DEVELOPMENT OF EFFICIENT METHODS FOR FACE RECOGNITION AND MULTIMODAL BIOMETRY

A THESIS

submitted by

ARPITA PATRA

for the award of the degree

of

Master of Science

(by Research)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

CHENNAI 600 036

June 2006

Certificate

This is to certify that the thesis entitled **"Development of Efficient Methods for Face Recognition and Multimodal Biometry"** submitted by **Arpita Patra** to the Indian Institute of Technology, Madras, for the award of the Degree of Master of Science is a record of bonafide research work carried out by her under my supervision and guidance. The contents of this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Madras 600 036 Research Guide

Date: (Dr. Sukhendu Das)

To

My Parents, Sisters

and

Dearest Friend Ashish

Acknowledgments

"Trust in the LORD with all your heart and lean not your own understanding; in all your ways, acknowledge Him, and He will make your paths straight."

First and foremost, I would like to thank my adviser Dr. Sukhendu Das for introducing me to this exciting area of Biometry. His attention to detail, quest for excellence, and love for perfection has inspired to give my best. I am deeply indebted to him for making my M. S. experience a memorable one. It is due to his faith in me that today I am submitting this thesis. He has always given me his time, encouragement, technical feedback and moral support that I needed during my research. It has been my privilege working with him and learning from him.

I am also extremely thankful to Prof. C. Pandu Rangan for his silent contribution. I still remember the day when my heart inundated with pleasure after getting a mail: "All doors are open for my special child". Indeed, I never hesitated to get any help from him and TCS lab. His wonderful lectures on Advanced Data Structure and Algorithm during my masters course are still evergreen in my memory. His dedication towards research really moves me.

I would like to express my gratitude to Prof. R. Kalyana Krishnan for his wonderful lectures during research methodology course. His criticism always inspired me to do something great. Everyone in this department may have a different opinion about him, but I am among those lucky persons who have discovered the presence of a touchy heart within a apparently harsh and tough person.

I would also like to thank my general test committee members, Prof. T. A. Gonsalves (Head of the Department), Dr. Deepak Khemani, Dr. Shrinivasa Chakrabarthy and Dr. Kamakoti for their critical but very valuable comments which has refined my thesis to a great extent.

My sincere thanks to all of our department office staffs Balu, Natarajan, Prema madam and Murali for helping me whenever I needed it. My special thanks to Prema madam for helping me with new books whenever I badly wanted it.

I like to express on my heartiest thanks to all my lab mates: Dyana, Vinod,

Mirnalinee, Shreyasee, Lalit, Manisha, Deepti, Surya, Abhilaash, Shivani, Sunando, Aakansha for all their helps and making our lab a very lively to do research. My sincere thanks to Poongudi for helping me in any official work whenever I needed it.

I would like to thank my parent and sisters for their incredible love, prayers, enthusiasm, and encouragement. My mother, the perfect dream-maker is the source of immense inspiration to me. "To do something great is not just to do well in study but to be a good human being": this is the lesson she taught me. I would also like to mention that I am the happiest person being blessed with such a nice father who has always given me the complete freedom in all respects. He is the first teacher in my life who taught me "Learn to live by struggle. Struggle is the color of life". The celestial smile on the faces of my sisters is my heaven. I can do anything to keep them intact.

Finally, I would like to mention about somebody who played a major role during my research. He is the most valuable gift that God has blessed me. I am extremely fortunate to have his company. He is a true friend and an excellent critic. I don't have any word to thank him. His perception for research redefined my concepts. I will never forget the words , he used to say me: "Do your work, don't think about result". His never-ending inspiration is still my powerhouse. His hard working attitude and his true love towards humanity really changed my view towards life. It is his delightful company which helped me to stand up and express my best after every failure. Every time I come across him I get a feeling of positive jealousy from him. He is none other than my dearest and beloved friend Ashish. May god bless him.

Arpita Patra

List of Abbreviations

AFR	Automatic Face Recognition
AFIS	Automatic Fingerprint Identification System
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
FLD	Fisher Linear Discriminant
2D-PCA	Two Dimensional Principal Component Analysis
2D-LDA	Two Dimensional Linear Discriminant Analysis
DCV	Discriminative Common Vectors
LFA	Local Feature Analysis
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
SVM	Support Vector Machine
RBF	Radial Basis Function
MLP	Multilayer Perceptron
SOM	Self-Organizing Map
PPCA	Probabilistic Principal Component Analysis
DIFS	Distance In Feature Space
DFFS	Distance From Feature Space
SVD	Singular Value Decomposition
ML	Maximum Likelihood
MAP	Maximum a posteriori Probability
ANN	Artificial Neural Network
HMM	Hidden Markov Model
FT	Fourier Transform
DCT	Discrete Cosine Transform
MCC	Multiple Classifier Combination
DP	Decision Profile
DT	Decision Template
DS	Dempster-Shafer Theory of evidence

LDC	Linear Discriminant Classifier
QDC	Quadratic Discriminant Classifier
LOG	Logistic Classifier
NN	Nearest Neighbor
FAR	False Acceptance Rate
FRR	False Rejection Rate
EER	Equal Error Rate

Table of Contents

List of Tables

List of Figures

Abstract

Biometrics is a rapidly evolving technology, which has been widely used in forensics such as criminal identification, secured access, and prison security. A biometric system is essentially a pattern recognition system that recognizes a person by determining the authenticity of a specific physiological and/or behavioral characteristic possessed by that person. Face is one of the commonly acceptable biometrics used by humans in their visual interaction. The challenges in face recognition stem from various issues such as aging, facial expressions, variations in the imaging environment, illumination and pose of the face.

In this thesis, we propose three novel techniques for extraction of facial features and recognition of faces from frontal and near-frontal face images. The principal objective of facial feature extraction is to capture certain discriminative features that are unique for a person. In the first face recognition technique, we propose a new method for representing faces, called as *subband face representation*. Subject-specific subband face extracts features that are invariant within a subject and at the same time distinguishable across different subjects. This method involves the process of selecting suitable subbands of a face, and then reconstructing it using Inverse Discrete Wavelet Transform (IDWT), based on certain criteria. *Subband face representation* has been integrated with recent linear subspace analysis techniques to obtain an efficient face recognition system. Other two proposed face recognition techniques deal with two subspaces, namely, range space and null space of within-class scatter which constitute the entire face space if combined. The range space holds the entire intra-class variations and the null space contains the intra-class commonalities present across samples containing variations in expression, illumination and pose. Two of the proposed methods of face recognition combine discriminative features from null space and range space to utilize the whole gamut of discriminative informations present in the face space, using feature fusion and decision fusion. Feature fusion based method uses Gramm-Schmidt Orthonormalization and covariance sum method to

combine the discriminative informations obtained from null space and range space for constructing a dual space. Then forward or backward selection technique is used to select optimal feature set from dual space. On the other hand, decision fusion based method constructs two different classifiers on the null space and range space separately and then combines them using decision fusion strategies. Results of all the three methods have been shown over three standard databases.

There are severe drawbacks and risks of using a unifeature biometric recognition system, specifically recognition based on fingerprint or face alone. A machine which has the ability to only analyze data, but has no power to determine if the data has been fed into by an authorized person or not, is vulnerable to such acts of impostors who are always on the lookout to create innovative means of breaking into a system. Multimodal biometry which uses multiple sources of information for decision-making is a first choice solution for the above defined problem.

In this thesis, a new approach for combining evidences from face and fingerprint classifiers at the decision level has been proposed. In this approach, each of the face and fingerprint classifier is separately exploited on the basis of availability of class-specific information to enhance combination performance. Results using face and fingerprint databases, show that the proposed methodology of using class-specific information at classifier's response outperforms the state-of-art fusion techniques.

Keywords: Face, dual space, subband face, decision fusion, feature fusion, eigenspace, multimodal, biometry.

CHAPTER 1

Introduction

The issues associated with identity usurpation are currently at the heart of numerous concerns in our modern society. Establishing the identity of an individual is considered as a fundamental requirement for the numerous operations of the state. Three approaches are available to prove a person's identity and to provide "the right person with the right privileges the right access at the right time" [133]. The identity proving approaches to establish the genuineness of the identity are:

- **Something you have:** The associated service or access is received through the presentation of a physical object like, keys, identity card, smart card, etc., in possession of the concerned person.
- **Something you know:** A pre-defined secret knowledge such as, password permits to access a service.
- **Something you are:** The access to a service can be achieved through the presentation of measurable biometric traits, such as biometric measures.

The third approach has some significant advantages over the other two. Without sophisticated means, biometrics are difficult to share, steal or forge and cannot be forgotten or lost. This latter solution provides thus a higher security level in identity prove.

Identity documents are tools that permit the bearers to prove or confirm their identity with a high degree of certainty. In response to the dangers posed by fraudulent use of identity documents, a wide range of biometric technologies is emerging, including e.g. face, fingerprint, iris, hand-geometry etc. Biometric identifiers which are conceptually unique attributes, are today portrayed as the panacea to verify someone's identity.

1.1 Biometric System

Biometrics is a term that encompasses "the application of modern statistical methods to the measurements of biological objects" [4]. Hence, biometric recognition refers to the use of distinctive physiological and behavioral characteristics (e.g. face, fingerprint, hand geometry, iris, gait, signature), called biometric identifiers or simply biometrics, for automatically recognizing a person. This has been used in several domains, such as person authorization examination in e-Banking and e-Commerce transactions or within the framework of access controls for security areas. Ideally the biometric characteristics used should satisfy the following properties:

- **Robustness:** This means that the biometric should be sufficiently invariant (*Permanence*) over a period of time and thus maintain a low *intra-class variability.*
- **Distinctiveness:** This indicates that biometric identifiers should differentiate (*Uniqueness*) any two persons and thus have large *inter-class variability.*
- **Availability:** Ideally, a biometric identifier should be possessed by every person (*Universality*).
- **Accessibility:** The characteristic should be easy to acquire (*Collectability*).

A biometric system is essentially a pattern-recognition system. Such system involves three aspects: data acquisition and preprocessing, data representation, and decision making. It can thus compare a specific set of physiological or behavioral characteristics extracted from a person with a template/model acquired beforehand, and recognize the individual. The digital representation recorded in a database as a description of a physical trait is defined as a template and is obtained by feature extraction algorithms. Among different traits the motivations behind using face and fingerprint for person authentication are manifold. To explain the impetus behind the use of any biometric system, some important applications are described in the following section.

1.1.1 Applications

Biometrics has been widely used in forensics applications such as criminal identification and prison security. It also has a strong potential to be widely adopted in civilian applications such as e-Banking, e-Commerce, and access control. Due to the huge increase in the use of electronic transactions, e-Banking and e-Commerce are becoming the most important emerging application of biometrics. Credit card and smart card security, ATM security, check cashing and fund transfers, online transactions and web access are some of the examples of the applications. The token-based authentication used in physical access control are gradually replaced by biometric systems. Similarly, knowledge-based authentication (e.g. password) for remote login and data access applications are substituted by biometrics. Other biometric applications include welfare disbursement, immigration checkpoints, national ID, voter and driver registration, and time and attendance.

1.1.2 Motivation behind Face-based Biometric System

The face is one of the most acceptable biometrics, and it has also been the most common method of recognition that human use in their visual interactions. The problem with authentication systems based on fingerprint, voice, iris and the most recent gene structure (DNA fingerprint) has been the problem of data acquisition. For example, for fingerprint the concerned person should keep his/her finger in proper position and orientation and in case of speaker recognition the microphone should be kept in proper position and distance from the speaker. But, the method of acquiring face images is nonintrusive and thus face can be used as a biometric trait for covert (where user is unaware that he is being subjected) system.

Face is a universal feature of human beings. Face recognition is important not only due to the capability of its lot of potential applications in research fields but also due to the capability of its solution which would help in solving other classification problems like object recognition.

1.2 Multimodal Biometric System

There are a number of problems involved in a unimodal biometric systems such as non-universality, spoof attacks, etc. These drive the research community to think about more robust and secured biometric system which will be more difficult to fool than a system based on a single biometric. The first choice is obviously a multimodal biometric system which uses multiple biometric traits to offer robust decision-making. The main motivations behind multimodal biometric system are the limitation of unimodal biometric systems in the following ways:

- **Noisy sensed data:** Unimodal biometric systems are very much sensitive to the trait they use. Noisy sensed data affects the performance of a system by getting incorrectly matched with templates in the database. Fingerprint with scar or voice altered by cold, accumulation of dirt or residual fingerprint on fingerprint sensor can result in noisy data.
- **Intra-class variations:** "Intra-class variation" can be explained by the fact that the biometric data acquired during verification will not be identical to the data used for generating user's template during enrollment due to various factors. In case of face, slight change in pose or ambient light and use of spectacles may cause the face to appear different from its template.
- **Lack of individuality:** Features from same biometric trait from different individuals can be quite similar. For face biometric the facial features like, eyes, nose, mouth are in right place for everybody. Moreover, twins or father and son can be identical in facial appearance due to genetic factors. This lack of uniqueness increases the False Acceptance Rate (FAR) of a biometric system.
- **Non-universality:** No biometric is truly universal. For example, fingerprints are supposedly universal, but there is a small fraction of the population that does not possess fingerprint due to hand-related disabilities that are unsuitable for automatic matching. Thus, there is a failure to enroll (FTE) rate associated with using a single biometric trait. It has been empirically estimated that around 4% of fingerprint images have poor quality ridges.

- **Spoof Attacks:** An impostor may attempt to spoof the biometric trait of a legitimate enrolled user in order to circumvent the system. This type of attacks are well-known in case of behavioral traits like signature and voice. Even it is possible to construct dummy fingers using lifted fingerprint impressions.

Multimodal biometry solves the above defined problems by combining the evidences obtained from different modalities with the help of an effective fusion scheme. An alternate use of multimodal biometry is to perform a search in an efficient and faster way by using a relatively simple and less accurate modality to prune the database before using the more complex and accurate modality on the remaining data to generate the final decision.

There are, however, a few disadvantages of using a multimodal biometric system. First, a multimodal biometric system is more expensive and requires more computational and storage resources than a unimodal system. Second, multimodal systems generally require more time for enrollment and verification, causing some inconveniences to the user. Furthermore, if the multiple modalities are not properly combined, the combination may actually degrade a system accuracy.

1.3 Brief Description of the Work Done

The contribution of the work presented in this thesis comes in terms of three novel face recognition approaches and one decision fusion technique for combining information from face and fingerprint classifiers for multimodal biometry.

- Among three face recognition approaches, the first approach presents an efficient method for frontal face recognition, using subject-specific subband face representation. The human face has certain visual features that are common among everybody and some others that exhibit the unique characteristics of an individual. Using the discrete wavelet transform (DWT), we extract these unique features from the face image for discriminating it from others. The face image is decomposed into several subbands to separate the common (approximation) and discriminatory (detail) parts. *Subband face* is reconstructed from selected wavelet subbands, in which a suitable approximation subband is

suppressed and a detail subband (in some cases) is eliminated. Reconstructing a face with an optimal selection of subbands enhances the performance of face recognition. We present four different criteria as cost functions to obtain an optimal subband face for each subject, and compare their performances. The performance of the subband face representation with several linear subspace techniques: PCA, LDA, 2D-PCA, 2D-LDA and Discriminative Common Vectors (DCV), on Yale, ORL and PIE face databases show that the subband face based representation performs significantly better than that proposed by Ekenel for Multiresolution Face Recognition [36] for frontal face recognition, in the presence of varying illumination, expression and pose.

- In the second approach, we propose a new face recognition technique by combining information from null space and range space of within-class scatter of a face space. The combination of information at feature level poses a problem of optimally merging two eigenmodels obtained separately from null space and range space. We use two different methods: 1) Covariance Sum and 2) Gramm-Schmidt Orthonormalization to construct a new combined space, named as dual space, by merging two different sets of discriminatory directions obtained from null space and range space separately. We employ forward or backward selection technique to select the best set of discriminative features from dual space and use them for face recognition. Experimental results on three public databases, Yale, ORL and PIE will show the superiority of our method over a face recognition technique called Discriminative Common Vectors (DCV) [20], based on only the null space of within-class scatter.

