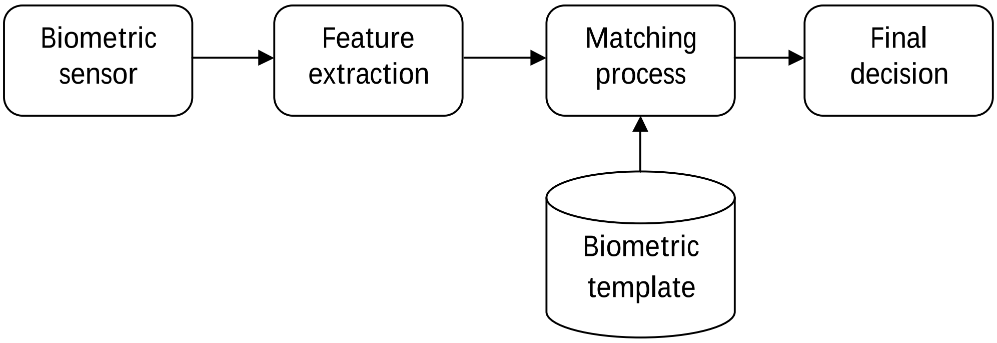


Figure 1.1: Basic task of a biometrics system

The sensor module captures biometric data from an individual. An example is a palmprint sensor that images the ridge and wrinkle structure of a user's palm. The feature extraction module then processes the biometric data acquired in order to extract a set of salient or discriminatory features. For example, the position and orientation of minutiae points in a palmprint image are extracted in the feature extraction module of a palmprint-based biometric system. The matching module subsequently compared the features extracted during recognition against the stored templates and generates matching scores. For example, in the matching module of a fingerprint-based biometrics system, the number of matching minutiae between the input and the template fingerprint images is determined and a matching score is reported. The matcher module also incorporates a decision making module, in which a user's claimed identity is confirmed (verification) or established (identification) based on the matching score. A system database module is used by the biometrics system to store the biometric templates of enrolled users. The enrolment module is responsible for enrolling individual biometrics information into the biometrics systems database. During enrolment, the biometric characteristic of an individual are first scanned by a reader to produce a digital representation of the characteristic. The capture of data during the enrolment process may or may not be supervised by a human, depending on the application. A quality check is generally performed to ensure that the sample acquired can be reliably processed in successive stages. In order to facilitate matching, the input digital representation is further processed by a feature extractor to generate a compact but expressive representation called a template. Depending on the application, the templates may be stored in the central database of the biometrics system or be recorded on a smart card issued to the individual. Usually, multiple templates of an individual are stored to account for variations observed in biometric traits, and the templates in the database may be updated over time.

1.3. Verification versus Identification

Biometric recognition systems generally consist of three different modes of enrolment, identification and verification, as shown in Figure 1.2 [1]. In the enrolment mode two general processes occur. The first is the acquisition of user biometric data, where biometric characteristic of an individual is first scanned by a biometric reader to produce a raw digital representation of the characteristic. The raw image is further processed by a feature extraction method to generate a compact representation containing rich information called a template. The second process concerns the storage of the biometric data for each user in a reference database. This can be in a variety of forms including a template or a statistical model generated using the raw data. Whichever method is used, the stored data is labelled according to the user identity in order to facilitate subsequent authentication. The second stage of the operation is a testing process where biometric data obtained from the user is

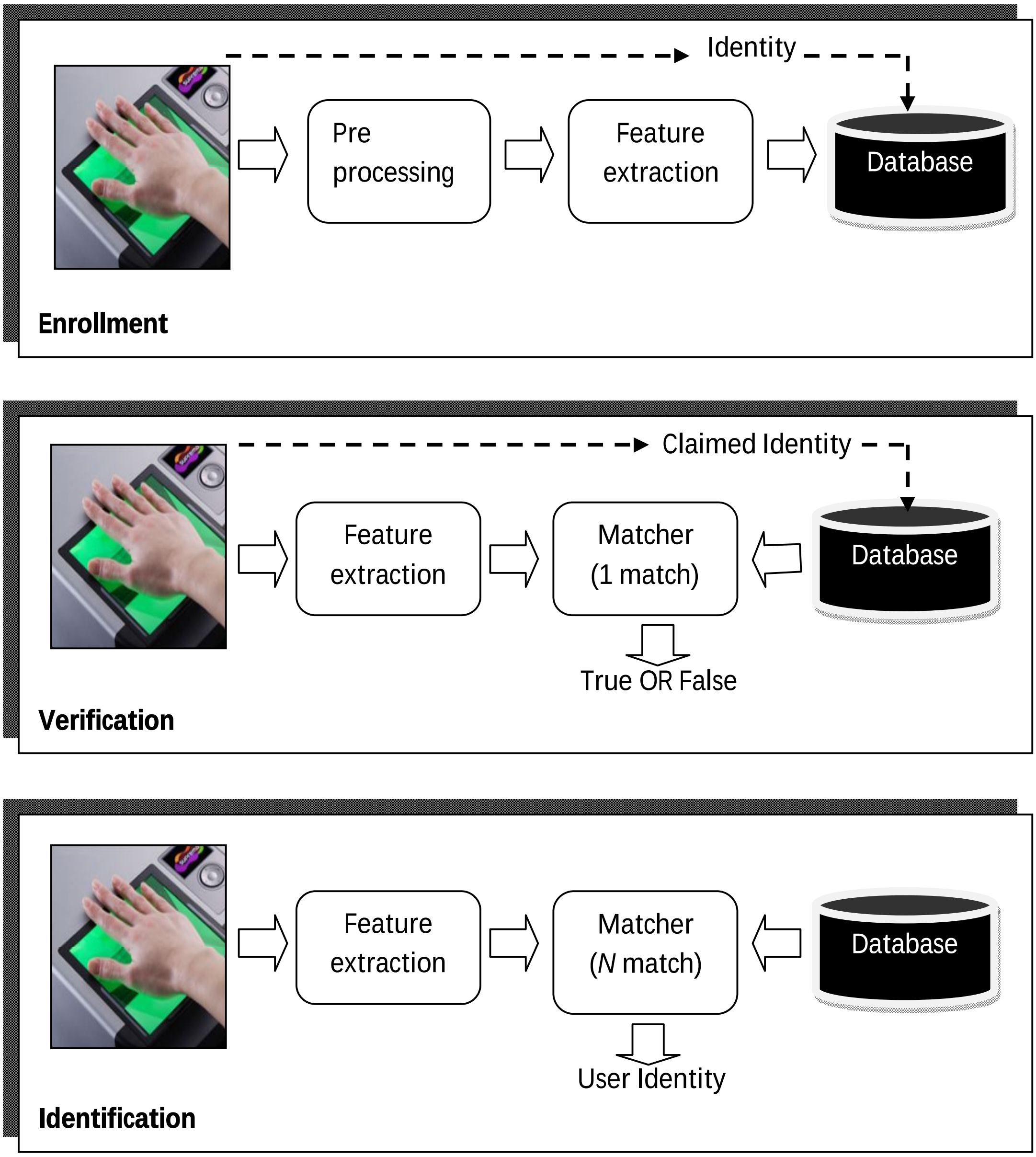


Figure 1.2: Block diagram of the basic process in biometric system consists of enrolment mode, verification mode and identification mode.

compared against the reference database for the purpose of recognition. The recognition process can be performed in the two modes of operation, verification and identification.

In the verification mode, the system validates a person's identity by comparing the captured biometric data with the relevant template stored in the database. In such a system, an individual who requests recognition first claims an identity, usually via a personal identification number (PIN), a user name, or a smart card, and the system conducts a one-to-one judgment to determine whether the claim is true or not. Identity verification is typically used for positive recognition, where the aim is to prevent more than one person from using the same identity. The verification problem can be explained as follows [2]: given an input feature X_{test} extracted from a test image, and a claimed identity X_i, determine if (X_{test}, X_i) belongs to class ω_1 or ω_2, where ω_1 indicates that the claim is true (a genuine user) and ω_2 indicates that the claim is false (an imposter). Typically, X_{test} is compared with the biometrics template corresponding to X_i in order to determine its category. Thus

$$(X_i, X_{test}) \;\in\; \begin{cases} \omega_1 & if\ S(X_i, X_{test}) \;\leq t \\ \omega_2 & otherwise \end{cases} \tag{1.1}$$

where S is a function that measures the similarity between feature vector X_i and X_{test}, where t is a predefined threshold. The value $S(X_i, X_{test})$ measures the similarity between the given biometric input of the user and the claimed identity, where the function can be Euclidean distance, Mahalonabis distance or Likelihood score. Parameter t is a predefined threshold which is compared with the function S, since the biometric traits of the same individual are never identical due to factors such as being taken at different times, with different poses for the facial images and with different types of sensor.

In the identification mode, the system recognizes an individual by searching the templates of all the users in the database for a match. Therefore, the system conducts a one-to-many comparison to establish an individual's identity without the subject having to claim an identity. Here, given a test feature vector X_{test}, determine the identity C_i, $i \in \{1, 2, \ldots, N, N+1\}$. In this case C_1, C_2,…C_N are identities enrolled in the system and C_{N+1} indicates rejection where no suitable identity can be recognised by the system. The identification problem can be summarised as follows:

$$X_{test} \in \begin{cases} C_i & if \ \max_k\{S(X_{test}, X_{C_i})\} \leq t, i = 1,2, \ldots, N \\ C_{N+1} & \text{otherwise} \end{cases} \tag{1.2}$$

where X_{C_i} is the biometric template corresponding to identity C_i, and t is the predefined threshold.

1.4 Limitations of Unimodal Biometrics

Although several advantages of biometrics system for both civilian and government authentication applications have been reported compared with conventional methods based on tokens and passwords, it is imperative that the vulnerabilities and limitations of these systems are considered when applied in real world applications. In real world scenarios where large numbers of users are involved, unimodal biometrics that use only single modalities will have various limitations. Some of the challenges commonly encountered by unimodal biometric systems are as follows [3]:

1) Noise or distortion in the data sensed. For example, fingerprints with the presence of scars or voices altered by illness will give different raw data during the testing and enrolment process. Noisy biometric data may be incorrectly matched with templates in the databases resulting in users being incorrectly rejected or identified.
2) Intra-class variations. The biometric data acquired from an individual during authentication may differ from the data that was used to generate the template during enrolment, thus affecting the matching process. There are several reason for these kinds of variations, such as users incorrectly interacting with the sensor or if the characteristics of the sensor are modified during the enrolment and verification process. Intra-class variations are more prominent in behavioural traits, since the varying psychological makeup of an individual might result in vastly different behavioural characteristics at different times.
3) Inter-class similarities refer to the development of feature spaces corresponding to multiple classes or individuals. Even when a biometric trait is expected to vary significantly between different persons, there may still be large inter-class similarities in the feature space that used to represent these traits. Thus, inter-class similarities will increase the rate of false match in the identification system if a large number of users are enrolled in the system. There is an upper boundary of the number of individuals that can be effectively discriminated amongst by any biometrics system [4]. This upper boundary on the number of distinguishable patterns indicates that the capacity of an identification system cannot be arbitrarily increased for a fixed feature set and matching algorithm.
4) Non-universality. The biometric system may not be able to obtain sufficient raw biometric data from a subset of users. A fingerprint biometric system, for

example, may extract incorrect minutiae features from the fingerprints of certain individuals, due to the poor quality of their fingerprint ridges. Thus, the system cannot enrol the modalities that have these kind of problems. Thus, there is a failure to enrol rate associated with using a single biometric trait such as reported in speaker recognition system [5].

5) Spoof attacks. Spoofing involves the manipulation of one’s biometric traits in order to avoid recognition. There is also possible to create artificial physical biometrics in order to assume the identity of another person. This type of attack is more relevant to behavioural traits such as signatures [6] and gait [7]. However, physical traits are also vulnerable to spoof attacks. For example, it is possible to construct artificial fingers or fingerprints to circumvent a fingerprint verification system [8].

1.5 Motivation for Multimodal Biometrics

Despite substantial advances in recent years, there are still severe challenges in obtaining reliable authentication through unimodal biometric systems. These are due to a variety of reasons. For instance, there are problems with enrolment due to the nonuniversal nature of relevant biometric traits. Non-universality implies the possibility that a subset of users do not possess the biometric trait being acquired. Equally worrying is biometric spoofing, which means that it is possible for unimodal systems to be fooled such as through the use of contact lenses with copied patterns for iris recognition. Moreover, the effect of environmental noise on the data acquisition process can lead to a lack of accuracy which may disable systems virtually from their inception [9]. Biometrics based on voice recognition, for instance, degrade rapidly in noisy environments. Similarly, the effectiveness of face verification depends strongly on lighting conditions and on variations in the facial image. Some of the limitations imposed by unimodal biometrics systems can be reduced by using multiple biometric modalities. The provision of multiple types of evidence through the acquisition of multimodal biometric data may focus on multiple samples of a single biometric trait, which is designated as multi-sample biometrics. It may also focus on samples of multiple biometric types, which is termed multimodal biometrics. Higher accuracy and greater resistance to spoofing are the basic advantages of multimodal compared to unimodal biometrics. Multimodal biometrics involves the use of complementary information as well as making it more difficult for an intruder to simultaneously spoof the different biometric traits of a registered user. In addition, the problem of non-universality is largely overcome, since multiple traits can ensure sufficient of the coverage population. Because of these advantages of multimodal biometrics systems they may be preferred over a single modality even though the storage requirements, processing time and computational demands are much higher.

The fusion of complementary types of information in multimodal biometric data has been an active research area, since it plays a critical role in overcoming certain important limitations of unimodal systems. Efforts in this area are mainly focused on fusing the information obtained from a variety of independent modalities for example in the combination of the face and palmprint modality to achieve the more reliable recognition of individuals. In such an approach, information from different modalities is used to provide complementary evidence about the identity of users. Most of the fusion of this information occurs at the matching score level. This is because the individual modalities provide different types of raw data, and involve different methods of classification to achieved discrimination. To date, a number of score-level fusion techniques have been developed for this task [10]. These range from the use of different weighting schemes that assign weights to information streams according to their information content, till the use of support vector machines which use the principle of obtaining the best possible boundary for classification according to the training data. Even though fusions at the matching score level achieves better results than those from unimodal biometrics, integrating information at the feature level is believed to be capable of increasing the richness of information in the feature space, thus making it possible to produce better results. However, fusion at the feature level has been implemented less often than at the matching score level due to the lack of fusion techniques which could combine features from two modalities.

Another issue in biometric systems is the effect of input data variation on recognition performance. Such variations are reflected in the corresponding biometric scores, and thereby can adversely influence the overall effectiveness of biometric recognition. Therefore, an important requirement for the effective operation of a multimodal biometrics system in practice is to minimise the effect of variations in the data obtained from the individual modalities deployed. This would allow the maximisation of recognition accuracy in the presence of variation, for example due to contamination, in some or all types of biometric data involved. However, this is a challenging requirement since data can vary due to a variety of factors and types of variations can have different characteristics. Another difficulty in multimodal biometrics is the lack of information about relative levels of variation in the different types of biometric data. The term data variation can be subdivided into two types. The first involves variation in each data type arising from uncontrolled operating conditions, and the second concerns variation in the relative degradation of data. The former can be due, for example, to the poor illumination of a user’s face in face recognition or background noise in voice biometrics or it can be generated by the user, such as uncharacteristic sounds from speakers or carelessness in using the sensor for providing palmprint samples [1]. Variation in the degradation of data, meanwhile, is due to the fact that in multimodal biometrics different data types are normally obtained through independent sensors and data capture apparatuses. Moreover, any data variation associated with operating conditions may in fact also result in variation in the relative degradation of the different biometric data deployed. Since, in practice, it may not be possible to fully compensate for degradation in all of the biometric data types involved, the relative degradation of data appears as another important consideration in multimodal biometrics. This thesis reports a number of contributions to increasing the accuracy of multimodal biometrics in the presence of variation. These are based on investigating methods which can be used to tackle the effects of data degradation and estimating the relative quality of different types of biometric data.

1.6 Current feature level fusion for face and palmprint

Currently most of the feature level fusion in multimodal biometric which combines face and palmprint modalities uses concatenation method to fuse the information extracted from both modalities. Concatenation method which serially combined the holistic feature vector produces a new fused feature vector which has a large feature dimensionality. The new fused feature vector may contain noise and redundant information that can affect the discrimination power of the feature vector. Furthermore, when limited number of training images is available, accurate estimation of model parameters for high dimensional feature vector will not be possible. In the previous feature level fusion, concatenation is performed after pre-processing method such as Gabor transform where the feature vector becomes larger than the original image. High dimensional fused feature vector is then reduced to smaller dimensionality by using linear projection method such as Principle Componnet Analysis (PCA) and Linear Discriminnat Analysis (LDA).

Most of the feature level fusion uses holistic features of face and palmprint images. While holistic features contain information of the whole image, they also include high frequency components that are not useful for fusion purpose. In contrast, fusion of low frequency components is expected to give higher degree of discrimination power, thus Discrete Cosine Transform (DCT) analysis can be used to extract the information into several frequency bands. Local region in the face and palmprint images may also hold important information to distinguish among different persons, therefore, fusion process using local features is expected to be superior to those using holistic features. Furthermore, information fusion using local features produces low dimensional fused feature vector compared to the holistic feature fusion. As a result, an accurate statistical model can be estimated using a large amount of fused feature vector extracted from each local region of the face and palmrint images.

Feature distribution of face and palmprint images are scattered in a non Gaussian form. Assuming a Gaussian distribution in the feature space produces a simple classification method called Euclidean distance classifier that will not utilize all the statistical information exists in the feature vector. This method only manipulates mean values of the feature space in order to perform classification. Non Gaussian information in the feature space requires a more complex model to represent the feature distribution and make use of all the statistical information existing in the face and palmprint images. Such model can be Gaussian mixture model (GMM) where the estimated models try to fit with the distribution of feature vector. Although this model is more complex when compared to the single Gaussian, the estimated model will use all statistical information exist in the feature vector thus fully utilize feature level fusion.

Existing verification analysis of multimodal biometric based on face and palmprint images that fuses the information at feature level use minimum value of distance classifier in order to compare with a threshold value during the verification process. Such method, only considers the information from a genuine user, thus is not effective in some cases such as when there are variations among the test and train images. Furthermore, this approach cannot include information from any imposter also trying to access the system. A reliable verification result can be obtained if we can also model the imposters that are trying to access the system. In such system, we should be able to model a genuine and imposter feature distribution. By using a specific statistical model, a degree of certainty for the input image that belongs to a genuine or an imposter user can be obtained.

1.7 Aim and Objectives

The aims of this thesis are to investigate the fusion of information at feature level in multimodal biometrics that use face and palmprint image as the biometric modalities. The investigation involves several stages relating to feature extraction, fusion and classification which focus on the fundamental theory, assumption made, and limitations. In addition, three novel frameworks are proposed according to which the feature extraction, fusion and classification processes can be tailored. Rigorous mathematical derivations and simulations are carried out to validate the effectiveness of the proposed methods. The objectives of this thesis are as follows:

1) To investigate a new feature fusion technique for face and palmprint images using matrix interleaved method to generate a new feature vector, which addresses the following issues:
    - Non-stationary information in the fused feature vector
    - Increased statistical properties of information in the fused feature vector
    - Increased discriminatory power in the feature vector
2) To develop a new types of compact local feature extracted from biometric images, which addresses the following issues:
    - Multiresolution feature extraction in the biometric image
    - Low frequency information which carries the most information about the image
    - Independent feature vectors of local regions in the image

3) To develop a learning method to capture the underlying statistical information in the fused feature vector, which addresses the following issues:
    - Estimation of class specific model parameters
    - Estimation of all user model parameters
4) To develop likelihood normalization in the classification process, which addresses the following issues:
    - Estimation and updating process for class specific model parameters
    - Computation of likelihood normalization among different numbers of users

1.8 Thesis Contributions

Multimodal biometric system has been continually designed and many methods of information fusion have been proposed by researchers. Most fusion methods focus on the matching score level, but less information can be gained at this level. Some types of existing feature level fusion use concatenation methods by using a global feature vectors which tend to be associated with dimensionality problems. This thesis has provided a novel methodology and pioneered a direction for research that will enable the development of a new feature fusion approach to multimodal biometrics capable of yielding better performance in identification and verification process. In addition, the proposed methods overcome the limitations associated with conventional feature fusion approaches in multimodal as well as unimodal biometrics. This thesis presents three novel methods representing significant improvements in performance in terms of identification and verification. The contributions of this thesis can be outlined in five areas:

i) The methodology of feature extraction developed in this thesis represents a new technique to extract local features that have low feature dimensionality while preserving low frequency components, which is important for discrimination analysis. This method uses multiresolution Gabor analysis, which is able to extract the information about the texture of images at different orientations and scales. The low frequency components are selected from the Discrete Cosine Transform (DCT) coefficients which can be used to reduce the dimensions of the feature vector. The extracted independent local feature vector suitable to be used in the feature fusion process or in unimodal recognition analysis.

ii) A new framework for feature fusion methods is based on the matrix interleaved using compact local feature representation. This is able to increase discrimination power in the new fused feature vector due to the rich information gained from feature fusion. The method combines low frequency information from two modalities and increases the size of the feature vector, and thus is able to give high discrimination power in the classification process.

iii) The learning method based on Gaussian Mixture Model (GMM) is able to capture the underlying statistical information which exists in the fused feature vector. Thus, a complex distribution of the fused feature vector can be represented by a probability density function. Whereas most biometric systems use the Euclidean classifier to conduct the classification process, in which assumptions of normal distribution are made concerning feature distribution, the method proposed here is able to give a new approach and break this assumption and is therefore more practical for real biometric data.

iv) Using likelihood normalization to compute a final likelihood score is able to compensate for the imposter scores that are trying to claim true identities. The likelihood normalization is computed based on the background model and is able to cover all imposters existing in a system.

v) The thesis helps to solve several major problems in the single modal biometrics, and offers ways for biometrics to be used for more secure and accurate applications such as in law enforcement, e-services and financial services.

1.9 Thesis Organisation

This research focuses on face and palmprint multimodal biometric fusion at the feature level. Three novel methods are proposed in this work covering feature extraction, information fusion and the classification process. These new method are discussed in Chapter 3, Chapter 4 and Chapter 5 respectively.

Chapter 2 first introduces multimodal biometrics systems and their advantages over single modal biometrics. Different levels of fusion and their implementation in existing method are briefly reviewed. Conventional feature level fusion in biometric images is discussed along with their method of classification. Then, an overview of learning method based on density estimation in parametric models is discussed.

Chapter 3 focuses on feature extraction and representation of the biometric images and establishes a new method for the compact local representation of face and palmprint images. Multiresolution image analysis and the Discrete Cosine

Transform (DCT) are briefly discussed and the concepts involved in the proposed compact local feature extraction in sub blocks of the image are introduced. Experiment analysis to validate the theory and evaluate the performance of the proposed method is then reported.

Chapter 4 presents the novel matrix interleaved feature fusion. The conceptual framework for the new fusion method is described and explanations provided of the estimation of the probability density functions in the fused feature vector. The recognition process performed using maximum likelihood values computed from the estimated model is also explained. The overall framework is validated using several experimental analyses which analyze recognition performance.

Chapter 5 introduces the likelihood normalization technique which is able to increase the performance of the verification system by incorporating imposter scores into the final computation of likelihood value. Two likelihood normalization methods are discussed, the Universal Background Model (UBM) and Cohort Background Model (CBM), which is able to compensate the likelihood score of imposters that are trying to access the system and reduce the variations in the input test images. Several comparisons and simulations are carried out to investigate the implementation of the proposed method.

Chapter 6 presents the overall conclusions, achievements and limitations of the thesis. It also addresses directions for further work.

Chapter 2

Multimodal Biometric Fusion and Classification

2.1 Introduction

This chapter reviews and discuss the common methods used for information fusion in multimodal biometrics systems. The advantages and disadvantages of each fusion level used are discussed in detail. This is followed by discussion of the types of classifier that can be used in multimodal biometric, estimation of parametric model and Bayesian learning methods.

2.2 Information Fusion

Multimodal biometrics can overcome some of the limitations of unimodal biometrics systems by using several biometric modalities to represent individual characteristics. Such systems are expected to be more reliable due to the presence of multiple and independent sources of evidence [11,12,13]. Multimodal biometric systems can overcome the problem of non-universality when multiple traits are used to ensure sufficient population coverage. Multimodal biometrics can also provide anti-spoofing measures by making it difficult for an intruder to simultaneously spoof the multiple biometric traits of a genuine user. Furthermore, by using multiple traits, more information can be gained which can eliminate the problem of inter-class similarity in the feature space and thus increase the performance of the recognition rate. The use of independent modalities is expected to have a very significant role in the improvement possible when feature from multiple biometric modalities are fused. A properly designed combination scheme that has been trained and tested on a large amount of data is expected to perform better than even the best single modality system. One of the challenges faced in reducing inter-class similarity is to find a method of integrating all of these traits. In a biometric recognition system, the amount of information available in the system will decrease when we move from front processing at the sensor level to end processing such as in decision modules. Information of the biometric traits can be merged at several different levels of fusion, as shown in Figure 2.1, either at the sensor and feature level; at matching score level; at the decision level [10]. At the early stages of information fusion, the integration of data from multiple biometric sources can be carried out at the sensor level or at the feature level.

Sensor level fusion involves combining raw data from sensors, and it can be achieved if the multiple sources represent samples of the same biometric trait obtained using single or different compatible sensors. In this method, the multiple modalities must be compatible with feature level in the raw data and must be known in advance, for example in the construction of 3D face images by integrating raw images captured from several cameras [14]. Jain [15] integrates information at the sensor level by forming a mosaic of multiple fingerprint impressions in order to construct a more elaborate fingerprint image. Feature level fusion (Figure 2.1a) refers to combining the features obtained from multiple modalities into a single feature vector. If the features extracted from multiple biometrics are independent of each other and involve the same type of measurement scale, it is reasonable to concatenate the two vectors into a single new vector.

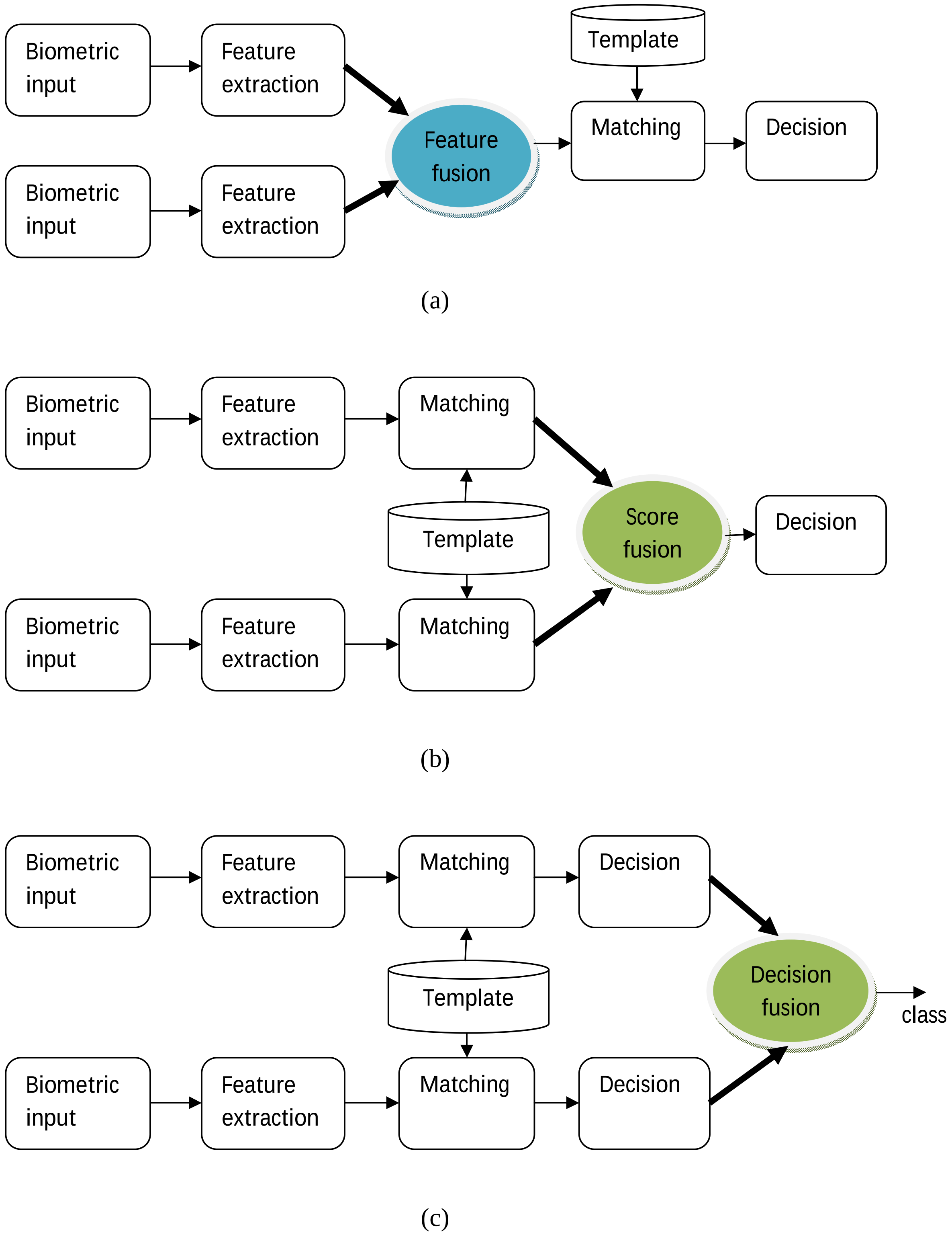


Figure 2.1: Different levels of fusion in multimodal biometrics systems. a) Feature level. b) Matching score level. c) Decision level

The new fused feature vector will have higher dimensionality and thus increase the discriminating power in feature space. Feature reduction techniques or feature selection schemes may then be employed to extract a small number of significant features from a larger set of features [16, 17, 18]. In some cases concatenation is not possible, such as when the feature sets are incompatible as in the case, for example, with fingerprint minutiae and eigen-face coefficients.

Fusion at matching score level (Figure 2.1b) refers to the combination of similarity scores provided by a matching module for each modality when the input features are compared against templates in the database [19]. This method is also known as fusion at the measurement or confidence level. The matched score output generated by biometrics matchers contain rich information about the input pattern after the feature extraction module. Fusion at matching score level can be categorised as involving two different approaches depending on how the matching score given by matching module is treated [20]. In the first approach, the fusion can be viewed as a classification problem where a feature vector is constructed using the matching score output by the individual matchers. Then, the constructed feature vector is classified into two of the classes whether to accept or reject the claim user. In the second approach, fusion is viewed as a combination approach where individual matching scores are combined to generate a single scalar score using normalization techniques and fusion rules. The new single scalar score is then used to make a final decision. The combination approach to the fusion of matching scores has been extensively studied, and Ross [10] concluded that it performs better than the classification approach. Fusing matching scores using the combination approach has some issues arise during computing a single fusion score given by different modalities. A normalization technique is required to transform matched scores into a common domain prior to combining them, since the matching scores generated from different modalities are heterogeneous [20]. Several different kinds of normalization technique have been proposed, such as min-max, median and sigmoid normalization.

Integration of the information at the decision level (Figure 2.1c) is performed when each of the individual biometric matchers decides the best match based on the input features presented to the matching module. Various methods such as majority vote [21], behaviour knowledge space [22], AND and OR rule [23] can be used to make the final decision. This kind of fusion uses binary information to derive a final decision, and thus fusion at decision level is not effective since only a limited amount of information is available at this level. Therefore, the integration of the information at feature and matching score level is generally preferred due to the richness of information available at the fusion stage.

Multimodal biometrics systems that fuse information at an early stage are believed to be more effective than those that integrate it at a later stage. This is due to the rich information which exists at feature level compared to that at matching and decision level. Since the features contain richer information about the input data, learning the distribution of the fused feature using statistical models to capture the relevant statistical properties is likely to give better classification performance. In most existing feature fusion methods, a concatenation feature vector is classified using a distance classifier, where an assumption is made of normal distribution in the fused feature vector. Such an assumption does not fully utilize the information available in the fused feature vector which has a non Gaussian feature distribution due to the non linear information in the biometric image. Thus, one of the aims in this thesis is to propose statistical method based on a Gaussian mixture model (GMM) to capture underlying statistical information which exists at the early stages of information fusion. To the best of the present author's knowledge, this is the first study of multimodal biometrics to propose the capture of underlying statistical information from the fused feature vector by using GMM, as well as the first to specifically use local features in multimodal biometrics based on face and palmprint images. Theoretically, GMM should be better to capture a complex distribution of features than a single Gaussian distribution due to the utilizing several normal distribution to form a mixture model.

2.3 Research on Multimodal Biometrics

Although biometrics technology for authentication has been studied for more than 30 years, multimodal biometrics that combines information from several different traits in the authentication process have received much attention in recent years. There are a number of advantages given by the combination of different traits, such as decreasing False Accept Rate (FAR) and False Reject Rate (FRR), making systems more robust in case of sensor failure and where the system is not able to read the input such as when scarring obscures in fingerprint. Most multimodal biometrics proposed in the literature focus on increasing the accuracy of recognition by measuring the error rates given by FAR and FRR. In 1995, Brunelli et al. [24] proposed a multimodal biometrics system that uses face and voice traits for identification. Their system performs the fusion of the matching scores given by five different matching modules computed from voice and face features to generate a single matching score used for identification. In 1997, Duc et al. [25] proposed a method to integrate information extracted from speech and face modalities by using a statistical framework based on Bayesian statistics. Hong et al. [26] proposed a multimodal biometrics that combines face and fingerprint information at the decision level for person identification. The identification for each modality is conducted separately by PCA based face and minutiae based fingerprint identification and then the decisions of each module are combined in the consolidation of multiple cues by associating different confidence measure with the individual biometric matchers. Their results showed a significant improvement in the accuracy of identification, where the multimodal approaches achieved 92% Genuine Acceptance Rate (GAR) at 0.01% FAR. On the other hand, single modal biometrics only achieved 86% and 45% GAR at 0.01% of FAR for face and fingerprint images respectively.

In 2000, Frischholz and Dieckmann [12] proposed a commercial multimodal product called BioID that uses voice, lip motion and face features of users for verifying their identity. Lip motion and face images are extracted from a video sequence and the voice is extracted from an audio signal. Their experimental results showed they achieved below 1% FAR when tested on 150 subjects. In 2003, Fierrez-Aguilar and Ortega-Garcia [27] then proposed a multimodal biometrics system based on face, fingerprint and signature data with a fusion method performed at the score level. Individual matching scores in each modality were computed separately based on global appearance face representations; minutiae based fingerprint representations and the Hidden Marcov Model (HMM) modelling of temporal functions for signature verification. Fusion at score level was then established by using a sum-rule and a support vector machine (SVM). Their results showed a significant improvement in terms of Equal Error Rate (EER), where the multimodal method achieved the best value of 0.05% EER compared to the single modal which is 3% EER. In the same year, Kumar el al. [28] proposed a multimodal approach based on palmprint and hand geometry by using a fusion method performed at the feature and matching score level conducted using concatenation and the max rule respectively. Their analysis showed that only fusion at matching score level outperformed the results given by a single modality system. Ross and Jain [10] then proposed a method to combine three modalities based on face, fingerprint and hand geometry. Three fusion methods at matching score level were considered. The sum rule, decision trees and a linear discrimination function were computed, and a normalization method was applied to the matching scores prior to combining them. Their results showed that sum rule fusion outperformed the other fusion strategies as well as a single modal system.