- The third approach of face recognition presents a novel face recognition technique by combining information from null space and range space of within-class scatter of a face space at decision level. Along with two classical decision fusion strategies sum rule and product rule, we employ our own decision fusion technique which exploits each classifier space separately to enhance combined performance. Our method of decision fusion uses Linear Discriminant Analysis (LDA) and nonparametric LDA on classifier's response to enhance class separability. This method is also evaluated using Yale, ORL and PIE database.

- A new approach for multimodal biometry has been proposed based on a decision fusion technique to combine decisions from face and fingerprint classifiers. This process of decision fusion exploits the individual classifier space on the basis of availability of class specific information present in the classifier output space. The class specific contextual information is exploited using LDA and nonparametric LDA at the classifier response level to enhance class separability. Eventhough, face classifier is observed to provide contextual information, fingerprint classifier often does not provide this information due to high sensitivity of available minutiae points, producing partial matches across subjects. The enhanced face and fingerprint classifiers are combined using sum rule. We also propose a generalized algorithm for Multiple Classifier Combination (MCC) based on our approach. Experimental results exhibit the superiority of the proposed method over other existing fusion techniques like sum, product, max, min rules, decision template and Dempster-Shafer theory.

1.4 Overview of the Thesis

In this thesis, the problem of face recognition has been attempted using three different techniques and then face and fingerprint information are combined to construct a multimodal biometric system. The rest of the thesis is organized in the following way.

Chapter 2: Literature Review - This chapter discusses the techniques and recent developments in the field of face, fingerprint, and multiple classifier combination. A few methods based on linear transformation for face recognition are also described in this chapter. We also provide a vivid description of a set of methods for Multiple Classifier Combination (MCC).

Chapter 3: An Efficient Method of Face Recognition using Subject-Specific Subband Faces - This chapter emphasizes on our proposed method of face recognition using *Subband face representation*. We propose four criteria and an algorithm for obtaining subject-specific subband faces and then integrate them separately with five linear subspace methods, namely, PCA, LDA, 2D-PCA, 2D-LDA and DCV. Results

are shown on three public databases.

Chapter 4: Dual Space Face Recognition using Feature and Decision Fusion - This chapter introduces dual space based face recognition using feature and decision fusion. The discriminatory information from null space and range space of within-class scatter are combined at feature and decision level separately to obtain better classification performance.

Chapter 5: Enhancing Decision Combination of Face and Fingerprint by Exploitation of Individual Classifier Space: An approach to Multi-modal Biometry - Decision fusion of face and fingerprint is discussed in this chapter. We exploit each classifier space to enhance performance of our combination strategy. We also provide an algorithm for general multiple classifier combination based on our approach.

Chapter 6: Conclusion - This thesis concludes with the summary of the work reported in this thesis, some remarks on our results, discussion on some of the open problems and the scope of further research in this area.

Appendix A: Fingerprint Recognition System - The steps involved in the fingerprint recognition system implemented by us is described in an appendix.

CHAPTER 2

Literature Review

This research work concentrates on face recognition problem as a part of unimodal biometric system and then moves towards the combination of face and fingerprint features to obtain a robust multimodal biometric system. In this chapter, we first present some of the previous work done on face recognition in Section 2.1. Section 2.2, provides a brief literature review on fingerprint recognition, and finally multiple classifier combination (MCC) techniques are reviewed in Section 2.3.

2.1 Recent Approaches to Face Recognition

Face recognition has been an active research area over last 30 years. This research spans several disciplines such as image processing, pattern recognition, computer vision, and neural networks. It has been studied by scientists from different areas of psychophysical sciences and those from different areas of computer sciences. Psychologists and neuroscientists mainly deal with the human perception part of the topic, whereas engineers studying on machine recognition of human faces deal with the computational aspects of face recognition. Face recognition has applications mainly in the fields of biometrics, access control, law enforcement, and security and surveillance systems.

The problem of face recognition can be stated as follows: Given still images or video of a scene, identifying one or more persons in the scene by using a stored database of faces [23]. The problem is mainly a classification problem. Training the face recognition system with images from the known individuals and classifying the newly coming test images into one of the classes is the main aspect of the face recognition systems.

This problem seems to be easily solved by humans where limited memory can be

the main problem; whereas the problems for a machine face recognition system are:

1. Facial expression change
2. Illumination change
3. Aging
4. Pose change
5. Scaling factor (i.e. size of the image)
6. Frontal vs. profile
7. Presence and absence of spectacles, beard, mustache etc.
8. Occlusion due to scarf, mask or obstacles in front.

The problem of automatic face recognition (AFR) is a composite task that involves detection of faces from a cluttered background, facial feature extraction, and face identification. A complete face recognition system has to solve all subproblems, where each one is a separate research problem. This research work concentrates on the problem of facial feature extraction and face identification.

Most of the current face recognition algorithms can be categorized into two classes, image template based and geometry feature-based. The template based methods [9] compute the correlation between a face and one or more model templates to estimate the face identity. Brunelli and Poggio [16] suggest that the optimal strategy for face recognition is holistic and corresponds to template matching. In their study, they compared a geometric feature based technique with a template matching based system and reported an accuracy of 90% for the first one and 100% for the second one on a database of 97 persons. Statistical tools such as Support Vector Machines (SVM) [92, 127], Principal Component Analysis (PCA) [114, 124], Linear Discriminant Analysis (LDA) [12], kernel methods [109, 136], and neural networks [50, 74, 97] have been used to construct a suitable set of face templates. Other than statistical analysis and neural network approach there are other approaches known as hybrid approaches which use both statistical pattern recognition techniques and neural network systems. Examples for hybrid approaches include the combination of PCA and Radial Basis Function (RBF) neural network [37, 122]. Among other methods, people have used

range [23], infrared scanned [137] and profile [79] images for face recognition. While templates can be viewed as features, they mostly capture global features of the face image. Facial occlusion is often difficult to handle in these approaches.

The geometry feature based methods analyze explicit local facial features, and their geometric relationships. Cootes et al. [72] have presented an active shape model in extending the approach by Yuille [139]. Wiskott et al. [131] developed an elastic bunch graph matching algorithm for face recognition. Penev et al. [95] developed PCA into Local Feature Analysis (LFA). This technique is the basis for one of the most successful commercial face recognition systems, FaceIt. The summary of approaches to face recognition is shown in Fig. 2.1.

2.1.1 Template based Methods

Template matching is conceptually related to holistic approach which attempts to identify faces using global representations [51]. These type of methods approach the face image as a whole and try to extract features from the whole face region and then classify the image by applying a pattern classifier. One of the methods used to extract features in a holistic system, is based on statistical approaches which are discussed in the following section.

2.1.1.1 Statistical Approaches

Images of faces, represented as high-dimensional pixel arrays, often belong to a manifold of intrinsically low dimension. Face recognition research has witnessed a growing interest in techniques that capitalize on this observation, and apply algebraic and statistical tools for extraction and analysis of the underlying manifold. The techniques that identify, parameterize and analyze linear subspaces are described below. Other than linear subspaces there are some statistical face recognition techniques which are based on nonlinear subspaces (like kernel-PCA and kernel-LDA), transformation (like DCT, DCT & HMM and Fourier Transform) and Support Vector Machine (SVM).

Appearance-based approaches for face recognition like PCA, LDA, and probabilistic subspace view a 2D face image as a vector in image space. A set of face images $\{x_i\}$ can be represented as a $M \times N$ matrix $X = [x_1, x_2, ..x_i, ..., x_N]$, where

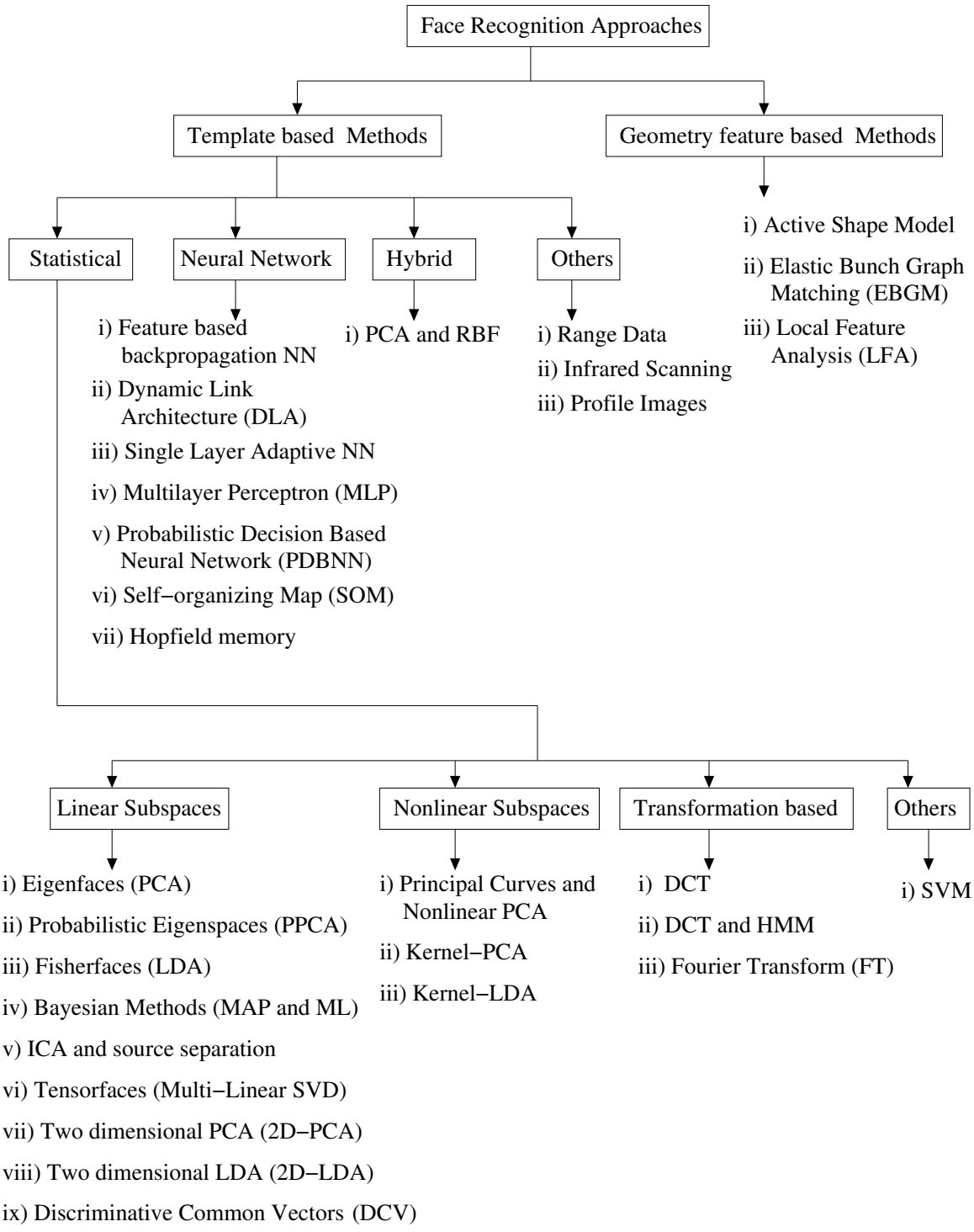
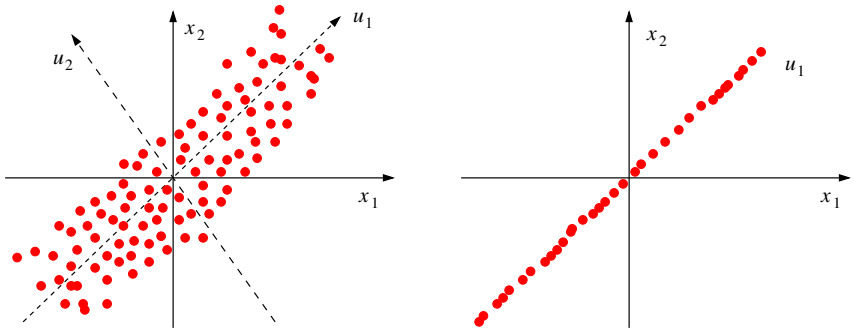


Figure 2.1: Summary of approaches to face recognition.



(a) PCA basis

(b) PCA reduction to 1D

Figure 2.2: The concept of PCA. (a) Solid lines: The original basis; Dashed lines: The PCA basis; Geometric interpretation of principal eigenvectors illustrated in 2D space. (b) The projection (1D reconstruction) of the data using the first principal component.

M is total number of pixels in the images and N is the total number of samples. Each of the face images x_i belongs to one of the C classes $\{1, 2, ..., C\}$.

- **PCA (Principal Component Analysis):** The key idea behind PCA [114, 124] is to find the best set of projection directions in the sample space that maximizes total scatter across all images. This is accomplished by computing a set of eigenfaces from the eigenvectors of total scatter matrix S_t, defined as:

$$S_t = \sum_{i=1}^{N} (x_i - m)(x_i - m)^T, \tag{2.1}$$

where m is the mean face of the sample set X. The geometric interpretation of PCA is shown in Fig. 2.2. For dimensionality reduction, K (where $K < M$) eigenvectors $U = [u_1, u_2, ..., u_K]$ corresponding to first K largest eigenvalues of S_t are selected as eigenfaces. Reduced dimension training samples, $Y = [y_1, y_2, ..., y_N]$ can be obtained by the transformation $Y = U^T X$. Now, when a probe image x_t is presented for identification/verification, it is projected on U to obtain a reduced vector $y_t = U^T x_t$. A response vector of length C, $R(x_t) = [r_1, r_2, \ldots, r_C]$ is calculated by measuring distances from the probe to the nearest training samples from each class. The distance function between

two vectors can be expressed in the following way:

$$d(y_i, y_j) = \|y_i - y_j\|^2. \tag{2.2}$$

The desired class label for the probe image can be obtained by minimum membership rule.

$$\mathcal{L}(x_t) = \arg\min_c r_c. \tag{2.3}$$

- **LDA (Linear Discriminant Analysis):** The objective of LDA is to find the subspace that best discriminates different face classes by maximizing between-class scatter, while minimizing the within-class scatter. The eigenvectors chosen by LDA provide the best separation among the class distributions, while PCA selects eigenvectors which provide best representation of the overall sample distribution. To illustrate the difference, Fig. 2.3 shows the first projection vector chosen by PCA and LDA for a two class problem. The eigenvectors for LDA can be obtained by computing the eigenvectors of $S_w^{-1}S_b$. Here, S_b and S_w are the between-class and within-class scatter matrices of training samples and are defined as:

$$S_w = \sum_{i=1}^{C} \sum_{x_k \in C_i} (x_k - m_i)(x_k - m_i)^T, \tag{2.4}$$

$$S_b = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T. \tag{2.5}$$

where m_i is the mean face for i^{th} class and n_i is the number of training samples in i^{th} class. LDA subspace is spanned by a set of vectors W, which maximizes the criterion, J, defined as:

$$J = \frac{tr(S_b)}{tr(S_w)}. \tag{2.6}$$

W can be constructed by the eigenvectors of $S_w^{-1}S_b$. In most of the image processing applications, the number of training samples is usually less than the dimension of the sample space. This leads to the so-called small-sample-size (SSS) problem due to the singularity of the within-class scatter matrix. To overcome SSS problem, the following approaches are attempted: a two stage PCA+LDA approach [120], Fisherface method [12] and discriminant component analysis [143]. In all cases the higher dimension face data is projected

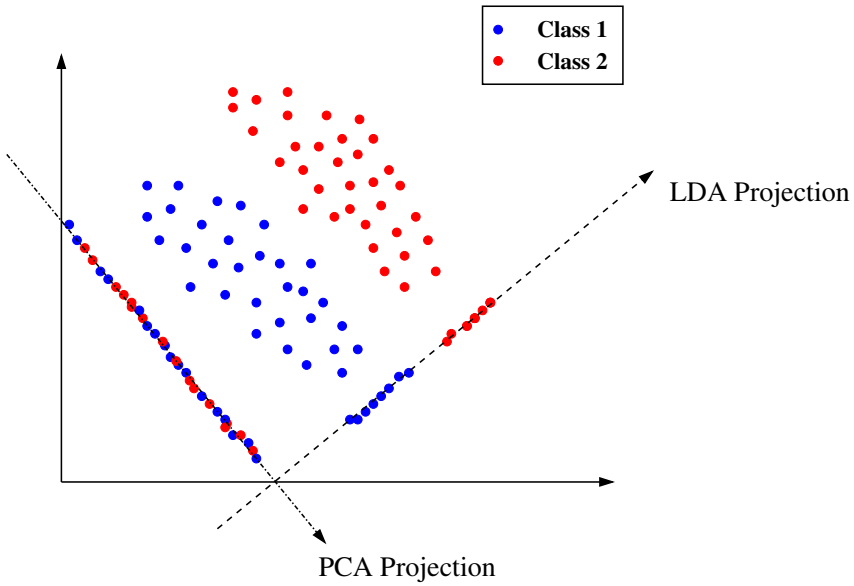


Figure 2.3: An example of PCA and LDA projection for a two class problem.

to a lower dimension space using PCA and then LDA is applied to this PCA subspace.

- **DCV (Discriminative Common Vectors Approach):** DCV [20] solves "small sample size problem" of LDA by optimizing a variant of Fisher's criterion. It searches for the optimal projection vectors in the null space of within-class scatter S_w (see equation 2.4), satisfying the criterion,

$$J(W_{opt}) = \arg \max_{|W^T S_w W|=0} |W^T S_b W| = \arg \max_{|W^T S_w W|=0} |W^T S_t W|. \qquad (2.7)$$

So, to find the optimal projection vectors in the null space of S_w, it projects the face samples onto the null space of S_w to generate common vectors for each class and then obtain the projection vectors by performing PCA on common vectors. A new set of vectors, called as discriminative common vectors, are obtained by projecting face samples on the projection vectors. Thus each class is represented by a single discriminative common vector. Among two algorithms to extract the discriminant common vectors for representing each person in the training set of face database, one algorithm uses within-class scatter matrix of the samples in the training set while the other uses the subspace methods and the Gram-Schmidt orthogonalization procedures to obtain the discriminative

common vectors. These discriminative common vectors are used for classification of new faces.

- **Probabilistic Eigenspace Method:** Probabilistic subspace method models intra-personal and extra-personal variations to classify the face intensity difference Δ as intra-personal variation (Ω_I) for the same class and extra-personal variation (Ω_E) for different classes. The MAP similarity between two images is defined as the intra-personal a posterior probability:

$$S(I_1, I_2) = P(\Omega_I|\Delta) = \frac{P(\Delta|\Omega_I)P(\Omega_I)}{P(\Delta|\Omega_I)P(\Omega_I) + P(\Delta|\Omega_E)P(\Omega_E)} \tag{2.8}$$

To estimate $P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$, the eigenvectors of intra-personal and extra-personal subspaces are computed from the difference set $\{(x_i - x_j)|\mathcal{L}(x_i) = \mathcal{L}(x_j)\}$ and $\{(x_i - x_j)|\mathcal{L}(x_i) \neq \mathcal{L}(x_j)\}$, respectively. The covariance matrices for intra-personal and extra-personal difference sets are defined as:

$$S_I = \sum_{\mathcal{L}(x_i)=\mathcal{L}(x_j)} (x_i - x_j)(x_i - x_j)^T, \tag{2.9}$$

$$S_E = \sum_{\mathcal{L}(x_i)\neq\mathcal{L}(x_j)} (x_i - x_j)(x_i - x_j)^T. \tag{2.10}$$

To estimate $P(\Delta|\Omega_I)$, the eigenspace of S_I is decomposed into intra-personal principal subspace F, spanned by the L largest eigenvectors, and its orthogonal complementary subspace $\bar{F}$, with dimension $M - L$. Then $P(\Delta|\Omega_I)$ can be obtained as the product of two independent marginal Gaussian densities in F and $\bar{F}$,

$$\begin{aligned} P(\Delta|\Omega_I) &= \left[\frac{exp(-\frac{1}{2}d_F(\Delta))}{(2\Pi)^{L/2}\prod_{i=1}^{L}\lambda_i^{1/2}}\right]\left[\frac{exp(-\varepsilon^2(\Delta)/2\rho)}{(2\Pi\rho)^{(N-L)/2}}\right] \\ &= \frac{exp\left[-\frac{1}{2}(d_F(\Delta) + \varepsilon^2(\Delta)/\rho)\right]}{\left[(2\Pi)^{L/2}\prod_{i=1}^{L}\lambda_i^{1/2}\right]\left[(2\Pi\rho)^{(N-L)/2}\right]}. \end{aligned} \tag{2.11}$$

Here, $d_F(\Delta) = \sum_{i=1}^{L}\frac{y_i^2}{\lambda_i}$ is a Mahalanobis distance in F and referred as "distance-in-feature-space" (DIFS). y_i is the principal component of Δ projecting to the i^{th} intra-personal eigenvector, and λ_i is the corresponding eigenvalue. $\varepsilon^2(\Delta)$, defined as "distance-from-feature-space" (DFFS), is the PCA residual (reconstruction error) in $\bar{F}$. ρ is the average eigenvalue in $\bar{F}$. $P(\Delta|\Omega_E)$ can be

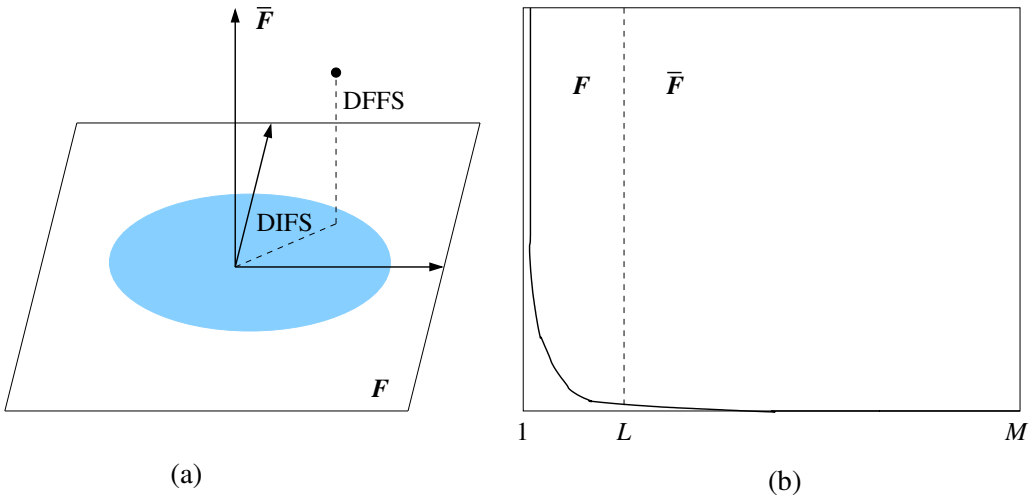


Figure 2.4: (a) Decomposition of $\mathcal{R}^{\mathcal{M}}$ into the principal subspace F and its orthogonal component $\bar{F}$ for a Gaussian density. (b) A typical eigenvalue spectrum and its division into the two orthogonal subspaces.

estimated in a similar way in extra-personal subspace computed from S_E. Fig. 2.4(a) shows the decomposition of Δ into DIFS and DFFS. An alternative maximum likelihood (ML) measure, using the intra-personal likelihood $S'(\Delta) = P(\Delta|\Omega_I)$ is as effective as MAP measure. In face recognition, all parameters in Equation 2.11 are same except $d_F(\Delta)$ and $\varepsilon^2(\Delta)$. So, it is equivalent to evaluate the distance,

$$D_I = d_F(\Delta) + \varepsilon^2(\Delta)/\rho. \tag{2.12}$$

- **2D-PCA (Two Dimensional Principal Component Analysis):** The 2D-PCA technique [141] is based on 2D image matrices instead of 1D vectors. The image covariance matrix is constructed directly from the original image. Let A denote the image with m rows and n columns. The image matrix is projected on to n-dimension column vector x as,

$$y = Ax. \tag{2.13}$$

where y is an m dimensional column vector called the projected feature vector of the image A. Let G_t is the image covariance matrix and is represented as,

$$G_t = E[(A - \bar{A})^T(A - \bar{A})] \tag{2.14}$$

where $E[.]$ is the expectation operator and $\bar{A}$ is the mean image of the entire training set. The column vector x is chosen such that the generalized total scatter criterion is maximized. This means that total scatter of the projected vectors is maximized. It is shown in [141] that the scatter criterion is satisfied when x is chosen as the orthonormal eigenvectors of the image covariance matrix G_t. For each eigenvector x_s, there exists a projected feature vector y_s. Therefore, the principal components of 2D-PCA is a vector unlike that of PCA which is a scalar. If S principal components are used (corresponding to the largest S eigenvalues of G_t), the principal component vectors obtained can be represented as, $B = [y_1, y_2, ..., y_S]$. Thus, B is the feature matrix of the image sample A. For recognition using the nearest neighbor strategy, the distance between any two feature matrices $B_i = [y_1^{(i)}, y_2^{(i)}, ..., y_S^{(i)}]$ and $B_j = [y_1^{(j)}, y_2^{(j)}, ..., y_S^{(j)}]$ is given as,

$$d(B_i, B_j) = \sum_{s=1}^{S} \|y_s^{(i)} - y_s^{(j)}\|_2 \tag{2.15}$$

where $\|y_s^{(i)} - y_s^{(j)}\|_2$ is the Euclidean distance between the two principal component vectors $y_s^{(i)}$ and $y_s^{(j)}$. A test image feature B_t is assigned to class C_k if, $d(B_t, B_l) = \min_j d(B_t, B_j)$, and $B_l \in C_k$. Unlike the conventional PCA, the 2D-PCA does not involve computation of a large correlation matrix and therefore is relatively less computation intensive. But on the other hand, it requires more memory for storing feature matrices.

Yang, Zhang et al. [141] tested 2D-PCA method on ORL, AR and Yale face databases. For the ORL database, the authors used two strategies for experiments: (a) 5 image samples per class for training and (b) leave one out strategy for observing the average performance. In case (a) the recognition rate is 96% and in case of (b) the same is reported to be 98.3% for ORL. Leave one out strategy was adopted for Yale database and maximum accuracy is reported to be 84.24%.

- **2D-LDA (Two dimensional Linear Discriminant Analysis):** In the recently proposed 2D-LDA [77], the image is not reordered as a column vector. The process of projection of an image using 2D-LDA is given as follows. An

image A of m rows and n columns is projected as $y = Ax$, where y is the projected vector and x is the projection vector. Optimal projection vectors are chosen when Fisher's criterion is maximized. The criterion is expressed as,

$$J(x) = \frac{x^T G_b x}{x^T G_w x} \tag{2.16}$$

where G_b and G_w are the *image between-class scatter matrix* and the *image within-class scatter matrix* as given below,

$$G_b = \sum_{k=1}^{C} n_k (\bar{A}^k - \bar{A})^T (\bar{A}^k - \bar{A}) \tag{2.17}$$

$$G_w = \sum_{k=1}^{C} \sum_{A \epsilon C_k} (A - \bar{A}^k)^T (A - \bar{A}^k) \tag{2.18}$$

where $\bar{A}^k$ is the mean image of class C_k and $\bar{A}$ is the global mean. The projection vector x is taken as the eigenvector of $G_w^{-1} G_b$. If the first S eigenvectors are used (corresponding to the largest S eigenvalues of $G_w^{-1} G_b$), the feature obtained can be represented as $B = [y_1, y_2, ..., y_S]$. The classification using nearest neighbor strategy is similar to the classification using 2D-PCA as discussed in the previous section.

Among those earliest to report the work on 2D-LDA was Liu et al. [80]. Kong, Teoh et al. [62] address the SSS (small sample size or undersampled) problem in LDA utilizing a 2D-FDA algorithm. The recognition performance was obtained by varying the number of training samples in the range: 2-9 in case of ORL with maximum accuracy 98%, 2-12 in case of Yale-B with maximum accuracy 92%. The latest of the works on 2D-LDA was by Li, Yuan [77] and Xiong, Swamy, Ahmad [134].

2.1.1.2 Neural Network based Approaches

Artificial Neural Network (ANN)[18, 113, 15, 84] is a powerful tool for pattern recognition problems. The use of neural networks (NN) in faces has addressed several problems: gender classification, face recognition and classification of facial expressions. One of the earliest demonstrations of NN for face recall applications is reported in Kohonen's associative map [61]. Using a small set of face images, accurate recall

was reported even when input image is very noisy or when portions of the images are missing. A few NN based face recognition techniques are discussed in the following.

- **Single Layer adaptive NN:** A single layer adaptive NN (one for each person) for face recognition, expression analysis and face verification was reported in [119]. A system named Wilke, Aleksander and Stonham's recognition devise (WISARD) was devised. It needs typically 200-400 presentations for training each classifier where the training patterns included translation and variation in facial expressions. One classifier was constructed corresponding to one subject in the database. Classification was achieved by determining the classifier that was giving the highest response for the given input image.
- **Multilayer Perceptron (MLP):** Much of the present literature on face recognition with neural networks present results with only a small number of classes (often below 20). In [33] the first 50 principal components of the images were extracted and reduced to five dimensions using autoassociative neural network. The resulting representation was classified using a standard multilayer perceptron (MLP).
- **Self-Organizing map (SOM):** In [73] Lawrence et al. presented a hybrid neural network solution which combines local image sampling, a self-organizing map (SOM) and a convolutional neural network. The SOM provides a quantization of the image samples into a topological space are also nearby in the output space, thereby providing dimensionality reduction and invariance to minor changes in the image sample. The convolutional neural network provides partial invariance to translation, rotation, scale and deformation. The recognizer provides a measure of confidence in its output. The classification error approaches zero when rejecting as few as 10% of the examples on a database of 400 images which contains a high degree of variability in expression, pose and facial details.
- **Hopfield memory model:** In [29], a Hopfield memory model for the facial images is organized and the optimal procedure of learning is determined. A method for face recognition using Hopfield memory model combined with the pattern matching is proposed. It shows better performance of database having

faces of 40 subjects.

- **Others:** A hierarchical neural network which is grown automatically and not trained with gradient descent was used for face recognition by Weng [130]. They reported good results for discrimination of ten subjects. The ability of the compression networks was demonstrated by Cottrell and Fleming in [27]. In [125] linear autoassociative networks, nonlinear autoassociative (or compression) and/or hetero-associative backpropagation networks are explored for face processing. In [78] Lin et al. proposed a face recognition technique based on Probabilistic Decision based Neural network (PDBNN). It adopts a hierarchical network structures with nonlinear basis functions and competitive credit assignment scheme. It demonstrated a successful application of PDBNN on FERET and ORL databases.

  Recently, Gutta et al. [44] described the application of mixtures of experts on gender and ethnic classification of human faces and pose classification and showed their feasibility on the FERET database. The mixture consists of ensembles of radial basis functions (RBFs). Inductive Decision Trees (IDTs) and SVMs implement the "gating network" components for deciding which of the experts should be used to determine the classification output and to restrict the support of the input space. Experimental results yield good results on gender, ethnic and pose classification, which can be effectively used in face recognition.

2.1.1.3 Hybrid Approaches

This type of approaches use both statistical pattern recognition techniques and neural networks.

- **PCA and RBF:** The method by Er et al. [37] suggests the use of RBF on the data extracted by discriminant eigenfeatures. They used a hybrid learning algorithm to decrease the dimension of the search space in the gradient method, which is crucial on optimization of high dimension problem. First, they tried to extract the face features by both PCA and LDA methods. Next, they presented a hybrid learning algorithm to train the RBF Neural Networks, so

the dimension of the search space is significantly decreased in the gradient method.

Thomaz et al. [122] also studied on combining PCA and RBF neural network. Their system for face recognition consists of a PCA stage which inputs the projections of a face image over the principal components into a RBF network acting as a classifier.