In 2003, Wang et al. [29] proposed a multimodal system that combined the scores given by PCA based face and iris verification systems. Their operated fusion strategy at the matching score level using weight sum rules, unweight sum rule, fisher discriminant analysis and a neural network. Toh et al. [30] also proposed the use of the weighted sum rule to fuse the matching scores given by hand geometry, fingerprint and voice modalities. In 2005, Snelick et al. [31] developed a multimodal biometrics system for face and fingerprint images with a fusion technique performed at matching score level. In their fusion framework, values of three fingerprint matching scores and one face matching score from a commercial system are used. They study the effect of using seven different score normalization techniques and different fusion strategies performed on the normalized scores, such as simple sum, min score and max score. The results showed all of the normalization and fusion approaches outperform a single modal biometric. Jain et al. [20] proposed a multimodal approach combining face, fingerprint and hand geometry using a fusion method at score level. The output of three matcher module is fused together from a similarity score for minutiae based matcher of fingerprint images, the Euclidean distance for PCA based algorithm of face images and the Euclidean distance for feature vector of hand geometry. They investigated several techniques for score normalization; namely, Min-Max, z-score, Median-MAD, Double sigmoid, Tanh and Parzen normalizations. The normalized techniques are tested using three different score fusion methods using the sum rule, max rule, and min rule. Their results showed all fusion methods outperformed the single modal biometric for all normalization methods except median MAD. At low FAR, the tanh and min-max normalization techniques outperformed the other techniques, while at higher FAR, the z-score normalization performed better than the other techniques.

From the above discussion it can be concluded that most types of multimodal biometric fusion proposed so far focus on fusion at the matching score and decision levels. Feature level fusion has been implemented less often for several reasons, such as the need to consider the presence of noise in component feature sets, and the fact that a new matching algorithm may be necessary to compare the fused feature sets. Thus this research work developed a new framework for multimodal biometric beginning with the extraction of important features in the face and palmprint modalities and the reduction of the noise present in the extracted features. Then, the information from each modality is integrated at feature level by using matrix interleaved method. Next, the fused feature vector is learned using a GMM where identification is computed based on maximum likelihood scores and verification is computed with the implementation of likelihood normalization.

2.4 Feature Level Fusion

Fusion is a popular method of increasing the performance of biometrics verification by consolidating information given by multiple modalities, such as face and palmprint, or iris and fingerprint data. Generally speaking fusion can be conducted at four different levels: at the sensor, feature, matching score or decision levels. Fusion at each level necessitates a different computational task due to the different nature of the input at that level. Figure 2.2 shows the amount of the information available for fusion at different levels. The raw data represents the richest source of information, whereas on the other hand, the final decision contains a single bit of information. However, although the raw data contains the richest information, it is corrupted by noise and exhibits high intra-class variation which needs to be reduced in the feature extraction module. Moreover, working on raw data entails high memory costs due to the large amount of information which needs to be stored in databases.

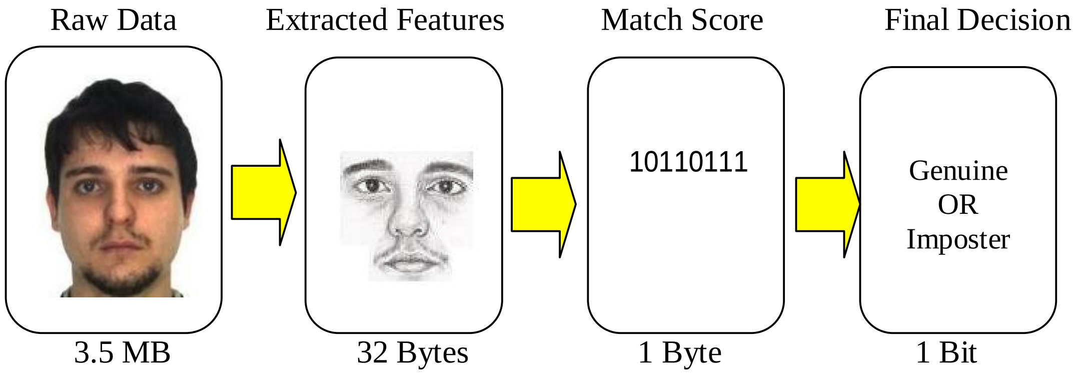


Figure 2.2: Amount of information from raw image to decision module.

Fusion at feature level will produce a new feature vector in high dimensional feature space given by $x_i \in \mathbb{R}^{m_i}, m_i \geq 1, i = 1,2, \dots N$. At matching score level, the feature vector is the output of each expert, and thus is reduced to a scalar value, $S_i \in \mathbb{R}, i = 1,2, \dots N$. At the decision level, the matching scores S_i are compared with threshold values in order to derive the binary decision $d_i \in \{1,0\}, i = 1,2, \dots, N$. Fusing at the matching score and decision level does not fully utilize important information in the feature vector because the integration is performed on scalar and binary value. Several rules for matching score fusion have been proposed in the literature, such as max, min and sum rules computed based on the output of each independent expert. Each modality is assumed to have its own statistical properties, but in reality they are related to each other. Thus integrating information at the feature level will combine not only the feature vectors but all of the statistical properties in each modality as well.

Most feature fusion methods utilise concatenation in order to integrate the information from different modalities. In this case, the types of features used for concatenation have an impact on the performance of the system. Some features might contain redundant information, and thus a feature selection method needs to be used in order to select the best features. In some cases, a fused feature vector might contain noise from the data, and this will degrade performance when fusion is performed on less discriminative data. In order to achieve good performance in the fusion of feature vectors, the type of feature vector used plays an important role in increasing discrimination power during the integration process. Thus a new feature extraction framework is proposed here that can be used in feature fusion methods for multimodal biometrics. This method is based on the compact energy representation of multiresolution independent local features extracted from 2D biometrics images. Previous studies have concatenated two types of features, where those that represent the whole image are known as holistic features, while those representing certain parts of images are known as local features.

The method proposed by Jing [32] used global features based on the Gabor transform to concatenate face and palmprint information at feature level. With different scales and orientations, a set of Gabor filters are used to extract the texture information existing in each modality. In their method concatenation was performed in a high dimensional feature space, and then the high dimensional fused feature vectors were reduced to a low dimensional feature space using kernel discriminative common vectors. Concatenation can also be performed on feature spaces with low dimensionality as suggested by Yao [33], where the extracted PCA features for face and palmprint images are fused using weighted values. In their method, important information existing in the face and palmprint images is extracted using the Gabor transform to produce a nonlinear feature vector. The assumption of normal distribution in the fused feature vector leads to a distance classifier being utilized, where in practice the distribution of the fused feature is scattered in a non Gaussian form due to the nonlinear properties of face and palmprint images. Another method for the fusion of face and palmprint modalities at feature level was proposed by Yan [34] using global features extracted from PCA in the fusion process. Each component of fused feature vector is derived by the addition of the inner products of low dimensional PCA features and a set of correlation filters using the 1D Fourier transform. Lu [35] proposed fusion at feature level using global features extracted from face and palmprint images using a set of Gabor filters. The information was fused at feature level using the summation of the weighted results of Independent Component Analysis (ICA) for each of the modalities. In this method, an appropriate weighting value must be chosen in order to achieve a good result. Rattani [36] proposed to fuse face and fingerprint features using the concatenation of local feature points obtained from two modalities, which is believed to have better discrimination power compared to the unimodal biometrics based on face and palmprint modalities. In this method, the fusion is performed on the local features of the face given by the eye, nose and mouth, whereas those for fingerprints are given by minutiae points and texture of palm image. Important information at each local region is extracted using the Gabor transform with different orientations and scales. Kong [37] proposed feature level fusion where multiple elliptical Gabor filters with different orientations were used to extract the phase information from palmprint images and then merged according to a fusion rule to produce a single fused feature called the Fusion Code. Wang [38] proposed a method to fuse palmprint and palm vein images using features extracted from a wavelet transform, and the fusion method is able to enhance the discrimination power of the images. These two modalities are fused using edge preserving and contrast enhancing wavelet fusion method where the modified multiscale edges of the palmprint and palm vein images are combined. Pajares [39] proposed image fusion based on wavelet decomposition, where images features in different multiscale are fused together. In this method, feature fusion is performed on the same group of wavelet multiscales of two different images. Feature fusion based on a wavelet transform has several advantages in that the multiscale approach is able to manage image resolution where the image information is preserved in different kind of wavelet decomposition. Ross [16] integrated 2D images of the hand and face at the feature level by using the concatenation of normalized feature vectors extracted from two modalities. In order to reduce high dimensional feature vectors and remove noisy features affected from the concatenation process, a feature selection process based on the floating search method proposed by Pudil [40] was used to choose minimal feature sets that would increase the performance of classification. Ravigaran [18] proposed to use feature selection algorithm using particle swarm optimization (PSO) in order to optimise the representation of feature after concatenating the log Gabor images of face and palmprint modalities. Implementing PSO to select dominant features will significantly reduce a large amount of feature vectors in the feature space. Then, the selected features are further reduced using kernel discriminant common vector (KDCV) method to select the most discrimination features from the fused feature vector. Another interesting fusion approach was proposed by Xie [41] where the local phase and magnitude of Gabor face images are fused together. In this method, the operator of XOR patterns is multiplied with Gabor image in the sub block window and then a block based Fisher Linear Discrimination (FLD) analysis is used to reduce the dimension of the concatenated histograms. Fusion of magnitude and phase is performed using the features extracted from block based FLD analysis. Ajay [42] proposed a new feature extraction method which can be used to integrate hand shape and palm texture information. This implies that only small subsets of features may be necessary to be fused together. For example, a feature selection method can be used to extract the hand shape feature, and compact energy features given by a DCT are used to extract the texture of the palmprint image.

2.5 Density Estimation in the Parametric Models

The distribution of the fused feature vector in the feature space can be represented as a density function. Given a class C_i and $i = 1,2,3, \dots c$, the densities $f_1, f_2, \dots f_c$ are known as the class conditional densities because they represent the probability density function of the fused feature vector in class C_i. The densities $f_1, f_2, \ldots f_c$ are usually unknown and need to be estimated from a given training feature vector from class C_i. Density estimation can be computed either by parametric or non-parametric methods [43]. In parametric density estimation techniques, the type of density function is known and only its parameters need to be estimated from the training data. Such density function distributions can be Gaussian or Normal density function, and the only parameters that need to be estimated from the training data are means and standard deviations. The Gaussian assumption used in representing class conditional densities will not be appropriate to represent the distribution of the fused feature vector due to the complexity of features in biometric data. More complex models, such as GMM, can be used to represent the class conditional density function of the fused feature vector. Same as the Gaussian density function, we only know its form; however, the details of parameters need to be estimated from the training data.

Parameter estimation can be computed in several ways, such as by using maximum likelihood and Bayesian estimation methods [44]. This thesis only considers density estimation by using maximum likelihood methods, where the unknown parameters are viewed as quantities with fixed but unknown values. The best estimates of parameter values are then defined as those which maximize the probability of obtaining the samples actually observed. Parameter estimation in the mixture model can be explained as follows. First, assume that a complete probability structure of mixture model is known except for the values of parameters that form their structure. Under this assumption, the form of the class conditional probability densities $p(\mathbf{x}|\omega_j, \boldsymbol{\theta}_j), j = 1,2, \ldots, c$ consists of a known number of c components and the prior probabilities $P(\omega_j)$, $j = 1,2, \ldots, c$ for each component are known. Such assumptions leave the problem of estimating the model parameters of the mixture models given by

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{j=1}^{c} p\left(\mathbf{x}|\omega_j, \boldsymbol{\theta}_j\right)P\left(\omega_j\right) \tag{2.1}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_c\}$. If the conditional densities $p\left(\mathbf{x}|\omega_j, \boldsymbol{\theta}_j\right)$ have a Gaussian or normal distribution the components will consist of means and covariance, $\boldsymbol{\theta}_i = \{\mu, \sum\}$. $P(\omega_i)$ is a mixing parameter and can also be included among the unknown parameters. The aim of the maximum likelihood method is to use samples drawn from the mixture models in Eq. (2.1) to estimate the unknown parameters $\boldsymbol{\theta}$. Once parameters $\boldsymbol{\theta}$ are known, the mixture can be decomposed into its components and use maximum a posterior classifier to perform the classification process. If a sample data given by $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots., \mathbf{x}_n\}$ consists of n samples drawn independently from the mixture model in Eq. (2.1), the objective is then to find a values of parameters $\widehat{\boldsymbol{\theta}}$ that maximizes $p(D|\boldsymbol{\theta})$, where $p(D|\boldsymbol{\theta})$ is the likelihood function. If $\mathbf{x}_i \in D$ are assumed to be independent, the joint parameter conditional probability density function (pdf) of the data D is:

$$p(D|\boldsymbol{\theta}) \equiv \prod_{k=1}^{n} p(\mathbf{x}_k|\boldsymbol{\theta}) \tag{2.2}$$

The goal is to maximize $p(D|\boldsymbol{\theta})$ with respect to the parameter vector $\boldsymbol{\theta}$, as illustrated in Figure 2.3 [44]. Thus, we need to differentiate the function $p(D|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and make them equal to zero as follows:

$$\nabla_{\boldsymbol{\theta}} p(D|\boldsymbol{\theta}) = 0 \tag{2.3}$$

One useful function is the natural logarithm function, that is, to work with $log[p(D|\boldsymbol{\theta})]$. By denoting the log of the likelihood function as:

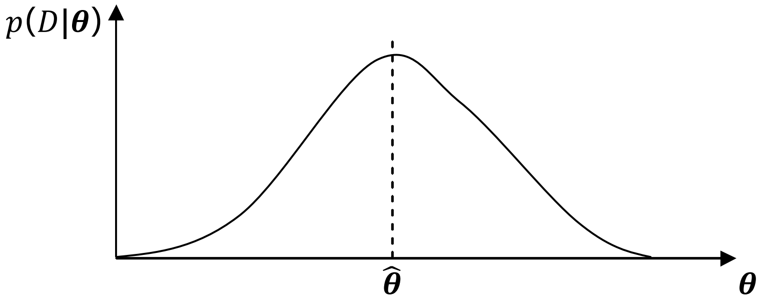
$$l(\theta) = log\{p(D|\boldsymbol{\theta})\} \tag{2.4}$$

then the independent assumption in Eq. (2.2) allows Eq. (2.4) to be re-written as:

$$l(\theta) = \sum_{k=1}^{n} log\{p(\mathbf{x}_k|\theta)\} \tag{2.5}$$

Using Eq. (2.3) yields the following:

$$\boldsymbol{\nabla}_{\theta} l = \sum_{k=1}^{n} \frac{1}{p(\mathbf{x}_k|\boldsymbol{\theta})} \boldsymbol{\nabla}_{\theta} \left[\sum_{j=1}^{c} p(\mathbf{x}_k|\omega_j, \boldsymbol{\theta}_j) P(\omega_j) \right] \tag{2.6}$$



$\hat{\theta}$ is the estimate that maximizes the likelihood function $p(D|\boldsymbol{\theta})$, that is, the ML estimate

Figure 2.3: Computation of maximum likelihood.

By using Bayes rule, posterior probabilities can be expressed as $P(\omega_i|\mathbf{x}_k, \boldsymbol{\theta}) = \frac{p(\mathbf{x}_k|\omega_i, \boldsymbol{\theta})P(\omega_i)}{p(\mathbf{x}_k|\boldsymbol{\theta})}$ and the gradient of the log likelihood can be written in a new form:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} l = \sum_{k=1}^{n} P(\omega_i|\mathbf{x}_k, \boldsymbol{\theta})\, \boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \log p(\mathbf{x}_k|\omega_i, \widehat{\boldsymbol{\theta}}_i) \tag{2.7}$$

Because the gradient must equal zero at the value of θ_i that maximizes *l*, the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_i$ must satisfy the conditions:

$$\sum_{k=1}^{n} P(\omega_i|\mathbf{x}_k, \boldsymbol{\theta})\, \boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \log p(\mathbf{x}_k|\omega_i, \hat{\boldsymbol{\theta}}_i) = 0 \quad for\ i = 1,2, \dots, c \tag{2.8}$$

The maximum value of $p(D|\boldsymbol{\theta})$extends over $\boldsymbol{\theta}$ and $P(\omega_i)$, subject to the constraints that $P(\omega_i) \geq 0$ for $i = 1,2, \dots, c$ and $\sum_{i=1}^{c} P(\omega_i) = 1$. If the likelihood function is differentiable and if $\hat{P}(\omega_i) \neq 0$ for any value of *i*, then $\hat{P}(\omega_i)$ and $\hat{\boldsymbol{\theta}}_i$ must satisfy:

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \hat{\boldsymbol{\theta}}) \tag{2.9}$$

and

$$\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \hat{\boldsymbol{\theta}})\boldsymbol{\nabla}_{\boldsymbol{\theta}_i} \ln p(\mathbf{x}_k|\omega_i, \hat{\boldsymbol{\theta}}_{\boldsymbol{i}}) = 0 \tag{2.10}$$

where

$$\hat{P}(\omega_i|\mathrm{x}_k, \hat{\boldsymbol{\theta}}) = \frac{p(\mathrm{x}_k|\omega_i, \hat{\boldsymbol{\theta}})\hat{P}(\omega_i)}{\sum_{j=1}^{c} p(\mathrm{x}_k|\omega_j, \hat{\boldsymbol{\theta}}_j)\hat{P}(\omega_j)} \tag{2.11}$$

Equations (2.9)–(2.11) above can be used to estimate the unknown value of the mixture parameters $\mu_i, \sum_i$ and $P(\omega_i)$. Let $x_p(k)$ be the *p*-th element of $\mathbf{x}_k$, $\mu_p(i)$ be the p-th element of μ_i, $\sigma_{pq}(i)$ be the *pq*-th element of $\sum_i$, and $\sigma^{pq}(i)$ be the *pq*-th element of $\sum_i^{-1}$. The following differentiation:

$$\ln p(\mathbf{x}_k|\omega_i, \boldsymbol{\theta}_i) = \ln\frac{\left|\sum_i^{-1}\right|^{1/2}}{(2\pi)^{d/2}} - \frac{1}{2}(\mathbf{x}_k - \mu_i)^t \textstyle\sum_i^{-1}(\mathbf{x}_k - \mu_i) \tag{2.12}$$

with respect to the elements of μ_i and $\sum_i^{-1}$will give:

$$\boldsymbol{\nabla}_{\mu_i} \ln p(\mathbf{x}_k|\omega_i, \boldsymbol{\theta}_i) = \textstyle\sum_i^{-1}(\mathbf{x}_k - \mu_i) \tag{2.13}$$

and

$$\frac{\partial \ln p(\mathbf{x}_k|\omega_i, \boldsymbol{\theta}_i)}{\partial \sigma^{pq}(i)} = \left(1 - \frac{\delta_{pq}}{2}\right)\left[\sigma_{pq}(i) - \left(x_p(k) - \mu_p(i)\right)\left(x_q(k) - \mu_q(i)\right)\right] \tag{2.14}$$

where δ_{pq} is the Kronecker delta. Substituting the result in Eq. (2.10) gives the following equations for the local-maximum likelihood estimates for $\hat{\mu}_i, \hat{\Sigma}_i$ and $\hat{P}(\omega_i)$ [43]:

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \widehat{\boldsymbol{\theta}}) \tag{2.15}$$

$$\hat{\mu}_i = \frac{\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \widehat{\boldsymbol{\theta}})\mathbf{x}_k}{\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \widehat{\boldsymbol{\theta}})} \tag{2.16}$$

$$\hat{\Sigma}_i = \frac{\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \widehat{\boldsymbol{\theta}})(\mathbf{x}_k - \hat{\mu}_i)(\mathbf{x}_k - \hat{\mu}_i)^t}{\sum_{k=1}^{n} \hat{P}(\omega_i|\mathbf{x}_k, \widehat{\boldsymbol{\theta}})} \tag{2.17}$$

where

$$\hat{P}(\omega_i|\mathbf{x}_k, \widehat{\boldsymbol{\theta}}) = \frac{p(\mathbf{x}_k|\omega_i, \widehat{\boldsymbol{\theta}}_i)\hat{P}(\omega_i)}{\sum_{j=1}^{c} p(\mathbf{x}_k|\omega_j, \widehat{\boldsymbol{\theta}}_j)\hat{P}(\omega_j)} \tag{2.18}$$

$$= \frac{\left|\hat{\Sigma}_i\right|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}_k - \hat{\boldsymbol{\mu}}_i)^t \hat{\Sigma}_i^{-1}(\mathbf{x}_k - \hat{\boldsymbol{\mu}}_i)\right] \hat{P}(\omega_i)}{\sum_{j=1}^{c}\left|\hat{\Sigma}_j\right|^{-1/2} exp\left[-\frac{1}{2}(x_k - \hat{\boldsymbol{\mu}}_j)^t \hat{\Sigma}_j^{-1}(\mathbf{x}_k - \hat{\boldsymbol{\mu}}_j)\right] \hat{P}(\omega_j)} \tag{2.19}$$

2.6 Expectation Maximization (EM)

The application of parameter estimation using the maximum likelihood derived in the previous section can be extended to permit the parameter estimation of the distribution of training features where some of its have missing features. The parameter estimation method is based on the expectation maximization (EM) algorithm which iteratively estimates the likelihood value that the given data is present and the iteration process continues until a certain criterion is achieved. Given a full sample $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ taken from a single distribution where some features are missing, and supposing that each data point consist of missing features, thus each of them can be written as $\mathbf{x}_k = \{\mathbf{x}_{kg}, \mathbf{x}_{kb}\}$, where $\mathbf{x}_{kg}$ is a good feature and $\mathbf{x}_{kb}$ is a bad feature. These two types of features can be separated into $\mathcal{D}_g$ and $\mathcal{D}_b$, where $\mathcal{D} = \mathcal{D}_g \cup \mathcal{D}_b$ is the union of good and bad features. So, the EM algorithm maximizes the expectation of the log likelihood function, conditioned on the observed samples and the current iteration estimate of $\boldsymbol{\theta}$. The EM algorithm consists of two steps as follows [43, 44].

- Expectation step (E-Step): Compute the expected value at the $(t + 1)th$ step of the iteration with a given parameter $\boldsymbol{\theta}(t)$:

$$Q(\boldsymbol{\theta};\ \boldsymbol{\theta}^i) = \varepsilon_{\mathcal{D}_b}[\ln p(\mathcal{D}_g, \mathcal{D}_b;\ \boldsymbol{\theta}) | \mathcal{D}_g;\ \boldsymbol{\theta}^i] \tag{2.20}$$

  where the use of the semicolon in $Q(\boldsymbol{\theta};\ \boldsymbol{\theta}^i)$ is a function of $\boldsymbol{\theta}$ with $\boldsymbol{\theta}^i$ assumed fix. The semicolon on the right hand side denotes that the expected value is over the missing features assuming that $\boldsymbol{\theta}^i$ are the true parameters describing the distribution.

- Maximization step (M-Step): Compute the next $(i + 1)th$ estimates of $\boldsymbol{\theta}$ by maximizing $Q(\boldsymbol{\theta}; \boldsymbol{\theta}(t))$, that is:

$$\boldsymbol{\theta}(i + 1): \frac{\partial Q(\boldsymbol{\theta}; \boldsymbol{\theta}(i))}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{2.21}$$

  The E-Step and M-Step processes will terminate if $\|\boldsymbol{\theta}^{i+1} - \boldsymbol{\theta}^i\| \leq \varepsilon$, where ε is a certain threshold value. The number of EM iterations depends on the value of ε. Small value of ε will produce many iterations and large values will give fewer number of iterations.

2.7 Summary

This chapter summarizes the state of the art concerning information fusion in multimodal biometrics and their advantages compared to single modal biometrics. Different levels of information fusion are briefly discussed and the relevant is explained in detail. From the literature review, it is clear that feature level fusion has been implemented less often even though it is believed to give superior results due to the rich information exist in the fused feature vectors. Thus, the following chapter investigates feature fusion methods and their effectiveness in the recognition and verification process. Next, the learning processes based on estimation methods which use parametric models to capture statistical information exist in the fused feature vectors are explained. Parametric models that use mixture densities can represent information using several statistical parameters such as weights, means and covariance matrices.

Chapter 3

Feature Extraction and Compact Feature Representation

3.1 Introduction

This chapter investigates global and local feature extraction techniques that can be used in feature fusion. Feature extraction is important for the success of the recognition and classification process, and should be able to extract more information while reducing noise and avoiding redundant data with fast computation. The features given by the extraction process can be used to gain statistical information using supervised and parametric learning techniques. In this chapter, a new framework for feature extraction is proposed which aims to represent local features in a multiresolution compact representation that can be used in the fusion method subsequently proposed in Chapter 4. The effectiveness of the proposed technique is then extensively examined using several types of experimental analysis and the results are compared with those of existing methods.

3.2 Holistic Features Representation

Holistic approaches directly compute the data from raw images and process the raw image as two dimensional holistic patterns. Linear projection method such as PCA and LDA are two traditional methods which have become default tools in holistic based approaches to reduce dimension of feature vectors and increase discrimination power in the feature space. Many methods for the extraction of face and palmprint features have been developed using projection methods, such as Eigenface [45], Eigenpalm [46], Fisherface [47,48], Fisherpalm [49] and their variants [50-54]. PCA performs linear projection method in the image space producing low dimensional feature vectors where the projection directions will maximize the total scatter across all images. During the projection process which maximizes the total scatter matrices, PCA also retains unwanted variations which exist in the images, such as those caused by differences in illumination and facial expressions which will not give the best results in terms of discrimination power. Given a set of *N* sample images in the one dimensional long vector $\{\mathbf{x}_1, \mathbf{x}_2, \ldots., \mathbf{x}_N\}$ with n-dimensional space, and assuming that each image belongs to one of C classes $\{X_1, X_2, \ldots, X_c\}$, the linear transformation mapping the original n-dimensional space into low m-dimensional space is given by $y_k = W^T\mathbf{x}_k$, where *k*=1, 2,…*N* and $W \in \mathbb{R}^{nxm}$ is a matrix with orthonormal columns which maximize the determinant of the total scatter matrix of the projected samples. The criterion that is used to optimize the matrix W is given by $W_{opt} = \arg\max_w |W^T S_T W|$, where S_T is a total scatter matrix is given by $S_T = \sum_{k=1}^{N}(\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T$ where *N* is the number of training images and $\mu \in \mathbb{R}^n$ is the mean image of all samples. A drawback of this approach is that it maximizes both between-class and within-class scatter, where within-class scatter is unwanted information in the discrimination process.

Belhumuer [47] proposed a supervised linear projection method called Fisher Linear Discrimination (FLD) to reduce the dimensionality of the feature space by incorporating a class label. FLD is a class specific method which tries to shape the scatter for features in the same class to be close to each other as possible and while separating the features in a different class to be as far as possible. This method is accomplished by choosing the optimal projection W_{opt} which maximizes the ratio of the determinant of the between-class scatter matrix to that of the within-class scatter matrix of the projected samples given by $W_{opt} = \arg\max_W \frac{|W^T s_B W|}{|W^T s_W W|}$. The between-class scatter matrix is given by $S_B = \sum_{I=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T$ and the within-class scatter matrix is defined by $S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (X_k - \mu_i)(X_k - \mu_i)^T$, where µ is the mean image of all images, μ_i is the mean image of class X_i and N_i is the number of samples in class X_i. Feature extraction based on PCA and LDA uses a linear projection subspace, which is not optimal when dealing with nonlinear information such as exists in biometric images associated, for example, with faces having different expressions and poses or palmprints with different effects of illumination effect and ageing. To overcome this limitation, a kernel method is proposed to extract nonlinear information which exists in biometric images by using kernel functions such as Gaussian function or Polynomial function where the interactions between elements of the features occur only through dot products [55, 56]. The kernel method is based on the transformation from an input space into a high dimensional feature space. Given the nonlinear mapping function Φ, the input data space R^n can be transformed into a new high dimensional feature space F as $\Phi: R^n \longrightarrow F, \mathrm{x} \longmapsto \Phi(x)$. The new feature space F could have an arbitrarily large, possibly infinite dimensionality and thus the direct computation of linear projections such as in PCA and LDA is very computationally costly. However a method called the kernel trick can be used to compute dot products between mapped patterns using kernel representations of the form $k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \bullet \Phi(\mathbf{y}))$. These allow the value to be computed of a dot product in F without the need to use the mapping function Φ. Several linear projection methods have been proposed for use with the kernel trick in order to extract complex and nonlinear information, such as the kernel PCA [57, 58] and kernel Fishers [59, 60].

Kernel PCA has been implemented to extract the information for the digital recognition of handwriting and the results show that it is able to extract nonlinear features and achieve high recognition accuracy [55]. Mika [61] applied a kernel trick to LDA and the experimental results of the proposed kernel LDA (KLDA) showed KLDA is able to extract the high discrimination features in the feature space. Yang [62] used kernel Eigenface and kernel Fisherface to extract global facial features for the recognition process, and the result showed that kernel methods give superior results compared to Eigenface and Fisherface. Several nonlinear techniques based on the kernel trick have been designed to extract features for palmprint images [63, 64]. Although higher recognition rates have been achieved using the kernel method, there are still many problems such as the image orientations, variations in pose and illumination in biometric images. To overcome these problems a feature design procedure is computed as a pre-processing method to the original image. This approach is the unsupervised method which does not required a training process.

The feature design procedure generate new features from 2D biometric images using signal processing tools to convert raw data into a frequency domain which is believed to enhance the information in the image. The analysis of biometric images in the frequency domain is a commonly used method in image representation and recognition, and some work has used frequency domain techniques to extract information for facial recognition. Hafed [65] proposed the transformation of features of facial images to the frequency domain using a Discrete Cosine Transform (DCT), and performance was evaluated in terms of the recognition rates.

In this method, the whole face image is transformed to the frequency domain and only low frequency coefficients are preserved for classification. The analysis showed that the DCT achieved equal performance of the Karhunen-Loeve (KL) transform in the compression of facial information. By manually selecting the frequency band of the DCT coefficients, this recognition method achieved similar recognition accuracy with the Eigenface method. Jing [66] uses DCT to extract the holistic features of face and palmprint images and then proposed a method called two dimensional separability judgements to select the DCT frequency band. The DCT coefficients in a selected frequency band of the whole image are learned using linear discrimination technique to reduce dimension of feature vector, and a nearest neighbour classifier is used for classification. Li et al [67] applied the Fourier transform to extract information from the spatial domain to the frequency domain in palmprint images. In order to perform the classification task, the features in the frequency domain are compared with the templates stored in a database. Lai [68] applied the Fourier transform to the wavelet transform of face images in order to generate a feature vector that invariant to translation, scale and rotation on the plane. A new representation called the holistic Fourier invariant features is computed from the wavelet sub band which corresponds to the low frequency components of both vertical and horizontal directions of the original image. Another method to represent features based on the Fourier transform is the use of phase information instead of values of magnitude, as suggested by Meraoumia [69]. The phase information given by the 2D-DCT of the palmprint image is used to calculate similarity to the templates stored in a database by using phase correlation function, where a sharp peak is achieved for a similar image. In order to increase the performance of the recognition process, the similarity measure from the phase information of the 2D-

DFT and feature representation using local 2D-DCT are fused at the matching score level.

Another approach to the representation of the biometric feature vector is to use a discrete wavelet transform where a multiresolution analysis of the signal is performed with localization in both time and frequency. Wavelet coefficients representing the contribution of wavelets in the function at different scales and orientations are computed by convolving the image with wavelet kernels such as the Haar and Daubechies wavelets. The wavelet decomposition technique is able to extract the intrinsic features and reduce the dimensions of data in the pixel images by dividing the original image into several sub images using low pass and high pass filters. Son [70, 71] proposed a feature fusion method for face and iris multimodal biometrics by concatenating the feature vectors extracted from the three levels of wavelet decompositions. Concatenation is performed using the lowest frequency components of the face and iris images which contain most of the information about the image. To further reduce the dimensionality of the fused feature vector and to enhance the discrimination power of the fused feature vectors, a linear projection method based on Direct Linear Discrimination Analysis (DLDA) is applied to the fused feature vector. Noore [72] fused multiple low frequency components of wavelet images from the four modalities of face, fingerprint, signature and iris to form a single composite multimodal biometric image. In this method, three levels of wavelet decomposition were computed separately for the different biometric traits and the low frequency components of each modality were then combined to form a fused feature image. Wang [73] also proposed feature fusion combining different biometric images from palmprints and palm veins to form a single fused image based on Mallat’s wavelet [74]. In order to preserve the ridges and veins in the palmprint and palm vein images, fusion was performed based on the wavelet maxima detection rule and then the wavelet maxima of both images were combined to produce fused wavelet maxima. The fused image can be reconstructed using on inverse wavelet transform to give a fused wavelet maxima.

Recently, the Gabor wavelet feature has been shown to be effective in feature level fusion. The Gabor wavelet representation of a biometric image is a convolution of the image with a family of Gabor kernels which are similar to the 2D receptive field profiles of the mammalian cortical simple cells which exhibit desirable characteristics of spatial locality and orientation selectivity and are optimally localized in the space and frequency domains. Yang [75] used the Gabor wavelet to extract the information from fingerprint and finger-vein data and then fused it at the feature level. The new fused feature vector is constructed based on supervised nonlinear correlation analysis which is a method used to find a relationship between two sets of feature vectors. Raghavendra [18] concatenated the features of face and palmprint data in a Log Gabor domain in order to integrate the information at the feature level. 2D Log Gabor face and palmprint images are rearranged into the one dimensional feature vectors and then serially combined to produce a high dimensional fused feature vector. A feature selection method based on Particle Swarm Optimization (PSO) is used to select the most dominant coefficients and reduce redundancy in the fused feature vector in order to project it into a kernel discriminative space using Kernel Direct Discriminant Analysis (KDDA). Another method of feature fusion by concatenating Gabor face and palmprint images was proposed by Jing [32]. In their method, Gabor images of faces and palmprints are vertically concatenated to produce a high dimensional feature vector. To reduce the dimensions of the fused feature vectors and capture nonlinear information, a kernel subspace method based on the discriminative common vector is used to project it into a new subspace. Yao [33] fused the information using PCA features extracted from Gabor face and palmprint images. The fused feature vectors are constructed by serially combining the face and palmprint feature vectors. Fu [34] also proposed to use the Gabor transform to extract important features in the face and palmprint, and then the high dimensional Gabor features are reduced using ICA. In order to combine the low dimensional features, a specific weighting value is introduced to the ICA features prior to combining then using concatenation.

3.3 Local Feature Representation

In contrast to the holistic features where an entire image is used to compute the feature representation, local feature extraction methods extract the information from diverse levels of locality and quantify them precisely. The general idea of local feature extraction technique is to divide the image into several parts and then the information is extracted each part individually. Another method is to locate several components of features such as the eye, nose and mouth in a facial image, and then classify them using several matching methods. Anila [76] proposed using the Gabor transform with local regions of a face image in order to construct an independent feature vector. The analysis shows that fusing the match scores from each local region gives better results compared to concatenating the local features and classifying them by using a single classifier. However, this method requires frontal face images with less variation in poses and the complexity increases due to the use of a multiple classifier for each local region. Sanderson [77] proposed representing local features by using modified 2D-DCT coefficients computed in 8 x 8 sub block windows. A modified DCT is computed using the DCT coefficients by introducing new coefficients for the first three original DCT coefficients where the new coefficients are computed from neighbouring blocks. By removing the first three DCT coefficients, it is believed that robustness to illumination changes increases, but a significant amount of discrimination information may be lost. Thus, to overcome the performance loss, it was suggested that the first three coefficients should be replaced with their proposed delta coefficients. Classification of the face image is conducting by using maximum likelihood values, where the GMM is used to learn statistical information concerning the feature vector of the specific class. In the same year, Cardinaux [78] proposed an extended local feature vector from a block based DCT which integrates a degree of spatial relations via embedded positional information from each sub block window. The embedded information existing in the local features is captured by using GMM and a pseudo 2D Hidden Markov Model (HMM) where the latter achieved better performance. However the best trade-off in terms of computational time, robustness and discrimination performance is achieved by extended local features using GMM.

Extracting local features can also be accomplished by choosing certain parts of a biometric image instead of dividing the image into several sub blocks as above. Lucey [79] proposed to use several shapes in face to extract local features, where regions of a face image such as the eye, nose, mouth and eyebrow are modelled in a single Gaussian distribution. Their results showed that using eight Gaussian distributions to model eight facial regions achieved the lowest EER compared to methods that divided the image into several sub block windows. However, the experiment was conducted using a sub block window of 16x16 pixels, which is larger than the best result reported by Sanderson [77] that using 8x8 pixels with a 50% pixel overlap. Liu [80] proposed a new flexible local representation called the

X-Y patch, and this method alleviates the requirement for pixel wise alignment and is robust to small spatial misalignment problems. The X-Y patch is a joint appearance and shape descriptor where the x-y coordinates are included in the patch. The information in the patch window is extracted by using the 2D DCT transform and then a new feature vector is constructed by concatenating the x and y coordinates with DCT coefficients. Statistical information in the new feature vector is then learned using GMM.

Another interesting method to extract and represent local features in the face [81] and palmprint [82] image use the local binary pattern (LBP), which is one of the types of image texture descriptor used in texture analysis. In this method, the image is divided into several regions from which the LBP feature distributions are extracted. The LBP method was originally used for texture description, and assigns a label to every pixel of an image by comparing the 3x3 neighbourhood of each pixel with a centre pixel value producing a result in binary number. Then, the concatenation histogram of the labels is used as a feature vector. Recently, Zhang [83] apply LBP in the magnitude of Gabor image to design a new object descriptor called the local Gabor Binary Pattern Histogram Sequence (LGBPHS). The image is convolved with a set of Gabor filters and each pixel value in the Gabor transform image is compared with its neighbours. The transformation image is then divided into non-overlapping rectangular regions and the histogram of each local region is concatenated to represent the original image. Zhang [84] also proposed the use of a Gabor phase to represent the local features based on a new local pattern called histogram of Gabor phase pattern (HGPP), which is able to encode the phase variations. The local Gabor phase pattern (LGPP) is able to encode the local neighbourhood variations by using a new local pattern called a local XOR pattern

(LXP) operator. As with the LBP extracted from Gabor magnitude, the transformation pattern of the LXP is divided into non-overlapping rectangular regions and then the spatial histograms are extracted and concatenated to represent the original image. Xie [85] proposed a Learned Local Gabor Patterns (LLGP) for face representation, which is a method based on Gabor features which defines cliques of features which appear frequently in Gabor features as the basic patterns. Compared to LBP, where the patterns are predefined and fixed, the new local patterns are learned from the patch set, which is constructed by sampling patches from the Gabor filtered image. In order to construct the representation of the image, each facial image is converted into multiple pattern maps and the histograms of each block image are concatenated together.

3.4 Multi-resolutions Gabor Filter

There has been proliferation of Gabor transform used as a feature extraction and image representation method in the biometric recognition system. Local feature extraction using multiresolution analysis involves the convolution of the image with a set of Gabor filters computed for a specific region of the image. The Gabor transform or Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of mammalian cortical simple cells, are widely used in image analysis due to their biological relevance and computational properties. The multiresolution structure in the frequency domain is similar to that of the wavelets, but without the orthogonal properties. Gabor features are considered to span a frame that has many beneficial properties such as spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. Gabor wavelets can be divided into an elliptical Gaussian and a complex plane wave defined as follows [86-88].

$$\Psi_{\mu,v} = \frac{\|k_{\mu,v}\|^2}{\sigma^2} e^{\left(-\|k_{\mu,v}\|^2\|z\|^2/2\sigma^2\right)}\left[e^{ik_{\mu,v}z} - e^{-\sigma^2/2}\right] \tag{3.1}$$

where μ and v are the orientation and scale of the Gabor kernels, $z = (x, y)$, and ||•|| denotes the norm operator. The wave vector $k_{\mu,v}$ is defined as follows:

$$k_{\mu,v} = k_v e^{i\emptyset_u} \tag{3.2}$$

where $k_v = k_{max}/f^v$and $\emptyset_\mu = \pi\mu/8$. k_{max} is the maximum frequency, and *f* is the filter tuning frequency, and the bandwidth of the filter is measured by variance σ corresponding to the two perpendicular axes of the Gaussian. The Gabor kernels are similar since they are generated from one filter with different scaling and orientation. Each kernel is a product of a Gaussian envelope and a complex sinusoidal wave. The Gabor wavelet representation of an image is the convolution of the image with a set of Gabor kernels as defined in Eq. (3.3). Let *I*(x,y) be a grayscale of an image, then convolution of an image *I* and the Gabor kernel $\Psi_{\mu,v}$ is defined as follows:

$$O_{u,v}(z) = I(z) * \Psi_{\mu,v}(z) \tag{3.3}$$

where $z = (x, y)$, * denotes the convolutive operator, and $O_{u,v}(z)$ is the convolution result corresponding to the Gabor kernel at orientation μ and scale *v*. By applying the convolutive theorem, each $O_{u,v}(z)$ can be derived from Eq. (3.4) using the Fast Fourier Transform (FFT):

$$F\{O_{u,v}(z)\} = F\{I(z)\}F\{\Psi_{u,v}(z)\} \tag{3.4}$$

and the Inverse Fast Fourier Transform (IFFT) is:

$$O_{u,v}(z) = F^{-1}\left\{F\{I(z)\}F\{\Psi_{\mu,v}(z)\}\right\} \tag{3.5}$$

Several parameters need to be considered in order to produce an image features which is related to the numbers and values of the Gabor filter frequencies and orientations of the Gabor kernels. The more frequencies and orientations are used, the better is the representation power of the Gabor features. By increasing these numbers, the shift sensitivity increases, thus allowing a more accurate determination of image texture information. However, the representation power is also affected by the effective areas of Gabor filters controlled by the bandwidth parameters. Generally, the bandwidth value can be set to 1.0 and the initial numbers of filter are 8 orientations and 4 scales, producing 32 sets of Gabor filter images as shown in Figure 3.1. Multiplying biometric image with a set of Gabor filter produces 32 sets of Gabor images as shown in Figure 3.2. The effect of changing parameter values can be evaluated experimentally based on the discrimination power of face and palmprint images. It is not necessary to compute features at all locations due to the enormous overlapping of the filters when they are sufficiently spaced.

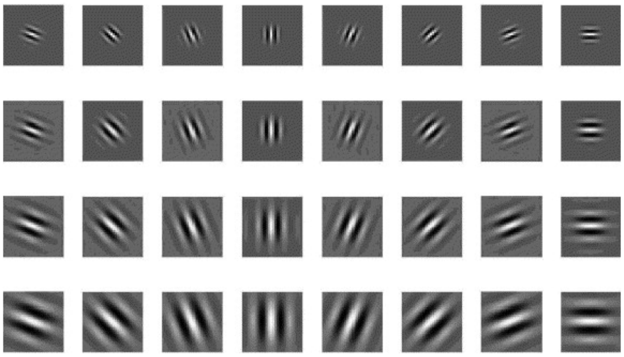


Figure 3.1: Gabor filter image with different orientations and scales.

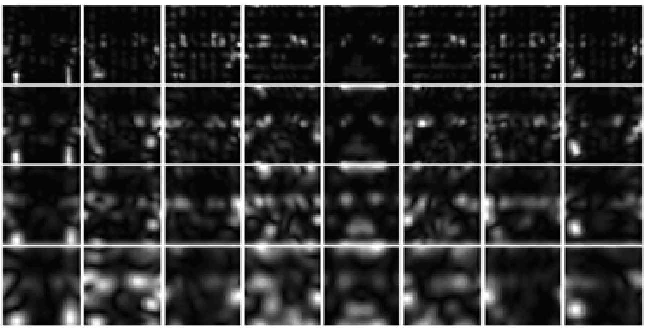


Figure 3.2: Gabor image of face modality.

3.5 2D DCT Feature Extraction

The preceding sections, illustrated the feature extraction method using multiresolution Gabor features, which concentrate primarily on feature analysis using different orientations and scales. The main purpose of this section is to transform images from the Gabor feature space to the frequency domain using DCT, where an image is decomposed into a combination of various and correlated frequency components. The advantage of DCT is that it is able to extract the features in the frequency domain to encode different texture details that are not directly accessible in the spatial domain [88, 90]. In the method proposed here, DCT feature extraction consists of two steps. Firstly the DCT is calculated for each block of 2D Gabor images, and secondly some of the coefficients are remained to construct a feature vector. DCT basis function in 8x8 windows is shown in Figure 3.4. Given M x N pixels image, where each image corresponds to a 2D Gabor feature, DCT coefficients are calculated as follows [91]:

$$F(u,v) = \frac{1}{\sqrt{MN}}\alpha(u)\beta(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) \times \cos\left(\frac{(2x+1)u\pi}{2M}\right) \times \cos\left(\frac{(2y+1)v\pi}{2N}\right) \quad u = 0,1,\ldots.M \quad v = 0,1,\ldots.N \tag{3.6}$$

where $\alpha(u)$ is defined as:

$$\alpha(\omega) = \begin{cases} \frac{1}{\sqrt{2}} & \omega = 0 \\ 1 & otherwise \end{cases} \qquad (3.7)$$

f(x,y) is the image texture and in our case, this consists of Gabor features. The DCT coefficient is given by *F(u,v)* and in general its values are divided into three bands of low, middle and high frequencies. The low frequency band correlates with illumination with an average value in the image. High frequency represents noise and small variations in details, while the middle frequency coefficients contain useful information in the construction of the basic structure of an image which is believed to have high discriminant features. Different DCT coefficients correspond to different spatial information. The upper left coefficient in the DCT domain encodes most of the energy in the image which corresponds to low frequency information as shown in Figure 3.3. When the DCT subsets are expanded towards the right and down words, more high frequency information is captured, and thus more details of an image are displayed in space. Several methods have been proposed to effectively choose the DCT coefficient in the DCT domain; namely, zig-zag, 2D separability judgements, energy probabilities and polynomial coefficients derived from the DCT. The present utilizes the zig-zag approach which is found to produce superior results when dealing with Gabor features.

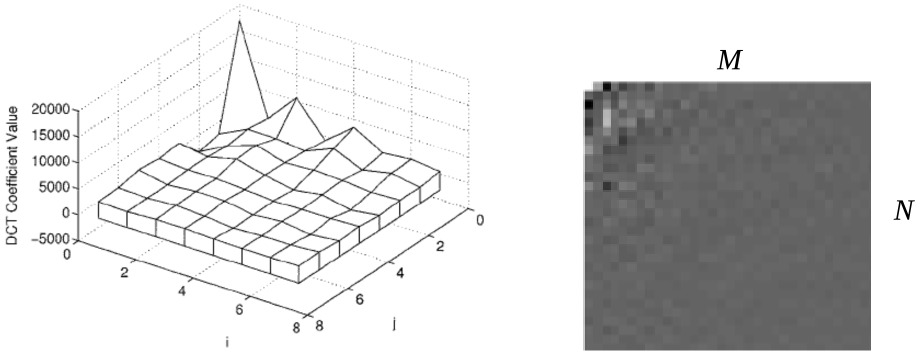


Figure 3.3: DCT coefficient showing most of the low frequency energy components concentrated in the top left corner of the image.

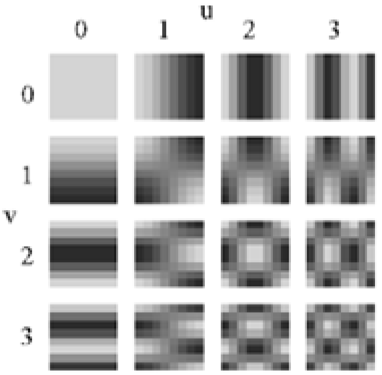


Figure 3.4: 2D basis function of DCT transform in 8x8 pixels sub block window.

3.6 Novel Compact Local Feature Representation

As mentioned in the Chapter 2, fusion at the feature level is believed to give better results by incorporating feature vectors which contain richer information which is nonlinear and captures different statistical properties. Therefore, a new feature extraction framework is proposed here to extract the compact local features that can be used in the matrix interleaved feature fusion framework proposed in Chapter 4. The proposed method focuses on the development of a new framework for feature extraction methods based on local frequency bands extracted at different scales and orientations. This method is able to extract important information in different orientations and scales, and is thus able to capture nonlinear information in face and palmprint images. In order to eliminate illumination and pose invariant problems, texture analysis in a sub-block window is utilized as suggested in [92, 93] but our method no longer implements window overlap in order to reduce the number of features. It is also shown in the experimental result that the verification performance of the proposed compact local features is higher than that of existing local feature extraction methods using window overlap. Furthermore, the compact representation of low frequency components is able to eliminate high frequency components which normally contain noise and do not offer any benefits for discrimination power. In fact, most low frequency components of face and palmprint images carry high discrimination power, such as the shapes of the mouth, nose, eyes and principle lines and wrinkles [66], [94]. The framework of novel compact feature extraction is shown in the front processing part in Figure 3.5. The new method of feature fusion based on matrix interleaved is discussed in Chapter 4 and the classification method based on the background model is discussed in Chapter 5. The new feature extraction method makes use of non-overlapping sub block windows to extract local features which contain independent information on a specific region of the image. Each sub block texture is pre-processed with a set of Gabor filters with different scales and orientations to capture the texture of an image which has different structures. A compact representation of the feature vector is computed using the Discrete Cosine Transform (DCT) for each Gabor image. By using the DCT, most of low frequency coefficients can be preserved by eliminating the high frequency coefficients using the zig-zag approach.

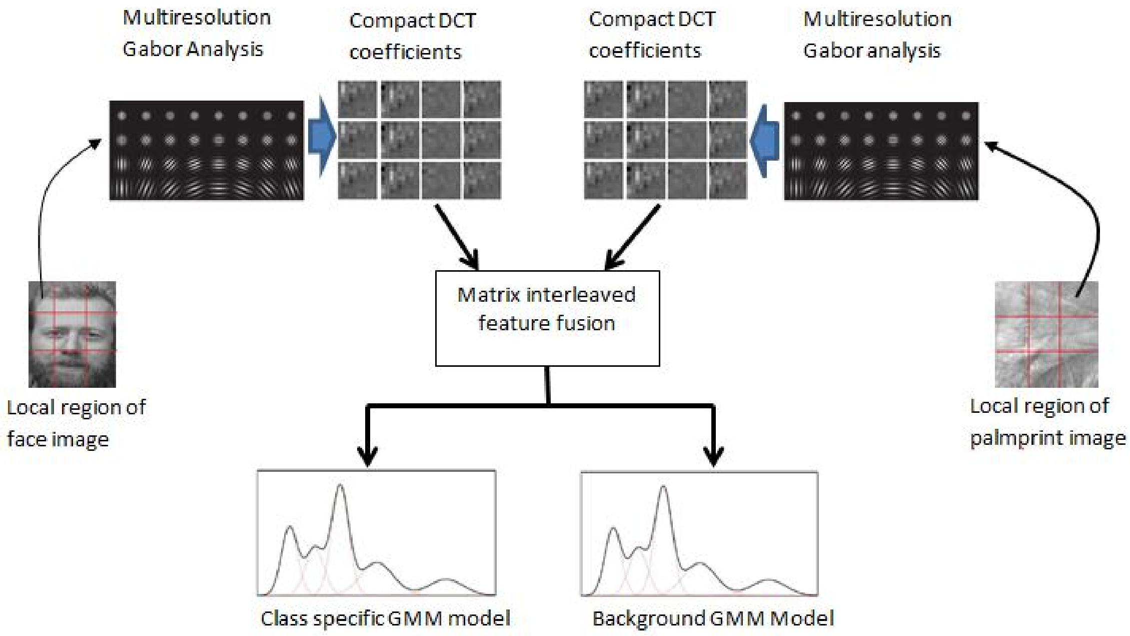

Multiresolution Gabor Analysis
Compact DCT coefficients
Compact DCT coefficients
Multiresolution Gabor analysis
Matrix interleaved feature fusion
Local region of face image
Local region of palmprint image
Class specific GMM model
Background GMM Model

Figure 3.5: Overall block diagram of the proposed method.

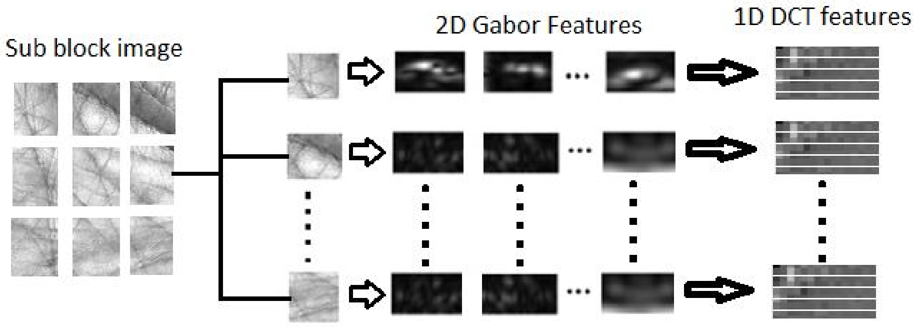


Figure 3.6: Compact local energy representation of local features.

3.7 Patch-based Compact Image Representation

Information fusion at the feature level is believed to give better performance for classification due to the utilisation of complementary and the most discriminative features existing in both modalities. However, the success of feature fusion depends on the information contained in both modalities. Thus, it is believed that, by using an effective method to extract the most important features, this could be beneficial to fusion itself. This section proposes patch based feature representation for feature level fusion utilising the frequency domain, which is able to extract multiresolution texture images. The proposed methods have three main beneficial features. Firstly the image is analysed in terms of local features in a process which can handle pose and illumination changes. Face and palmprint images are subdivided into several blocks and each block is treated independently. Independent features in each block are then used to estimate the model parameter. Secondly, each sub block image is analysed using Gabor kernels which have different orientations and scales. By using appropriate number of Gabor parameters, important information can be extracted from the spatial images, including wrinkles from the texture of the palm surface with unique orientations and the shape of face in which the nose, eyes and mouth all have different scales. Thirdly, most of the energy in the Gabor features is encoded by using DCT features, and only the energy with the most discriminative information is preserved. This technique not only reduces the dimensions of the feature vector but also helps in the separation of feature vectors during the classification process since high frequency components have already been removed. Given an image with i regions $i = 1,2, \dots 9$, as shown in Figure 3.6, the Gabor response for each region is given by $G = \{R^1, R^2, \dots, R^9\}$, where each region R^i consists of 32 sets of Gabor images as shown in Eq. (3.8)-(3.10):

$$R^1 = \begin{pmatrix} \Psi(x_0, y_0; f_0, \theta_0) & \Psi(x_0, y_0; f_0, \theta_1) & \bullet\bullet\bullet & \Psi(x_0, y_0; f_0, \theta_7) \\ \Psi(x_0, y_0; f_1, \theta_0) & \Psi(x_0, y_0; f_1, \theta_1) & \bullet\bullet\bullet & \Psi(x_0, y_0; f_1, \theta_7) \\ & \bullet & & \\ & \bullet & & \\ \Psi(x_0, y_0; f_3, \theta_0) & \Psi(x_0, y_0; f_3, \theta_1) & \bullet\bullet\bullet & \Psi(x_0, y_0; f_3, \theta_7) \end{pmatrix} \tag{3.8}$$

$$R^2 = \begin{pmatrix} \Psi(x_0,y_1;f_0,\theta_0) & \Psi(x_0,y_1;f_0,\theta_1) & \bullet\ \bullet\ \bullet & \Psi(x_0,y_1;f_0,\theta_7) \\ \Psi(x_0,y_1;f_1,\theta_0) & \Psi(x_0,y_1;f_1,\theta_1) & \bullet\ \bullet\ \bullet & \Psi(x_0,y_1;f_1,\theta_7) \\ & & \bullet & \\ & & \bullet & \\ \Psi(x_0,y_1;f_3,\theta_0) & \Psi(x_0,y_1;f_3,\theta_1) & \bullet\ \bullet\ \bullet & \Psi(x_0,y_1;f_3,\theta_7) \end{pmatrix} \tag{3.9}$$

$$R^9 = \begin{pmatrix} \Psi(x_3,y_3;f_0,\theta_0) & \Psi(x_3,y_3;f_0,\theta_1) & \bullet\ \bullet\ \bullet & \Psi(x_3,y_3;f_0,\theta_7) \\ \Psi(x_3,y_3;f_1,\theta_0) & \Psi(x_3,y_3;f_1,\theta_1) & \bullet\ \bullet\ \bullet & \Psi(x_3,y_3;f_1,\theta_7) \\ & & \bullet & \\ & & \bullet & \\ \Psi(x_3,y_3;f_3,\theta_0) & \Psi(x_3,y_3;f_3,\theta_1) & \bullet\ \bullet\ \bullet & \Psi(x_3,y_3;f_3,\theta_7) \end{pmatrix} \tag{3.10}$$

$$\left.\begin{array}{l} \left.\begin{array}{l} DCT_1^{f_0\theta_0} = \left[F_{x_0y_0}^{f_0\theta_0}(3), F_{x_0y_0}^{f_0\theta_0}(4), F_{x_0y_0}^{f_0\theta_0}(5), \ldots., F_{x_0y_0}^{f_0\theta_0}(15)\right] \\ DCT_2^{f_0\theta_1} = \left[F_{x_0y_0}^{f_0\theta_1}(3), F_{x_0y_0}^{f_0\theta_1}(4), F_{x_0y_0}^{f_0\theta_1}(5), \ldots., F_{x_0y_0}^{f_0\theta_1}(15)\right] \\ \\ DCT_{32}^{f_4\theta_8} = \left[F_{x_0y_0}^{f_3\theta_7}(3), F_{x_0y_0}^{f_3\theta_7}(4), F_{x_0y_0}^{f_3\theta_7}(5), \ldots., F_{x_0y_0}^{f_3\theta_7}(15)\right] \end{array}\right\} \begin{array}{l}\text{DCT coefficient} \\ \text{in a sub block} \\ \text{window}\end{array} \\ \\ DCT_{32}^{f_4\theta_8} = \left[F_{x_2y_2}^{f_3\theta_7}(3), F_{x_2y_2}^{f_3\theta_7}(4), F_{x_2y_2}^{f_3\theta_7}(5), \ldots., F_{x_2y_2}^{f_3\theta_7}(15)\right] \\ \\ DCT_{32}^{f_4\theta_8} = \left[F_{x_3y_3}^{f_3\theta_7}(3), F_{x_3y_3}^{f_3\theta_7}(4), F_{x_3y_3}^{f_3\theta_7}(5), \ldots., F_{x_3y_3}^{f_3\theta_7}(15)\right] \end{array}\right\} \begin{array}{l}\text{DCT} \\ \text{coefficient} \\ \text{in the whole} \\ \text{image}\end{array}$$

where x_i and y_i represent the sub block image in grey scale and f_i and θ_i are the frequency and orientation of the Gabor image. The compact representation of the feature vector is computed by using the DCT transform and the most discrimination coefficient is selected by implementing a zig-zag approach. In the experimental work, it is demonstrated that using the first 15 coefficients is sufficient to produce the best performance, due to the removal of high frequency components and noise in the DCT coefficients when the number of coefficients is greater than 15. The compact frequency feature vector for each sub block image has 32x15 feature vectors denoted as $M_{32\text{x}15}$. Each two dimensional image has 9 sub-block windows, and thus vertical concatenation produces $M_{32\text{x}15\text{x}9} = [M_{32\text{x}15} \cup M_{32\text{x}15} \cup M_{32\text{x}15} \cup \ldots M_{32\text{x}15}]$ feature vectors. This preserves contains important information from both modalities, such as statistical properties and nonlinear information. It is believed that the fusion of this information will increase the discriminative power and have benefits for model estimation due to the increasing number of independent data points.

Feature fusion can be constructed by using concatenation methods as implemented in [16, 95-97]. Instead of using the concatenation process, a new framework of feature fusion based on matrix interleaved as proposed in Chapter 4 can be applied to increase the statistical information in the fuse feature vector. In our method, compact local features of both modalities are concatenated and interleaved both of them to form a new feature vector. The proposed feature fusion method has the advantage of using different scales and orientations in the local feature vector whereas the method in [78, 92, 93] depends on local features extracted from 50% overlapping windows on the spatial image, where the relationship of each sub-block window is important to gain better discriminative power. The proposed method produces sufficiently independent feature vectors, which is important in estimating model parameters using the GMM. Previous studies have proven GMM to be an effective tool to capture the underlying statistical information in the single modal biometric feature vectors based on face and voice traits [99 - 101].

3.8 Experimental Analysis

In this section, the effective of the local feature extraction method is tested using a FERET face dataset [102, 103] and a PolyU palmprint dataset [104, 105] by virtually combining them to form multimodal biometrics. Both datasets have been released to the public and are widely employed as a benchmark in biometrics recognition analysis. This combination is acceptable since for each person these two modalities can be considered to be independent. The experimental setting is briefly described below and the performance analysis evaluated for verification and recognition tasks. Then, the proposed method is evaluated with different parameters in the multiresolution Gabor features. The size of sub-block windows is also examined to produce the best feature representation as well as to conduct a comparison with several existing local feature methods, some of which suggest the use of window overlapping in order to gain information about relationships between neighbouring block windows. Following that, the size of pixel overlap that gives the best feature representation is also examined. The number of DCT coefficients is also examined in order to find the best number that can be used to represent the face and palmprint images. Finally, a detailed comparison is made between the approach suggested here and some existing multimodal approaches using these two datasets.

The FERET database is widely used as a benchmark to measure performance in facial recognition algorithms. The proposed method randomly selects face images from fa, fb, rc and rd which have different levels of diversity in terms of gender, pose, illumination, hairstyle, expression and ageing. The image selection is acquired without any restrictions imposed on facial expression and with at least two frontal images shot at different times during the same photographic session. In this experiment, 1000 frontal face images corresponding to 200 subjects are randomly selected such that each subject has six images of size 256 x 384 pixels with 256 gray scale values. From six images per subject, three images are randomly selected for training and three for testing. The centre of the eyes in an image is manually detected, and then scaling transformations align this position to predefined locations. The facial region is manually cropped to the size of 128 x 128 pixels and further normalized to zero mean and unit variance. Figure 3.7 shows some example of the facial regions used in the analysis which have been resized to 100 x 90 pixels.

PolyU datasets are publicly available and widely used as a benchmark to evaluate palmprint recognition performance. The palm images are collected in two separate sessions with an interval of around two months between them. Each subject consists of left palm and right palm images taken under different light sources and with a different focus of a CCD camera. 200 subjects are randomly selected and each subject consists of six images. Three randomly selected images are used for training and three for testing. The original size of an image is 384 x 284 pixels and the image is pre-processed to obtain the region of interest as suggested in [104]. Then, the image is normalized to zero mean and unit variance. Figure 3.7 shows six palmprint images of one subject taken in two sessions. In order to construct the virtual multimodal biometric datasets of the face and palmprint, each virtual person is associated with six samples of face and palmprint produced randomly from the face and palmprint samples of two persons in the respective databases FERET and polyU. Thus, the resulting virtual multimodal biometric dataset consists of 200 subjects such that each subject has six samples of images.

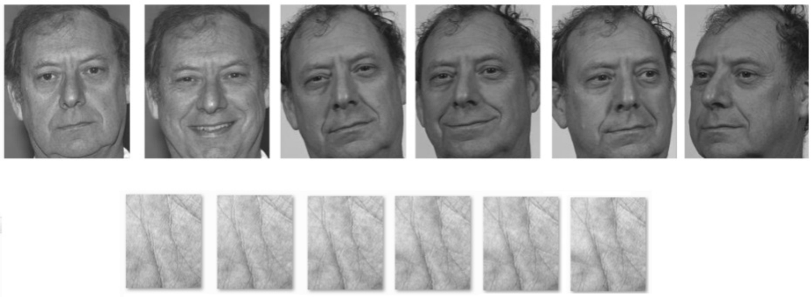


Figure 3.7: Samples of face and palmprint images for one subject.

3.8.1 Analysis of Different Scales and Orientations

The number of scales and orientations of Gabor filters is important in capturing the information existing in the texture of images. By using fewer orientations the classification performance may become degraded as insufficient features will be extracted from the texture of the images. Meanwhile, using too many orientations may increase data redundancy and noise. The number of scales also influences feature representation, and thus an appropriate number must be used for each patch window to have more access to features with high discriminative power. In order to demonstrate the effectiveness of multiple scales and orientations of Gabor features in extracting important information from the patch image, eight sets of orientations are used separately and each orientation is analysed with four different scales. Set 1 consists of $I_{0,90,180,270}$, set 2 consists of $I_{0,45,90,\ldots 360}$ and set 3 consists of $I_{0,15,30,\ldots 360}$. All of these orientations are combined with four different scales. Different numbers of scales gives different sizes of Gaussian envelop, and therefore the use of the proper size of scales is crucial. The method used previously [106, 107] suggests four scales and eight orientations to extract the face images, which will thus produce 32 sets of Gabor images. In these methods, 32 sets of Gabor filters are used to extract global features which cover the whole of facial images. The proposed method in this study focuses on local features in small patch windows as well as different features from two modalities. Using the same setting of Gabor filters as used in unimodal biometrics may produce redundant features whose analysis might be time consuming and not useful in training the GMM.

In order to examine the best number of orientations and scales which produce superior performance, Figure 3.8 shows the analysis of EER for all scales and orientations. The lowest EER is 0.6% achieved using 8 orientations with 3 and 4 scales. However, when the number of orientations is increased to 12, the EER will increase by 0.1% for 3 and 4 scales. This shows that using many orientations will not give the advantage to the extracted features. The performance is degraded when 4 scales and 12 orientations are used to extract the sub block image. This is due to redundant information and noise being extracted, which will give less discrimination power. However, using fewer of scales and orientations, such as 2 scales with 4 orientations, produces higher EER of 1.2% due to the fact that less information can be extracted from the image. In the present analysis, the best EER could be achieved by using 3 scales and 8 orientations. The analysis of recognition rate is shown in Figure 3.9, and the highest recognition rate of 97% was achieved using 3 scales and 8 orientations.

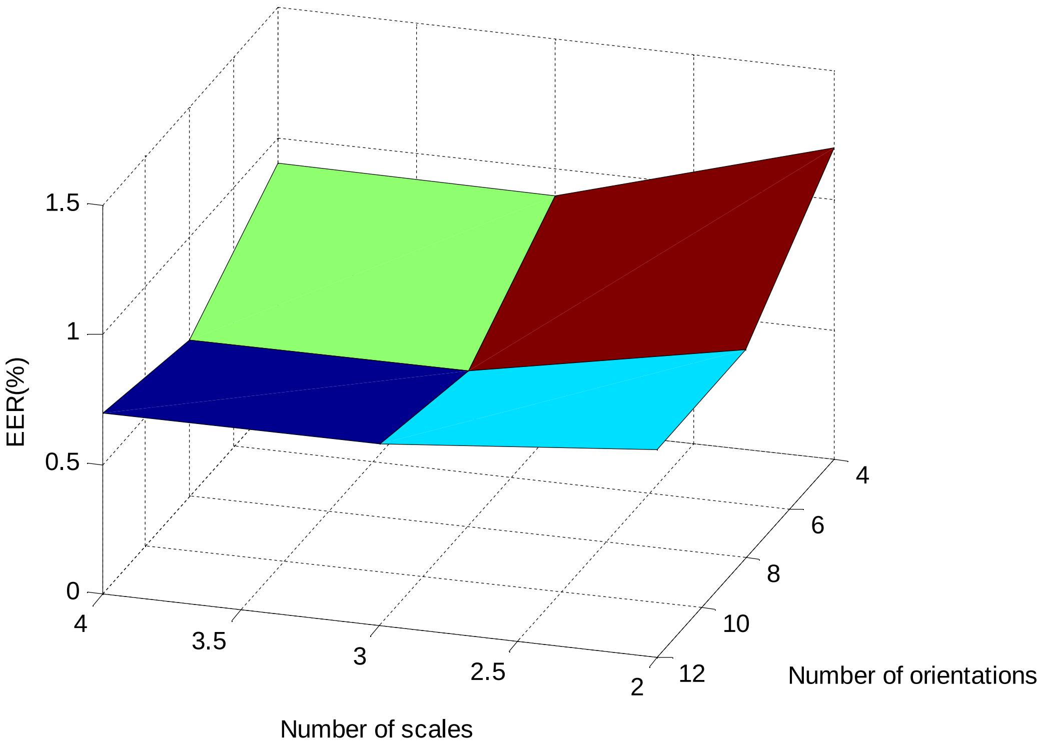


Figure 3.8: Performance analysis of EER (%) with different number of scales and orientations of Gabor filter.

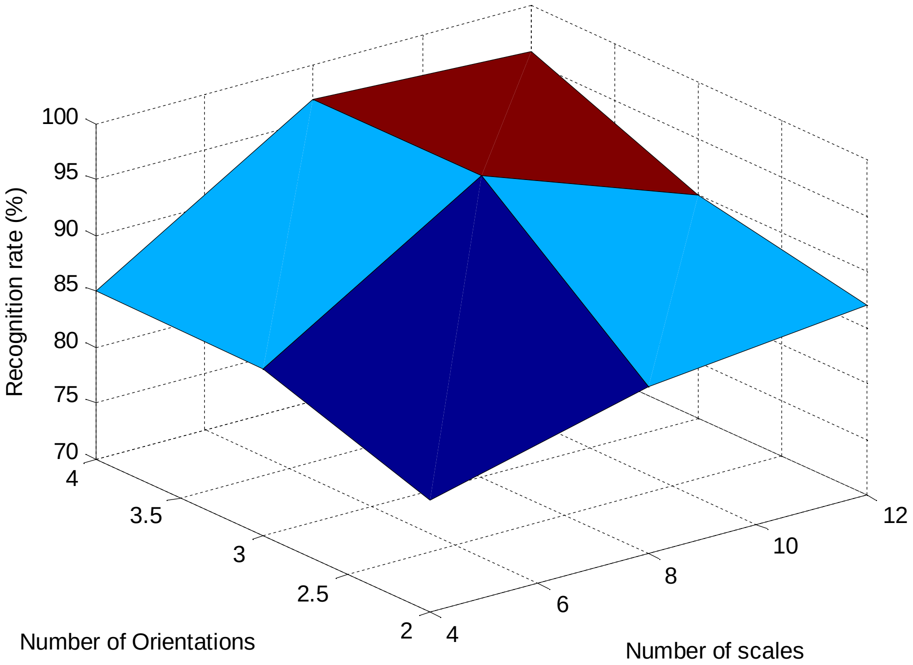


Figure 3.9: Recognition rate of different scales and orientations.