2.1.1.4 Other Approaches

- **Range Data:** One of the different methods used in face recognition task is using the range images. In this method data is obtained by scanning the individual with a laser scanner system. This system also has the depth information so the system processes 3-dimensional data to classify face images [23].
- **Infrared Scanning:** Another method used for face recognition is scanning the face image by an infrared light source. Yoshitomi et al. [137] used thermal sensors to detect temperature distribution of a face. In this method, the front-view face in input image is normalized in terms of location and size, followed by measuring the temperature distribution, the locally averaged temperature and the shape factors of face. The measured temperature distribution and the locally averaged temperature are separately used as input data to feed a neural network and supervised classification is used to identify the face. The disadvantage of visible ray image analysis is that the performance is strongly influenced by lighting condition including variation of shadow, reflection and darkness. These can be overcome by the method using infrared rays.
- **Profile Images:** Liposcak and Loncaric [79] worked on profile images instead of frontal images. Their method is based on the representation of the original and morphological derived profile images. Their aim was to use the profile outline that bounds the face and the hair. They take a gray-level profile image and threshold it to produce a binary image representing the face region. They normalize the area and orientation of this shape using dilation and erosion. Then, they simulate hair growth and haircut and produce two new profile silhouettes. From these three profile shapes they obtain the feature vectors. After nor-

malizing the vector components, they use the Euclidean distance measure for measuring the similarity of the feature vectors derived from different profiles.

2.1.2 Geometry Feature based Methods

Geometry feature based methods uses the facial feature measures such as distance between eyes, ratio of distance between eyes and nose etc., but it is significantly different from the feature-based techniques that it constructs the topological graph using the facial features of each subject.

- **Graph Matching based Methods:** In [69] Lades et al. presented a dynamic link architecture for distortion invariant object recognition which employs elastic graph matching to find the closed stored graph. Objects were represented with sparse graphs whose vertices were labeled with geometrical distances. In this system, individual faces were represented by a rectangular graph, each node labeled with a set of complex Gabor wavelet coefficients, called a jet. Only the magnitudes of the coefficients were used for matching and recognition. When recognizing a face of a new image, each graph in the model gallery was matched to the image separately and the best match indicated the recognized person. They presented good results with a database of 87 subjects and test images composed of different expressions and faces turned 15 degree. The matching process was computationally expensive, taking roughly 25 seconds to compare an image with 87 stored objects when using a parallel machine with 23 transputers.

  Wiskott et al. [131] extended this system to handle larger galleries and larger variations in pose and to increase the matching accuracy. Firstly, they use the phase of the complex Gabor wavelet coefficients to achieve an accurate location of the nodes and to disambiguate patterns which would be similar in the magnitudes of the coefficient. Secondly, they employ object adapted graphs, so that nodes refer to specific facial landmarks, called fiducial points. The correspondences between two faces can be found across large viewpoint changes. Thirdly, a new data structure called the bunch graph was introduced which serves as generalized representation of faces by combining jets of a small

set of individual faces. This allows the system to find the fiducial points in one matching process, which eliminates the need for matching each model graph individually. This also reduces computational effort significantly. It offers good performance of about 98% for FERET databse. But the drawback in this feature matching approach is that it requires manual intervention to select the fiducial points in the facial image and it requires precise location of those points.

- **Feature based PCA:** Cagnoni and Poggi [19] suggested a feature based approach instead of a holistic approach to face recognition. They applied the eigenface method to sub-images (eye, nose and mouth). They also applied a rotation correction to the faces in order to obtain better results.

2.2 Recent Approaches to Fingerprint Recognition

Among all biometric traits, fingerprint has one of the highest levels of reliability [13] and has been extensively used by forensic experts in criminal investigations [17]. A fingerprint refers to the flow of ridge patterns in the tip of the finger. The ridge flow exhibits anomalies in local regions of the fingertip, and it is the position and orientation of these anomalies that are used to represent and match fingerprints. Although the word "fingerprint" is popularly perceived as synonymous with individuality, uniqueness of fingerprints is not an established fact but an empirical observation. With the stipulation of widespread use of fingerprints, there is a rightfully growing public concern about the scientific basis underlying individuality of fingerprints. Automated fingerprint matching systems do not use the entire discriminatory information in the fingerprints, but only a parsimonious representation extracted by a machine unsupervised by human fingerprint experts.

However, fingerprints are believed to be unique across individuals and across fingers of the same individual [93]. Even identical twins having similar DNA, are believed to have different fingerprints [54]. The manual method for fingerprint identification is a tedium and monotonous work because of the demand imposed by the painstaking attention to visually match the fingerprint of varied qualities. All these prompted for the initiation and development of Automatic Fingerprint Identification Systems

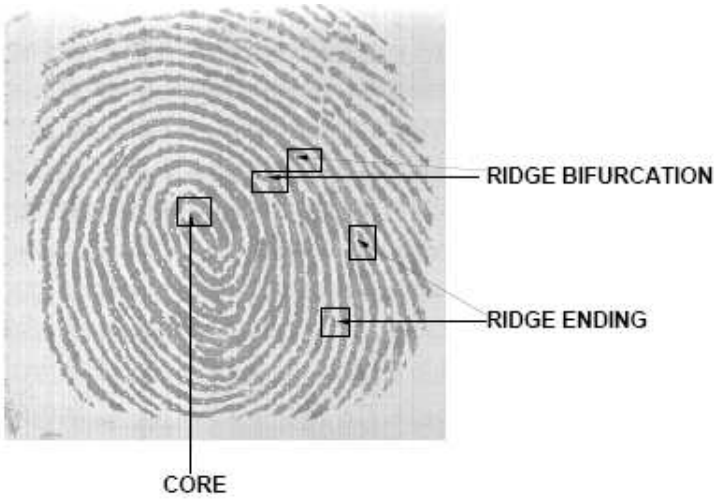


Figure 2.5: A fingerprint image with the core and four minutiae points marked on it.

(AFIS). The task of fingerprint identification is very complex and can be decomposed into the following stages:

- Preprocessing
- Feature extraction
- Matching

2.2.1 Preprocessing

In this stage the fingerprint is processed prior to feature extraction in order to eliminate noise and to enhance the features used in next stages. Fingerprint preprocessing involves,

- Normalization
- Image segmentation
- Image enhancement

2.2.1.1 Image Segmentation

Normalization is a contrast enhancement process and is used prior to image segmentation. The term segmentation is generally used to denote the separation of fingerprint

area (foreground) from the image background. Separating the fingerprint area is useful to avoid extraction of features in noisy areas of the fingerprint and background. It is desirable that the background and foreground regions be identified at the earliest possible stage so that the subsequent processing can effectively concentrate on the foreground region of the image. Thus segmentation prior to other steps saves processing time and cost. Furthermore, it enhances the performance of feature extraction modules.

Because fingerprint images are striated patterns, using global or local thresholding technique [43] does not allow the fingerprint area to be effectively isolated. In fingerprint, what really discriminates foreground and background is not the average image intensities but the presence of a striped and oriented pattern in the foreground and of an isotropic pattern (which does not have a dominant direction) in the background. In practice, the presence of noise (due to the dust on the surface of live scan fingerprint scanners) requires more robust segmentation techniques.

Mehre et al. [89] isolated the fingerprint area according to local histograms of ridge orientations. Ridge orientation is estimated at each pixel and a histogram is computed for each 16×16 block. The presence of a significant peak in a histogram denotes an oriented pattern, whereas a flat or near-flat histogram is characteristic of an isotropic signal. The above method fails when a perfectly uniform block is encountered (e.g. a while block in background) because no local ridge orientation may be found. To deal with this case Mehre and Chatterjee [88] proposed a composite method that, besides histograms of orientations, computes the gray-scale variance blocks and, in the absence of reliable information from the histograms, assigns the low-variance blocks to the background.

Ratha, Chen, and Jain [101] assigned each 16×16 block to the foreground or the background according to the variance of gray-levels in the orthogonal direction to the ridge orientation. They also derive a quality index from the block variance. The underlying assumption is that the noisy regions have no directional dependence, whereas regions of interest exhibit a very high variance in a direction orthogonal to the orientation of ridges and very low variance along ridges.

Bazen and Gerez [10] proposed a pixel-wise segmentation technique, where three

features (gradient coherence, intensity mean, and intensity variance) are computed for each pixel, and a linear classifier associates the pixel with the background or the foreground. A supervised technique is used to learn the optimal parameters for the linear classifier for each specific acquisition sensor. A final morphological post-processing step [43] is performed to eliminate holes in both the foreground and background and to regularize the external silhouette of the fingerprint area. Their experimental results showed that this method provides accurate results; however, it's computational complexity is markedly higher that most of the previously described block-wise approaches.

2.2.1.2 Image Enhancement

The goal of an enhancement algorithm is to improve the clarity of the ridge structures in the recoverable regions and mark the unrecoverable regions as too noisy for further processing.

The most widely used technique technique for fingerprint image enhancement is based on contextual filters. In contextual filtering, the filter characteristics change according to the local context. In fingerprint enhancement, the context is often defined by the local ridge orientation and local ridge frequency. The methods proposed by O' Gorman and Nickerson [90, 91] was one of the first to use contextual filtering. They defined a mother filter based on four main parameters of fingerprint images at a given resolution; minimum and maximum ridge width, and minimum and maximum valley width. The local ridge frequency is assumed constant and therefore, the context is defined only by the local ridge orientation.

Sherlock, Monro, and Millard [110, 111] performed contextual filtering in the Fourier domain. The filter defined in the frequency domain is the function:

$$H(\rho, \theta) = H_{radial}(\rho).H_{angle}(\theta), \tag{2.19}$$

where H_{radial} depends only on the local ridge spacing $\rho = 1/f$ and H_{angle} depends only on the local ridge orientation θ. Both H_{radial} and H_{angle} are defined by bandpass filters.

Hong, Wan, and Jain [49] proposed an effective method based on Gabor filters. Gabor filters have both frequency-selective and orientation-selective properties and

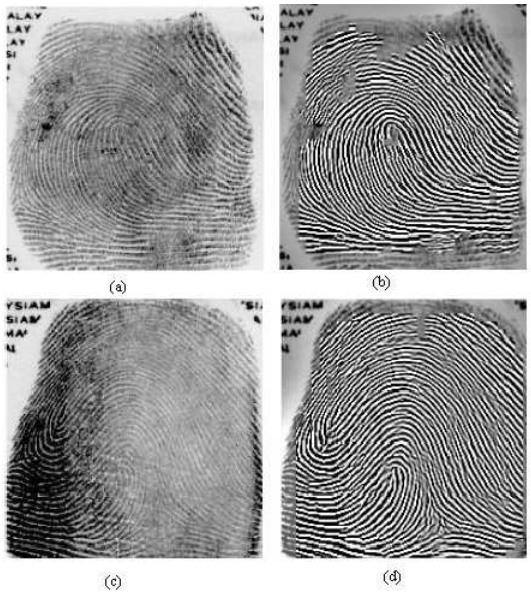


Figure 2.6: (a) and (c) are Input Images; (b) and (d) are enhanced recoverable regions superimposed on corresponding input images.

have optimal joint resolution in both spatial and frequency domains [32]. Examples of fingerprint enhancement with Gabor filtering as proposed by Hong, Wan , and Jain are shown in Fig 2.6.

2.2.2 Feature Representation and Feature Extraction

A good fingerprint representation should have the following two properties: saliency and suitability. Saliency means that a representation should contain distinctive information and suitability means that the representation can be easily extracted, stored in a compact fashion, and be useful for matching. The fingerprint pattern, when analyzed at different scales, exhibit different features.

- **At global level:** The ridge line flow, singular points (loop and delta), external fingerprint shape, orientation image, frequency image.
- **At local level:** Minutiae (ridge termination and ridge bifurcation).
- **At very-thin level:** Intra-ridge details (sweat pores).

An elegant and practical method based on the Poincare index was proposed by Kawagoe and Tojo [58]. Let G be a vector field and C be a curve immersed in G.

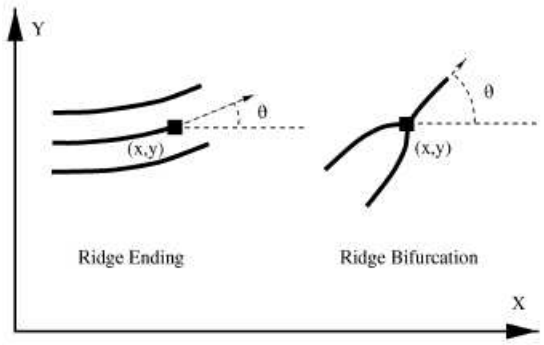


Figure 2.7: Examples of minutiae; A minutia can be characterized by position and orientation.

Then the Poincare index $P_{G,C}$ is defined as the total rotation of the vectors of G along C. An interesting implementation of Poincare method for locating singular points was proposed by Bazen and Gerez [11]. According to Green's Theorem, closed line integral over a vector field can be calculated as a surface integral over the rotation of this vector field. In practice, instead of summing angle differences along a closed path, they compute the "rotation" of the orientation image and then perform a local integration over a small neighborhood of each element. Based on the observation that only a limited number of singularities can be present in a fingerprint, Karu and Jain [57] proposed to iteratively smooth the orientation image until a valid number of singularities is detected by Poincare index. Srinivasan and Murthy [117] extract singularities according to the local histogram of the orientation image. Koo and Kot [63] used multi-resolution approach to determine the singularities with single pixel accuracy.

Most automatic systems for fingerprint comparison are based on minutiae matching. Hence, reliable minutiae extraction is an extremely important task and lot of search has been devoted to this topic. An example of minutiae is given in Fig 2.7. In some approaches minutiae are extracted on a thinned image which is obtained after binarization process. Some others have proposed minutiae extraction approaches that work directly on the gray-scale images without binarization and thinning. After obtaining binary skeleton, a simple image scan allows the pixel corresponding to minutiae to be selected. Pixels corresponding to minutiae are characterized by a crossing

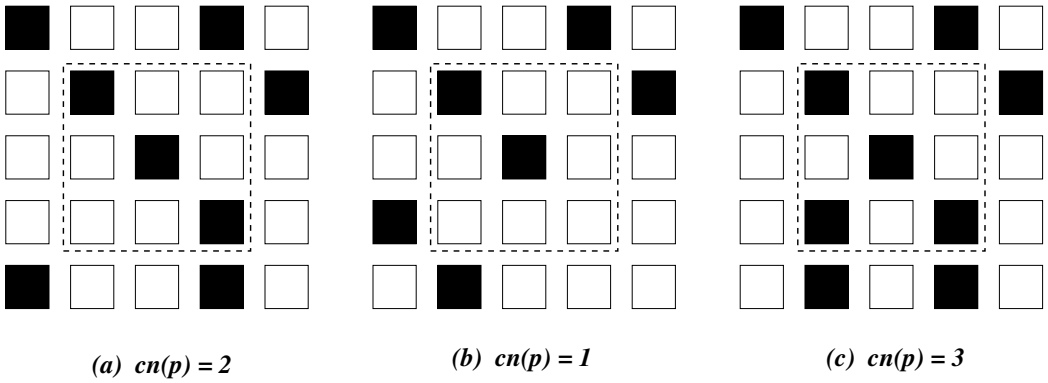


Figure 2.8: (a) Intra-ridge pixel; (b) Termination minutia; (c) Bifurcation minutia

number different from 2. The crossing number [6] $cn(p)$ of a pixel p in a binary image is defined as half the sum of the differences between pairs of adjacent pixels in the 8-neighborhood of p:

$$cn(p) = \frac{1}{2} \sum_{i=1,..,8} |val(p_{i \bmod 8}) - val(p_{i-1})|. \tag{2.20}$$

where $p_0, \ldots, p_7$ are the pixels belonging to an ordered sequence of pixels defining the 8-neighborhood of p and $val(p) \in \{0, 1\}$ is the pixel value. It is simple to note (Fig. 2.8) that for a pixel p with $val(p) = 1$, it:

- is an intermediate ridge point if $cn(p) = 2$;
- corresponds to a termination minutia if $cn(p) = 3$;
- defines a more complex minutia (bifurcation, crossover, etc.) if $cn(p) \geq 3$

Leung et al.'s [76] method extracts the minutiae from thinned binary images but, instead of using the crossing number, they use a three-layer perceptron neural network. In some approaches, minutiae are detected from binary images without any intermediate thinning. Weber [129] first performed an image enhancement by band-pass filtering in the frequency domain and then did binarization via local threshold. Finally, the minutiae are detected from the thick binary ridges through a ridge tracking algorithm. Szekély and Szekély [121] developed a minutiae detection technique based on the computation of the orientation image divergence.

Miao and Maltoni [82] proposed a direct gray-scale minutiae extraction algorithm, where the basic idea was to tract the ridge lines in a gray-scale image by "sailing",

according to the local orientation of the ridge pattern. From a mathematical point of view, a ridge line is defined as a set of points that are local maxima relative to a section orthogonal to the ridge direction. By connecting the consecutive maxima, polygonal approximation of the ridge line can be obtained.

2.2.3 Fingerprint Matching

A fingerprint matching algorithm compares two given fingerprints and returns either a degree of similarity (without loss of generality, a score between 0 and 1) or binary decision (matched/non-matched). Matching fingerprint images is an extremely difficult problem, mainly due to the large variability in different impressions of the same finger (i.e. large intra-class variations). The main factors responsible for intra-class variations are:

- Displacement
- Rotation
- Partial overlap
- Non-linear distortion
- Pressure and skin condition
- Noise

The large number of approaches to fingerprint matching can be coarsely classified into three families: 1) Correlation-based, 2) Minutiae-based and 3) Ridge feature-based. In pattern recognition literature, the minutiae matching problem has been generally addressed as point pattern matching problem. Even though a small difference exists due to the presence of a direction associated with each minutia point, the two problems may be approached analogously. For point pattern matching there are a family of approaches: relaxation methods, algebraic and operational research solutions, tree-pruning approaches, energy-minimization methods, Hough transforms, and so on.

The relaxation approach [104, 99] iteratively adjusts the confidence level of each corresponding pair of points based on its consistency with other pairs until a certain criterion is satisfied. Hong and Tan [47], and Sprinzak and Werman [116] proposed

algebraic methods for point pattern matching. The Hough transform-based approach [7] converts point pattern matching to the problem of detecting peaks in the Hough space of transformation parameters.

Ratha et al. [102] proposed a Hough transform-based minutiae matching approach where the transformation parameters are displacement, rotation, and scale. Two quite atypical fingerprint matching approaches, classifiable as local minutiae matching, have been introduced by Maio and Maltoni [81] and by Kovacs-Vajna [64]. Both these techniques operate asymmetrically. Fingerprint enhancement and accurate minutiae extraction are performed only on the template fingerprint at enrollment time, resulting in the minutiae set **T**. During testing, the existence of a correspondence for each minutia in **T** is checked by locally searching the input fingerprint.

2.3 Recent Approaches to Multiple Classifier Combination

In real-world applications, some limits of monomodal biometric systems have already been reported. Indeed some biometrics have only little variation over the population, have large intra-class variability over time, or/and are not present in all the population. To fill these gaps, the use of multimodal biometrics is a first choice solution. Multimodal biometric systems increase robustness and are more reliable. Multimodal approaches provide appropriate measures to resist against spoof attacks, as it is difficult to counterfeit several modalities at the same time, to circumvent a system. They also provide an adopted solution to the limitations of universality, as even if a biometrics is not possessed by a person, the other(s) modality(ies) can still be used.

Multimodal biometric systems that have been proposed in literature can be classified based on four main parameters: (1) Architecture or operational mode, (2) Fusion scenarios, (3) Level of fusion and (4) Methods for integrating multiple cues.

2.3.1 Architecture or Operational Mode

Architecture of multimodal biometric system refers to the sequence in which the multiple cue are acquired and processed. Multimodal biometric systems can operate in three different modes: (a) Parallel, (b) Serial and (c) Hierarchical (see in Fig. 2.9).

- **Parallel Mode:** This operational mode consists in completing the combination of the modalities simultaneously. Different modalities operate independently and their results are combined using an appropriate fusion scheme.
- **Serial Mode:** This operational mode consists in completing the combination of the modalities one after the other, as it permits to reduce the population at each stage before the following modality is used. The decision could thus be taken before all the remaining biometrics are acquired, reducing considerably the processing duration. Here the outcome of one modality affects the processing of the subsequent modalities.
- **Hierarchical Mode:** This operational mode consists in completing the combination of the modalities in a hierarchical scheme, like a tree structure, when the number of classifiers is large. This architecture can also allow the user to decide which modality he/she would present first. Finally, if the system is faced with the task of identifying the user from a large database, it can utilize the outcome of each modality to successively prune the database, thereby making the search faster and more efficient. An example of a cascaded multibiometric system is the one proposed by Hong and Jain in [48].

Most proposed multibiometric systems have a parallel architecture [115, 106]. But the choice of system architecture depends on the application. User-friendly and less security like ATMs can use hierarchical mode. On the other hand parallel mode are well suited for applications where security is of paramount importance (military installations).

2.3.2 Fusion Scenarios

Multimodal biometric systems overcome some of the limitations of unimodal biometric systems by consolidating the evidence obtained from different sources (see Fig. 2.10). The sources are given below:

- **Single Biometry, Multiple Sensors:** The same biometric is acquired by different sensors and combined to complete and improve the recognition process (e.g. optical and solid-state fingerprint sensors). The use of multiple sensors

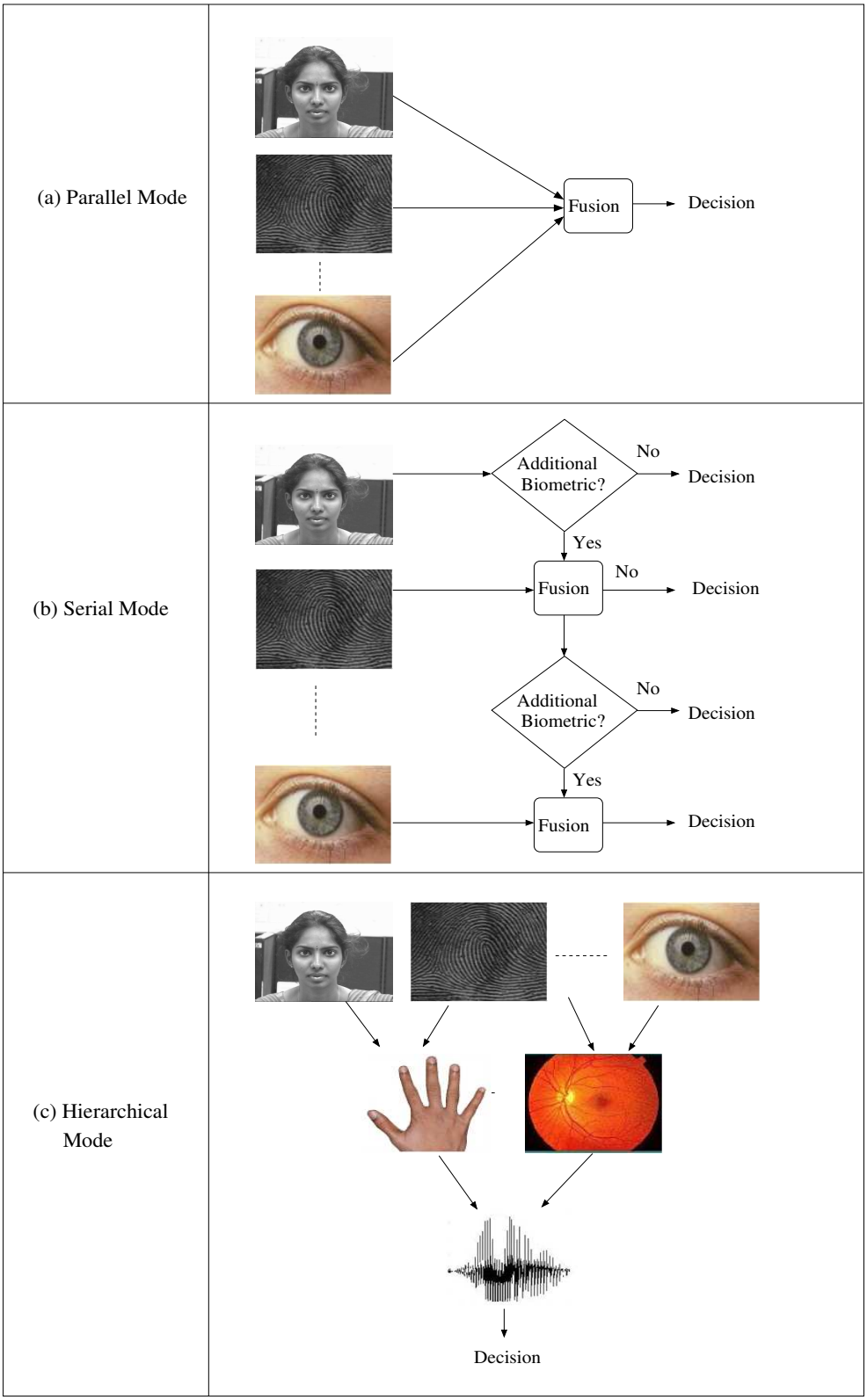


Figure 2.9: Different architectures of multimodal biometric system; (a) Parallel, (b) Serial and (c) Hierarchical.

can address the problem of noisy sensor data, but all other potential problems associated with unimodal biometric systems remain.

- **Single Biometry, Multiple Instances:** The same biometric unit is acquired several times by a same sensor and combined to complete and improve the recognition process (e.g. multiple face images of a person obtained under different pose/lighting conditions).
- **Single Biometry, Multiple Units:** The same biometric, but different units are acquired and combined to complete and improve the recognition process (e.g. left and right iris images). This is a recognition system that works on multiple units of the same biometric measurements (e.g. left middle finger followed by a right thumb).
- **Single Biometry, Multiple Representations:** The same biometric is acquired once by a single sensor and different approaches of feature extraction and matching are combined to complete and improve the recognition process (e.g. multiple face matcher like PCA and LDA).
- **Multiple Biometrics:** Different biometrics of the same person are acquired and combined to complete and improve the recognition process (e.g. face, fingerprint and iris). This approach is the only well-used multimodal biometric fusion scenario.

Although the first four methods improve the recognition performance of a system, they still suffer from some of the problems faced by unimodal systems. A multimodal systems based on different traits seems to be more robust to noise, address the problem of non-universality, provide reasonable protection against spoof attacks and also improve matching accuracy.

2.3.3 Fusion Levels and Methods of Integration

Fusion in multimodal biometric systems can take place at three major levels, namely, sensor level, feature level and decision level. These three levels can be broadly classified into fusion prior to matching and fusion after matching [108]. Decision level fusion can be divided into three classes based on the type of the output of biometric

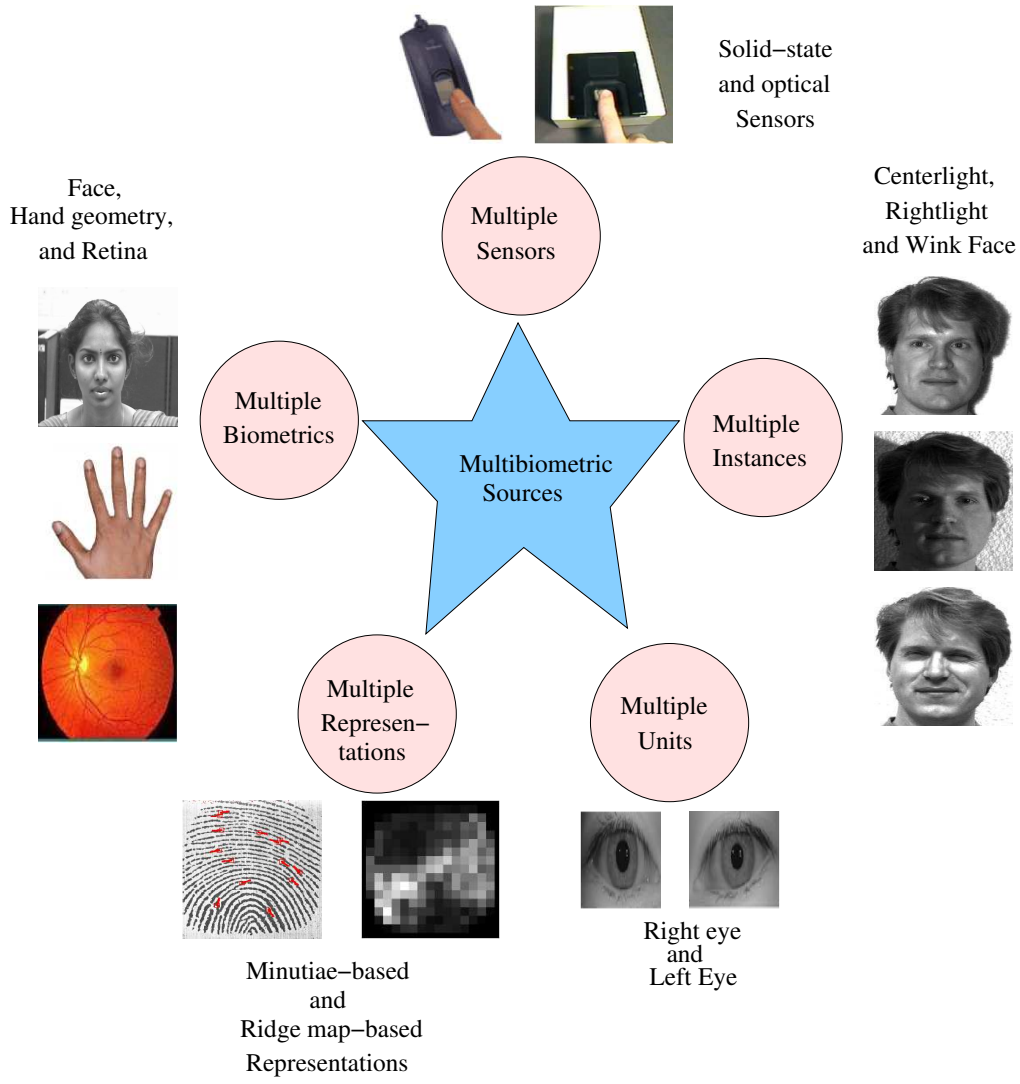


Figure 2.10: Sources of multiple evidence in multimodal biometrics.

matchers: (a) Abstract level, (b) Rank level and (c) Measurement level. Fig. 2.11 shows the summary of approaches to information fusion in biometry.

2.3.3.1 Fusion Prior to Matching

- **Sensor level:** Integration of information at sensor level is performed prior to matching. The raw data from the sensor(s) are combined in sensor level fusion [52]. This type of fusion can be done only if the multiple cues are either instances of the same biometric trait obtained from multiple compatible sensors or multiple instances of the same biometric trait obtained using a single sensor.

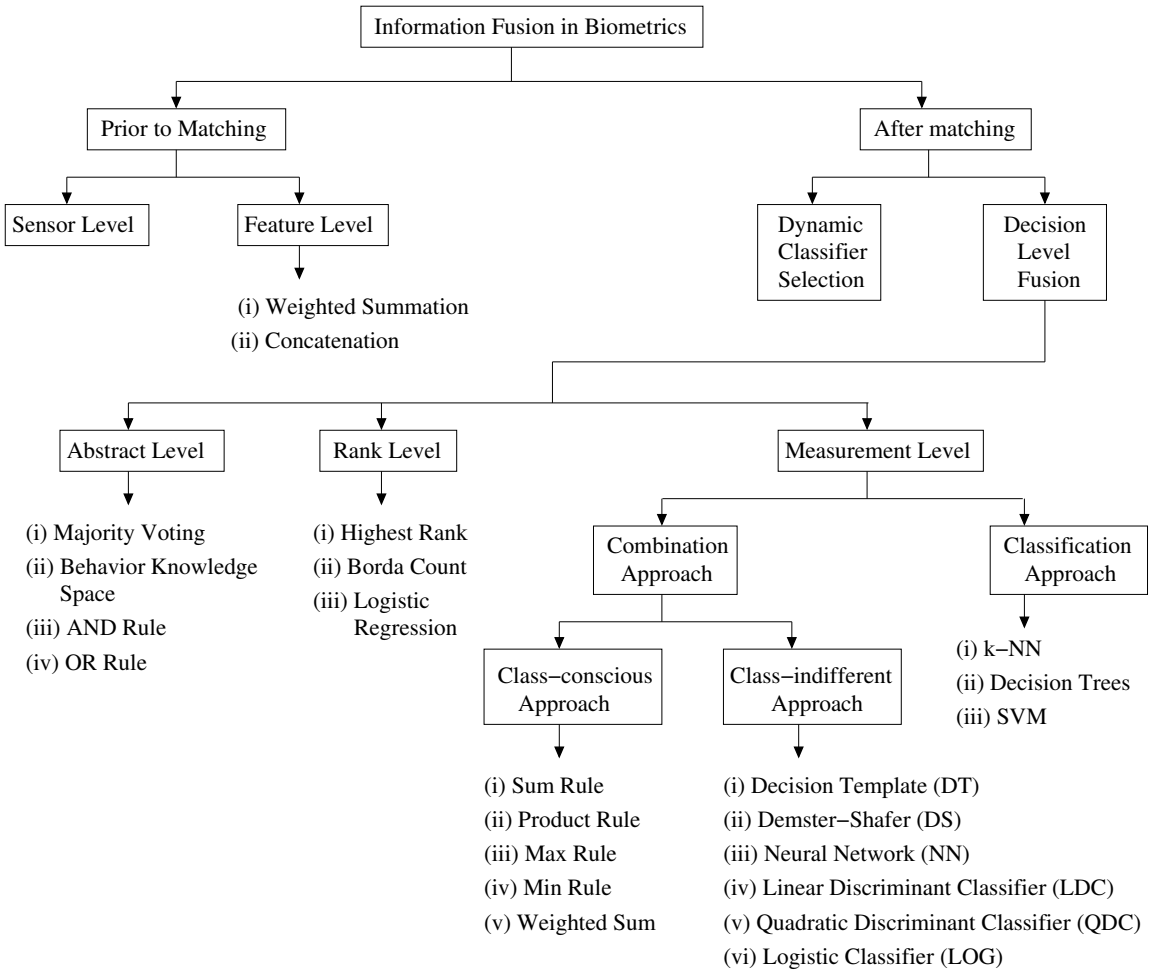


Figure 2.11: Summary of approaches to information fusion in biometric systems.