3.8.2 The Effect of Size Patch Window

Local features are extracted from several sub-block windows from an image which are independent of each other. This analysis examines the effectiveness of the size of patch windows in yielding the best results. Most local features extracted from face images are analysed in several sub-windows, and the method used in [77] suggests the use 8x8 pixel size whereas in [92] suggests the use 50% pixel overlaps to incorporate the relationship between neighbouring blocks. In the analysis here, the size of window was varied from 40x40, 20x20, 16x16, 10x10 and 8x8 pixels, and multiresolution Gabor images computed from the original image in each patch. The existing methods of local features use the DCT energy from each patch window as a feature vector for the training process. However, for the proposed method it is believed that more information can be gained by using multiresolution Gabor analysis followed by computation of DCT coefficients from the patch window. Using non-overlapping windows for each patch will produce strongly independent features, and thus there will be less effect of illumination and pose invariants. We also demonstrated that almost similar performance can be obtained when using overlapping windows. This shows that multiresolution Gabor analysis in patch window can provide strong independent features where the relationships between blocks can be ignored without sacrificing good results. Figure 3.10 shows the analysis of EER with different sizes of sub block windows and different percentages of pixel overlapping. The results show that 40 x 40 pixels sub block windows give the lowest EER for all degrees of pixel overlap, and the lowest EER is 0.6% achieved by using non overlapping windows. This shows that the proposed method does not depend on pixel overlapping due to the rich information which can be obtained in the features extracted from the proposed framework here that uses multiresolution analysis and low frequency information.

In the following analysis, non-overlapping windows are used for all sizes of sub block windows due to the encouraging results given in Figure 3.10. Figure 3.11 shows the performance in terms verification rate given as Receiver Operating Characteristic (ROC) curve with different sizes of sub block windows. The best result is 98% GAR at 0.1% FAR achieved by using 40x40 pixel sub block windows. The performance is degraded when the size of sub block window is smaller than 40x40 pixels. Using patch windows which are too small degrades the performance since information can be lost when convolving with the Gabor filter. This might be because some of the important parts of faces and palmprints are bigger than sub-block windows, such as the nose, eyes and palm wrinkles. Using too large sub block window would involve too much information that could be lost when compressed with the DCT transform, since only limited numbers of DCT coefficients are preserved for the training process. Thus, the proposed method suggest to use 40x40 pixels sub block windows to obtain the best representation of local features.

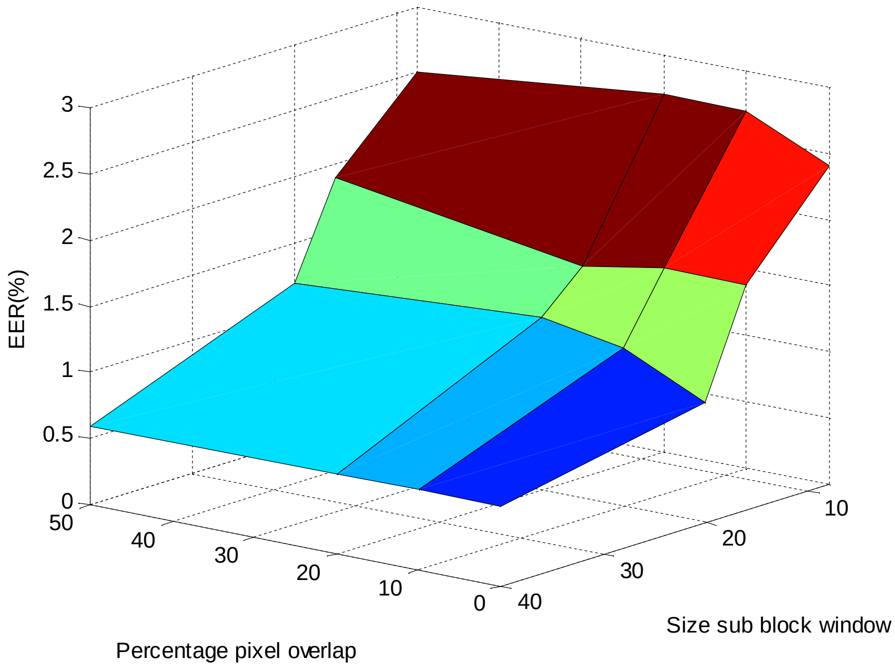


Figure 3.10: Performance analysis of EER (%) with different sizes of sub block window and pixel overlap.

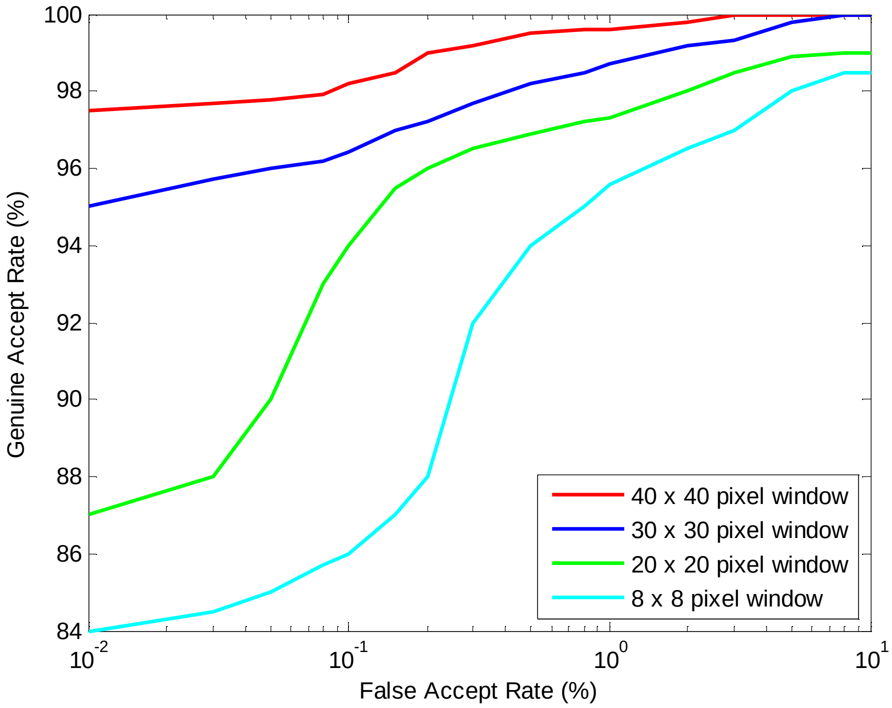


Figure 3.11: Analysis of the effect of different patch size on genuine acceptance rates.

3.8.3 The Effect of Different Number of DCT Coefficients

This analysis compares the performance of the system when local features are represented with different numbers of DCT coefficients. Again, the extracted local features are fused in a matrix interleaved method as discuss in Chapter 4. For each local patch window, 32 sets of Gabor images are extracted and the energy of each Gabor image is computed using the DCT transform. DCT coefficients consist of DC value, low frequency components, and high frequency components. Most discrimination features such as the shapes of the nose, eye and texture of palmprints exist in the low frequency components, while high frequency components contain noise and details of image texture. In order to employ the most valuable information in the DCT coefficients during the classification process, high frequency components need to be discarded and only low frequency components are preserved. This analysis examines the most effective size of DCT coefficients that will give the best results in terms of verification and identification rates. The DCT coefficients are varying from 5 to 40 coefficients and analysed in different sizes of sub block windows. Figure 3.12 shows that the lowest EER is 0.6% achieved when using 30 DCT coefficients extracted from 40x40 pixels sub block windows. When the number of DCT coefficients decreases less than 30 coefficients, the EER will increase as well due to insufficient information being available to discriminate between each class. Meanwhile, when the number of DCT coefficients increased larger than 30 coefficients, the EER will increase due to the effect of noise or high frequency components on the discrimination power. The effect of recognition rates when different numbers of DCT coefficients are used to represent the feature vector is also examined. In the recognition process, the system must be able to discriminate between subjects and thus it is more challenging than the verification analysis. The highest recognition rate of 97% was achieved by using 30 DCT coefficients as shown in Figure 3.13.

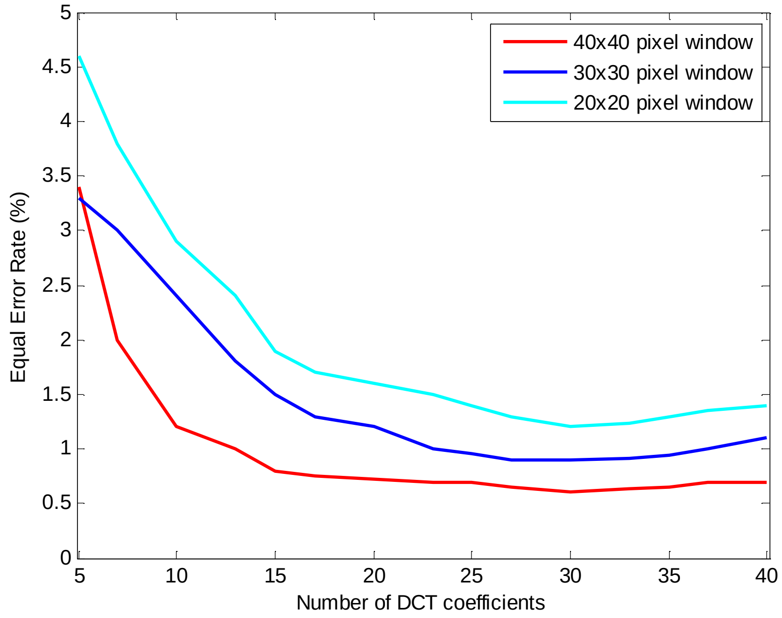


Figure 3.12: Verification performance analysis (EER rates) with different number of DCT coefficients and different size of sub windows. The analysis is tested on FERET-PolyU dataset.

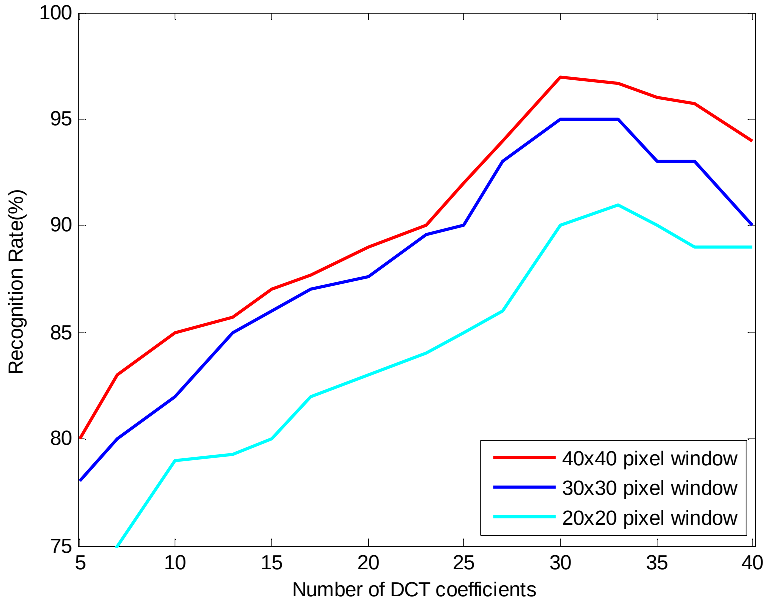


Figure 3.13: Recognition rates with different number of DCT coefficients extracted from different size of sub window. The analysis is tested on FERET-PolyU dataset.

3.8.4 Comparison of Local DCT Coefficients

The proposed local feature extraction method is also compared with several existing local feature extraction techniques based on the number of DCT coefficients, as shown in Figure 3.14 and Figure 3.15. The proposed method achieved the lowest EER of 0.6% and the highest recognition rate of 97% using 30 DCT coefficients. The three other local feature extraction methods used in the comparison are the DCTmod2 [93], DCT_modified [78] and DCT overlap [93], which are used to extract facial features and then modelled as feature distributions using GMM. All of these methods use sub-block windows to extract local DCT coefficients from the original image without implementing multiresolution analysis which is a powerful tool for texture analysis. Compared to the method proposed here that use multiresolution analysis based on Gabor transforms, more information can be gain when using Gabor analysis on each patch window. In addition, it is found that by using Gabor analysis on each patch window, the use of window overlapping can be avoided, which is one of the steps that must be conducted in existing local features methods. These methods compute the feature vector based on a 50% window overlap in the small 8x8 pixels windows, and also suggest, that 8x8 patch windows give the best local feature representation that can be used to train using GMM and HMM. However, in the method suggested here it was found that by using multiresolution analysis local features can be extracted in larger sub block windows producing feature vectors that are independent of each other due to the non-overlapping of sub windows. The DCTmod2 [93] method, replaces the first three coefficients with modified delta coefficients calculated from neighbouring blocks in order to reduce the effect of illumination on the first DCT coefficient, and incorporates vertical and horizontal neighbour information. Meanwhile, DCT_modified [78] uses the same delta coefficient method and incorporates the x-y coordinates of the patch window. However in the present method, the feature vector is extracted independently and thus is suitable for training using GMM to learn the distribution of features because in the GMM analysis each feature vector is treated independently.

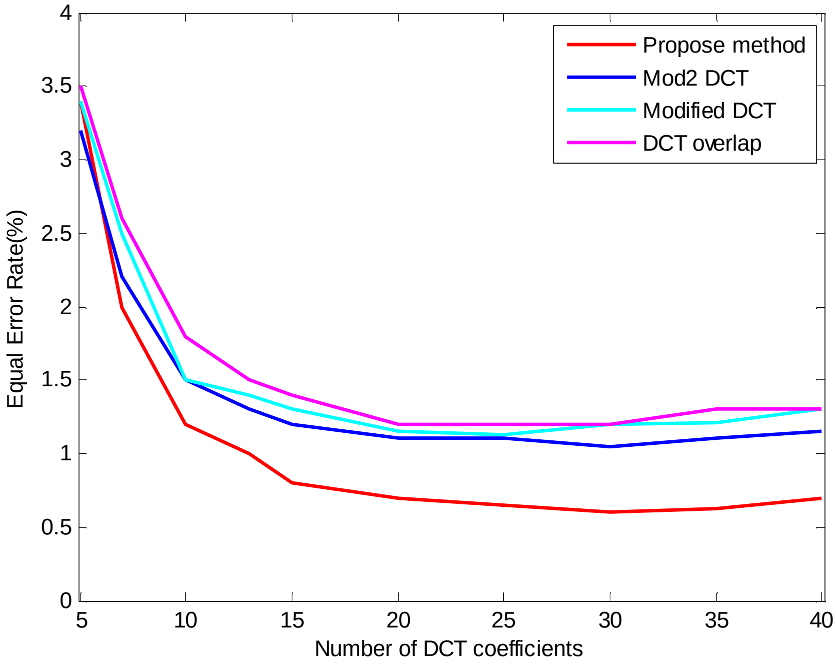


Figure 3.14: Comparison of the proposed method with three existing local feature extraction method based on DCT transform. The performance is compared against the number of feature vector and verification rates (EER).

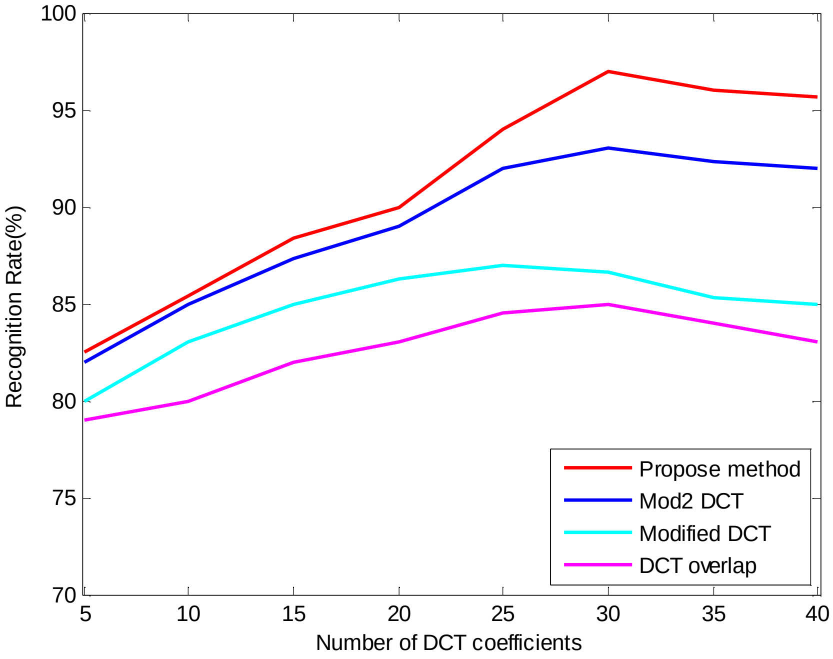


Figure 3.15: Comparison the effect of number DCT coefficient on the recognition rate (%) of the proposed local feature extraction method with existing DCT based local feature extraction. The analysis is tested on FERET-PolyU dataset.

3.9 Summary

A new method to extract local features in face and palmprint images is presented in this chapter. The proposed method is based on multiresolution analysis using the Gabor transform to extract information in several sub block windows and then the DCT compact energy representation is computed. The extracted features vector has low dimensionality and contains low frequency information from the image, and thus is suitable to be used in the feature fusion framework proposed in chapter 4. The proposed feature extraction method has several advantages. Firstly, the extracted local features in the sub block windows are independent, and thus suitable to be trained using GMM. Secondly, the compact feature representation is able to preserve low frequency information and discard high frequency information that contains noise, and thus it has high discrimination power in the classification process. Finally, the feature vector contains low dimensional features and thus the estimation of GMM components does not require a large number of data points and the estimated GMM parameter has a small number of coefficients, which is more efficient in terms of storage in database. From the experiments, the proposed local feature method achieves significant improvements in the recognition and verification analysis of multimodal biometrics compared to existing local feature methods.

Chapter 4

Feature Fusion and Parameter Estimation

4.1 Introduction

This chapter discusses the proposed new framework for the matrix interleaved feature fusion method and the determination of the statistical properties of the fused feature vector. The matrix interleaved method is different from the conventional concatenation method and it is able to increase the statistical information compare to the information given by single modal biometrics. Several existing feature fusion methods based on conventional concatenation are also contrasted with the proposed method. Then the effectiveness of using GMM in capturing the underlying statistical properties of the new fused feature vector is discussed. The classification process according to maximum likelihood used to measure the effectiveness of the proposed matrix interleaved method in recognition analysis is also demonstrated. The feasibility of the proposed method has been successfully tested using two multimodal datasets which are face and palmprint biometrics from FERET-PolyU and ORL-PolyU multimodal dataset. The effectiveness of the proposed method is demonstrated by comparing its performance against that of other multimodal biometrics using feature level fusion [32, 34, 35, 96, 108]. Finally, conclusions are given concerning the integration of face and palmprint images in the new framework and the increased performance of the recognition system.

4.2 Feature Concatenation

Most previous research into the fusion of biometrics data at the feature level uses concatenation methods. Concatenation can be performed after feature extraction, such as in the concatenation of magnitude information of Gabor images [11, 32, 37]. The Gabor image representations capture salient visual properties such as those relating to spatial localisation and orientation and thus superior performance is reported. A major drawback of this method is the high dimensionality of fused feature vectors and therefore computational complexity, due to the use of a Gabor filter bank with different scales and orientations. Fused Gabor images that contain nonlinear information are further processed to extract important information as well as to reduce feature dimensions using linear [33] and nonlinear dimensional reduction methods [32]. The nonlinear methods [55] achieve better performance than linear methods [45], [47], but are computationally demanding since the biometric images contain too much nonlinear information.

Instead of concatenating information after feature extraction, several methods concatenate features after dimensional feature reduction using approaches such as PCA, LDA and ICA. Yao [33] and Ahmad [108] have used PCA and LDA to extract information from two modalities and merged the feature vectors to form a long 1D vector. However, these linear feature reduction methods do not fully utilize the nonlinear information inherent that exists in the modalities used, and thus produce a merged feature vector with less discriminatory power. Concatenation can also be undertaken with raw data (i.e. without pre-processing) using a feature extraction method [96], and here all the raw data is normalized to the same size and sorted into a long vector prior to the application of a subspace selection method to construct the transformation matrix. A new subspace selection method is used, based on linear discrimination algorithms. This method offers better computational speed, but is limited to problems where the raw data is highly nonlinear. To overcome the nonlinearity problem in raw data, a kernel trick is adopted to map raw data to high dimensional feature space.

To date, most feature level fusion methods use a concatenation process which serially combines features from two or more modalities to form a long vector. In order to extract important information and reduce dimensionality, linear and nonlinear feature extraction methods are used prior to and after the fusion process. By using a concatenation process that does not take into account the distribution of data in both modalities, some data may become redundant and overlap on each others. In reality, some of modalities have nonlinear distributions of features such as face images with different poses and expressions, or palm-print images with different levels of aging and thus utilizing linear subspace reduction techniques cannot fully exploit the information contained in these modalities.

4.3 Framework of the Proposed Method

This section proposes and discusses a new methodology that uses non-stationary feature fusion to efficiently integrate significant information from biometrics modalities, especially those such as face and palmprint images having different statistical properties. The novelty of this new framework of matrix interleaved can be summarised as follows:

- A new structure for fusion using a concatenation method is introduced. The new structure will change the distribution of the original feature vectors by interleaving the extracted face and palm features to produce new features with a different statistical distribution. Compared to conventional methods which apply concatenation methods to merge feature vectors in order to form a single vector for classification purposes, the present method is designed to model the data distribution of the concatenation features and determine their various statistical properties. Concatenation using this new method has the benefit of using double data points compared to the conventional that use of only single data points. Higher discrimination power can thus be expected due to the presence of extra data points in modelling distribution.
- The proposed method makes use of the new compact local features discussed in Chapter 3 to represent local features in the fusing process. This is more robust in terms of pose, expression and illumination compared to the use of global features.
- The GMM is used to capture non-stationary information from the fused features vector. The advantage of GMM is that different types of fused data distributions can be represented as a convex combination of several normal distributions with their respective means and variances. Various mixture component parameters such as weight, mean and variance, of the data distribution are estimated using maximum likelihood (ML) algorithm [43, 109]. The Expectation Maximization (EM) algorithm is used to find the maximum likelihood estimation of the parameters of an underlying distribution from a given dataset when the data is incomplete or has missing values.

In the new framework of feature fusion, local features are utilized based on spatial frequency extracted using a DCT method. Local features are superior to global features since they are more robust to changes in illumination and pose invariant. Even though DCT based features have been widely used for single modal biometrics, the feature fusion of face and palmprint using DCT coefficients has not been accomplished before. Furthermore, the proposed method uses only 15 DCT coefficients, where the first coefficient of DCT is removed since it is an average pixel value. Feature fusion is performed based on these DCT coefficients and the results show a significant improvement in the recognition rates compared to the conventional fusion process that uses the same or a greater number of DCT coefficients. The fusion of DCT coefficients in the new framework increases recognition rates by incorporating extra statistical characteristics in the fused feature vector. In the context of multimodal biometrics, existing methods using GMM classifiers are used in two different ways. Either GMM is used to learn the statistical properties of biometric in each single modal biometrics, and fusion is then performed using the likelihood scores given by each modality [110, 111]; or it is used to learn the distribution of matching score values given for each modality [112]. In this study however, GMM is used at an early stage of the fusion process, and thus most of the existing statistical properties during feature fusion can be captured. Instead of determining the distribution of scores in two modalities using GMM and combining the likelihood value given by two different GMMs, the proposed method can give better results because the learning process is undertaken with the fused feature vectors which contains more discrimination information required in the recognition process. The proposed fusion method for the integration of face and palmprint data at the feature level using the new concatenation structure is shown in Figure 4.1.

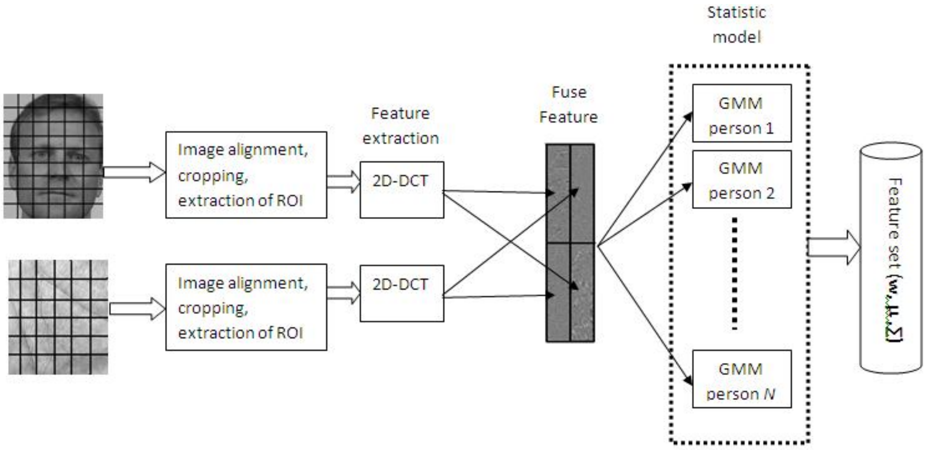


Figure 4.1: Overall block diagram of the new framework of the information fusion in multimodal biometrics system.

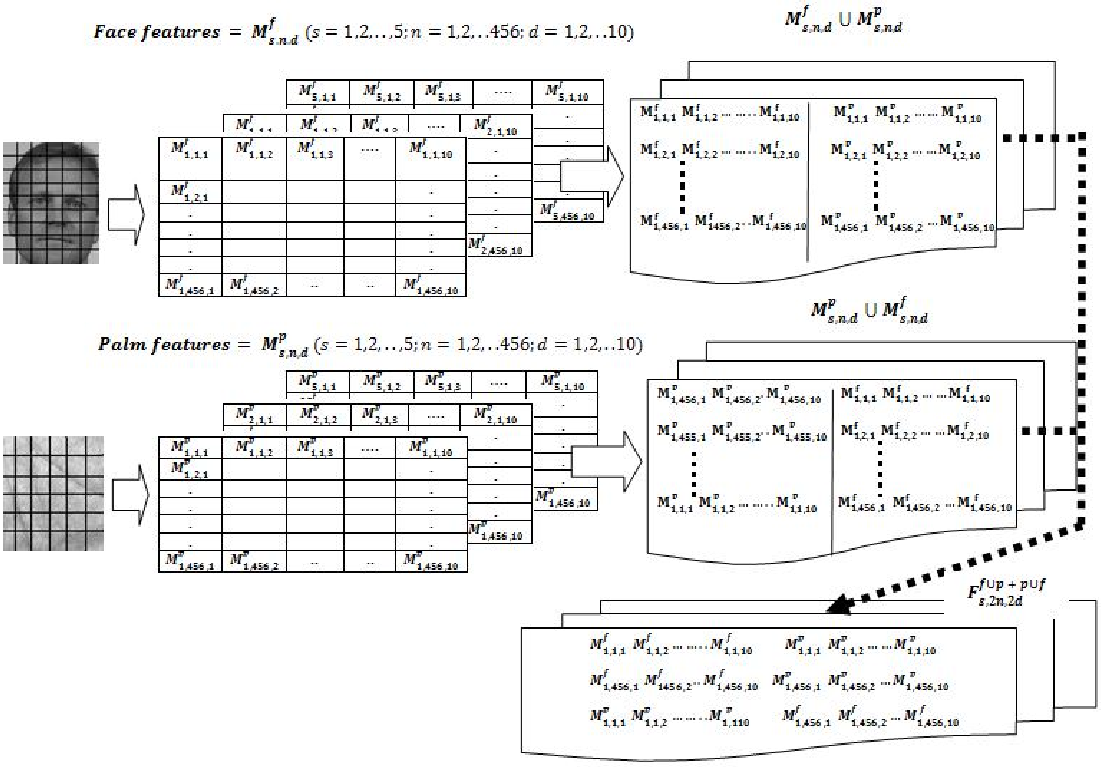

$Face\ features = M^f_{s,n,d}\ (s = 1,2,\ldots,5; n = 1,2,\ldots 456; d = 1,2,\ldots 10)$
$M^f_{s,n,d} \cup M^p_{s,n,d}$
$Palm\ features = M^p_{s,n,d}\ (s = 1,2,\ldots,5; n = 1,2,\ldots 456; d = 1,2,\ldots 10)$
$M^p_{s,n,d} \cup M^f_{s,n,d}$
$F^{f\cup p + p\cup f}_{s,2n,2d}$

Figure 4.2: Local feature extraction and matrix interleaved fusion method.

Suppose that there are S training images per subject for the face and palmprint, where $F_1, F_2, \ldots F_S$ represent face images while $P_1, P_2, \ldots P_S$ represent palmprint images. Each of the F_i and P_i is subdivided into m sub block windows with 8x8 pixels, and every sub block window are overlaps by 4 pixels which corresponds to 50% of the size of 8x8 pixels. Let F and P become the $i \times j$ pixel images, and thus the total number of sub block window with 4 pixels overlapping for each 8x8 pixels window is given by $\left(2\left(\frac{j}{8}\right) - 1\right) \mathrm{x} \left(2\left(\frac{i}{8}\right) - 1\right)$ block windows. The important information in each patch window is extracted using the DCT transform, which produces a feature vector $M^{f}_{s,n,d}(s = 1,2,..S; n = 1,2,\ldots N; d = 1,2,\ldots D)$ for face data and $M^{p}_{s,n,d}(s = 1,2,..S; n = 1,2,\ldots N; d = 1,2,\ldots D)$ for palmprint data respectively. S is the number of training images, N is the number of block windows and D is the dimension of the feature vector. If we have 5 training images with 456 sub block (100x80 pixels image, 8x8 window, 4 pixels overlap) and each patch has 10 dimensions, the extracted features for the face can be rearranged as:

$$M^{f}_{s,n,d} = \left\{M^{f}_{1,1,10}, M^{f}_{1,2,10}, \ldots, M^{f}_{1,456,10}, \ldots, M^{f}_{5,456,10}\right\} \tag{4.1}$$

and the feature vector for the palmprint is given by:

$$M^{p}_{s,n,d} = \left\{M^{p}_{1,1,10}, M^{p}_{1,2,10}, \ldots, M^{p}_{1,456,10}, \ldots, M^{p}_{5,456,10}\right\} \tag{4.2}$$

Fusion is performed by concatenation in two ways: 1) concatenating face features followed by palm features as given by$\left[M^{f}_{s,n,d} \cup M^{p}_{s,n,d}\right]$, and 2) concatenating palm features followed by face features as in $\left[M^{p}_{s,n,d} \cup M^{f}_{s,n,d}\right]$. Both of these feature vectors represent the same subject but with different statistical properties. These two groups of feature vector are then stacked on top of each other to produce a fused feature vector as follows:

$$F^{f\cup p\, +\, \mathrm{p}\cup f}_{s,2n,2d} = \left\{M^{f}_{s,n,d} \cup M^{p}_{s,n,d}\, , M^{p}_{s,n,d} \cup M^{f}_{s,n,d}\right\} \tag{4.3}$$

where $M^{f}_{s,n,d} \cup M^{p}_{s,n,d}$ for the 10 dimension is given by:

$$M^{f}_{s,n,d} \cup M^{p}_{s,n,d} = \{M^{f}_{1,1,10} \cup M^{p}_{1,1,10}, M^{f}_{1,2,10} \cup M^{p}_{1,2,10}, M^{f}_{1,3,10} \cup M^{p}_{1,3,10}, \cdots$$
$$\cdots, M^{f}_{1,456,10} \cup M^{p}_{1,456,10}\} \tag{4.4}$$

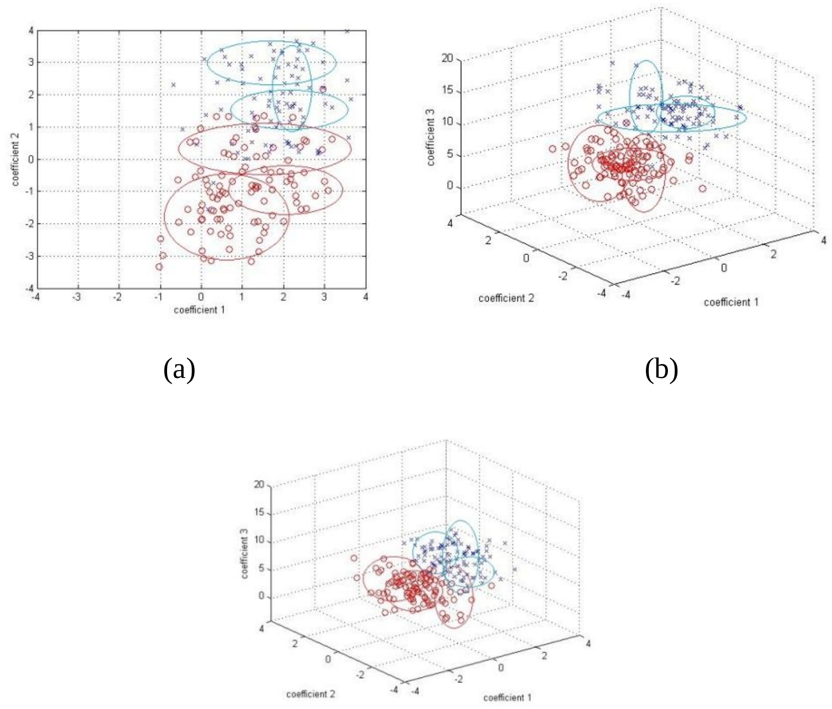
and $M^{p}_{s,n,d} \cup M^{f}_{s,n,d}$ is provided by:

$$M^{p}_{s,n,d} \cup M^{f}_{s,n,d} = \{M^{p}_{1,1,10} \cup M^{f}_{1,456,10}, M^{p}_{1,2,10} \cup M^{f}_{1,455,10}, M^{p}_{1,3,10} \cup M^{f}_{1,454,10}, \cdots$$
$$\cdots, M^{p}_{1,456,10} \cup M^{f}_{1,1,10}\} \tag{4.5}$$

The combination of $M^{f}_{s,n,d} \cup M^{p}_{s,n,d}$and $M^{p}_{s,n,d} \cup M^{f}_{s,n,d}$ is called non-stationary feature fusion. This method can be visualised in Figure 4.2. The fused feature vector is expressed in Eq.(4.3), where $F^{f\cup p\,+\,p\cup f}_{s,2n,2d}$ has an extra dimension that will increase the discriminant power of the feature vector and facilitate the learning process.

The proposed fusion method based on matrix interleaved is able to capture different statistical properties when the DCT features from face and palmprint modalities are fused together. The feature spaces for face and palmprint modalities have their own statistical properties, such as means and variances. By using single modalities, some of these statistical properties may still overlap with each other even between different subjects. When two different modalities are fused together using the proposed matrix interleaved method, the statistical properties that exist in each modality are changed to a new form. The new feature space from the fusion process contains more reliable information retaining more of the statistical properties of the data than that from a single modality. For this reason the proposed matrix interleaved feature fusion is better at capturing different statistical properties of the data. Figure 4.3 shows the Gaussian mixture distribution of the single modality of facial features (Figure 4.3a) and the new Gaussian mixture distribution after the fusion process (Figure 4.3b and Figure 4.3c). From the plot of the Gaussian mixture distribution, it is clear that when two feature vectors are fused together, the different distributions produced contain more statistical information and have higher discriminative power. Gaussian distributions in a two dimensional feature space still overlap with each other, while the new Gaussian distribution in three dimensional feature space is clearly separated.

The determination of complex feature distributions is a crucial part of achieving an accurate model to represent the correct distribution in the feature space. Figure 4.3 shows the fused feature vector scatter for several groups or clusters. Using a single Gaussian distribution to estimate statistical information such as means and covariances can be used to reduce the complexity of the estimation process. However, as can be seen in the feature distribution, a single Gaussian distribution is not able to capture all of the statistical information which exists in the fused feature vector. Furthermore, using a single Gaussian estimation will lead to a simple types of classification such as Euclidean or Mahalanobis distance classifier. To overcome the limitations of single Gaussian models in estimating the fused feature distribution, it is proposed to use GMM to model the complex distribution of data in fused feature space. GMM has been used in single modal biometrics to model speech features in speaker recognition [115, 116] and facial features in face recognition systems [114]. GMM contains an infinite number of Gaussian functions and is suitable to capture the underlying statistical properties in each group or cluster in the distribution of fused features. Furthermore, the Bayes’ classifier based on GMM is calculated based on maximum likelihood values, which is better than using a distance classifier. Maximum likelihood values give a higher degree of certainty that the observed feature vector belongs to a certain group of subjects.