For example, the face images obtained from several cameras can be combined to form a 3D model of the face.

- **Feature level:** Feature level fusion refers to combining different feature vectors that are obtained from one of the following sources: multiple sensors for the same biometric trait, multiple instances of the same biometric trait, multiple units of the same biometric trait or multiple biometric traits. When feature vectors are homogeneous, a single resultant feature vector can be calculated as a weighted average of the individual feature vectors. In case of non-homogeneous features, we can concatenate them to form a single feature vector which is not possible for incompatible feature sets. Attempts by Kumar et al. [66] in combining palmprint and hand-geometry features and by Ross and Govindarajan

[105] in combining face and hand-geometry features have met with only limited success.

Since the features contain richer information about the input biometric data than the matching scores obtained at decision level, integration at feature level is believed to be more effective than decision level fusion. But feature fusion is often not possible for the unknown or incompatible relationship between different feature spaces, and often simple concatenation of features may cause "curse of dimensionality problem".

2.3.3.2 Fusion After Matching

Information fusion after matching can be divided into two main categories: dynamic classifier selection and decision fusion.

A dynamic classifier selection technique chooses the result of that classifier which is most likely to give correct decision for the specific input pattern [132]. This is well-known as winner-take-all approach and the device that performs the selection is known as associative switch [24].

Information fusion at decision level can be at abstract, rank and measurement level based on the type of matcher's output.

- **Abstract level:** The biometric matcher individually decides on the best match based on the input presented. Methods like majority voting [71], behavior knowledge space [70], weighted voting based on the Dempster-Shafer theory of evidence [135], AND and OR rules [31], etc. are use to obtain the final decision.
- **Rank level:** The output of each biometric matcher is a subset of possible matches sorted in decreasing order of confidence. Ho et al. [46] described three methods to combine the ranks assigned by different matchers. In the highest rank method, each possible match is assigned the highest (minimum) rank as computed by different matchers. Final decision is made based on the combined ranks and ties are broken randomly. The Borda count method uses the sum of the ranks assigned by the individual matchers to calculate the combined ranks. The logistic regression method is a generalization of the Borda count method where weighted sum of the individual ranks is calculated and the weights are

determined by logistic regression.

- **Measurement level:** The output of the biometric matchers is a set of possible matches along with the matching scores. Measurement level output contain richest information about the input pattern. Here, the scores must be transformed to a common domain to ensure a meaningful combination from different modalities. As this is the most common approach used for fusion, we discuss it in more detail in the following section.

2.3.3.3 Measurement level

Measurement level fusion can be approached in two ways: (1) Classification approach and (2) Combination approach. Wang et al. [128] consider the matching scores resulting from face and iris recognition modules as a two-dimensional feature vector. Fisher's discriminant analysis and a neural network classifier with radial basis function are then used for classification. Verlinde and Cholet [126] combine the scores from two face recognition experts and one speaker recognition expert using three classifiers: k-NN classifier using vector quantization, decision tree based classifier and a classifier based on a logistic regression model. Chatzis et al. [21] use fuzzy k-means and fuzzy vector quantization, along with a median radial basis function neural network classifier for the fusion of the scores obtained from biometric systems based on visual (facial) and acoustic (vocal) features. Ross and Jain [106] use decision tree and linear discriminant classifiers for combining the scores of face, fingerprint, and hand-geometry.

In a traditional multiple classifier system, a feature vector x is classified into one of the C classes using L classifiers $\{D_1, D_2, ..., D_L\}$, each using the feature vectors $x_l, l = 1, 2, ..., L$, respectively. Measurement-level (also called response vector level) combination strategies give final decision by fusing the response vectors from multiple classifiers. Formally, for a feature vector x, response vectors from multiple classifiers

can be organized as a matrix called decision profile (DP):

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \dots & d_{1,j}(x) & \dots & d_{1,C}(x) \\ \dots & \dots & \dots & \dots & \dots \\ d_{i,1}(x) & \dots & d_{i,j}(x) & \dots & d_{i,C}(x) \\ \dots & \dots & \dots & \dots & \dots \\ d_{L,1}(x) & \dots & d_{L,j}(x) & \dots & d_{L,C}(x) \end{bmatrix}$$

We denote i^{th} row of the above matrix as $D_i(x) = [d_{i,1}(x), ..., d_{i,C}(x)]$, where $d_{i,j}(x)$ is the degree of support given by classifier D_i to the hypothesis that x belongs to class j. $D_i(x)$ is the response vector of classifier D_i for a sample x. The task of any combination rule is to construct $\tilde{D}(x)$, the fused output of L classifiers as:

$$\tilde{D}(x) = \mathcal{F}(D_1(x), ..., D_L(x)) \tag{2.21}$$

Some fusion techniques known as class-conscious [68], do column-wise class-by-class operation on *DP(x)* matrix to obtain $\tilde{D}(x)$. Example of this type of fusion techniques are: sum, product, min, max, etc [60]. Another fusion approach known as class-indifferent [68], use entire *DP(x)* to calculate $\tilde{D}(x)$.

Class-conscious Methods

Given $DP(x)$, class-conscious methods operate class-wise on each column of $DP(x)$. The architecture of class-conscious methods is demonstrated in Fig. 2.12.

- **Sum Rule:** Sum Rule computes the soft class label vectors using

$$\tilde{d^j}(x) = \sum_{i=1}^{L} d_{i,j}, \;\; j = 1, ..., C \tag{2.22}$$

- **Product Rule:** Product Rule computes the soft class label vectors as

$$\tilde{d^j}(x) = \prod_{i=1}^{L} d_{i,j}, \;\; j = 1, ..., C \tag{2.23}$$

- **Min Rule:** Min Rule computes the soft class label vectors using

$$\tilde{d^j}(x) = min(d_{1,j}, d_{2,j}, ..., d_{L,j}), \;\; j = 1, ..., C \tag{2.24}$$

- **Max Rule:** Max Rule computes the soft class label vectors using

$$\tilde{d^j}(x) = max(d_{1,j}, d_{2,j}, ..., d_{L,j}), \;\; j = 1, ..., C \tag{2.25}$$

Class-indifferent Methods

- **Decision Template (DT):** Let $Z = \{z_1^1, z_2^1, ...z_M^1, z_1^2..., z_M^2, ..., z_M^C\}, z_j^i \in \mathcal{R}^n$ be the crisply labeled validation set which is a disjoint from training and testing set. M is the number of validation samples per class. DP^Z is the set of DP's corresponding to the samples in Z. Hence DP^Z is a 3-dimensional matrix of size $L \times C \times N$ where $N = M * C$.

  1. Decision template DT_i of class i is the $L \times C$ matrix and is computed from DP^Z as

  $$DT_i = \frac{1}{M} \sum_{j=(i-1)*M+1}^{i*M} DP^Z_{.,.,j},\ i = 1, ..., C \tag{2.26}$$

  The decision template DT_i for class i is the average of the decision profiles of the elements in the validation set Z labeled as class i.

  2. When a test vector, $x \in \mathcal{R}^n$ is submitted for classification, $DP(x)$ is matched with $DT_i,\ i = 1, ..., C$ and produces the soft class label vector

  $$\tilde{d}^j(x) = S(DT_i, DP(x)),\quad i = 1, ..., C. \tag{2.27}$$

  where S is a function which returns similarity measure between it's arguments. It can be euclidean distance or any fuzzy set similarity measure.

- **Dempster-Shafer Combination (DS):** Dempster-Shafer algorithm performs the following steps:

  1. Let DT_j^i denotes the i^{th} row of the decision template for class j. Calculate the proximity Φ between DT_j^i and $D_i(x)$ for every class *j=1,....,C*, and for every classifier *i=1,...,L*. The proximity [103] is calculated using

  $$\Phi_{j,i}(x) = \frac{(1 + \|DT_j^i - D_i(x)\|^2)^{-1}}{\sum_{k=1}^{C}(1 + \|DT_k^i - D_i(x)\|^2)^{-1}}, \tag{2.28}$$

  where $\| * \|$ is any matrix norm.

  2. For each class, *j=1,....,C* and for every classifier, *i=1,....,* L, calculate *belief degrees*

  $$b_j(D_i(x)) = \frac{\Phi_{j,i}(x) \prod_{k \neq j}(1 - \Phi_{k,i}(x))}{1 - \Phi_{j,i}(x)[1 - \prod_{k \neq j}(1 - \Phi_{k,i}(x))]}, \tag{2.29}$$

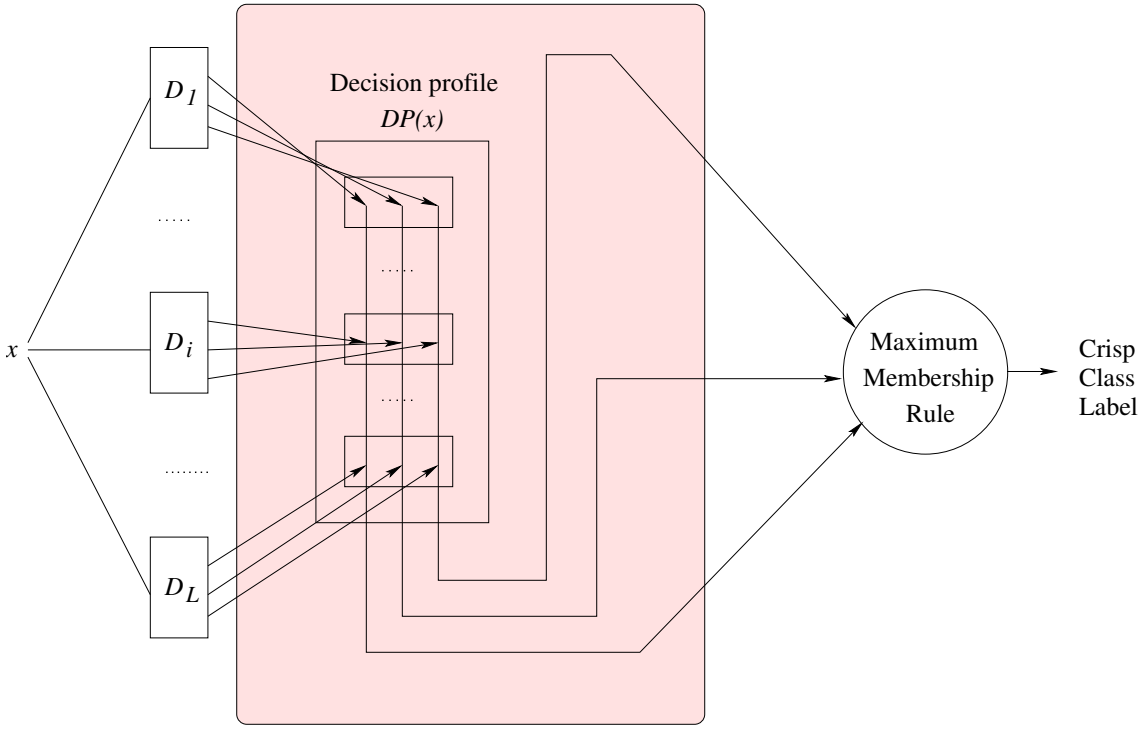


Figure 2.12: Operation of class-conscious methods.

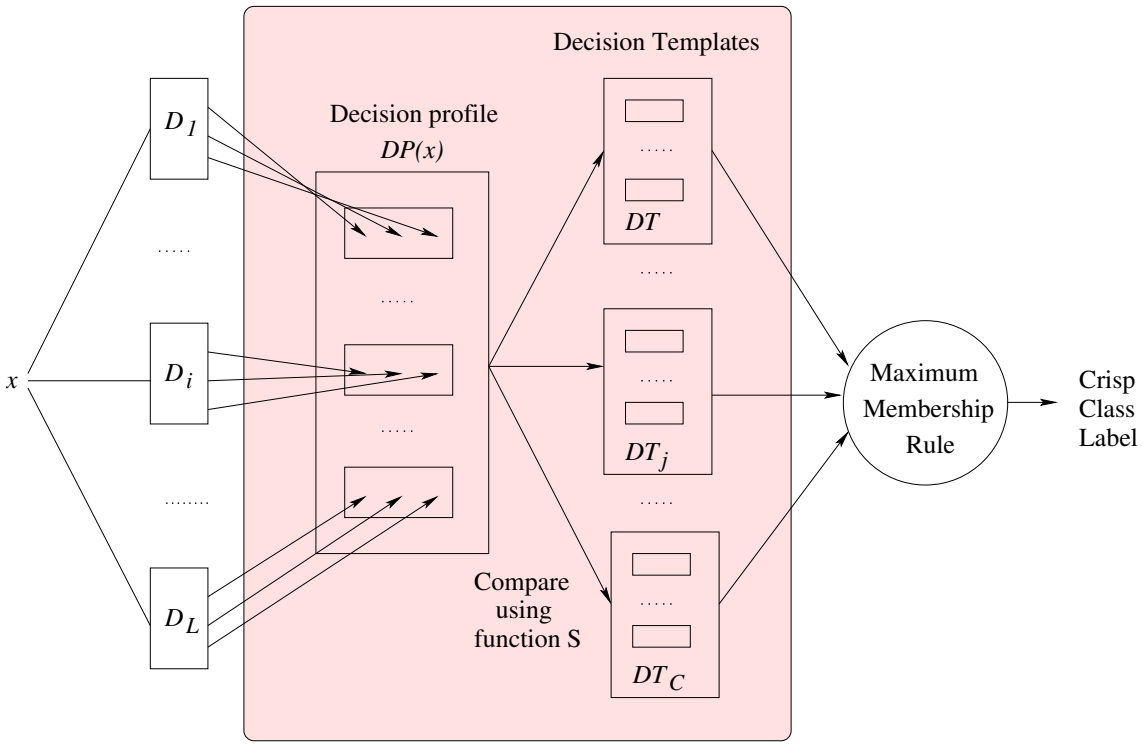


Figure 2.13: Operation of class-indifferent methods.

3. The final DS soft vector is calculated as

$$\tilde{d}^j(x) = K \prod_{i=1}^{L} b_j(D_i(x)),\ j = 1, ..., C \tag{2.30}$$

where K is a normalizing constant.

Architecture for a class-indifferent method is shown in Fig. 2.13.

CHAPTER 3

An Efficient Method of Face Recognition using Subject-Specific Subband Faces

This chapter presents an efficient method for frontal face recognition, using subject-specific subband face representation. The human face has certain visual features that are common among everybody and some others that exhibit the unique characteristics of an individual. Using the discrete wavelet transform (DWT), we extract these unique features from the face image for discriminating it from others. The face image is decomposed into several subbands to separate the common (approximation) and discriminatory (detail) parts. *Subband face* is reconstructed from selected wavelet subbands, in which a suitable approximation subband is suppressed and a detail subband (in some cases) is eliminated. Reconstructing a face with an optimal selection of subbands enhances the performance of face recognition. We present four different criteria as cost functions to obtain an optimal subband face for each subject, and compare their performances. The performance of the subband face representation on several linear subspace techniques: PCA, LDA, 2D-PCA, 2D-LDA and Discriminative Common Vectors (DCV) with Yale, ORL and PIE face databases shows that the subband face representation performs significantly better than that proposed by Ekenel for multiresolution face recognition [36] for frontal face recognition, in the presence of varying illumination, expression and pose. The rest of the chapter is organized as follows. Section 3.1 provides the motivation and a brief introduction to our proposed method along with a concise overview on wavelet based face recognition techniques. The Wavelet Decomposition and the method to reconstruct a subband face is described in section 3.2. Section 3.3 highlights over the criteria used as cost functions and also the algorithm to select optimum subject-specific subbands. Section 3.4 discusses the experimental results of the proposed subband face representation tested

on three face databases and compared with the state-of-art closest competitor [36] of it. Section 3.5 concludes the chapter.

3.1 Introduction and Motivation

How do humans identify individuals with remarkable ease and accuracy? This question has haunted psychologists, neurologists and, recently, engineers in biometry for a long time. The human face is different from any other natural or man-made objects, but has similar structural features across different races, sex and regions. The subtle variations in the face structure are captured by the human brain and help in discriminating one face from the another. The human brain is able to filter out the common visual features of a face and retain only those suitable to exhibit the unique characteristics (discriminatory evidence) of an individual. An efficient face-based biometric system must possess these properties (as in those in the retina and visual cortex), to perform efficiently like a human being.

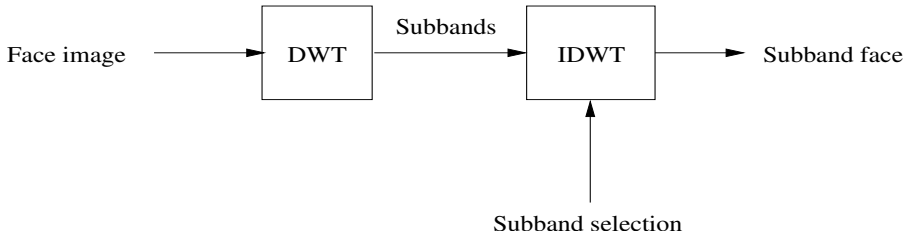
Wavelet-based features have already been used to obtain a better face representation in [36, 40, 26, 140]. The use of wavelet packet for face recognition was reported in [40]. Here, 2D-DWT is used to fully decompose the face image and simple statistical features such as mean and variance are extracted from the decomposed coefficients, and used as a feature vector for representation. The use of wavelet subband representation along with kernel associative memory for face recognition was reported in [140]. Here wavelet decomposed faces are used to build an associative memory model for each class and kernel methods are used to exploit higher order relations which cannot be captured by linear transformations. Face images with illumination variations can be modeled by low-dimensional linear subspaces. The existence of single light source directions that are effective for face recognitions was discussed in [75]. Ekenel [36] et al. proposed multiresolution face recognition with fusion at data, feature and decision levels to test on face images that differ in expression or illumination separately, obtained from CMU PIE, FERET and Yale databases. Significant performance gains are reported against illumination perturbations. They selected a number of subbands on the basis of performance on testing set and termed them as successful subbands. Data, feature and decision level fusion are performed on these successful subbands to

improve the accuracy further. They obtained a maximum accuracy of 96.67% using ICA2 on a database (for expression variations) comprised of 272 images from CMU PIE and rest 328 from FERET. For finding subbands insensitive to illumination they experimented on a database consisting of 272 images from CMU PIE and remaining 60 images from Yale database and reported a maximum accuracy of 77.71%. The main drawback of this method lies in the adhoc method of selecting successful subbands on a testing set which involves over-tuning to attain maximum possible accuracy.

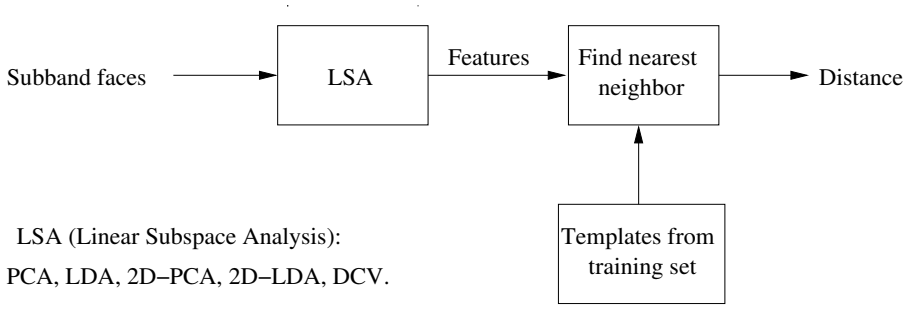
None of these approaches have attempted to exploit the fact that, if some of the common visual features of a face exist in a certain subband and are suppressed during reconstruction, we obtain a subband face with only the discriminatory information. The retina of the human visual system has been observed to perform local multi-resolution, multi-channel processing of the signal using a bank of tuned band-pass filters [65, 85, 98]. The visual cortex uses second order relational differences in structure with respect to an average face for perception [41, 56]. Hence, a unified computational model consisting of DWT/IDWT (signal processing) and PCA/LDA (statistical processing) will be able to closely imitate the human visual system, more effectively than what statistical processing does alone. These studies motivated us to explore the use of IDWT to obtain a subband face, by suppressing an approximation and retaining some of the discriminating subbands in the process of image reconstruction, and then use the subband face for recognition using statistical approaches.

In this chapter, we propose the *subband face* as a new representation for the face recognition task. Only the discriminatory information of a face is retained or captured in a subband face. Discrete wavelet transform is used to decompose the original face image into approximation and detail subbands. We perform multi-level dyadic decomposition of a face image using the Daubechies filters [30]. The subband face may be reconstructed from selected subbands by suppressing the approximation at a suitable higher level and retaining the details. An inherent information fusion is being performed by the reconstruction process which retains only the inter-class discriminatory informations and discards the inter-class common informations. The information of a face in the details at lower levels of decomposition, usually contains noise and redundant pixels which do not constitute any discriminatory information

for a face. This was illustrated by experimentation in [94] where the reconstruction of a face image from subbands upto a certain level, below that of the original face image at the root, provides better performance for face recognition. It is thus often necessary to eliminate the details for the representation of a face, and in such cases the subband face is obtained by partial reconstruction. In this chapter, we provide an algorithm to obtain optimum subject-specific subbands on the basis of minimization of four different criteria. Results of our experimentation presented in section 3.4 will validate a hypothesis that *"subband face representation provides efficient face recognition"*. Our approach is different from that suggested in [36, 40, 26, 140] where the individual subbands of the face image have been used as features separately or unified by subspace analysis techniques for face recognition. None of the past approaches deals with a method to reconstruct a subband face using IDWT (synthesis filter bank).



(a)



(b)

Figure 3.1: (a) Block diagram for generating subband face. The face image is decomposed using Discrete Wavelet Transform (DWT) into subbands. Selected subbands are used to reconstruct the subband face using inverse DWT. (b) Subband faces are projected onto a lower dimension featurespace using PCA or LDA (or their variants) and matched using the nearest neighbor strategy.

Fig. 3.1 shows the block diagram of the proposed method. Fig. 3.1 (a) shows the generation of the subband face. The face image is decomposed into several subbands using the analysis filter of 2D Discrete Wavelet Transform (2D-DWT). The 2D-DWT is successively applied to the approximations to obtain a dyadic decomposition [38]. The subbands thus obtained can be used to reconstruct the original face using the inverse DWT (by synthesis filters). The subband face is obtained from selected subbands by partial reconstruction of the original face. Fig. 3.1 (b) shows that the subband face is transformed from image-space to feature-space by applying subspace methods (PCA , LDA, 2D-PCA, 2D-LDA and DCV). The details of the subspace methods used for integrating with *subband face representation* are provided in section 2.1.1.1. The features thus obtained are matched using the nearest neighbor strategy.

3.2 Subband Face Representation

A set of suitable wavelet subbands are selected to generate the *subband face* using IDWT, for recognition using PCA, LDA, 2D-PCA, 2D-LDA and DCV. The method of reconstructing a subband face is described in the following.

3.2.1 Wavelet Decomposition

The wavelet transform [38] is expressed as a decomposition of a signal $f(x) \in L^2(R)$ into a family of functions which are translations and dilations of a mother wavelet function $\psi(x)$. The 2D wavelet transform uses a family of wavelet functions and its associated scaling function to decompose the original image into different subbands, namely the low-low (LL), low-high (LH), high-low (HL) and high-high (HH) subbands, which are also known as A (approximations), H (horizontal details), V (vertical details), D (diagonal details), respectively. Fig. 3.2 shows the 2D-DWT performed using low pass $h(.)$ and high pass $g(.)$ filters [83]. The decomposition process can be recursively applied to the low frequency channel (LL) to generate dyadic decomposition at the next level.

Fig. 3.3 shows a face image, and its level-1 and level-3 dyadic decompositions using the 4-tap Daubechies [30] wavelet filter. Lowest level approximation subband

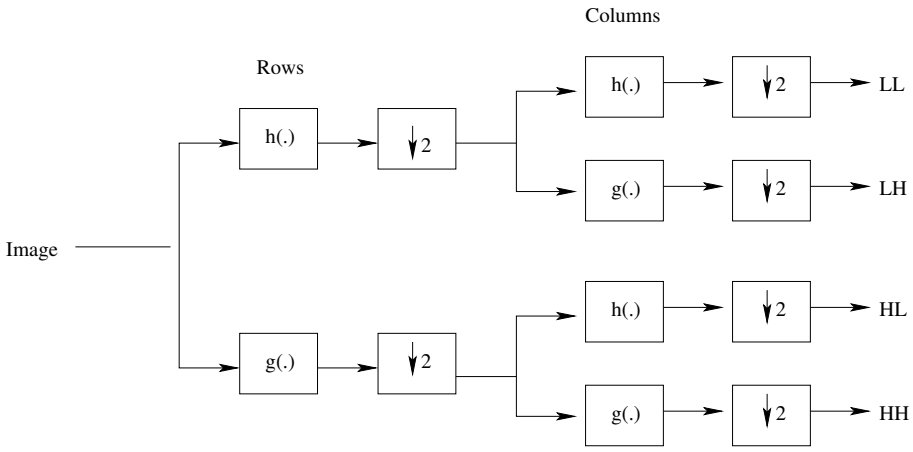
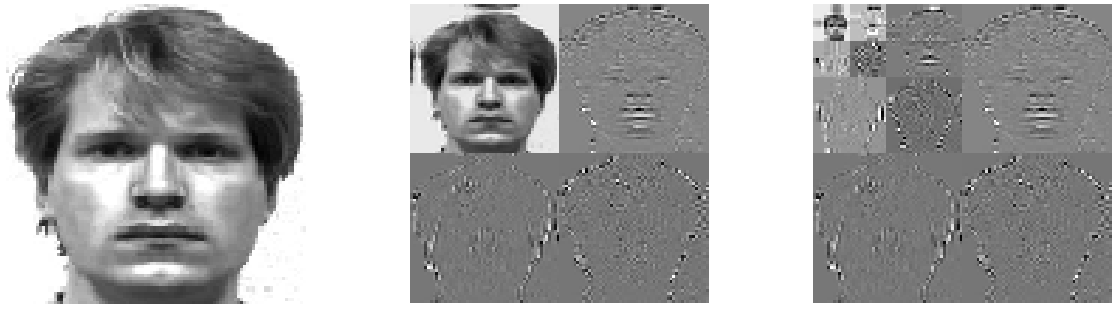


Figure 3.2: Two dimensional discrete wavelet transform (DWT) for an image: The low pass filter $h(.)$ and the high pass filter $g(.)$ are applied along the rows initially followed by downsampling by a factor of two. This is followed by filtering using $h(.)$ and $g(.)$ along the columns and downsampled by factor of two.

of the wavelet decomposition can be further decomposed using the DWT, giving a multi-level dyadic subband decomposition of the face. In the following subsection, the method of reconstructing a subband face using selective wavelet-decomposed subbands is discussed.

3.2.2 Subband Face Reconstruction

A face image of a person contains common (approximations) as well as discriminatory (details) information with respect to faces of all other persons. The discriminatory information is due to structural variations of the face which are acquired as intensity variations at different locations of the face. The location and degree of intensity variations in a face of an individual are unique features which discriminate one from the rest of the population. The similarity of a face with respect to another is in the global appearance and structure of the face. These information (similar and discriminatory) are segregated at different subbands using different levels of decomposition of the face image. Wavelet decomposition helps to split the features of a face in different subbands with "approximations" containing the common (smooth) parts of the face and "details", at certain levels of decomposition, containing the discriminatory (vari-



(a) (b) (c)

Figure 3.3: (a) Original face, (b) the level-1 wavelet decomposition of the face image into subbands A, H (top) and V, D (bottom) and (c) level-3 dyadic wavelet decomposition. All subband signals are normalized individually.

ations) information. Since level-1 decomposition may not be adequate to effectively isolate these pair of visual features, it is necessary to explore multi-resolution, multi-channel representation at higher levels to obtain a suitable isolation. With higher levels of decomposition, the approximation will have increasing effect of smoothening (with lesser subband bandwidth) on the face image, and the details will have more of the discriminatory information. At a certain higher level of decomposition, the entire discriminatory information will remain in the details and approximation will contain the common (or similar) structural information. In order to eliminate the similarity and retain the discrimination, the approximation is suppressed (by replacing with zeros) and the face image is reconstructed using IDWT.

The details at lower levels (1 or 2) of decomposition also do not contain any useful information. This can be realized from the fact that humans do not need an image at a very large resolution to identify an individual. Too low a resolution, on the other hand, is not adequate to hold all of the discriminatory information. Hence there exists an optimal resolution at which the image contains most of the discriminatory and less (or no) redundant pixels (information). Thus it is often necessary to eliminate certain details at lower levels of decomposition to discard redundant image pixels or noise if present in the signal. Thus the subbands which only contain the discriminatory information for face recognition are selected for representation and the others are discarded. Based on the explanation presented above, we propose an efficient method of face recognition using subband face representation which provides an improvement

in the accuracy of face recognition. We will verify the efficiency of subband face representation using experimental results, which will be discussed in section 3.4.

Fig. 3.4 shows the block diagram of IDWT that is used to reconstruct the original image from different subbands at level-1 decomposition. Fig. 3.5 shows the block diagram of the subband face reconstruction from level-2 subbands, by suppressing (replacing the coefficients by zeros) only the approximation coefficients at level-2. The subband face thus obtained will hence not posses the spectral energies in the $[0 - \Pi/4]$ subband which correspond to a smoothed version (i.e. average structure) of the face image.

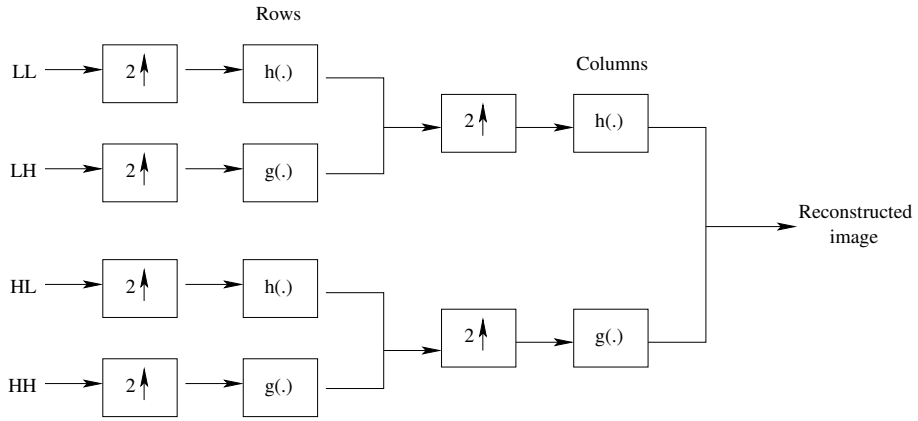


Figure 3.4: Two dimensional inverse discrete wavelet transform (IDWT) for an image. The rows are initially upsampled by factor of two and filtered using low pass $h(.)$ and high pass $g(.)$ filters. This is followed by upsampling along columns by a factor of two and applying the low pass $h(.)$ and high pass $g(.)$ filters along the columns to get the reconstructed image.