(a) (b)

(c)

Figure 4.3: GMM distribution in two and three dimensional feature space. In a two dimensional feature space (a), some of the feature points of different subjects still overlap. Each subject has their own statistical properties in terms of means and variances. When coefficients 1 and 2 are concatenated with coefficient 3, a new feature space as in (a) and (b) can be obtained. In the new feature space, the feature points of subjects 1 and 2 are well separated, and they also have different statistical properties compared to most in (a).

4.3.1 Gaussian Mixture Models

The fused feature vectors developed above can then be classified using a classifier or pattern recognition method to identify a specific class to which the given test features belongs. In general, a statistical method is preferable since this will provide an interpretable form of the certainty of information along with tractable mathematical background [43]. Statistical methods also offer reliable basis for decision making, and thus a low risk option can be chosen. The training process for the feature vectors includes supervised learning, in which a group of known feature vectors in a specific class is introduced to assist in determining its statistical properties. Then, these statistical properties are used to classify unknown observations and to estimate confidence values of the classification process.

The information in a specific class is represented in low dimensional features in terms of its probability density function (PDF) rather than using high dimensional features. Finding the appropriate PDF will lead to successful classification. A simple model, such as one using Gaussian distribution with only a single set of statistical properties can efficiently represent features that have normally distributed features, but more sophisticated models such as finite mixture models must be used to approximate the complex PDFs of features distributed in a non-Gaussian form. It is generally understood that a finite mixture model can estimate a wide variety of PDFs and become an interactive solution when a single Gaussian form fail to model a complex data distributions. The mixture can be formed using any type of basic distribution function such as Gamma and Beta, but multivariate Gaussian distributions is well known and widely use to model the PDFs. The M-dimensional random variable $\boldsymbol{y} = \{\boldsymbol{y}_i\}_{i=1}^{N}$ follows a Gaussian mixture distribution when its PDFs $p(y_i|\theta)$ can be represented by a weighted sum of normal distributions:

$$p(y_i|\theta) = \sum_{i=1}^{K} \omega_i p(\boldsymbol{y}_i|\boldsymbol{\theta}_i) \tag{4.6}$$

$$= \sum_{i=1}^{K} \omega_i \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \times exp\left\{-\frac{1}{2}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i\right)^T \Sigma_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{\mu}_i\right)\right\} \tag{4.7}$$

where $0 \leq \omega_i \leq 1$ and $\sum_{i=1}^{K} \omega_i = 1, i = 1,2, \ldots K$, K is the number of Gaussian components, and $\omega_1, \ldots \omega_K$ are the priori probabilities of each component K. The whole set of parameters that describe each component is $\theta = [\theta_1, \theta_2, \ldots \theta_K]$ with $\theta_i = [\omega_i, \mu_i, \Sigma_i]$ where μ_i and Σ_i are the mean and covariance matrix of each component respectively. Obtaining an optimal set of the parameter θ is usually defined in terms of maximizing the log-likelihood of the PDFs to be estimated, using the existing independent identically distributed samples $= [y_1, \dots y_N]$. The log-likelihood function of i-th component is:

$$\mathcal{L}(y|\theta_j) = \sum_{i=1}^{N} \ln\left(\omega_j p(y_i|\theta_j)\right) \tag{4.8}$$

$$= \sum_{i=1}^{N} \ln\left(\omega_j \frac{1}{(2\pi)^{M/2}\left|\Sigma_j\right|^{1/2}}\right) \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{\mu}_j\right)^T \Sigma_j^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_i)\right\} \tag{4.9}$$

Due to the missing label information in the label $\{\boldsymbol{y}_i\}_{i=1}^{N}$, the maximum likelihood (ML) estimation cannot be determined analytically but it can be estimated using expectation-maximization (EM), which is widely used in applications involving tasks with incomplete data sets [119].

The EM algorithm consists of the two steps of expectation step and maximization step. The EM algorithm generates a sequence of estimations of the set of parameters $\{\theta(t), t = 1,2, \dots\}$ by alternating expectation step and maximization until convergence reaches a certain threshold value. The expectation of Eq. (4.8) is;

$$Q(\theta; \theta(t)) = E\left[\sum_{i=1}^{N} \ln\left(\omega_j p(y_i|\theta_j)\right)\right] \tag{4.10}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{J} P(j|\boldsymbol{y}_i; \boldsymbol{\theta}(t)) \ln\left(\omega_j p(\boldsymbol{y}_i|\boldsymbol{\theta}_j)\right) \tag{4.11}$$

where:

$$P(j|\boldsymbol{y}_i; \boldsymbol{\theta}(t)) = \frac{p\left(\boldsymbol{y}_i|\boldsymbol{\theta}_j(t)\right) P_j(t)}{p(\boldsymbol{y}_i; \boldsymbol{\theta}(t))} \tag{4.12}$$

$$= \frac{p\left(\boldsymbol{y}_i|\boldsymbol{\theta}_j(t)\right)P_j(t)}{\sum_{j=1}^{K} p\left(\boldsymbol{y}_i|\boldsymbol{\theta}_j(t)\right)P_j(t)} \tag{4.13}$$

The maximizing step (M-step) is calculated by taking the derivatives of Eq. (4.11) with respect to $\boldsymbol{\theta}$ and $\boldsymbol{P}$ respectively, and setting them equal to zero:

$$\frac{\partial Q(\theta, \theta(t))}{\partial \theta} = 0 \tag{4.14}$$

$$\frac{\partial\left[Q(\theta, \theta(t)) + \leftthreetimes \left(1 - \sum_{j=1}^{K} P_j\right)\right]}{\partial x} = 0 \tag{4.15}$$

Substituting Eq. (4.11) into Eq. (4.15), gives a new updated parameter as shown in Eq. (4.16) to Eq. (4.18). Initialization of the parameter in the first iteration (t=0, $\boldsymbol{\theta}(0)$) is obtained by using K-means clustering. New parameters are then estimated in the iteration process. If $\theta(t+1) - \theta(t) > \mathcal{E}$, the algorithm will repeat the E-Step followed by the M-Step, where $\mathcal{E}$ is the threshold value:

$$w_j = \frac{1}{N}\sum_{i=1}^{N} f(j|\mathbf{x}_i, \boldsymbol{\theta}^s) \tag{4.16}$$

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} \mathbf{x}_i f(j|\mathbf{x}_i, \boldsymbol{\theta}^s)}{\sum_{i=1}^{N} f(j|\mathbf{x}_i, \boldsymbol{\theta}^s)} \tag{4.17}$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} f(j|\mathbf{x}_i, \boldsymbol{\theta}^s)\left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)\left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)^T}{\sum_{i=1}^{N} f(j|\mathbf{x}_i, \boldsymbol{\theta}^s)} \tag{4.18}$$

4.3.2 Recognition Process using Maximum Likelihood

During the classification process, the aim is to assign a class of test feature vectors to the model parameter that has the highest value of log-likelihood [115, 116]. Classification process in GMM can be seen as the maximum likelihood rule (ML), where the given test feature vector suppose to produce the highest values of likelihood. Let a group of S feature vectors belong to S class given by X = $\{X_1, X_2, \ldots X_s\}$ and each of the X_i has a non-stationary fused feature vector given by $x_t = \{x_1, x_2, \ldots, x_T\}$. The S class is represented by a Gaussian mixture model using S model parameters given by $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_s$ where each model parameter $\boldsymbol{\theta}_i = \{\mathbf{w}_k, \boldsymbol{\mu}_k, \sum_k\}$, and $k = 1,2, \ldots M_c$ where M_c is the number of GMM components used to model the distribution. The objective of classification is to find the model $\boldsymbol{\theta}_i$ which has the maximum a posteriori probability of a given observation feature vector X_i. This criterion can be written as:

$$\hat{S} = \arg\max_{1 \leq k \leq s} p(\boldsymbol{\theta}_k | X) \tag{4.19}$$

By using Bayes's rule, $p(\boldsymbol{\theta}_k|X)$ can be written as:

$$\hat{S} = \arg\max_{1 \leq k \leq s} \frac{p(X|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k)}{p(X)} \tag{4.20}$$

Since $p(X)$ is constant for all class models, and $p(\boldsymbol{\theta}_k)$ is a priori probability of the k-th model which is equally likely for all classes ($p(\boldsymbol{\theta}_k) = \frac{1}{S}$), therefore the maximum of posteriori $p(\boldsymbol{\theta}_k|X)$ can be changed to the maximum likelihood of $p(X|\boldsymbol{\theta}_k)$ since $p(X)$ and $p(\boldsymbol{\theta}_k)$ are constant parameters. This is shown as follows:

$$\hat{S} = \arg\max_{1 \leq k \leq s} p(X|\boldsymbol{\theta}_k) \tag{4.21}$$

Using log-likelihood and assuming the independence of observations, the identification process can be written as:

$$\hat{S} = \arg\max_{1 \leq k \leq s} \sum_{t=1}^{T} \log p(\boldsymbol{x}_t|\boldsymbol{\theta}_k) \tag{4.22}$$

where $p(\boldsymbol{x}_t|\boldsymbol{\theta}_k)$ is a mixture model given by Eq. (4.7).

4.4 Experimental Analysis and Discussion

The proposed method is evaluated using two sets of virtual multimodal datasets based on face and palmprint images. The construction of virtual multimodal datasets from the two separate face and palmprint datasets can be designed since the modalities are independent of each other. Several analyses of multimodal biometrics have also been evaluated based on virtual multimodal datasets [32, 33]. In the present analysis two sets of virtual multimodal datasets are constructed from the ORL-PolyU and FERET-PolyU dataset. Multimodal ORL-PolyU datasets contain 800 images corresponding to 40 subjects acquired at different times, poses and facial expressions, while the FERET-PolyU datasets consist of 1200 images that correspond to 100 subjects with different poses, expressions and ageing.

4.4.1 Performance Analysis using the ORL-PolyU Dataset

The ORL datasets developed at the Olivetti Research Laboratory, Cambridge, consist of 400 greyscale images for 40 subjects with 10 different images for each subject. From 10 images per subject, 5 images are randomly chosen for training and the remaining 5 for testing. A greyscale image in ORL dataset has a resolution of 112 x 92 pixels and all images are aligned in terms of their geometric position. Some of the images have an angle of $\pm 15^0$ with varying poses and expressions. The facial expressions include smiling and not smiling, eyes open and closed and wearing glasses or not. The image is pre-processed by reducing the size to 100 x 80 pixels and the important features are extracted using the DCT in a sub block

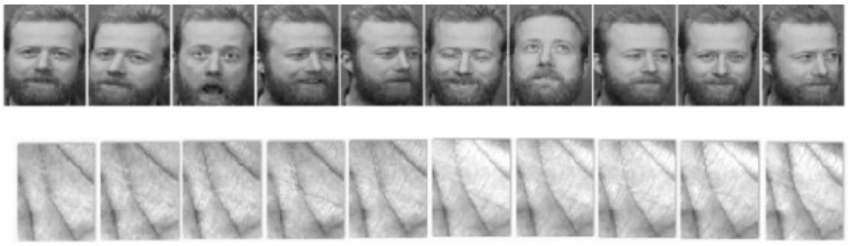


Figure 4.4: Sample of virtual multimodal datasets for 1 subject constructed from ORL face dataset and PolyU palmprint dataset.

window. In order to create a multimodal ORL-PolyU dataset which has the same dimensions as the face images, the palmprint images are cropped to 100 x 80 pixels. The selection of the region of interest (ROI) is explained in [107]. To ensure the reliability of the virtual datasets, a computer program is used to randomly select and pair the face and palmprint images. This computer program has two main tasks: to randomly pair the face and palmprint images of different subjects and to randomly select the images used for training and testing. The experiment is repeated 100 times and the average performance is calculated. Figure 4.4 shows the combination of face and palmprint images. The face images show variations in ageing, pose style and angle, illumination, expression and hairstyle, while palmprint images show variations in line orientation, illumination and ageing. In order to validate the proposed method, 5 images are used for training and the other 5 are used for testing.

The first analysis examines the number of mixture components required to accurately model the fused feature distribution and the relationship between the local features to give the best statistical representation of the most important features. The images from both datasets are aligned at the same size. In this analysis, the number of mixture component is varied from 3 to 8 components while pixel overlap for the 8x8 DCT sub block window is changed from 0 to 2, 4 and 6 pixels. Figure 4.5 shows that the highest recognition accuracy is achieved by using at least

5 GMM components and a 4 pixel overlap in the sub block window. The use of more than 5 GMM components makes no difference to recognition accuracy, but the computational complexity involved and the time requires to estimate statistical distribution increases. On the other hand, using less than 5 components decreases performance because a limited number of normal distributions is available in the statistical model to capture the underlying statistical properties of the fused features. The selection of the number of mixture components depends on the complexity of the fused feature vector distribution. It should be noted that an appropriate number of mixture components should be carefully chosen to avoid both the problem of over fitting which arises from using too many GMM components and failure to determine distribution with insufficient GMM components. In this analysis, it was experimentally found that by using 5 GMM components to model the fused features extracted from a 4 pixel overlap in DCT sub block windows gives the highest performance. The statistical properties represented by 5 components each of which consists of a weight, mean and covariance are stored in the database as a low dimensional feature vector. The effect of pixel overlap on local features was also examined by varying the degree of pixel overlap in the horizontal and vertical directions. The extraction of local features assumes that each region in the face and palmprint image is independent. In practice, some of the regions are related to each other, such as the positions of the eyes and nose, and thus implementing pixel overlap incorporates information about features extracted from the neighbouring sub block windows. A previously used method [93] suggests that using pixel overlap can eliminate the effect of illumination in facial images. Another advantage of local feature extraction using overlapping pixels is its capability to generate an adequate

number of feature vectors in order to determine their statistical properties using a method such as GMM.

The overlapping pixels approach to local features has been proven to be effective in feature extraction [92, 93]. From the analysis shown in Figure 4.5, the best result is achieved by using an overlaps of 4 pixels. When their number is reduced to less than 4, performance dramatically decreases because fewer feature vectors are available in proportion to the number of sub block windows. If pixel overlap is increased, the number of sub block windows also increases, thus generating a sufficient number of feature vectors to train the statistical model. Based on the results in Figure 4.5, recognition accuracy decreases when pixel overlap is increased to 6 pixels. This is due to the extraction of redundant features when the degree of overlap covers more than 50% of the block size. When 6 pixels overlap are used, a higher number of mixture components are needed in order to increase the recognition rate.

The second analysis examines the effect of different numbers of fused feature vectors to the discrimination power, which will affect recognition rates. 64 DCT coefficients will be extracted from 8x8 sub block window, but not all of these lead to superior discrimination power. In order to select a smaller number of coefficients, a zig-zag method is utilized where only several coefficients are retained which correspond to low frequency information. The discrimination power of the fused feature vector is examined by using vectors varying from 5 to 20 dimensions. The results in Figure 4.6 show that the best recognition accuracy of 99.7% can be achieved by using 15 DCT coefficients modelled using 5 GMM components. Note that using 20 DCT coefficients will not improve discrimination power but will increase computational complexity and the number of feature vectors stored in the database. Some high dimensional features extracted from the DCT transform using zig-zag patterns consist of high frequency information, or noise, that affects the parameter estimation process. High frequency information is not suitable for discriminating face and palmprint images, whereas low frequency information represents the basic appearance of the faces and palmprints.

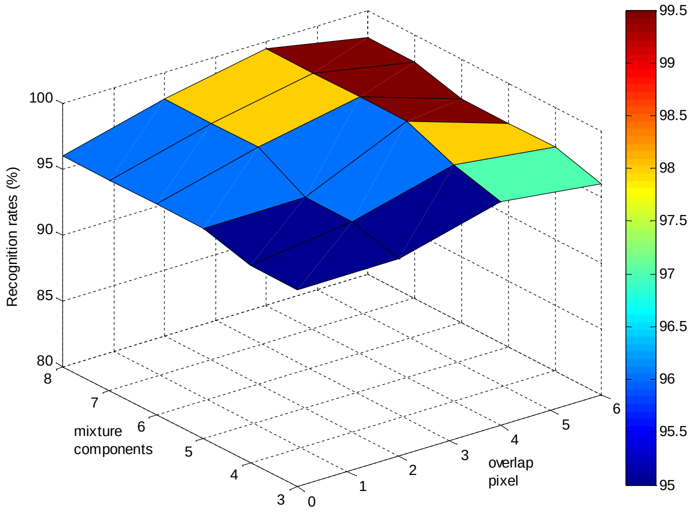


Figure 4.5: Recognition accuracy of the proposed method with different number of GMM components and overlapping pixels in local features. The best recognition accuracy is achieved using 5 GMM components and 4 pixels of overlap.

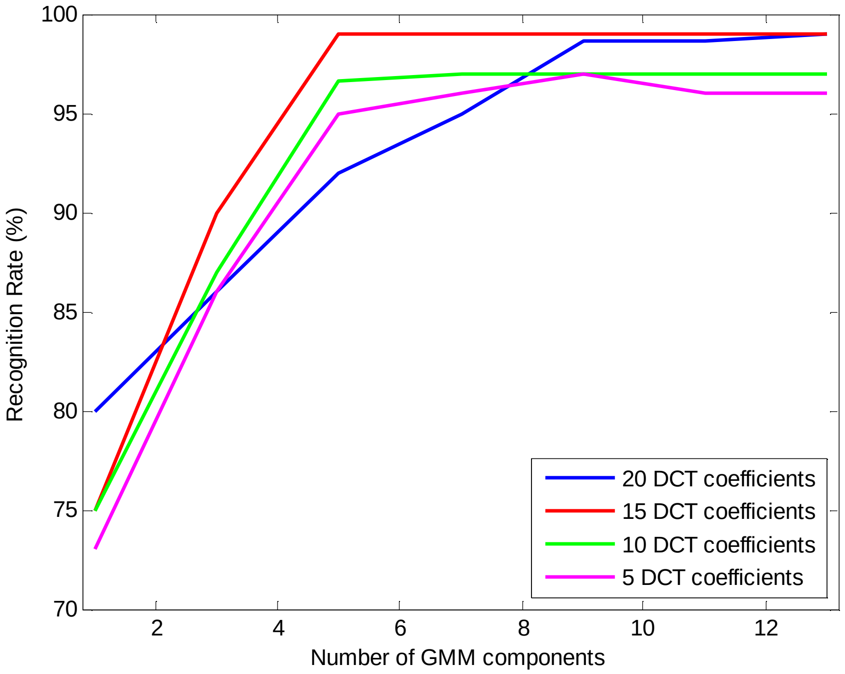


Figure 4.6: Analysis of recognition rate (%) with different numbers of DCT coefficients.

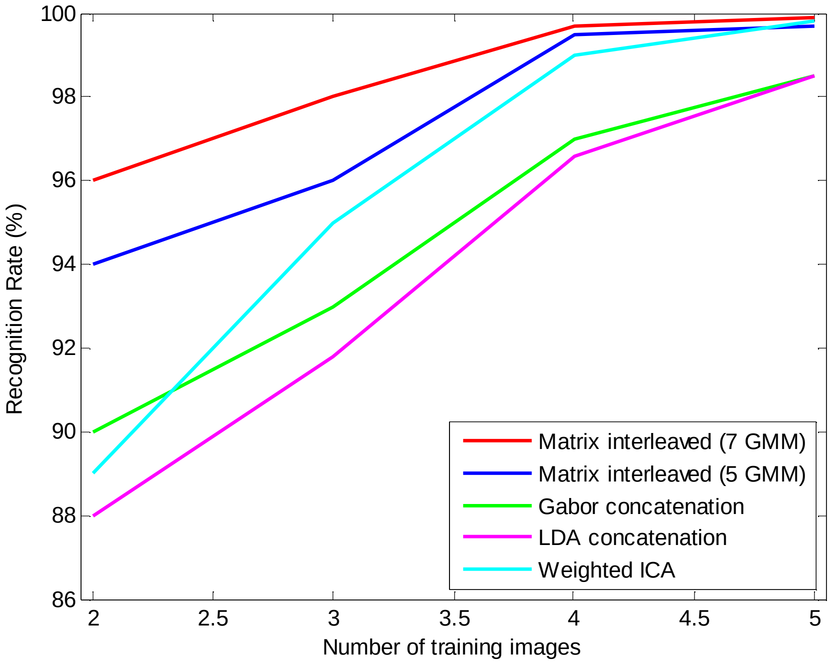


Figure 4.7: Comparison of the proposed matrix interleaved method with several existing feature fusion methods where the number of training images varies from two to five. The experiment is run on a ORL-PolyU dataset.

The third analysis compares the proposed fusion method with several existing fusion methods that combine information at feature level. The number of training images is varied from 2 to 5 images in order to observe the effect of incorporating less statistical information during the training process. Three existing feature fusion techniques are evaluated: 1) the concatenation of Gabor images from face and palmprint data [32]; 2) the concatenation of LDA features extracted from Gabor face and Gabor palm images [108]; and 3) the concatenation of weighted ICA features extracted from Gabor face and Gabor palm images [34]. The results show that the proposed method only requires 4 training images to achieve the best result of 99.7% recognition accuracy. This compares to existing methods that needing at least 5 training images to achieve 99.5% recognition accuracy. From the analysis, it is clear that the proposed method could give better results by using fewer training images. The results shown in Figure 4.7 demonstrate that the matrix interleaved method can preserve more statistical information compared to the concatenation method. This can be clearly seen when two training images are used where the performance of conventional concatenation methods is 6% less than the proposed matrix interleaved method.

In order to validate the proposed method and to conduct a fair comparison, the results are compared with those from several existing methods which perform fusion at the feature level using the same datasets. All of these methods use 10 images per subject and 5 images each are used for training and for testing. When fusion at feature level is performed by concatenating the weighted features extracted using ICA, 99% recognition accuracy is achieved [34]. This method used the wavelet transform to extract global features from the face and palmprint images. The highest recognition accuracy of 99.2% was achieved using feature fusion based on user specific weighted rules [35]. In this method, a specific user weight is applied during the concatenation of the feature vector. Finally, a recognition rate of 99.5% was achieved when fusion is performed by concatenating features extracted by LDA from both modalities [108]. These three concatenation methods are totally different from the method proposed here in terms of the number of data points used since they do not take into account non-stationary information. The present method fuses the matrices and interleaves them to increase the number of data points, and moreover this method has the ability to learn non-stationary information. Table 4.1 shows the recognition accuracy of the proposed method compared to existing methods which made use of Gabor filters to extract information and thus are more computationally demanding. In addition, the existing methods did not fully utilize the nonlinear information which exists in the different modalities when Gabor images are pre-processed by linear projection methods such as ICA, LDA and PCA. The proposed method also has the advantage that the low feature vector has small dimension and requires only 155 (5+ 5x15 + 5x15) feature coefficients, comprising 5 for weight, 5x15 for means and 5x15 for diagonal covariance when modelled by using 5 GMM components.

Table 4.1: Comparison of the proposed method with the existing method in terms of recognition rates (%) using ORL-PolyU datasets.

Method	**Top recognition rate**
Yao et al’s method [34]	99.2%
Lu et al’s method [35]	99.0%
Ahmad et al’s method[108]	99.5%
Proposed method	99.7%

4.4.2 Performance Analysis using FERET-PolyU Dataset

In this experiment 100 subjects were randomly selected, each with six frontal images of which three each are used for training and testing. The facial images in this datasets were captured under various levels of illumination, different facial expressions and poses ranging from angles of ±15^o, to ±60^o. Each image has 384x256 pixels size with 256 grayscales. Most of the images have different size of frontal image including the background and body chest region. The six images per subject used in this experiment are selected among the "fa", "fb", "rb" and "rc" which belong to different pose, aging, expression and pose angle.

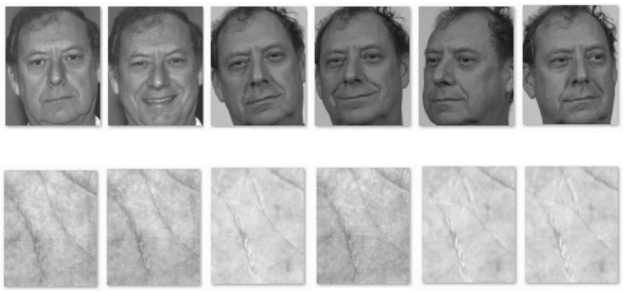


Figure 4.8: Samples of virtual FERET-PolyU multimodal datasets with variations in illumination, pose, expression and ageing.

In order to construct virtual multimodal datasets for the face and palmprint images, 100 subjects are randomly chosen from the FERET face and PolyU palmprint datasets; with 6 images per subject. A computer program was used to randomly select and pair the face and palmprint images in order to construct a virtual multimodal dataset. The computer program first randomly paired different subject of face and palmprint images; and secondly randomly selected images to be used for training and testing. The experiment was also repeated 100 times in order to get a reliable results and overall performance was computed based on the average value. For each experiment, different subjects from face and palmprint images are paired together and different images used for training and testing. The facial images varied in terms of illumination, ageing, expressions and pose style and angle, while the palmprint datasets varied in terms of illumination, line orientation and ageing. 6 images from face and palmprint datasets for each subject are paired together as shown in Figure 4.8.

In this experiment all facial images were manually resized and cropped using centre of eyes to a size of 100x80 pixels. The images were then normalized using histogram equalization. The first analysis investigated the best possible number of GMM components to capture statistical information in the fused feature vectors. The FERET-PolyU multimodal dataset contain images that are different in terms of ageing, hairstyle, pose and illumination, and thus the distributions of features are expected to be highly nonlinear and non-Gaussian. The results in Figure 4.9 show that the highest recognition accuracy of 97% was achieved when using 9, 11, 13 and 15 GMM components. From this experiment with the FERET-PolyU datasets, it is suggested that 9 GMM components are required to accurately model the fused feature in order to reduce computational complexity and memory usage. In this analysis, fused feature vectors were extracted using an 8x8 DCT block window with different degree of pixels overlap varying from 0 to 6 pixel and the best results were given by the 4 pixel overlap. By convolving the 100x80 pixel image using 8x8 DCT block window with 4 overlapping pixels, 456 feature vectors were produced per image. A training process was carried out using 3 training images, and thus the total number of feature vectors used to train the GMM was 1368 (456x3). If the feature dimension of DCT coefficients is limit to the first 15 coefficients, the estimated parameters that are required to represent 9 GMM components are reduced to 279 ($\omega = 9, \mu = 9 \times 15$ and $\sum = 9 \times 15$) feature vectors. This is only 20.4% the original 1368 feature vectors and thus requires relatively smaller storage space in the database. The size of the feature vectors can be further decreased by reducing the number of GMM components, but performance would then be degraded as the number of single Gaussians is insufficient to model the fused feature distribution. This can be seen in Figure 4.9, when the number of mixture is reduced to 5 components, and the estimation of parameter (feature matrix) is reduced to 159 ($\omega = 5, \mu = 5 \times 15$ and $\sum = 5 \times 15$) feature vectors which is 6.8% of the original image size of 2280. However, the recognition rate dropped to 85%, and hence, there is a trade-off between recognition accuracy and the number of mixture components. The results in Figure 4.9 also show that when pixel overlap is reduced to 0 or 2 pixels, performance also declines, for several reasons. Firstly, the number of the feature vectors extracted from the 8x8 DCT block window reduces with a reduction in pixel overlap, and therefore the training process does not involve an adequate number of feature vectors to capture the statistical characteristics. Secondly, pixel overlap takes into account the relationships amongst neighbouring blocks. If pixel overlap is reduced, some important information about the relationship between two neighbouring blocks may be lost, thus reducing the performance of the system. A higher pixel overlap of 6 in Figure 4.9 produces similar performance to that shown by 4 pixels. The number of feature vectors is increased with an increase in pixel overlap, but is increases redundant information and is computationally costly. From the experimental results, it is suggested that 4 pixel overlap is sufficient to produce good results using the proposed fusion framework.

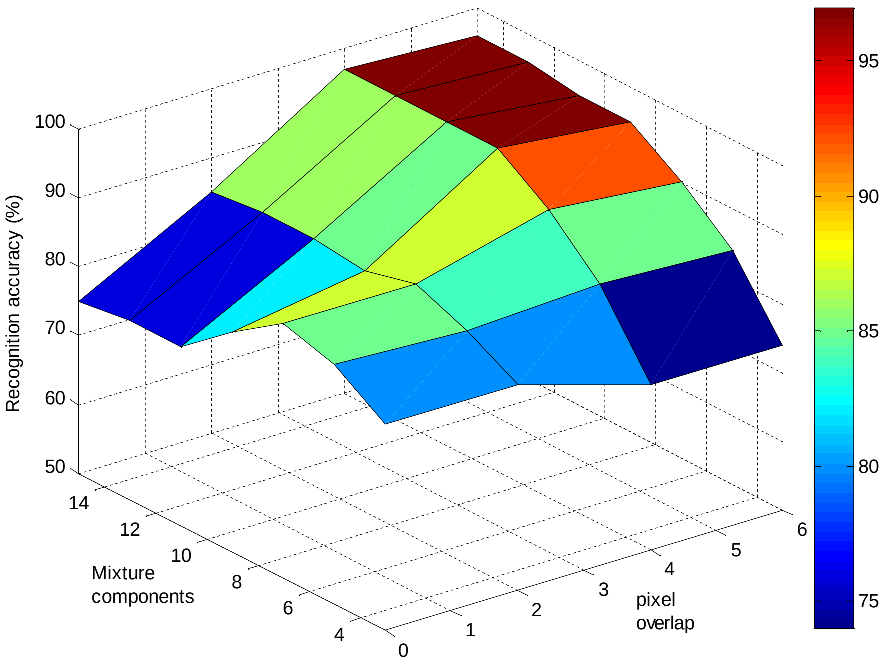


Figure 4.9: Analysis of recognition rate of the proposed method with different number of pixels overlapping in 8x8 DCT block windows with different numbers of GMM mixture components. The best recognition accuracy is achieved using 9 mixture components and 4 pixels overlap.

The second analysis investigated the effect of numbers of features extracted using various sizes of sub block windows. This was carried out to find the best size of DCT windows from which the most discrimination features can be derived. The different sizes of DCT block windows used are 8x8, 12x12, 16x16 and 24x24 pixels. Table 4.2 gives a summary of the recognition rates achieved for each size of DCT block window. The highest recognition rates of 97% achieved by using 8x8 sub block window. Using a smaller sub block window is advantageous in reducing the computational time required to transform the image into the DCT domain. It was found that the highest recognition rate could be achieved using 23% of the coefficients in each of the sub block windows. By using larger sub block window, more DCT coefficients are produced, and thus large amount of feature vectors are needed to accurately estimate the GMM components. The effect of a small number of feature vectors to the different size of sub block window can be seen in Table 4.2 where the performance is degraded in the larger size of sub block windows due to an insufficient number of feature vectors to estimate the GMM mixture components. For example, a 16x16 sub block produces a 128 features with 30 dimensions, while 8x8 sub block produces a 456 feature vectors with 30 dimensions which is triple compared to 16x16 sub block window.

Table 4.2: Analysis of recognition rates (%) given by different size of sub block window model by using different number of GMM components.

Sub block window size	**Mixture components**					
	3	***5***	***7***	***9***	***11***	***13***
8x8	86 %	96%	97%	97%	97%	97%
12x12	85%	84%	87%	87%	86%	85%
16x16	84%	85%	83%	86%	81%	73%
24x24	80%	88%	87%	70%	70%	68%

The third analysis again examined the effect of varying dimension of DCT coefficients on the performance of the proposed method. Varying this number will produce different sizes of component parameters which need to be stored in the database. It is known that not all DCT coefficients contain high discrimination power, thus by removing several coefficients; an effective size of feature vector can be achieved without scarifying the performance of the system. In order to identify the best dimensions of DCT coefficients to produce the best results, 4 different dimensions of DCT coefficients are examined such as 5, 10, 15 and 20 coefficients. It must be noted that, using GMM to capture the statistical information in a limited numbers of feature vectors requires small dimensions of feature vectors in order to accurately estimate the model parameters. Figure 4.10 shows that 15 dimensions of DCT coefficients can give recognition rates up to 97% when the distribution of fused features is modelled using 7 GMM components which require 225 feature vectors. Using more than 15 dimensions of DCT coefficients does not improve the recognition rate except when a smaller small number of GMM mixture components are used. Reducing the dimension of DCT coefficients to 5 also has no significant effect on recognition rate. The highest recognition rates are achieved by using 15 or 20 dimensions of DCT coefficients where 15 dimensions is preferable in order to reduce memory storage.

In the fusion process, only a few DCT coefficients are preserved using the zig-zag method, and low frequency components that contain most of the discrimination power are used while high frequency components that do not contribute to the classification process are removed. The proposed method uses sub block window of 8x8 pixels to extract local DCT features. Each block window will produce 64 DCT coefficients and the fusion process only uses 23% of the total number of DCT coefficients (15 DCT coefficients) in order to achieve the best results. In addition, concatenating selected DCT coefficients from two different modalities will further increase the discrimination power in the feature space.

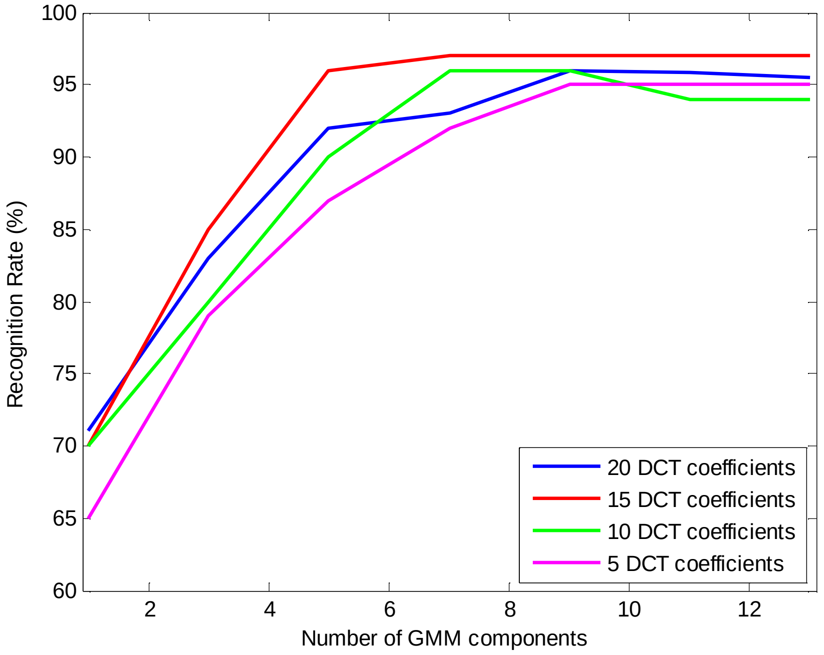


Figure 4.10: Recognition rate of the proposed method when fusion is performed using different numbers of DCT coefficients and numbers of GMM components vary from 1 to 13.

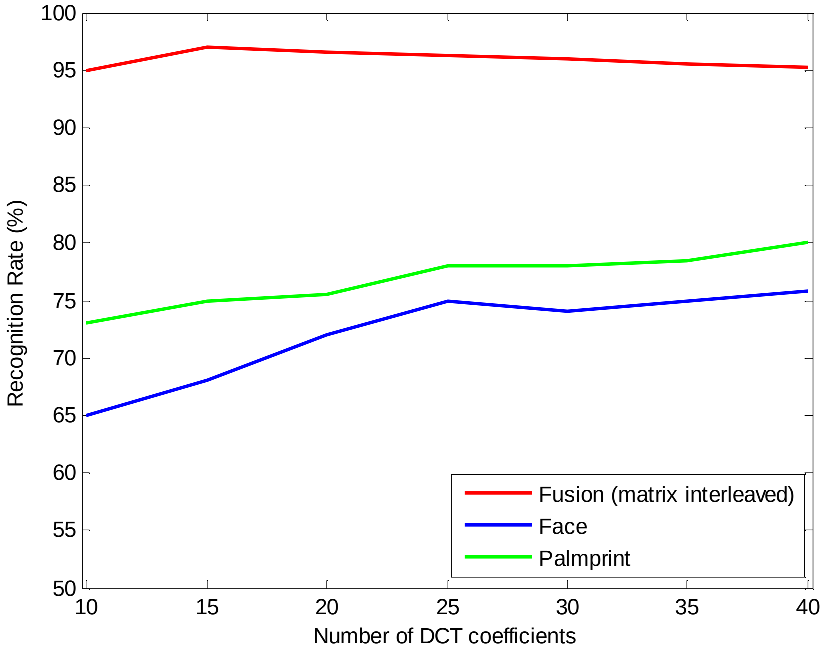


Figure 4.11: Analysis of recognition rates with different numbers of DCT coefficients using FERET-PolyU multimodal datasets. The highest recognition rate is given by matrix interleaved method using 15 DCT coefficients for each subject and the information is captured using 7 GMM mixture components.

Combining DCT coefficients from two different modalities can better represent the data in the feature space compared to the use of DCT coefficients derived from single modalities. The fusion of low frequency components reduces the dimensions of the feature vectors while increasing the statistical information, and therefore can eliminate the problem of high dimensional feature vectors. The use of single modalities needs high dimension of DCT coefficients to increase performance, whereas the fusion method here only uses small dimension of DCT coefficients to achieve the best results. Figure 4.11 shows the performance of the proposed method compared to the use of a single modality when different numbers of DCT coefficients are used in the fusion process. The proposed method achieves the best result of a 97% recognition rate when 15 DCT coefficients are used during the fusion process. Meanwhile, the use of a single modality requires 40 DCT coefficients to achieve the best results of 80% accuracy for palmprint and 75% accuracy for face images but these recognition rates are 17% or more lower than that of the proposed method. Based on this analysis, the proposed fusion method can reduce the dimension of feature vectors while increasing the performance of the system. This is because matrix interleaved produces more statistical information during the feature combination process.

The performance of the proposed method was compared with several existing methods that use a single modality, of face or palmprint in order to investigate the effectiveness of multiple modalities. The proposed matrix interleaved method combining multiple modalities was found to produce higher recognition rates even when the feature dimension of the single modality is higher than the multiple modalities. This is due to the ability of the fusion method to capture more important information during the fusion process. The results in Figure 4.12 show that feature fusion using only 15 DCT coefficients, which results in a vector with 30 coefficients can produce the highest recognition rate of 97%. This is higher than the rates achieved with single modalities using 40 DCT coefficients, with an accuracy for face images is 69% and for palmprint is 79% when the statistical information is captured using 7 GMM components. The proposed method has also been tested with different numbers of DCT coefficients such as 5, 10, 15 and 20, and the best result was given by 15 DCT coefficients of 97% recognition accuracy. It appears that when the number of GMM components is increased to more than 7, there are no significant improvements in recognition rates. This suggests that the proposed method could achieve optimum recognition accuracy using 7 GMM components and 15 DCT coefficients. It has also been demonstrate that the proposed matrix interleaved fusion method is better than conventional feature fusion using concatenation methods. The conventional concatenation method only achieved 89% recognition accuracy when 7 GMM components and 15 DCT coefficients were used, while the proposed method gives 97% recognition accuracy. The analysis shows that feature fusion using matrix interleaved always gives a better result even though concatenation is performed on smaller feature vectors. This analysis shows that the proposed matrix interleaved feature fusion is able to preserve the most important statistical information which results in the most discriminating power. It is also shown that GMM is a suitable statistical tool to capture the underlying statistical characteristics of the fused feature vectors.

In order to validate the proposed method, the results were compared with those of several other multimodal methods using the same FERET and PolyU datasets. The comparison is based on the highest recognition accuracy achieved in previous studies [32], [96] as shown in Table 4.3. The method in [96] used three modalities in order to increase the recognition accuracy to 93.7%, while the method in this study uses only two modalities to achieve 97% recognition accuracy. Another study [32] made use of a kernel method to extract the nonlinear information embedded in the fused feature vector to produce a recognition rate of 92%. The method proposed here makes use of statistical modelling based on GMM and is able to model complex fused data and capture nonlinear information from both modalities, thus producing a better result of 97%.

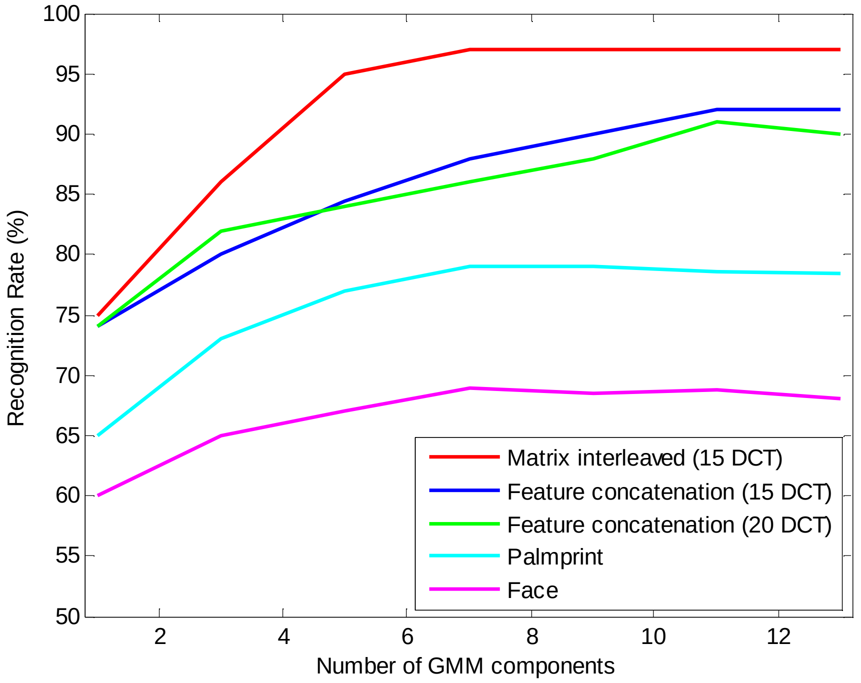


Figure 4.12: Comparison of the proposed method with several existing methods where the number of GMM mixture components is varied from 1 to 13. The highest recognition rate is given by matrix interleaved with 15 DCT coefficients and learnt using 7 GMM mixture components. The experiment is tested using FERET-PolyU dataset.

Table 4.3: Comparison of the highest recognition rates (%) of the propose method and those of existing multimodal fusion methods using face and palmprint images.

Methods	Modalities	Recognition rates
Method by Zhang [96]	FERET-PolyU-USF	93.7%
Method by Jing [32]	FERET-PolyU	92%
Proposed method	FERET-PolyU	97%

4.5 Summary

A new method to fuse the information at feature level is presented in this chapter. The proposed method is based on matrix interleaved feature fusion where local feature vectors of face and palmprint images are combined to form a new fused feature vector. The proposed matrix interleaved feature fusion has several advantages compared to conventional concatenation methods. The number of data points is larger than in conventional methods, and thus GMM is able to accurately estimate the model parameters. The fusion method also produces richer statistical information in the new fused feature vector due to the two ways concatenation method applied in the feature vectors. The experimental analysis using FERET-PolyU and ORL-PolyU datasets shows that the proposed matrix interleaved feature fusion give better recognition results than the conventional concatenation methods.

Chapter 5

Likelihood Score Normalization

5.1 Introduction

This chapter discusses and implements a likelihood normalization method using the likelihood score given by a background model in order to increase the performance of the verification process. The propose method makes use of all of the training data in the database to train the background model, and then its likelihood score is computed to represent the imposter score for each claimed subject. Background model has been used in speaker recognition [115, 117] and faces recognition [78] with very promising results. However, the implementation of a background model to compute the imposter likelihood score in the feature level fusion of multimodal biometrics is new, and this method is expected to be able to offer superior results in the verification process. Thus, this chapter discusses the motivation for using a background model to model the imposter feature distribution. There are two approaches to compute the background model. Firstly, all subjects are represented in a single model called the universal background model [119]. Secondly, each claimed user will have a group of their own background models called cohort background model [120]. Finally, several experimental analyses are conducted to validate the propose method by the comparison of its effectiveness with a baseline method that does not use score normalization.

5.2 Likelihood Score Normalization

A common problem associated with multimodal or single modal biometric recognition is the effect of undesired variation in the input data. Such variation is caused by the effects of data capturing devices and various non-ideal operating conditions such as background noise in face image and sensory noise in palmprint image. Such variations affect biometric matching scores due to the differences between the input and template. This problem can strongly influence the overall usefulness of the biometric recognition process. Thus, an important requirement for the effective operation of a multimodal biometric system is the presence of the capability to minimise the effect of variations in the biometric feature vector. This will then increase recognition accuracy despite the presence of variation caused by contamination in the biometric data involved. This chapter presents an investigation of the effect on the verification accuracy of multimodal biometrics of introducing likelihood score normalization using a background model framework. The main purpose of implementing likelihood score normalization in this research is to differentiate whether the likelihood score value given by a claimed user is known or unknown.

In biometric verification systems the purpose of likelihood score normalization is to separate test feature vectors whether belong to genuine or imposter [117]. This verification is more difficult compared to the identification task even though only a binary decision to accept or reject is required. The verification system will decide if the input features originated from the claimed user, with a well-defined model, or not the claimed user, which is ill-defined. The general hypothesis to be tested states that, for a given input feature Y and a claimed identity, the choice is between H_0 and H_1 such that [115]:

$$H_0\!: Y\ is\ from\ the\ claimed\ user$$

$$H_1\!: Y\ is\ NOT\ from\ a\ claimed\ user$$

The optimum test to decide between H_0 and H_1 required a likelihood ratio test as shown below:

$$\frac{p(Y|H_0)}{p(Y|H_1)}\begin{cases}\geq & \theta \quad accept\ H_0\\ < & \theta \quad reject\ \ H_0\end{cases} \tag{5.1}$$

where $p(Y|H_0)$ is the probability density function for the hypothesis H_0(genuine user) computed for the given test feature vector Y. Meanwhile, $p(Y|H_1)$ is the probability density function for the hypothesis H_1 (imposter user) evaluated for the given test feature vector Y. These two probability density functions are also refer as likelihood of the hypothesis H_i for a given test feature vector. The likelihood score of these hypothesis is then compared with the pre-defined decision threshold θ whether to accept or reject H_0. The crucial task in the successful of the verification process based on likelihood ratios or likelihood score normalization is to determine the appropriate techniques to model and compute the two likelihoods $p(Y|H_0)$ and $p(Y|H_1)$.

The basic components in the verification system based on likelihood score normalization are shown in Figure 5.1. The input to this stage is typically a sequence of independent feature vectors representing important information existing in the biometric modalities. Normally, specific types of pre-processing method are used to generate low dimensional feature vectors and these depend on the type of modality such as 2D biometric image or 1D signal of voice. The pre-processing method used to produce the input feature vector $X = \{x_1, x_2, \dots x_T\}$ must be the same as that used to train the model. This is to make sure that the trained and test feature vectors come from the same feature space.

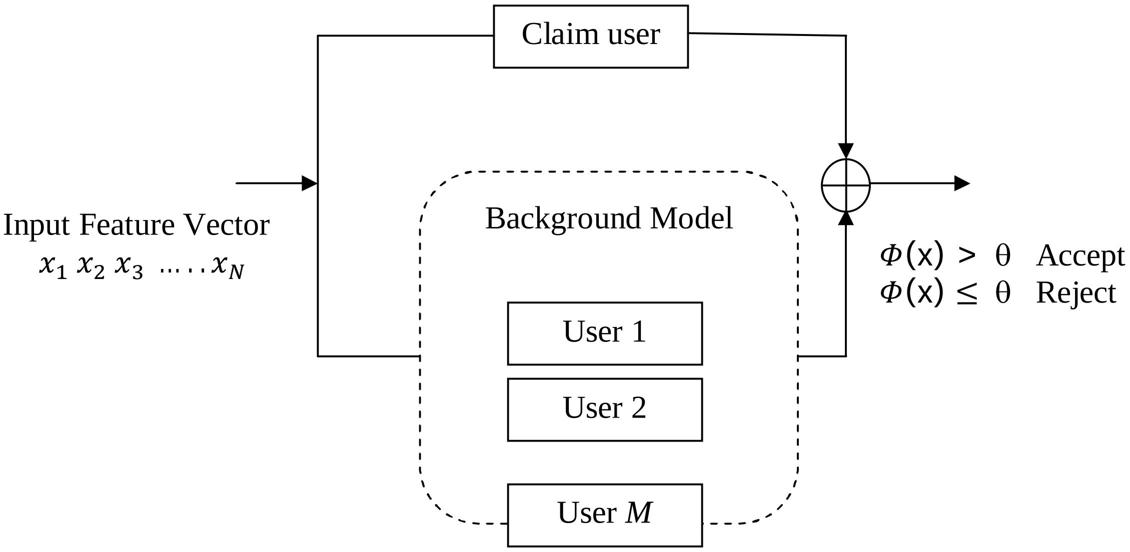

Claim user
Input Feature Vector
$x_1\, x_2\, x_3\, \dots\dots x_N$
Background Model
User 1
User 2
User M
$\Phi(\text{x}) > \theta$ Accept
$\Phi(\text{x}) \leq \theta$ Reject

Figure 5.1: Basic components in the verification system that use likelihood score normalization computed from a user background model.

The input feature vectors in Figure 5.1 are used to calculate the likelihood score of a claimed user whether it is a genuine user (H_0) or not from a claimed user (H_1). Generally, hypothesis H_0 can be represented by a model given by λ_{claim} that characterizes the genuine user in the feature space of *x*. On the other hand, hypothesis H_1 is represented by a model $\lambda_{\overline{claim}}$ which belongs to a degree of certainty that a claimed user is not a genuine user. Such model that can be used to represent λ_{claim} and $\lambda_{\overline{claim}}$ can be a single Gaussian function consisting of the parameter mean and covariance matrix(μ, Σ). Another type model that is able to represent more complex distributions of λ_{claim} and $\lambda_{\overline{claim}}$ is a Gaussian mixture model (GMM), where the model parameters are denoting as weight, mean and covariance matrix$\{\omega_i, \mu_i, \Sigma_i\}_{i=1}^{C}$. By using two of these models, the likelihood ratios of genuine and imposter users can be expressed as $p(X|\lambda_{claim})/p(X|\lambda_{\overline{claim}})$. The likelihood ratio in the logarithm form can be expressed as follows:

$$\Phi(X) = \log p(X|\lambda_{claim}) - \log p(X|\lambda_{\overline{claim}}) \tag{5.2}$$

The likelihood ratio $\Phi(X)$ is then compared to a threshold θ and the claimed user is accepted if $\Phi(X) > \theta$ and rejected if $\Phi(X) \leq \theta$. The likelihood score normalization will measure the degree of similarity between a claimed user likelihood score for a given test feature vectors and a likelihood score of non-claimant model. The success of likelihood score normalization depends on the accuracy of λ_{claim} and $\lambda_{\overline{claim}}$. Given a set of training feature vectors, the model λ_{claim} can be accurately estimated. However, the model for $\lambda_{\overline{claim}}$ is less well defined, it must subsequently represent the entire space of possible alternatives to the claimed user.

In the recent years, there have been several studies on likelihood score normalization method using the Bayesian framework to reduce intra-class variations in biometric verification systems. The general problem in biometric verification is to minimise overlapping features in the distributions of genuine users and impostors, so that it is possible to verify or reject a claimed user with a high degree of confidence using a certain threshold value. To date, a number of normalization techniques using background models have been established, mainly with the aim of attempting the problem of genuine and imposter matching scores overlapping during the verification process. In general, this technique has been successfully applied in the research of speaker recognition [115, 119], and has recently been extended to face recognition [78, 114]. The original approach used this in method, which developed for speaker recognition, is based on the concept that if inconsistent events in the test utterance cause a speaker's scores against the claim model to degrade, then the scores obtained for the same speaker against certain other background models will also be affected in the same way. This will cause the ratio of the score for the target model to a static of scores for the considered background models remains relatively unchanged. Currently, two established types of background model have been developed, which are the universal background model and cohort background model. There are discussed in the next section.

5.3 Universal background model Normalization

This method involves the estimation of parameters in the background model $p(X|\lambda_{\overline{claim}})$ using a pool of features from all training images. This kind of model is normally referred to as a world model [118] or the universal background model [119]. Important information in the background model is captured by using GMM to represent the probability density functions of the distribution of features for all users in the system. In practice, the number of background users should be as large as possible in order to better model imposter distribution. Parameter estimation for a user model can be adapted from the parameters of universal background model which has been established. Figure 5.2 shows the framework of parameter estimation for a user model adapted from the background model.

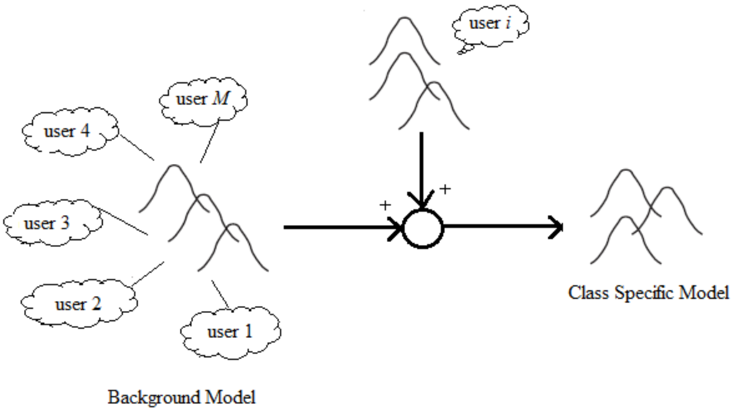


Figure 5.2: Parameter estimation of the user model adapted from the background model.

There are several approaches that can be used to derive the background model when fused feature vectors from all training users are available. The basic approach is to simply pool all features to train the background model using the EM algorithm. Using this method, pooled data must be balanced over the populations of all users so that there is a balance between male and female feature distributions. Lack of balance between male and female users would produce a final model that would be biased towards a dominant population. Another approach is to train the background models based on different subpopulations, such as one for males and one for females and then to pool the subpopulation models. This approach is believed to have advantages when the data in the training population is unbalanced. Two of these approaches are tested under speaker verification, for example when the difference between male and female voices is very significant [115]. In the present method, face and palmprint images from different genders is not a crucial issue, and thus the background model is trained by using pooled data from both genders. The GMM model parameters for a specific class can be derived by adapting the parameters of the background model using the training feature vector from that class. The parameters of class specific model are derived by updating the well trained parameters in the background model via adaptation process. This method provides a tighter coupling between the class specific model and the background model, thus giving a more rapid estimation process which can avoid parameter initialization. In addition, this method is also suitable when there is limited number of feature vectors available to train the class specific model, as shown in the experimental analysis.

The parameter adaptation of the class specific model from the background model parameter consists of two steps. The first is the expectation step in the EM algorithm, where the statistical parameters of the GMM components are estimated from the given class specific feature vectors. In the second step, the new parameters are then combined with the background model parameters using data dependent mixing coefficient. The adaptation process can be explained as follows. Given that a training vector for a specific class is $X = \{x_1, x_2, \ldots . x_T\}$ and parameters of background model $\{\boldsymbol{\omega}_i, \boldsymbol{\mu}_i, \sum_i\}_{i=1}^{C}$, we determined the probability of the training feature vectors to the background model GMM components. The probability of mixture-i in the background model is computed as:

$$Pr(i|x_t) = \frac{\omega_i p_i(\boldsymbol{x}_t)}{\sum_{j=1}^{M} \omega_j p_j(\boldsymbol{x}_t)} \tag{5.3}$$

Then, by using a given feature vectors, the mixture parameters for weight, mean and variance can be estimated by using $Pr(i|\boldsymbol{x}_t)$ as follows:

$$n_i = \sum_{t=1}^{T} Pr(i|x_t) \tag{5.4}$$

$$E_i(\boldsymbol{x}) = \frac{1}{n_i}\sum_{t=1}^{T} Pr(i|x_t)\boldsymbol{x}_t \tag{5.5}$$

$$E_i(\boldsymbol{x}^2) = \frac{1}{n_i}\sum_{t=1}^{T} Pr(i|x_t)\boldsymbol{x}_t^2 \tag{5.6}$$

The new parameter estimated from the training feature vector is then used to update the parameters in the background model. The adaption of class specific model parameter using background model parameters for mixture *i* is given as:

$$\widehat{\omega}_i = [\alpha_i^w n_i / T \ + (1 - \alpha_i^w)\omega_i]\varsigma \tag{5.7}$$

$$\hat{\mu}_i = \alpha_i^m E_i(x) + (1 - \alpha_i^m)\mu_i \tag{5.8}$$

$$\hat{\sigma}_i^2 = \alpha_i^v E_i(x^2) + (1 - \alpha_i^v)(\sigma_i^2 + \mu_i^2) - \hat{\mu}_i^2 \tag{5.9}$$

The scale ς is computed for all parameter of weight in order to make sure the weight sum to unity. The new parameter $\{\widehat{\omega}_i,\ \hat{\mu}_i, \widehat{\sigma}_i^2\}$ to represent the class specific model is adapted by using the old parameter $\{\omega_i, \mu_i, \sigma_i^2\}$ which is first computed from the background model.

The adaptation method based on a background model has been successfully applied in speaker and face recognition [78, 115]. In speaker recognition, the single background model has been trained using pooled of speaker users in the database and the class specific speaker model is computed by using the adaptation method shown in Eq. (5.7)-(5.9). Meanwhile, the adaptation process applied in face verification in [78] only use parameter mean to update the class specific parameter model, while the other parameters are taken from the background model. The classification process in the framework of GMM classifier for both speaker and face modalities are computed using a difference of likelihood scores between class specific model and background model. However, to date, no investigation has been reported into the use of universal background model to compute likelihood score normalization in multimodal biometrics where information is fused at the feature level specifically in face and palmprint multimodal biometrics. Thus the aim of this chapter is to explore the potential usefulness of likelihood score normalization computed using universal background model to increase the accuracy of the verification process. The proposed framework for likelihood score normalization computed using universal background model is shown in Figure 5.3.

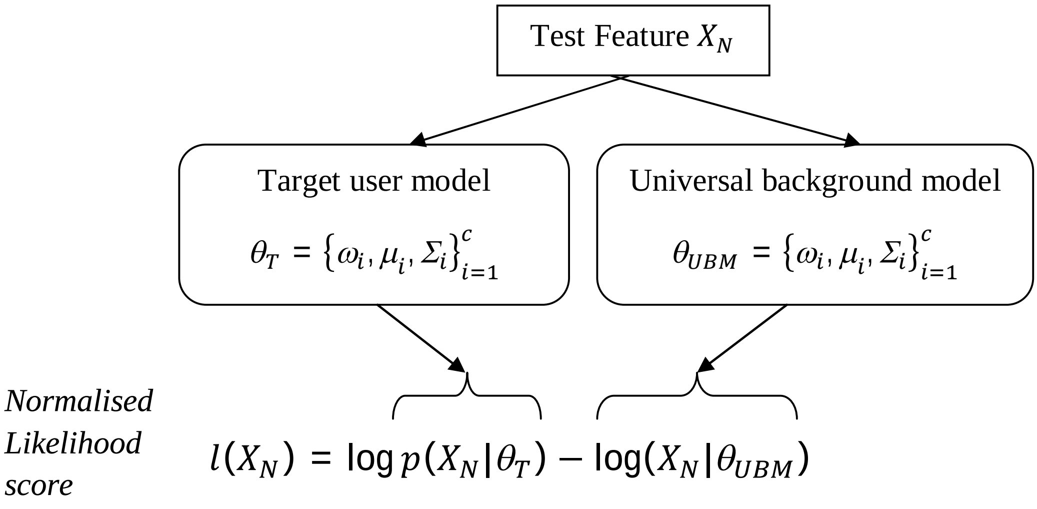

Test Feature X_N
Target user model
$\theta_T = \{\omega_i, \mu_i, \Sigma_i\}_{i=1}^{c}$
Universal background model
$\theta_{UBM} = \{\omega_i, \mu_i, \Sigma_i\}_{i=1}^{c}$
Normalised Likelihood score
$l(X_N) = \log p(X_N|\theta_T) - \log(X_N|\theta_{UBM})$

Figure 5.3: Normalised likelihood score in the universal background model framework.

This method is able to suppress or reduce the likelihood score of an imposter trying to claim true user identity. This is because the total likelihood score computed from the universal background model is high, due to the existing of imposter statistical information in the universal background model. Meanwhile, the likelihood score computed from a claim user model is low due to different statistical information exist in a claim user model. Thus, this situation will produce a low total likelihood score as shown in Eq. (5.2). Using a decision rule in Eq. (5.1), the likelihood scores given by an imposter will be rejected when they are less than a threshold value. On the other hand, when a genuine user trying to claim a true identity, the likelihood score given by a claim user model is high due to the same statistical information exists in both of them. Meanwhile, likelihood score computed from universal background model also high due to the existing of claim user information in the universal background model. However, the likelihood score computed from user model is higher than those computed from the background model, and thus the total likelihood score will be relatively higher than the threshold value. Thus, the system will accept a claimed user as a genuine user by using a decision rule in Eq. (5.2). The next section investigates a different approach to implement likelihood score normalization using cohort of user in the background model.

5.4 Cohort Background Model

This technique works by selecting a cohort of users in background model using likelihood score that are close to that of the likelihood score of claim user model. The selection of cohort user that has a similarity of likelihood score with the claim user was conducted during the testing process. The degree of similarity between the claimed user model and the cohort of user in background model is measured using the likelihood score computed from the test feature vectors. An average likelihood score for the cohort background model is computed using the selected number of background model shown as follows:

$$p(x) \approx \left[\prod_{k=1}^{K} p(x|\lambda_k)\right]^{1/K} \tag{5.10}$$

where $p(x|\lambda_k), k = 1,2, \ldots., K$ is the probability density function with the highest K likelihood scores computed using test feature vector from a set of M (M > K) background models. Thus, the highest likelihood scores of K background models are called competing models and their statistical properties are nearly similar to those of the claimed user model.

The average likelihood score in Eq. (5.10) can be expressed in the log domain in order to simplify the calculation. The likelihood score of background models for cohort background model to represent $p(X|\lambda_{\overline{claim}})$ in Eq. (5.2) can be expressed as follows

$$p_{CBM}(X|\lambda^{ML}, K) = \frac{1}{K}\sum_{k=1}^{K} \log p(X|\lambda_{(k)}) \qquad (5.11)$$

where $X(i) \neq X(j)$ if $i \neq j$ and $\lambda_1, \lambda_2, \ldots, \lambda_K$ are the cohort of user in the background model belongs to the highest K likelihood score after the likelihood score given by the claim user model. The average likelihood score of K potential competing background models are selected from their closeness or similarity with claimed user model by using a given test feature vectors as shown in Eq. (5.11). The advantage of using cohort background model to normalize the likelihood score of the claimed user in multimodal biometric verification is the possibility that it can assist in distinguishing between the scores of genuine users and imposters. This is due to the suppression of total likelihood score when there is an impostor that is trying to claim a true identity. The reason for this is that, for a given type of biometrics and an adequately large set of background models, an impostor targeting a particular client model is more likely going to match one or a few background models. As a result, likelihood score normalization using the cohort background model method is able to suppress the total likelihood score when the claimed user is an imposter instead of a genuine user.

The verification process in the cohort background model framework is depicted in Figure 5.4. For a given test feature vector, the probability of likelihood score of the claimed user is computed using GMM model parameters that are belong to the claimed user model. Then, the likelihood score of the claimed user model is normalized using the average likelihood score computed from a set of K competitive background models. Compared to the universal background model discussed in the previous section, this approach does not need to utilize parameter adaptation from the background models. The model parameters for each class are estimated independently without inference from background model parameters, thus it requires large amount of training feature vectors in order to accurately estimate model parameters. In real applications, biometrics systems always have a problem with small numbers of training images, and thus this method will suffer to accurately estimate the model parameters for each class. The drawbacks of using cohort background model with a small number of training images are discussed in the experimental analysis at the end of this chapter.

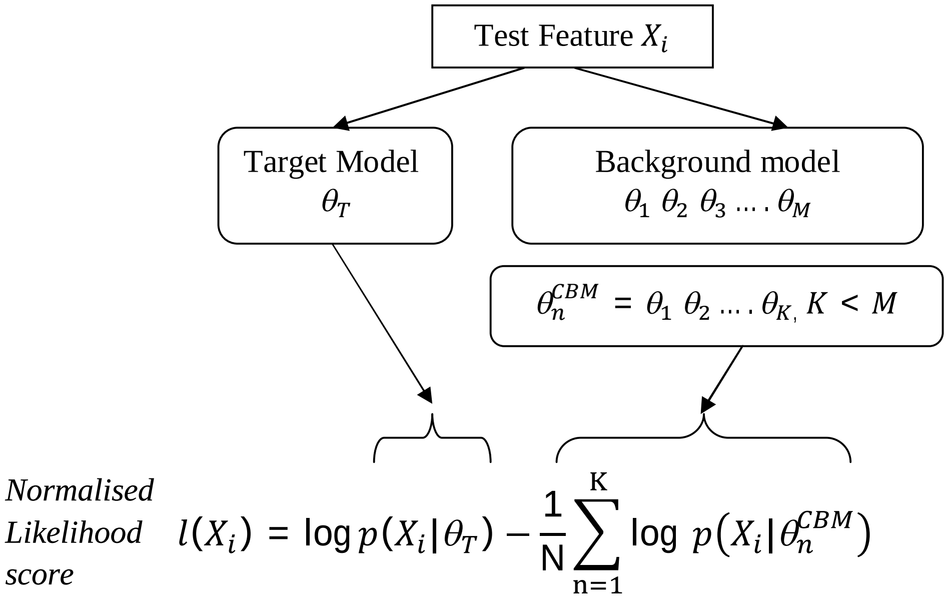

Test Feature X_i
Target Model θ_T
Background model $\theta_1\ \theta_2\ \theta_3 \dots .\ \theta_M$
$\theta_n^{CBM} = \theta_1\ \theta_2 \dots .\ \theta_{K,}\ K < M$
Normalised Likelihood score
$l(X_i) = \log p(X_i|\theta_T) - \frac{1}{N}\sum_{n=1}^{K} \log\ p(X_i|\theta_n^{CBM})$

Figure 5.4: Likelihood scores normalization using cohort background model.

5.5 Experiment Analysis

The performance of the verification system to evaluate the advantage of using likelihood score normalization computed from the background model are measured in terms of EER and the curve of receiver operating characteristic (ROC). The ROC curve is a two dimensional plot showing the percentage of false acceptance rate

(FAR) against false rejection rate (FRR). Meanwhile, the EER is the percentage of errors when FAR is equal to FRR. Thus, a good verification system will try to minimize the value of EER. However, in practice there are several difficulties in reducing EER in real biometrics applications, such as the effect of variations in train and test data and the limitation of classification algorithms to capture the important features. This section evaluates the effectiveness of using a background model to calculate the likelihood score normalization in the classification process to reduce the EER values. A series of experimental studies has been conducted using two sets of virtual multimodal datasets. The ORL-PolyU multimodal dataset consisting of virtual combinations of ORL face images and PolyU palmprint images; and the FERET-PolyU multimodal datasets consisting of virtual combinations of FERET face and PolyU palmprint dataset. These are two benchmark face datasets that are commonly used to evaluate face recognition systems. Meanwhile, the PolyU palmprint dataset [104] is a benchmark dataset that is used in the analysis of palmprint recognition systems.

The virtual multimodal dataset constructed using ORL face and PolyU palmprint dataset consists of 800 images of face and palmprint modalities. These images belong to 40 subjects with 10 images for each subject. In order to construct multimodal dataset, a computer program is used to randomly choose 2 images for training and 8 images for testing from the face and palmprint datasets. Then, these images are paired together to form multimodal datasets. ORL face images consist of frontal face images with several variations in pose, expression and hairstyle. All images are resized to 100 x 90 pixels and the important features are extracted using the technique discussed in Chapter 3. Palmprint images contain middle image of palm area extracted using the ROI method proposed in [104] and the important information in the middle palm texture is extracted using the feature extraction method proposed in Chapter 3. Palmprint images have several variations such as noise, ageing, and different orientations. On the other hand, the virtual multimodal FERET-PolyU dataset is constructed from the FERET face and PolyU palmprint dataset consists of 2400 images from 200 subjects. For each subject, 2 images are used for training and another 4 images are used for testing. Face images in FERET dataset have several variations in terms of pose, expression, ageing, hairstyle and illumination.

5.5.1 Performance Analysis of the Universal Background Model

In this analysis, the effect of using the universal background model to compute the likelihood score normalization to the verification performance is investigated. The experimental work is tested using ORL-PolyU multimodal dataset and FERET-PolyU multimodal dataset. The GMM parameters used to represent universal background model are estimated using the training fused feature vectors for all subjects. Meanwhile, the estimation of GMM parameters to represent each class model is adapted from the universal background model parameters. In this section a series of analyses are performed to investigate the effect of parameter adaptation in GMM components, the effect of using different numbers of training images and the effect of using different numbers of GMM components to estimate the mixture models.

In the first analysis, the effect of using GMM parameter adaptation to estimate the model parameters for each class is investigated. The three parameters in the GMM component are adapted from the background model parameters are weight, mean and covariance matrix. The model parameters for each class are estimated using adaptation framework which used GMM parameters in the universal background model to compute new model parameters. Parameter adaptation can be computed in two separate groups of model parameters. First, only the mean parameter is adapted while the remaining parameters are taken from universal background models, and second, all parameters are adapted using universal background model parameters. This analysis examines two of these adaptation approaches that are able to achieve the highest verification performance. Parameter adaptation in the GMM is an iterative process based on the expectation and maximization steps in the EM algorithm and requires initial parameters in the first step of iterations. In this case, initial parameters are taken from the background model parameters. Figure 5.5 shows verification performance in terms of GAR versus FAR for the different approaches of parameter adaptation. The verification performance of parameter adaptation using means does not give a significant difference compared to the parameter adaptation using mean, weight and covariance. However, the highest verification performance of 97% GAR at 0.01% FAR was achieved using the fusion method utilizing mean parameter adaptation. In this approach, only mean parameters are adapted, while the other parameters are taken from the universal background model parameters. Compared to a baseline method which does not use likelihood score normalization, the verification performance is lower than the propose method. Thus, this shows that likelihood normalization computed from background models in the feature level fusion of multimodal biometrics is able to improve the verification process. The proposed method is also compared with single modal biometrics based on face and palmprint modalities, and the results show that the fusion method achieved superior results. Figure 5.6 shows the same analysis tested using the ORL-PolyU multimodal datasets. The results show that the adaptation of mean parameters gives the highest GAR of 94.5% at 0.01% FAR. Meanwhile, the baseline method which does not use likelihood normalization achieved the lowest verification rates.

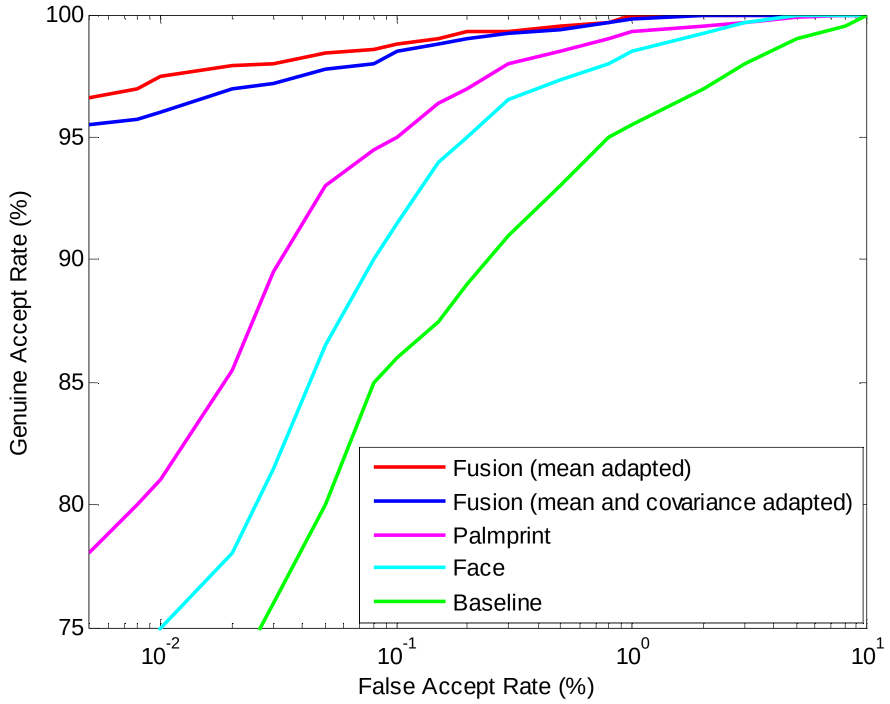


Figure 5.5: ROC curve shows the verification rates (%) when the model parameters are adapted using weight, mean and covariance from the universal background model. The analysis is tested using FERET-PolyU multimodal datasets.

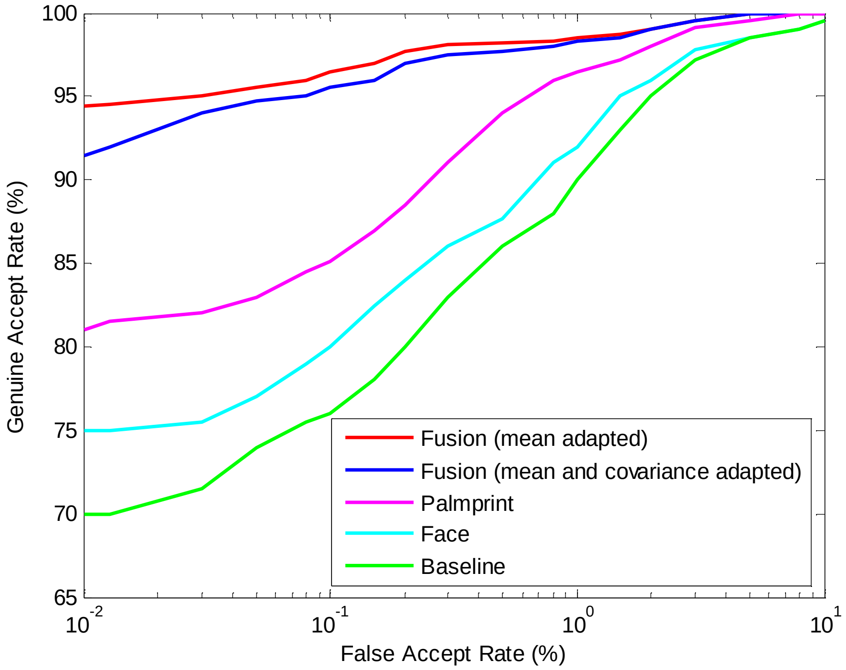


Figure 5.6: ROC curve shows the verification rates (%) when the model parameters are adapted using weight, mean and covariance from the universal background model. The analysis is tested using ORL-PolyU multimodal datasets.

The previous analysis shows that the best verification rates are achieved using the adaptation of mean parameters, and thus in the following analysis only the adaptation of mean parameters is used in order to estimate the model parameters for each class. This analysis investigates the effect of varying the number of GMM components on verification rates. The GMM components vary from 1 to 18 components and performance is measured in terms of EER. This analysis determines the number of GMM components to best able capture the underlying statistical information in the fused feature vectors. The proposed method is compared with a single modal biometrics, baseline method and conventional concatenation method. The result in Figure 5.7 shows that the proposed method requires 10 GMM components to achieve 0.6% EER, which is the lowest of EER of all of the methods. Meanwhile, a baseline method that does not utilize likelihood score normalization achieved higher rates of EER for all numbers of GMM components. From this analysis, likelihood score normalization using the universal background model is able to reduce the imposter likelihood scores if there are enough GMM components. The best verification rate in single modal biometrics is 0.9% EER achieved by using 7 GMM components; however the EER values are still higher than those for the proposed method. A conventional concatenation method achieved 0.7% EER when parameter estimation was performed using 8 GMM components. The proposed method requires 10 GMM components to obtain the best result due to the richness information exists in the fused feature vector. Thus, extra GMM components are required to capture the underlying statistical information in the fused feature vector. Figure 5.8 shows the verification analysis in terms of EER with different numbers of GMM components when tested using the ORL-PolyU multimodal dataset. The lowest EER of 1.5% is achieved using 5 GMM components. The experimental work using this dataset only required 5 GMM components to achieve the best results because the ORL-PolyU dataset is smaller than the FERET-PolyU dataset, for which 10 components were required in order to achieve the best results. Both of these analyses show that likelihood score normalization using the universal background model is able to achieve lower values of EER in both small and large multimodal datasets if sufficient GMM components are used to capture the underlying statistical properties in the fused feature vectors.

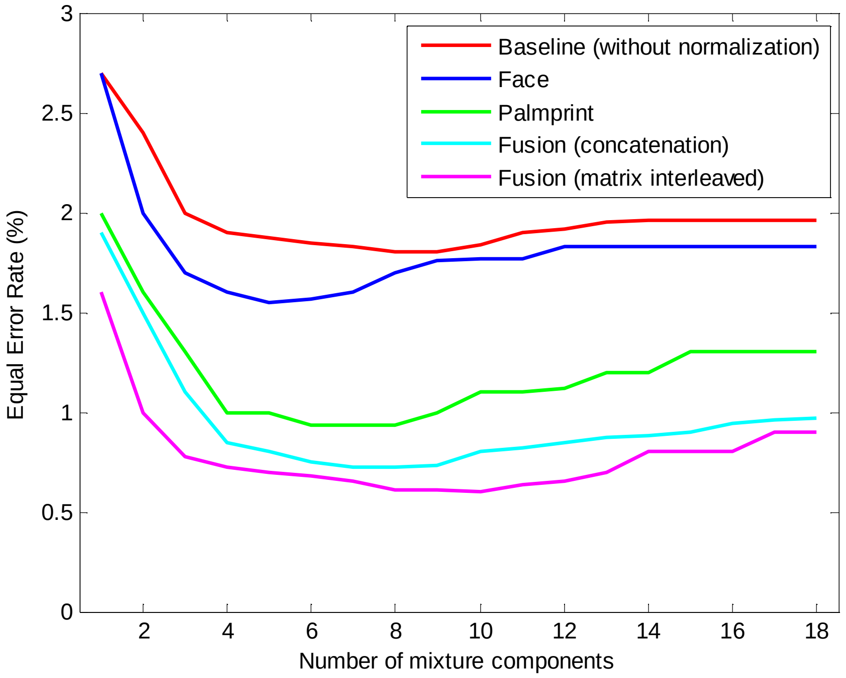


Figure 5.7: Analysis of the effect of the use of universal background model on values of EER (%) for different numbers of GMM mixture components when tested using the FERET-PolyU multimodal dataset.

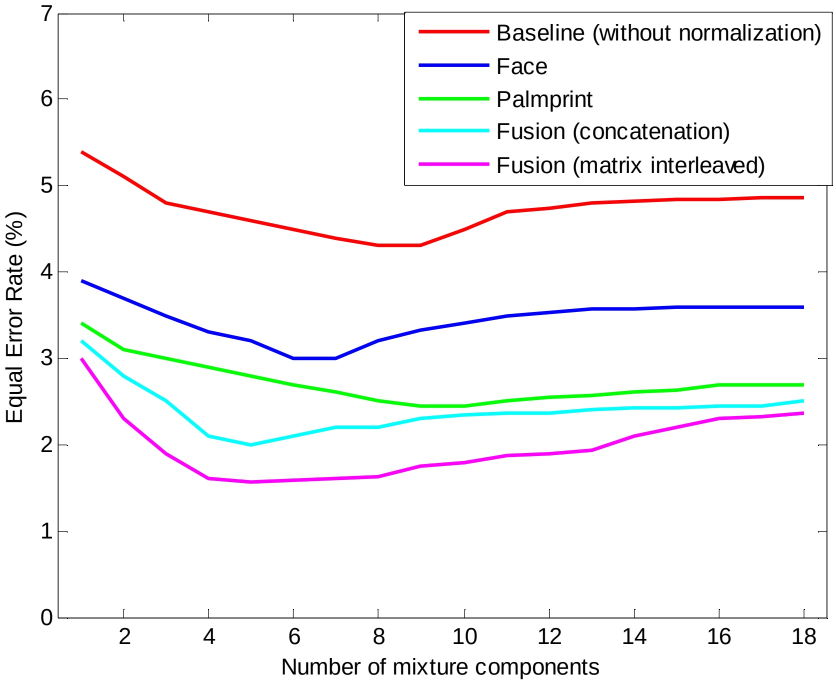


Figure 5.8: Analysis the effect of universal background model in terms of EER (%) for a different number of GMM mixture components when tested using ORL-PolyU multimodal dataset.

The effect on verification performance when the estimation process is performed using different numbers of training images is investigated next. Reducing the number of training images for each subject will decrease the variations in images used to train the background model and user model parameters. These variations contain important information which is important in distinguishing between different persons. Theoretically, parameter estimation using less information should affect the accuracy of the model constructed due to the smaller amount of information available during the training process. However, the estimation of the parameter of user models are affected to a lesser extent due to the parameter adaptation approach utilized in the universal background model. The performance of verification system would be affected by reducing number of training images. The effect on EER values of varying numbers of training images from 1 to 5 on verification rates is investigated. The results in Figure 5.9 show that the proposed method achieved a constant rate of EER of 0.6% when training was performed using different numbers of training images. Moreover, the highest performance was achieved when the estimation was conducted using at least 2 training images for each subject. However, a baseline method using linear projection and a Euclidean distance classifier achieved the lowest EER of 1.1 % when using 5 training images for each subject. The proposed method was also compared with a conventional concatenation method where the concatenated feature vectors are modelled using GMM and a classification process is performed using normalized likelihood score. The best verification rate achieved by conventional concatenation method was 0.75% EER which can only be achieved when more than 4 training images are used. Figure 5.10 shows the values of EER when the training process is performed using different numbers of training images with the ORL-PolyU multimodal dataset. The results are identical to the analysis using the FERET-PolyU dataset, where the proposed method is able to achieve the lowest EER of 1.5% using 2 training images for each subject. The results from both datasets shows that likelihood score normalization using the universal background model consistently achieves the lowest EER when model parameters are estimated using fewer training images. This is due to the adaptation method applied during the estimation process where the background model parameters estimated from a large number of users are not affected by the small number of training images. In the adaptation process the estimation of model parameters for each user model is initiated from a stable background model parameter. Therefore by a given fused feature vectors from a specific class of user, a probability density function for user model will try to fit to an appropriate shape using statistical information exist in the fused feature vector.

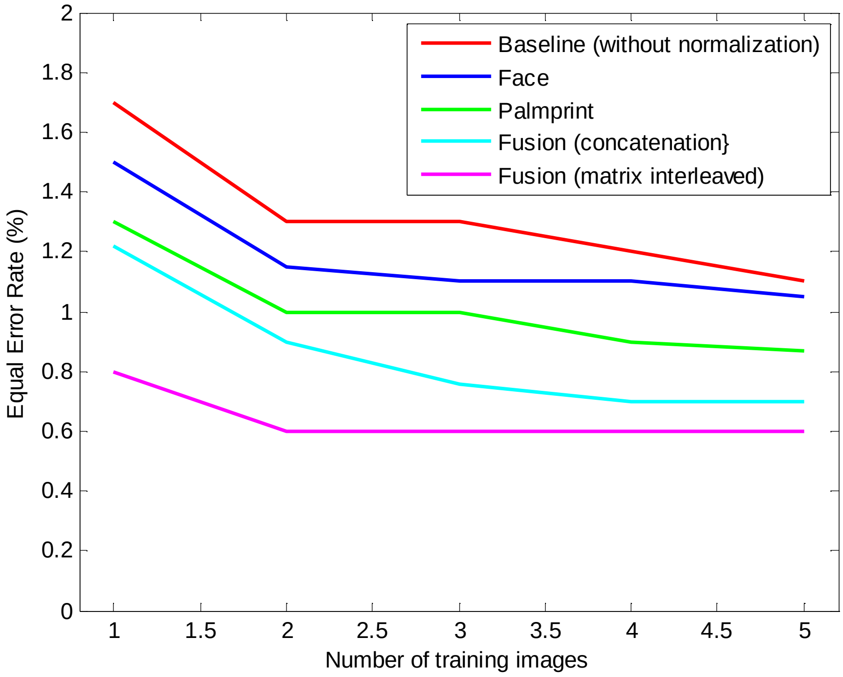


Figure 5.9: Analysis the effect of universal background model in terms of EER (%) for a different number of training images tested using FERET-PolyU multimodal dataset.

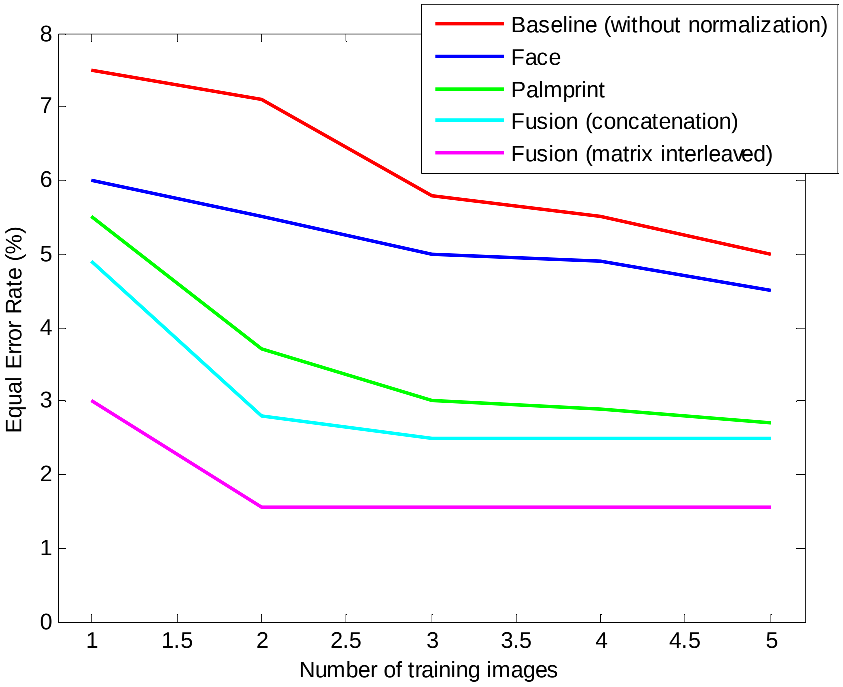


Figure 5.10: Analysis the effect of universal background model in terms of EER (%) for a different number of training images when tested using ORL-PolyU dataset.

5.5.2 Performance Analysis using Cohort Background Model

The effect on verification rates when the background model is constructed using cohort background model is investigated in this section. Likelihood score normalization using cohort background model are computed using average likelihood score from several users in the background models that is believed to have similar statistical information to the claim user. Model parameters in the cohort background model framework are estimated independently using the fused feature vectors for each class of user. Unlike the universal background model, the cohort background model does not require a parameter adaptation process, and thus it is less complex to perform. However, cohort background model has several limitations in terms of verification rates when only a small number of training images is available for the training process. This experimental work analyses the effectiveness of background models using cohort background model.

The first analysis investigates the effect on verification performance of different cohort sizes in the cohort background model. There are several numbers of cohort users used to compute the average likelihood score of background model in order to increase the effectiveness of the system to suppress the imposter likelihood score. In practice, an imposter trying to claim a user’s identity supposed to present in the background model, thus by using several number of cohort users might be able to compute imposter likelihood scores. This can be achieved using several number of the highest likelihood score computed from the background model to compute the average likelihood scores. This analysis determines the cohort size able to achieve the lowest rate of EER. The effectiveness of the proposed method is compared against that of single modal biometrics, the baseline method without likelihood scores normalization and the conventional fusion method using concatenation. The results in Figure 5.11 show that the proposed method needs a cohort size of 10 to achieve the lowest EER which is 0.6% when tested using the FERET-PolyU multimodal dataset. Increasing cohort size then has no further significant effect on the value of EER. Meanwhile, reducing cohort size to less than 10 will increase value of EER. The baseline method that uses likelihood score without likelihood scores normalization in the background model produces higher EERs compared to the proposed method. The conventional concatenation method requires a cohort sizes of 8 to achieve 0.8% EER, which is 0.2% higher than the proposed method. Figure 5.12 show the analysis of EER using different sizes of cohorts tested using the ORL-PolyU multimodal datasets. Likelihood score normalization using the cohort background model requires 7 cohorts to achieve the lowest EER of 1.5%.

Meanwhile, the baseline method that does not use likelihood normalization achieved the highest rates of EER compared to the other methods. This suggests that likelihood normalization using the cohort background model is able to increase the performance of verification in multimodal biometrics systems that use fused feature vectors.

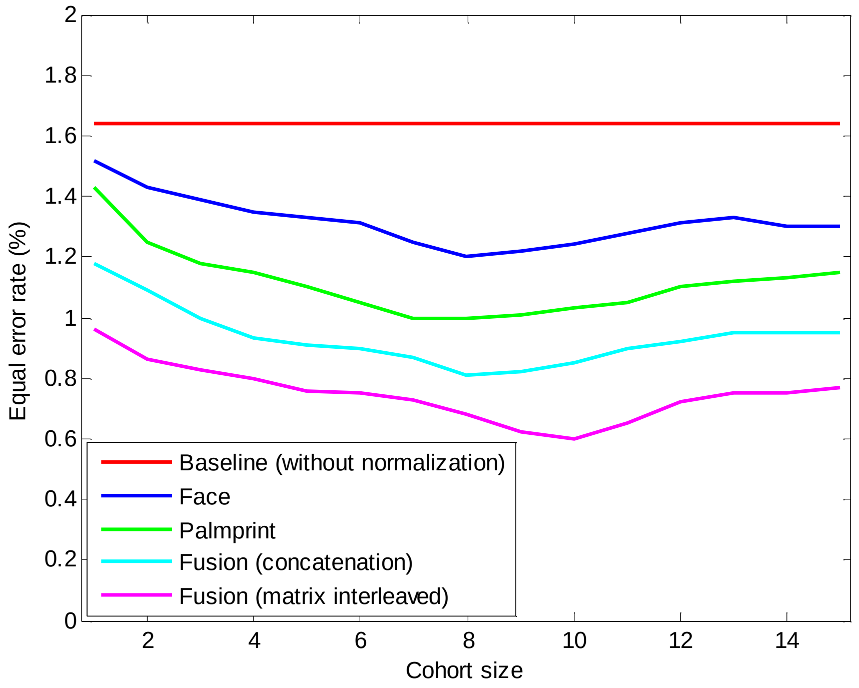


Figure 5.11: Analysis of the effect of cohort background model on EER (%) for a different number of cohort sizes tested using FERET-PolyU multimodal datasets.

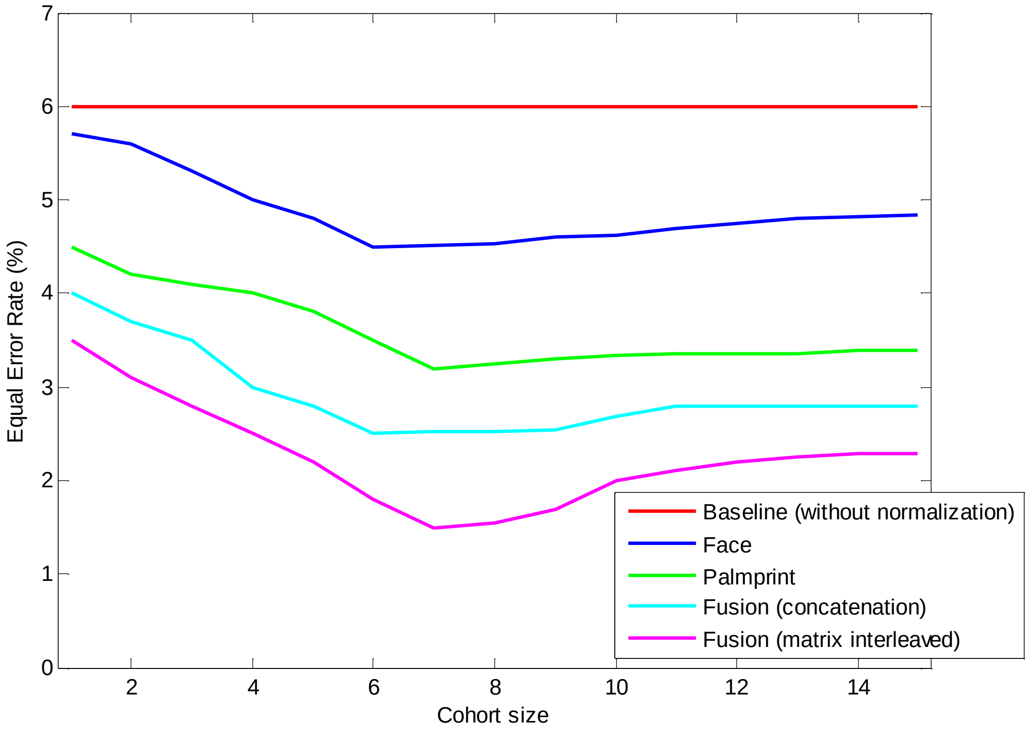


Figure 5.12: Analysis of the effect of cohort background model on EER (%) for a different number of cohort sizes tested using ORL-PolyU multimodal datasets.

The previous, Figure 5.11, analysis found that 10 cohorts is the best size to calculate average likelihood score for the background model in order to achieve the lowest EER. Thus in the following experimental work, 10 cohorts is used in a further analysis. The performance of GMM is investigated using different numbers of mixture components. Using too many of these during the estimation process will increase the complexity of the GMM due to the increasing numbers of parameters in the Gaussian function that need to be estimated. Modelling a probability density function using an appropriate numbers of GMM mixture components should allow the model parameters to be accurately estimated. Meanwhile, using fewer GMM mixture components will produce a probability density function with less discriminative information due to the smaller numbers of Gaussian functions available to model the distribution of the fused feature vectors. This experimental work investigated the best number of GMM mixture components needed to estimate the model parameters which is able to achieve the lowest rates of EER. The number of GMM mixture components is varied from 1 to 18 components and the performance in terms of EER of each number is measured. The proposed method is compared with the baseline method, conventional concatenation and a single modal biometrics system. The result in Figure 5.13 shows that the proposed method achieved 0.6% EER when the estimation of model parameters was performed using 10 GMM mixture components. Increasing the number of GMM mixture components had no significant effect on EER, and thus 10 GMM components is the best number in accurately estimating the model parameters. The lowest EER for the conventional concatenation method was 0.75%, which was achieved using 10 GMM mixture components. However, the baseline method without likelihood normalization achieved the lowest EER of 1.6% using 8 GMM mixture components. The lowest rate of EER for single modal biometrics using face images was 1.3% when the model parameters were estimated using 10 GMM components. Meanwhile, single modal biometrics using palmprint images achieved 1% EER when the model parameters were estimated using 10 GMM mixture components. Single modal biometrics using face and palmprint images required fewer GMM mixture components compared to the proposed method because less statistical information exists in the feature vector. Larger numbers of GMM mixture components are required to accurately estimate the probability density function of the fused feature vectors due to the richness of statistical information which results from the fusion process. The effect of using different numbers of GMM components was also tested using the ORL-PolyU multimodal dataset and the results are shown in Figure 5.14.

In this analysis, 6 GMM mixture components were required to accurately estimate the model parameters in order to achieve 1.5% EER.

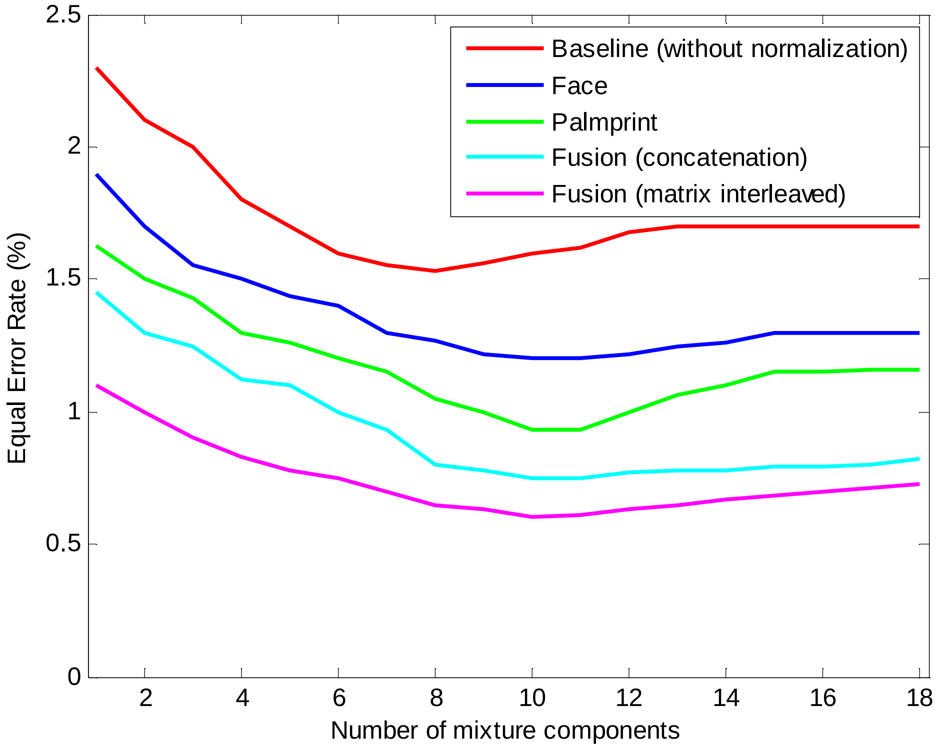


Figure 5.13: Analysis of the effect of cohort background model in terms of EER(%) for a different numbers of GMM mixture components tested using FERET-PolyU multimodal dataset.

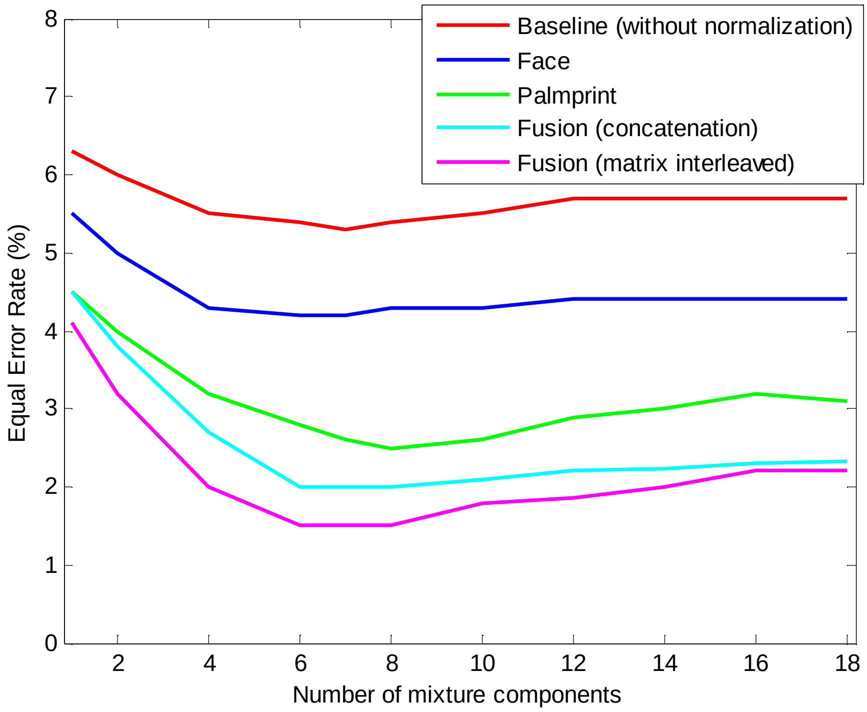


Figure 5.14: Analysis of the effect of cohort background model in terms of EER (%) for a different numbers of GMM mixture components tested using ORL-PolyU multimodal dataset.

Biometric recognition systems sometimes have only limited number of training images during the training process. For some cases, only one image is available during the enrolment process. By using different numbers of training images, the performance of the verification system will be affected, for example, large numbers of training image may produce robust verification whereas a small number of training images will degrade performance. Model estimation using small numbers of training images produces an inaccurate probability density function due to less statistical information existing in the training data. In this analysis, the numbers of training images are varied from 1 to 5 images for each subject and the performance of the system is analysed in terms of EER. In order to validate the effectiveness of the proposed method, the results are compared with those of a baseline method using PCA and a Euclidean distance classifier, conventional concatenation, and single modal biometrics. The results in Figure 5.15 show that the proposed method achieved the lowest EER of 0.6% when training was conducted using 5 training images for each subject. The rates of EER from other methods were higher irrespective of numbers of training images used. When number of the training images is reduced to less than 5 for each subject, the resulting EER, dramatically increase for all methods. This suggests, that performance for all methods depends on the number of training images used. However, the method using the universal background model discussed in the previous section exhibits robust performance in terms of numbers of training images. However, the matrix interleaved fusion method still gives a better result compared to the baseline method and a conventional concatenation with a smaller number of training images. The effect on EER of different numbers of training images using the ORL-PolyU multimodal dataset was also tested. Figure 5.16 show that the lowest EER was achieved using 5 training images for each subject. This result is identical to that of the analysis using the FERET-PolyU multimodal dataset shown in Figure 5.15. The rates of EER are significantly increased when fewer training images for each subject are used. Therefore, we can conclude that the performance of verification systems using cohort background model depends on the number of training images.

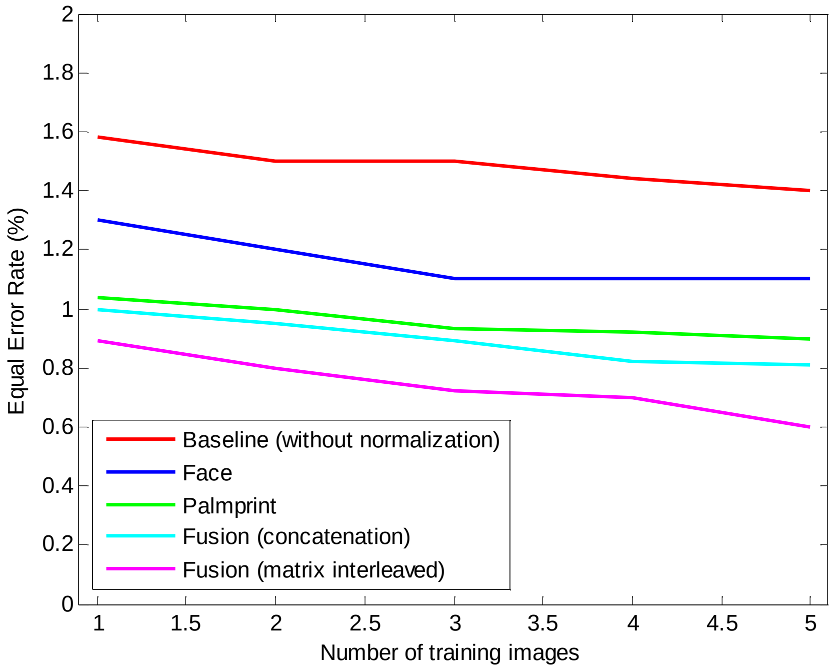


Figure 5.15: Analysis of the effect of cohort background model on EER (%) for different numbers of training images tested using FERET-PolyU dataset.

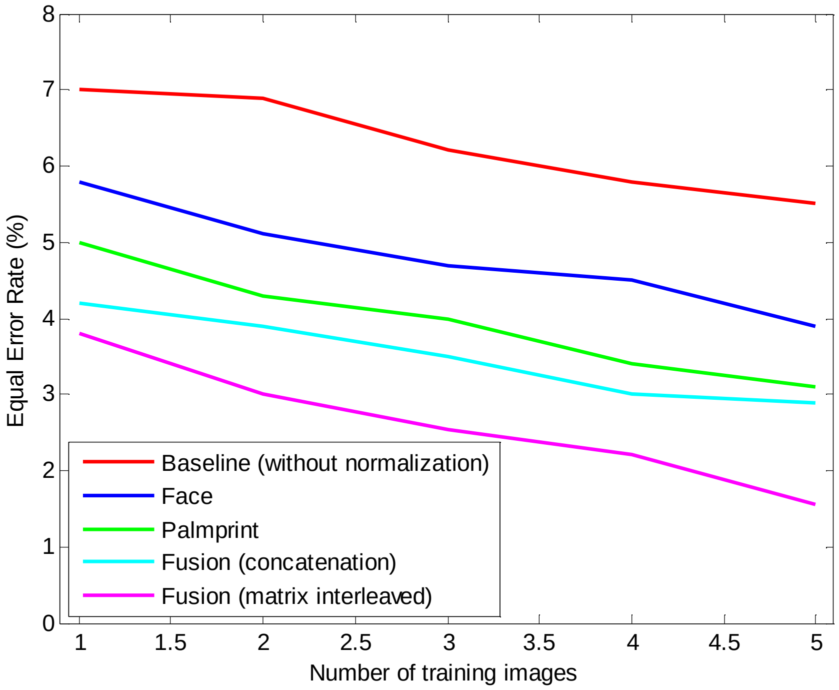


Figure 5.16: Analysis of the effect of cohort background model on EER (%) for different numbers of training images tested using ORL-PolyU dataset.

5.5.3 Comparison between Universal and Cohort Background Model

The previous section has analysed and discussed the advantages of likelihood normalization computed from a background model using the two different approaches of the universal background model and the cohort background model. An extensive experimental analysis has demonstrated that both are able to increase verification rates in multimodal biometrics systems which fuse information at feature level. In other words, both background model methods are able to suppress imposter scores by using the underlying statistical properties which result from the richness of information in the fused feature vectors. However, both of these approaches use different techniques for model parameter estimation. Thus, in this analysis, these background models are compared in order to find the best types of likelihood score normalization to implement in multimodal biometrics systems. A series of experimental studies are conducted using the ORL-PolyU and FERET-PolyU multimodal datasets.

The first analysis evaluates the verification performance for both background models using ROC curves, which plot the rate of FAR versus GAR. To validate the effectiveness of the background models against existing methods, a comparison is made using a baseline method that does not involve likelihood score normalization. Figure 5.17 shows the verification rates of the universal background model, cohort background model and the baseline method when using the FERET-PolyU multimodal dataset. The result show that both background models achieved 23% better verification rates at 0.01% FAR than the baseline method. However, the verification rates for the universal background model and cohort background model are approximately equivalent to each other with the highest rates of GAR of 97% at

0.01% FAR. Figure 5.18 shows the same analysis tested using the ORL-PolyU multimodal dataset. The results are identical to those of the previous analysis using the FERET-PolyU multimodal dataset, with no significant different found in verification rates between the universal background model and the cohort background model.

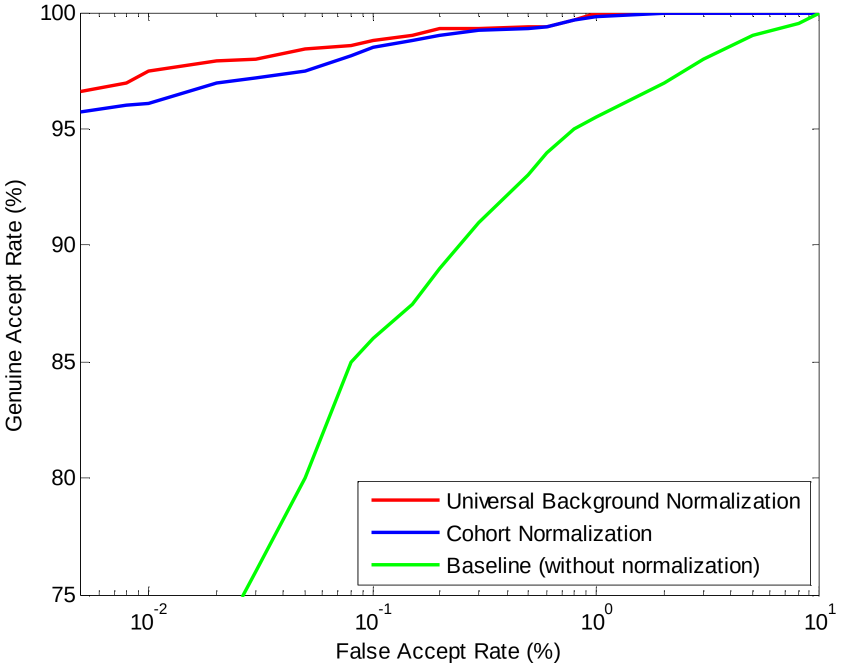


Figure 5.17: Comparison of the verification rate of the universal background model, cohort background model and baseline method in terms of GAR (%) and FAR(%) tested by using FERET-PolyU dataset.

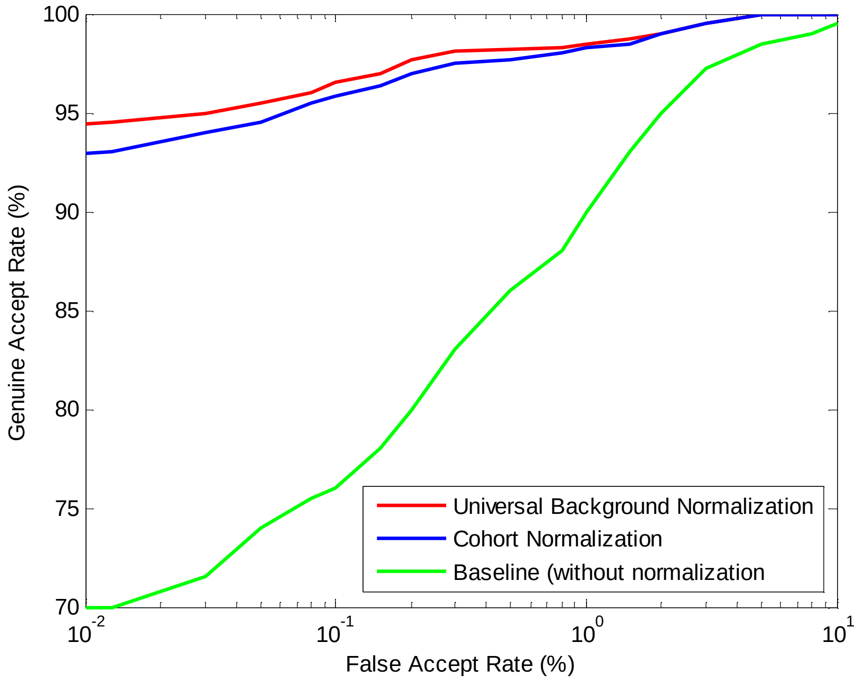


Figure 5.18: Comparison of the verification rates of the universal background model, cohort background model and baseline methods in terms of GAR (%) and FAR(%) tested using ORL-PolyU dataset.

The performance of verification in terms of EER is next investigated using different numbers of training images to estimate the model parameters for the universal background model and cohort background model. Although both background models use the same estimation technique involving the EM algorithm, they use a different approach to choose a background model which is related to the number of training images used for each subject. The universal background model makes use of all users in the training images to train the background model parameters, whereas the cohort background model uses a group of user models to represent these parameters. Because these approaches employ different types of the selection of background models, it is expected that using different numbers of training images will affect their verification rates. This analysis aims to determine which is the best type of background model in terms of being more robust with a limited number of training images. Figure 5.19 shows the verification rates in terms of EER when tested using the FERET-PolyU multimodal dataset with the numbers of training images varied from 1 to 5 images for each subject. The EER of the universal background model is 0.4% lower than the cohort background model when parameter estimation is conducted using 2 training images. However, when using 5 training images for each subject, the performance of cohort background model is equivalent to that of the universal background model of 0.6% EER. From this analysis, it is concluded that parameter estimation in the universal background model is more robust than the cohort background model with smaller number of training images. Figure 5.20 shows the same analysis tested using ORL-PolyU multimodal datasets. The results are identical to those of the previous analysis where the universal background model achieves the lowest EER of 1.5% using 3 training images for each subject. The universal background model provides better verification rates with a small number of training images due to its inclusion of all users in the training of a single background model and then applying adaptation to estimate the model parameters for each user. However in the cohort background model approach, the background model is developed from a specific class user model where the parameters are directly estimated from the training images for each subject. Thus, less statistical information is used to estimate modal parameters with a smaller number of training images.

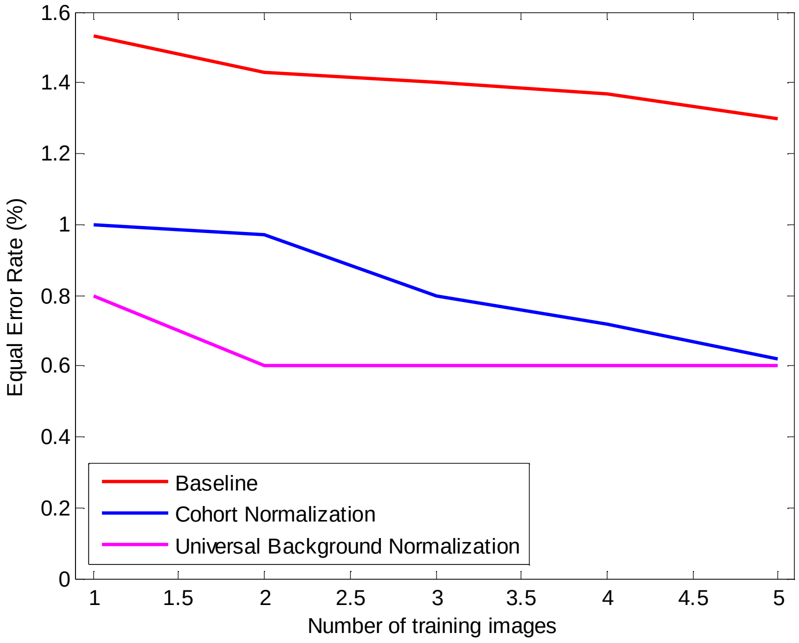


Figure 5.19: Comparison of cohort background model and universal background model in terms of EER (%) with different numbers of training images tested using FERET-PolyU multimodal dataset.

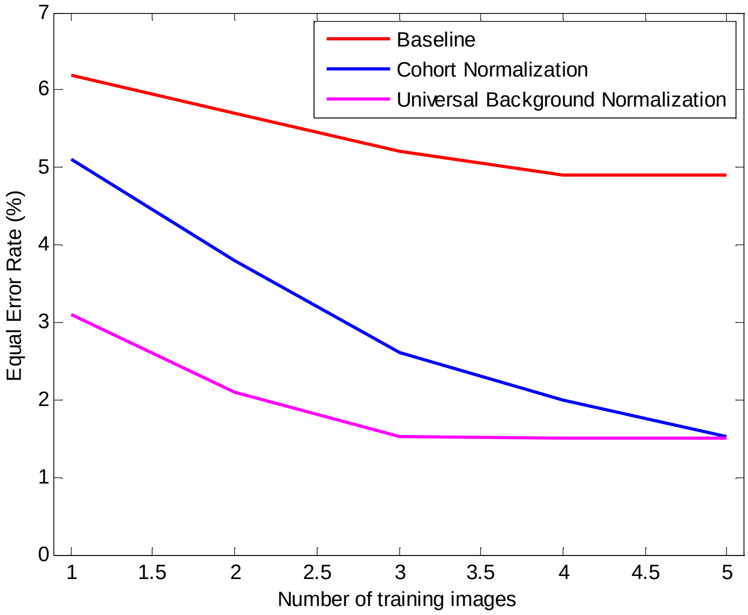


Figure 5.20: Comparison of cohort background model and universal background model in terms of EER (%) with different numbers of training image tested using ORL-PolyU multimodal dataset.

5.6 Summary

A method to increase the verification performance of multimodal biometrics systems that use feature fusion is presented. The framework is developed using likelihood score normalization computed from two types of background models, the universal background model and the cohort background model. Likelihood score normalization approaches show improvement in verification rates due to the ability of the system to suppress the imposter likelihood score. Likelihood score normalization also reduces the effect of variations in the test feature vectors. A series of experimental analyses using the ORL-PolyU and FERET-PolyU multimodal dataset has been conducted to further validate the proposed method. Likelihood score normalization computed from the universal background model is robust compared to the cohort background model with a smaller number of training images. However, with a large number of training images, both background models give superior results compared to the baseline method that does not use likelihood normalization.

Chapter 6

Conclusion and Future Work

This chapter summarise the proposed framework for multimodal biometric fusion at the feature level for face and palmprint modalities. The contributions made in each chapter are first briefly discussed and summarized. However, there still have several questions to be addressed in future work, and recommendations are made concerning potential research directions towards achieving more efficient feature fusion techniques in multimodal biometrics. Overall the work in this thesis has fulfilled the aim and objectives mentioned in Chapter 1.

6.1 Summary and Contribution

Multimodal biometrics is able to achieve better performance than single modal biometrics by consolidating multiple traits in the recognition process. As discussed in Chapter 1, several of the limitations of single modal biometrics can be solved by using multimodal biometrics due to the presence of extra modalities in the recognition system. All of the objectives and aims set out in Chapter 1 have been fulfilled in this thesis. In Chapter 2 the general concepts of multimodal biometric fusion were explained. Information fusion can be carried out at several levels, such as feature level, matching score level, and decision level. The richest information is given by feature level fusion because here integration is performed at an early stage of information fusion. As a result, the thesis investigates a novel method of feature level fusion based on matrix interleaved framework combining face and palmprint features. The feature level fusion produces new feature vectors which contain richer statistical information, and thus an appropriate statistical model is required to capture and estimate model parameters. Thus, a brief discussion was given of the density estimation method based on a parametric model. Model parameter estimation and adaptation is then discussed based on the GMM framework.

In Chapter 3, the feature extraction method for face and palmprint images based on global and local features is discussed. A novel compact local representation of face and palmprint images is proposed where important information in the image is extracted using multiresolution analysis and compact energy representation computed using the DCT transform. The new local features extracted in sub block windows produce independent local feature vectors suitable for the estimation of GMM parameters. It was found that the highest performance of the recognition system was achieved when using only 23% (15 of 64) of DCT coefficients. Moreover, the proposed method does not need to apply pixel overlap to the sub block windows, which has previously been required when local features based on the DCT transform are extracted in sub block windows. The compact feature representation of low frequency components in the DCT transform image only appear in a small number of DCT coefficients. By removing high frequency component in DCT coefficients, the proposed feature extraction method can solve the problem of high dimensionality when two feature vectors are concatenated. The experimental results tested using FERET-PolyU dataset achieved the highest recognition rates of 97% respectively. Meanwhile, the verification analysis shows that the FERET-PolyU analysis achieved an EER of 0.6%.

In Chapter 4 a novel method of feature fusion based on matrix interleaved is proposed to integrate the feature vectors given by face and palmprint images. The proposed method is able to increase the statistical information used compared to conventional concatenation methods when two feature vectors are concatenated and then interleaved. The increased statistical information in the fused feature vector increases the discrimination power of the fused feature vector. Thus, more powerful discrimination can be achieved by this fusion process compared to single modal biometrics and conventional concatenation methods. The method also has advantages during the estimation of the model parameters, due to the existence large number of data points in the training set, which gives more accurate parameter estimation. Even though, matrix interleaved fusion increases the number of data points in the feature space, the information stored in the database for use in the recognition process only consists of the model parameters which includes weight, mean and covariance matrices. The experiment analysis using the FERET-PolyU and ORL-PolyU datasets shows that the proposed matrix interleaved method achieved 97% and 99.7% recognition accuracy, which is higher than the best achieved by conventional concatenation method. A comparison with the use of single modals shows that the proposed method achieved 10% improvement compared to unimodal face and palmprint biometrics respectively.

In chapter 5 the implementation of likelihood score normalization using background model is proposed for the calculation of final likelihood scores in order to increase the performance of the verification process. Likelihood score normalization is constructed by using two different approaches of background model based on unconstraint cohort normalization and universal background model. The advantage of using likelihood normalization in the verification process arises from the ability of the system to suppress the imposter likelihood score with the assumption that the imposter must exist in the large population of the users. Likelihood normalization can also reduce the effect of data variations in the test feature vector. It was found that, with a small number of training images, likelihood normalization based on the universal background model gives better verification results compared to those from CBM due to the existing large number of users to estimate parameters of the background model. Moreover, universal background model use parameter adaptation approach where a new model parameter for a specific class user is adapted from the background model parameters. Parameter estimation of the GMM required many feature vectors to accurately estimate the model parameter, thus in a small number of training images, UBM which depend on feature vectors from a specific class to estimate the model parameter will not have enough information to accurately estimate the model parameter. However, in a large number of training images, both of background model achieve superior result due to the enough information to estimate the model parameters. Finally, from the experimental analysis in this chapter it was found that likelihood score normalization is able to achieve better results compared to verification process without implementing likelihood score normalization in both small and large size of multimodal datasets.

To conclude, this thesis has presented and explained a novel method of multimodal biometric fusion, which cover a new method of feature extraction, a new framework of feature fusion and implementation of likelihood score normalization using background model. The proposed method is able to increase the performance of recognition and verification of biometrics system compared to that of single modal biometrics and conventional concatenation methods. The thesis has accomplished and fulfilling the aims and objective of information fusion of face and palmprint modalities by integrating information at feature level using a new type of compact local feature representation. Using statistical learning method such as GMM is able to capture the underlying statistical properties which exist in the fused feature vector, thus permitting the use of maximum likelihood to measure a degree of certainty that can be used in the classification process. The performance of the verification process is further enhanced by introducing the likelihood score normalization method which can suppress the likelihood score given by an imposter trying to access the system. The proposed method achieved the best recognition rates of 99.7% and 97% when tested on ORL-PolyU and FERET-PolyU datasets. On the other hand, verification analysis shows that the proposed method achieved 1.5% EER when tested on ORL-PolyU datasets. The ROC curve shows that the system achieved 96% GAR at 0.1% FAR when tested on ORL-PolyU datasets. The best EER for FERET-PolyU datasets was 0.6% EER, while the ROC curve shows that the highest GAR of 99% can be obtained at FAR 0.1%.

6.2 Future Work

Based on the work presented in this thesis, there are several possible investigations on the future work that can be initiated. Feature level fusion in multimodal biometric can be extended by several ideas in terms of feature extraction and combination.

This thesis computed the local features in all regions of face and palmprint images for the fusion process. However, some of the region may consist low frequency information thus will have high discrimination power, while the others may contain high frequency information thus not effective for discrimination. Additionally it has been shown that the feature extraction method based on local features extracted from sub block window can be fused to form a new fused feature vector. In the future work, we can pre-process the information from each region to select the region that has high discrimination power and eliminate the region with redundant features. This mechanism perhaps will increase the information in the fused feature vector when the less informative region is removed from the feature vector. Moreover, some regions may also contain intra-class similarity, and removing them could therefore reduce this effect.

The proposed feature extraction and fusion method is designed to deal with grayscale images. The input image is converted to grayscale and then the important information is extracted from the image. This framework could be further extended in the future to deal with colour images that might contain extra information. Fusing information extracted from the red, green and blue components of an image might produce a better fused feature vectors which contains richer information than that in grayscale images.

Another future study could extend this work to other biometric traits such as those of irises, fingerprints and gait. The feature fusion framework discussed in chapter 4 can be generalized so as to be employed with different biometric traits. Some of the biometrics images require similar processing and feature extraction technique as discussed in this thesis. However, other biometric traits such as fingerprint and gait require a different method to transform the features to a compatible form for fusion process such as by generating a statistical map of the extracted feature points distribution in the biometric image. The new feature vectors generated from these types of transformation could then be used in the fusion process.

REFERENCES

1. A. K. Jain, A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4-20, 2004.
2. S. Prabhakar, S. Pankanti, A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security and Privacy,* vol. 1, pp. 33-42, 2003.
3. A. Ross, K. Nandakumar, A. K. Jain, *Handbook of Multibiometrics*, Springer, 2010.
4. M. Golfarelli, D. Maio, D, Maltoni, "On the error-reject trade-off in biometric verification systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 786-796, 1997.
5. J. Fàbregas and M. Faundez-Zanuy, "On-line signature verification system with failure to enrol management," *Pattern Recognition,* vol. 42, pp. 2117-2126, 2009.
6. Y. Shen, J. Bi, J. Wu, Q. Liu, "A two-level source address spoofing prevention based on automatic signature and verification mechanism," *Proceeding of IEEE Symposium on Computers and Communications,* pp. 392-397, 2008.
7. D. Gafurov, E. Snekkenes, P. Bours,"Spoof attacks on gait authentication system," *IEEE Transactions on Information Forensics and Security,* vol. 2, pp. 491-502, 2007.
8. T. Matsumoto,H. Matsumoto, K. Yamaha, S. Hoshino, "Impact of artificial "gummy" fingers on fingerprint systems," *Proc. of The International Society for Optical Engineering,* pp. 275-289, 2002.
9. M. Faundez-Zanuy, "Data fusion in biometrics," *IEEE Aerospace and Electronic Systems Magazine,* vol. 20, pp. 34-38, 2005.
10. A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters,* vol. 24, pp. 2115-2125, 2003.
11. R. Snelick, M. Indovina, J. Yen, A. Mink, "Multimodal biometrics: Issues in design and testing," *Proc of International Conference on Multimedia Interfaces,* pp. 68-72, 2003.
12. R. W. Frischholz and U. Dieckmann, "BioID: A multimodal biometric identification system," *Computer,* vol. 33, pp. 64-68, 2000.

13. M. Faundez-Zanuy, "Data fusion in biometrics," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, pp. 34-38, 2005.

14. A. K. Jain and A. Ross, “Fingerprint mosaicking,” *Proc. Of IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 4064–4067, vol. 4, 2002.

15. A. Jain and A. Ross, "Fingerprint mosaicking," *Proc of International Conference on Acoustics, Speech and Signal Processing,* pp. 4064- 4067, 2002.

16. A. Ross and R. Govindarajan, "Feature level fusion using hand and face biometrics," *Proc of SPIE Conference on Biometric Technology for Human Identification II*. vol. 5779, pp. 196-204, 2005.

17. R. Raghavendra, B. Dorizzi, A. Rao, G. K. Hemantha,"Particle swarm optimization based fusion of near infrared and visible images for improved face verification," *Pattern Recognition,* vol. 44, pp. 401-411, 2011.

18. R. Raghavendra, B. Dorizzi, A. Rao, G. K. Hemantha, "Designing efficient fusion schemes for multimodal biometric systems using face and palmprint," *Pattern Recognition*, vol. 44, pp. 1076-1088, 2011.

19. G. L. Marcialis and F. Roli, "Fingerprint verification by fusion of optical and capacitive sensors," *Pattern Recognition Letters*, vol. 25, pp. 1315-1322, 2004.

20. A. Jain, K. Nandakumar, A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, pp. 2270-2285, 2005.

21. L. Lam and C. Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, vol. 27, pp. 553-568, 1997.

22. L. Lam and C. Y. Suen, "Optimal combinations of pattern classifiers," *Pattern* Recognition Letters, vol. 16, pp. 945-954, 1995.

23. J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, pp. 226-239, 1998.

24. R. Brunelli, D. Falavigna, T. Poggio, L. Stringa, "Automatic person recognition by acoustic and geometric features," *Machine Vision and Applications,* vol. 8, pp. 317-325, 1995.

25. B. Duc, E. S. Bigun, J. Bigun, G. Maitre, S. Fisher, "Fusion of audio and video information for multi modal person authentication," *Pattern Recognition Letters*, vol. 18, pp. 835-843, 1997.

26. L. Hong and A. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1295-1307, 1998.

27. J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia, J. Gonzalez, “A comparative evaluation of fusion strategies for multimodal biometric verification,” *Lecture Notes in Computer Science*, vol. 2688, pp. 830-837, 2003.

28. A. Kumar, D.C.M. Wong, H.C. Shen, A.K. Jain, “Personal verification using palmprint and hand geometry biometric” *Lecture Notes in Computer Science,* vol. 2688, pp. 668-678, 2003.

29. Y. Wang, T. Tan and A. K. Jain,“Combining face and iris biometrics for identity verification,” Lecture Notes in Computer Science, vol. 2688, pp. 805-813, 2003.

30. K. A. Toh, X. Jiang, W. Y. Yau, "Exploiting global and local decisions for multimodal biometrics verification," *IEEE Transactions on Signal Processing,* vol. 52, pp. 3059-3072, 2004.

31. R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 450-455, 2005.

32. X. Y. Jing, Y. F. Yao, D. Zhang *et al.*, “Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometric recognition,” *Pattern Recognition,* vol. 40, no. 11, pp. 3209-3224, 2007.

33. Y. F. Yao, X. Y. Jing, and H. S. Wong, “Face and palmprint feature level fusion for single sample biometrics recognition,” *Neurocomputing,* vol. 70, no. 7-9, pp. 1582-1586, 2007.

34. Y. Fu, Z. Z. Ma, M. Qi *et al.*, “A Novel User-Specific Face and Palmprint Feature Level Fusion,” *Proc. of International Symposium on Intelligent Information Technology Application,* vol. 3, pp. 296-300, 2008.

35. Y. H. Lu, Y. Fu, J. S. Li *et al.*, “A Multi-modal Authentication Method Based on Human Face and Palmprint,” *Proceedings of Second International*

Conference on Future Generation Communication and Networking, Vol.1 and 2, pp. 689-692, 2008.

36. A. Rattani, D. R. Kisku, M. Bicego, M. Tistarelli, "Feature level fusion of face and fingerprint biometrics," *IEEE Conference on Biometrics: Theory, Applications and Systems,* art. no. 4401919, 2007.
37. A. Kong, D. Zhang, and M. Kamel, "Palmprint identification using feature-level fusion," *Pattern Recognition,* vol. 39, pp. 478-487, Mar 2006.
38. H.-G. Wang, W.-Y. Yau, A. Suwandy, and E. Sung, "Person recognition by fusing palmprint and palm vein images based on "Laplacianpalm" representation," *Pattern Recognition,* vol. 41, pp. 1514-1527, 2008.
39. G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition,* vol. 37, pp. 1855-1872, 2004.
40. P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters,* vol. 15, pp. 1119-1125, 1994.
41. S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition," *IEEE Transactions on Image Processing,* vol. 19, pp. 1349-1361, 2010.
42. A. Kumar and D. Zhang, "Personal recognition using hand shape and texture," *IEEE Transactions on Image Processing,* vol. 15, pp. 2454-2461, 2006.
43. R. Duda, P. Hart, and G. Stork, *Pattern Classification*. Wiley, 2001
44. Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition,* Academic Press, 1990.
45. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience,* vol. 3, pp. 71-86, 1991.
46. G. Lu, D. Zhang, K. Wang, "Palmprint recognition using eigenpalms features," *Pattern Recognition Letters,* vol. 24, pp. 1463-1467, 2003.
47. P. N. Belhumeur, J.P. Hespanha, D.J. Kriegman,"Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 711-720, 1997.
48. X. Y. Jing, D. Zhang, Y.Y. Tang, "An improved LDA approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 34, pp. 1942-1951, 2004.
49. X. Wu, D. Zhang, K. Wang, "Fisherpalms based palmprint recognition," *Pattern Recognition Letters,* vol. 24, pp. 2829-2838, 2003.

50. D. Zhang and Z. H. Zhou, "(2D)2 PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing,* vol. 69, pp. 224-231, 2005.

51. X. Y. Jing, H. S. Wong, D. Zhang "Face recognition based on 2D Fisherface approach," *Pattern Recognition,* vol. 39, pp. 707-710, 2006.

52. M. Li and B. Yuan, "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recognition Letters,* vol. 26, pp. 527-532, 2005.

53. P. Nagabhushan, *et al.*, "(2D)2 FLD: An efficient approach for appearance based object recognition," *Neurocomputing,* vol. 69, pp. 934-940, 2006.

54. J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, pp. 131-137, 2004.

55. B. Schölkopf, A. Smola, K.R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation,* vol. 10, pp. 1299-1319, 1998.

56. B. Schölkopf, S. Mika, C.J.C. Burges *et al.,*"Input space versus feature space in kernel-based methods," *IEEE Transactions on Neural Networks,* vol. 10, pp. 1000-1017, 1999.

57. K. I. Kim, K. Jung, H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Processing Letters*, vol. 9, pp. 40-42, 2002.

58. J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, Z. Jin,"KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 230-244, 2005.

59. Q. Liu, H. Lu, S. Ma, "Improving Kernel Fisher Discriminant Analysis for Face Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 42-49, 2004.

60. J. Yang, Z. Jin, J.Y. Yang, D. Zhang, A.F. Frangi, "Essence of kernel Fisher discriminant: KPCA plus LDA," *Pattern Recognition,* vol. 37, pp. 2097-2100, 2004.

61. S. Mika, G. Ratsch,"Fisher discriminant analysis with kernels," *Proc. of IEEE workshop Neural Networks for Signal Processing,* pp. 41-48, 1999.

62. J. Yang, A. F. Frangi, J. Y. Yang, "A new kernel Fisher discriminant algorithm with application to face recognition," *Neurocomputing,* vol. 56, pp. 415-421, 2004.

63. M. Ekinci and M. Aykut, "Gabor-based kernel PCA for palmprint recognition," *Electronics Letters,* vol. 43, pp. 1077-1079, 2007.

64. M. Ekinci and M. Aykut, "Palmprint recognition by applying wavelet-based kernel PCA," *Journal of Computer Science and Technology,* vol. 23, pp. 851-861, 2008.

65. Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform," *International Journal of Computer Vision,* vol. 43, pp. 167-188, 2001.

66. X. Y. Jing and D. Zhang, "A face and palmprint recognition approach based on discriminant DCT feature extraction," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 34, pp. 2405-2415, 2004.

67. W. Li, *et al.*, "Palmprint identification by Fourier transform," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 16, pp. 417-432, 2002.

68. J. H. Lai, P.C. Yuen, G.C. Feng, "Face recognition using holistic Fourier invariant features," *Pattern Recognition,* vol. 34, pp. 95-109, 2001.

69. A. Meraoumia, S. Chitroub, A. Bouridane, “Efficient person identification by fusion of multiple palmprint representations,” vol. 6134 LNCS, pp. 182-191, 2010

70. B. Son and Y. Lee, "The fusion of two user-friendly biometric modalities: Iris and Face," *IEICE Transactions on Information and Systems,* vol. E89-D, pp. 372-376, 2006.

71. B. Son and Y. Lee, "Biometric authentication system using Reduced Joint Feature Vector of iris and face," *Lecture Notes in Computer Science,* vol.3546, pp. 513-522, 2005.

72. A. Noore, R. Singh, M. Vatsa, "Robust memory-efficient data level information fusion of multi-modal biometric images," *Information Fusion,* vol. 8, pp. 337-346, 2007.

73. J. G. Wang, W.Y. Yau, A. Suwandy, E. Sung, "Person recognition by fusing palmprint and palm vein images based on "Laplacianpalm" representation," *Pattern Recognition,* vol. 41, pp. 1531-1544, 2008.

74. S. G. Mallat, "Theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 11, pp. 674-693, 1989.

75. J. Yang and X. Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification," *Pattern Recognition Letters,* vol. 33, pp. 623-628, 2012.

76. S. Anila, N. Devarajan, “Global and local classifiers for face recognition,” European Journal of Scientific Research, vol. 57, pp. 556-566, 2011.

77. C. Sanderson and K. K. Paliwal, "Features for robust face-based identity verification," *Signal Processing,* vol. 83, pp. 931-940, 2003.

78. F. Cardinaux, C. Sanderson, S. Bengio, "User authentication via adapted statistical models of face images," *IEEE Transactions on Signal Processing,* vol. 54, pp. 361-373, 2006.

79. S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation," P*roceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 855-861, 2004.

80. M. Liu, S. Yan, Y. Fu, T.S. Huang, "Flexible X-Y patches for face recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing* **-** *Proceedings*, pp. 2113-2116, 2008.

81. T. Ahonen, A. Hadid, M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 2037-2041, 2006.

82. X. Wang, H. Gong, H. Zhang, B. Li, "Palmprint identification using boosting local binary pattern," *Proceedings of International Conference on Pattern Recognition*, pp. 503-506, 2006.

83. W. Zhang, *et al.*, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition,"*Proceedings of the IEEE International Conference on Computer Vision*, pp. 786-791, 2005.

84. B. Zhang, S. Shan, X. Chen, W. Gao, "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 57-68, 2007.

85. S. Xie, S. Shan, X. Chen, X. Meng, W. Gao,"Learned local Gabor patterns for face representation and recognition," *Signal Processing,* vol. 89, pp. 2333-2344, 2009.

86. J. G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” *Vis. Res.,* vol. 20, pp. 847–856, 1980.

87. C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing,* vol. 11, pp. 467-476, 2002.

88. L. Wiskott, J. M. Fellous, N. Krüger, and C. D. Von Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 775-779, 1997.

89. W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 36, pp. 458-466, 2006.

90. C. Podilchuk and X. Zhang, "Face recognition using DCT-based feature vectors," Proceedings of *IEEE International Conference on Acoustics, Speech and Signal Processing* , pp. 2144-2147, 1996.

91. R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992

92. S. Eickeler, S. Müller, and G. Rigoll, "Recognition of JPEG compressed face images based on statistical methods," *Image and Vision Computing,* vol. 18, pp. 279-287, 2000.

93. C. Sanderson and K. K. Paliwal, "Fast features for face authentication under illumination direction changes," *Pattern Recognition Letters,* vol. 24, pp. 2409-2419, 2003.

94. G. Heusch and S. Marcel, "Face authentication with salient local features and static Bayesian Network," *Lecture Notes in Computer Science* LNCS vol. 4642, pp. 878-887, 2007

95. S. Ribaric and I. Fratric, "A biometric identification system based on eigenpalm and eigenfinger features," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 1698-1709, 2005.

96. T. Zhang, X. Li, D. Tao, and J. Yang, "Multimodal biometrics using geometry preserving projections," *Pattern Recognition,* vol. 41, pp. 805-813, 2008.

97. X. Zhou and B. Bhanu, "Feature fusion of side face and gait for video-based human identification," *Pattern Recognition,* vol. 41, pp. 778-795, 2008.

98. P. Kenny*, et al.*, "Speaker and session variability in GMM-based speaker verification," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 15, pp. 1448-1460, 2007.

99. F. Cardinaux,C. Sanderson, and S. Marcel, “Comparison of MLP and GMM classifiers for face verification on XM2VTS,” *Proc. Audio and Video Based Biometric Person Authentication (AVBPA)*, vol. 2688, pp. 911-920, 2003.

100. C. Seo*, et al.*, "GMM based on local PCA for speaker identification," *Electronics Letters,* vol. 37, pp. 1486-1488, 2001.

101. D. E. Sturim*, et al.*, "Speaker verification using text-constrained Gaussian mixture models," *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP,* pp. I/677-I/680, 2002

102. P. J. Phillips, H. Wechsler, J. Huang, P.J. Rauss,"The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing,* vol. 16, pp. 295-306, 1998.

103. P. J. Phillips, H. Moon, S.A Rizvi, P.J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 1090-1104, 2000.

104. D. Zhang, W.K. Kong, J. You, M. Wong, "Online palmprint identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, pp. 1041-1050, 2003.

105. A. Kong, D. Zhang, M. Kamel, "Palmprint identification using feature-level fusion," *Pattern Recognition,* vol. 39, pp. 478-487, 2006.

106. C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing,* vol. 11, pp. 467-476, 2002.

107. L. Wiskott*, et al.*, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 775-779, 1997.

108. M. I. Ahmad, W. L. Woo, S. S. Dlay, “Multimodal biometric fusion at feature level: Face and palmprint,” *Proc. of International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP) , pp*. 801-804, 2009.

109. A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via EM algorithm,” *Journal of the Royal Statistical Society Series B-Methodological,* vol. 39, no. 1, pp. 1-38, 1977.

110. F. Alsaade, A.M. Ariyaeeinia, A. S. Malegaonkar, S. Pillay, "Enhancement of multimodal biometric segregation using unconstrained cohort normalisation," *Pattern Recognition,* vol. 41, pp. 814-820, 2008.

111. F. Alsaade, A.M. Ariyaeeinia, A. S. Malegaonkar, S. Pillay, "Qualitative fusion of normalised scores in multimodal biometrics," *Pattern Recognition Letters,* vol. 30, pp. 564-569, 2009.

112. K. Nandakumar, Y. Chen, S.C. Dass, A.K. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, pp. 342-347, 2008.

113. R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 450-455, 2005.

114. G. Heusch and S. Marcel, "A novel statistical generative model dedicated to face recognition," *Image and Vision Computing,* vol. 28, pp. 101-110, 2010.

115. D. A. Reynolds, T.F. Quatieri, R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing: A Review Journal,* vol. 10, pp. 19-41, 2000.

116. D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing,* vol. 3, pp. 72-83, 1995.

117. D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech Communication,* vol. 17, pp. 91-108, 1995.

118. R. Auckenthaler, M. Carey, H. Lloyd, "Score normalization for text-independent speaker verification systems," *Digital Signal Processing: A Review Journal,* vol. 10, pp. 42-54, 2000.

119. D. A. Reynolds, “Comparison of background normalization methods for text-independent speaker verification”, *Proc. of Eurospeech’97,* pp. 963-966, 1997

120. A. E. Rosenberg, “The use of cohort normalized scores for speaker verification”. *Proc. of International Conference on Spoken Language Processing (ICSLP’92),* vol. 1, pp. 599-602, 1992.

121. A. M. Ariyaeeinia, “Analysis and comparison of score normalization methods for text-dependent speaker verification”. *Proc. of Eurospeech’97*, pp.1379-1382, 1997.