Fig. 3.6 shows the original face images along with the subband faces generated by suppressing the approximations at different levels. The face image in Fig. 3.6 (a) is illuminated from the left and the corresponding subband faces shown in Figs. 3.6 (a1-a3) have the effect of illumination reduced or almost eliminated. This can also be observed for the face image in Fig. 3.6 (b) which is illuminated from the right, and its corresponding subband faces in Figs. 3.6 (b1-b3) do not posses similar amount of illumination variation. Fig. 3.6 (c) shows a face with frontal illumination, where the intensity variations (edges) of the subband faces in Figs. 3.6 (c1-c5) are stronger than the gray level image. It is also noticeable that the suppression of approximation

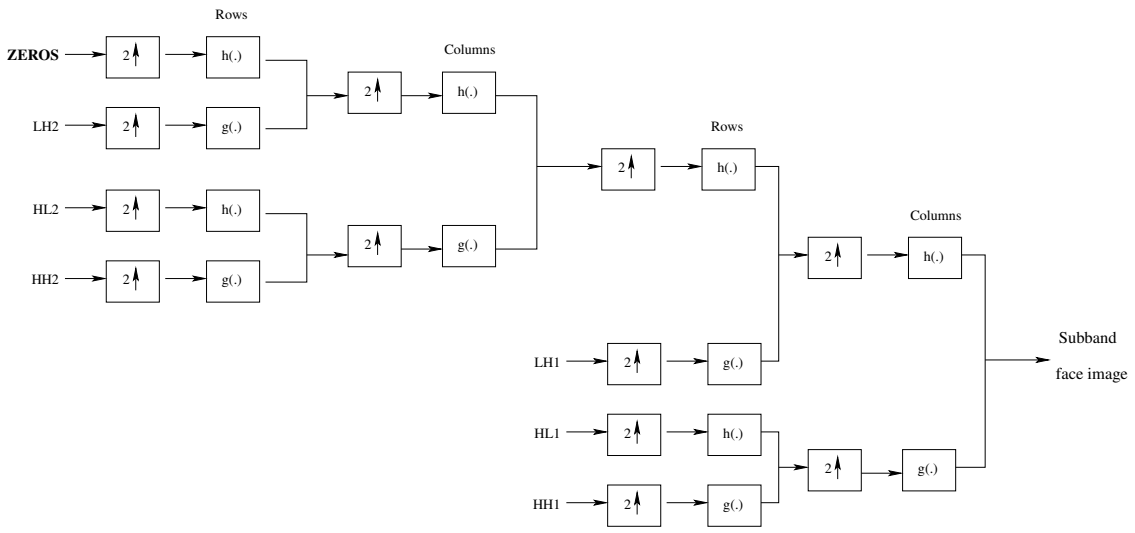


Figure 3.5: Reconstruction of a subband face using two dimensional Inverse Discrete Wavelet Transform (IDWT) from level-2 subbands. The approximation at level-2, LL2, is suppressed (by replacing with zeros) while reconstructing the approximation at level-1, LL1. This is further used to generate the subband face at level-0.

subband at a larger level of decomposition leads to reconstruction of a face that is similar to the original gray level image.

Let the original face image, at the root of the decomposition tree, be A_0 and the approximations at different levels of decomposition be denoted by A_1, A_2 and so on. The subband face that is reconstructed with the approximation suppressed at level-i be denoted as $(A_0 - A_i)$. In [94], it was shown that the subband face reconstructed (using Haar transform) with the approximation suppressed at level-5, $(A_0 - A_5)$, gives the best performance for PCA based face recognition for Yale database. In this chapter, we propose a method of implementation to obtain a subband face, $(A_l - A_k)$, which is reconstructed to the level-l, after suppressing the approximation at level-k, where $0 \leq l < k \leq \log N$, $N * N$ being the resolution of the image. Note that this eliminates the details: $V_i, H_i, D_i,$ for levels $i = 1, 2, \ldots, l$ when $l > 0$, from the original face image. It was found that for optimal performance of face recognition, the levels of decomposition l and k depend on two factors:

- Subject under consideration,
- Linear subspace method used for face recognition

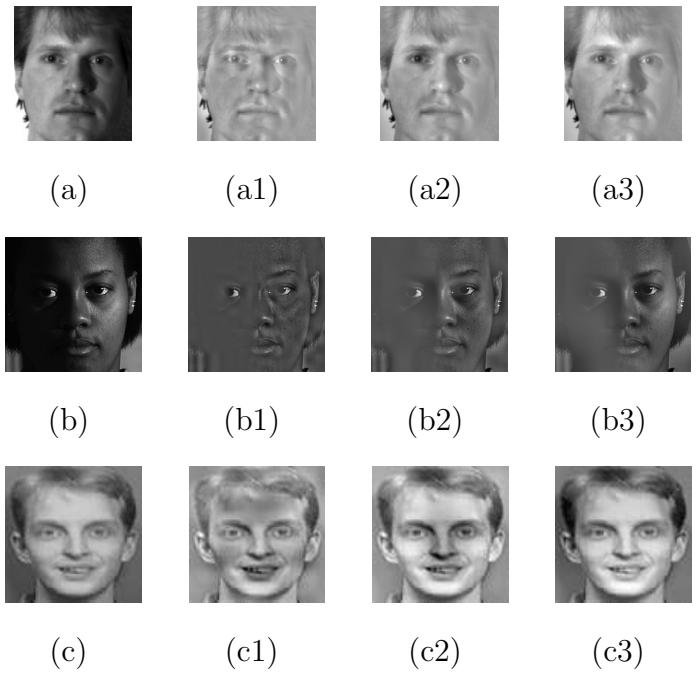


Figure 3.6: (a), (b), (c): Typical examples of sample face images from the Yale, PIE and ORL face databases. The respective subband faces (a1-c1), (a2-c2) and (a3-c3) with the approximations at level-4, level-5 and level-6 suppressed.

In the following section, we explain and describe the algorithm for selecting subject-specific optimal subbands and the criteria used for searching optimal values of l and k to generate the best subband face.

3.3 Proposed Methods for Subject-Specific Subband Selection

In this section, we will describe four different criteria and the algorithm used for subject-specific subband selection. As every person has a unique facial structure with respect to others, self-shadowing and other factors such as outline, appearance etc. provide a strong basis for the existence of a separate subband $(A_{l_i} - A_{k_i})$ for each person. Precisely, the output of subband selection algorithm will be C $(l_i,\ k_i)$ pairs, where $i = 1, 2..., C$, C being the number of subjects. We calculate two measures called *goatishness* and *lambishness* [34] for each subject and for each l and k pair where l and k varies from $l = 0, 1, ..., L$ and $k = (l + 1), ..., K$. The

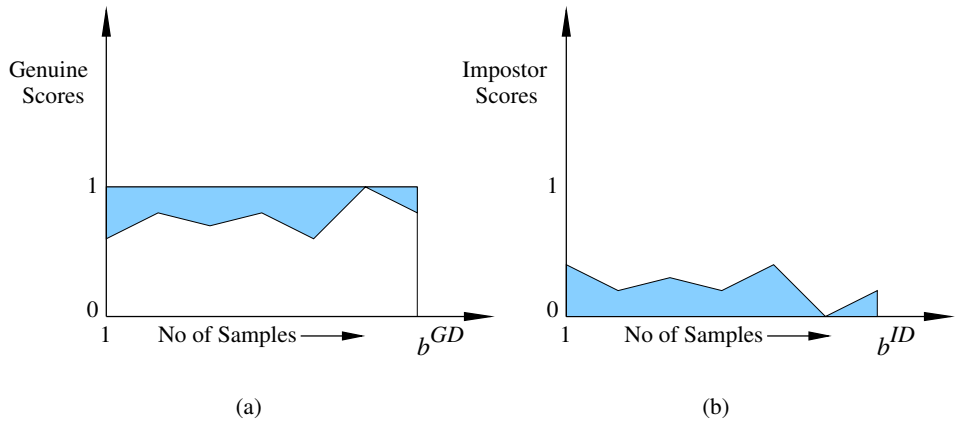
maximum values of l and k (L and K, respectively) depend on original image size and the subspace method used.

The terms *Goat* and *Lamb* originated from speaker verification [34] and then propagated in the field of biometry [107]. *Goats* are the group of subjects that are hard to authenticate and generate the majority of false rejects. The group of subjects that are easy to imitate and cause false acceptance are regarded as *Lambs*. Appropriate subband for a person gathers all common intra-class detail features specific to a subject as well as eliminates the common inter-class structural features present across subjects. So, it can be stated that a proper subband face has the power to extract features that are invariant within a subject and at the same time discriminant across different subjects. *Goatishness* and *labmishness* measures for each person are calculated from the genuine and impostor distributions obtained from the confusion matrix on a validation set.

Formally, let n and p be the number of samples per subject used for training and validation, respectively. The columns and rows of a confusion matrix correspond to the training and validation samples for different classes arranged sequentially. So a confusion matrix (CM) can be organized as follows,

$$CM = \begin{bmatrix} m_{(11)(11)} & \dots & m_{(11)(1j)} & \dots & m_{(11)(1n)} & \dots & m_{(11)(ij)} & \dots & m_{(11)(Cn)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ m_{(1s)(11)} & \dots & m_{(1s)(1j)} & \dots & m_{(1s)(1n)} & \dots & m_{(1s)(ij)} & \dots & m_{(1s)(Cn)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ m_{(1p)(11)} & \dots & m_{(1p)(1j)} & \dots & m_{(1p)(1n)} & \dots & m_{(1p)(ij)} & \dots & m_{(1p)(Cn)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ m_{(rs)(11)} & \dots & m_{(rs)(1j)} & \dots & m_{(rs)(1n)} & \dots & m_{(rs)(ij)} & \dots & m_{(rs)(Cn)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ m_{(Cp)(11)} & \dots & m_{(Cp)(1j)} & \dots & m_{(Cp)(1n)} & \dots & m_{(Cp)(ij)} & \dots & m_{(Cp)(Cn)} \end{bmatrix} \tag{3.1}$$

An element $m_{(rs)(ij)}$ represents the similarity measure (score value) between s^{th} validation sample from class r and j^{th} training sample from class i and takes value in the range of $[0, 1]$. Now the genuine and impostor score sets for class i, denoted as



(a) (b)

Figure 3.7: Area difference between genuine and impostor distribution with their ideal counterparts.

GD_i and ID_i, can written as,

$$GD_i = \{m_{(rs)(ij)} \mid i = r\ , j = 1, 2, ..., n \text{ and } s = 1, 2, ..., p\} \quad (3.2)$$

$$ID_i = \{m_{(rs)(ij)} \mid i \neq r,\ r = 1, 2, ..., C,\ j = 1, 2, ..., n \text{ and } s = 1, 2, ..., p\} \quad (3.3)$$

We have formulated four different criteria for measuring goatishness and lambishness for a person. Ideally, all elements of the confusion matrix for a genuine subject should be one (1) whereas, elements for impostor subjects should be zero (0). We have tried to measure the amount of deviations of genuine and impostor distributions from their ideal counterparts. The criteria need to be minimized for choosing appropriate subbands. The criteria are described below:

3.3.1 Criteria used for measuring Goatishness and Lambishness

- *First Criterion* (C_1): Let the number of elements in the genuine and impostor sets be $b^{GD} = n * p$ and $b^{ID} = n * (C - 1) * p$, respectively irrespective of any subject. Taking the elements along x-axis, if we plot the scores of the genuine samples we obtain a multimodal hill-shaped curve (genuine distribution). The difference in area between this and the ideal curve gives the measure of deviation for genuine distribution from its ideal counterpart. For impostor distribution, the measure is the area under impostor curve (see Figure 3.7).

  So goatishness and lambishness measures for i^{th} subject denoted as G_i and L_i,

are expressed as,

$$\begin{aligned} G_i &= b^{GD} - area(GD_i), \\ L_i &= area(ID_i). \end{aligned} \tag{3.4}$$

where $area(Y)$ is a function which returns area under curve Y. The lower values of G_i and L_i reduce the chance of false rejection and false acceptance, respectively.

- *Second Criterion* (C_2): Second criterion utilizes the deviation of mean and variance of genuine and impostor distributions from their ideal distributions. Ideally, genuine distribution should have mean as one (1) and variance as zero (0). Similarly, for impostor distribution mean and variance should be zero ideally. Here we have considered additive effect of the deviations of mean and variance from ideal values. So goatishness and lambishness measure for i^{th} subject is written as,

$$\begin{aligned} G_i &= (1 - \mu(GD_i)) + \sigma(GD_i), \\ L_i &= \mu(ID_i) + \sigma(ID_i). \end{aligned} \tag{3.5}$$

where μ and σ provides the mean and variance of their argument, respectively.

- *Third Criterion* (C_3): Multiplicative effect of the deviations of mean and variance from their ideal values is considered as the third criterion. Here, the goatishness and lambishness measures for i^{th} subject are formulated as,

$$\begin{aligned} G_i &= (1 - \mu(GD_i))\, exp(\sigma(GD_i)), \\ L_i &= \mu(ID_i)\, exp(\sigma(ID_i)). \end{aligned} \tag{3.6}$$

- *Fourth Criterion* (C_4): Fourth criterion uses the concept of performance metrics used in biometry. *ZeroFRR* is the lowest FAR for which no false rejection occurs. The threshold (t) corresponding to *ZeroFRR*, denoted by $\mathcal{T}_{ZeroFRR}$, is the minimum score among the genuine scores, as FRR becomes non-zero if the operating threshold is set as greater than $\mathcal{T}_{ZeroFRR}$. Let $ID_i^{\mathcal{T}_{ZeroFRR}}$ be the set of impostor scores having values above $\mathcal{T}_{ZeroFRR}$ and can be defined as,

$$ID_i^{\mathcal{T}_{ZeroFRR}} = \{x \in ID_i \mid x > \mathcal{T}_{ZeroFRR}\}. \tag{3.7}$$

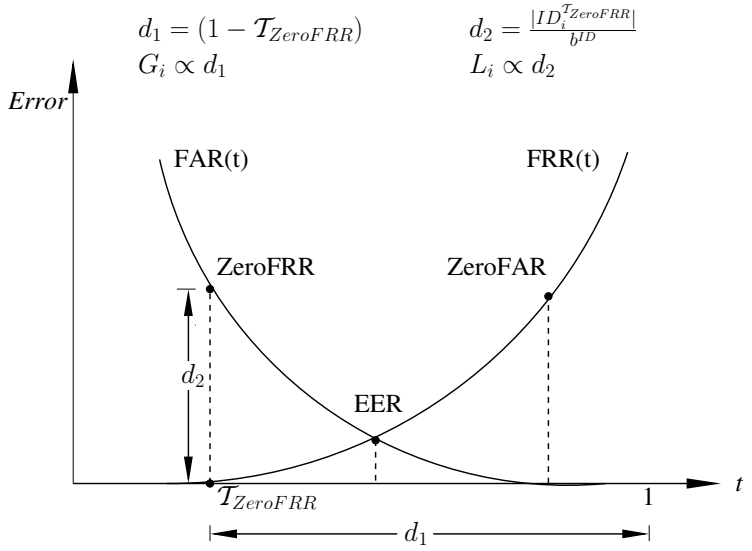


Figure 3.8: *ZeroFRR* and $\mathcal{T}_{ZeroFRR}$ are illustrated using hypothetical curves of *FAR(t)* and *FRR(t)*.

Hence, goatishness and lambishness can be expressed as,

$$\begin{aligned} G_i &= exp\left(1 - \mathcal{T}_{ZeroFRR}\right), \\ L_i &= exp\left(\frac{|ID_i^{\mathcal{T}_{ZeroFRR}}|}{b^{ID}}\right). \end{aligned} \tag{3.8}$$

where ID_i is the set of impostor scores for i^{th} subject and $|.|$ denotes the cardinality of a set. The significance of G_i and L_i are explained using a diagram in Figure 3.8.

3.3.2 Subject-Specific Subband Selection Algorithm

The above defined criteria are used in a subject-specific subband selection algorithm. As described in Section 3.2.2, each (l, k) pair corresponds to a subband face ($A_l - A_k$) which is obtained by reconstructing from subbands upto the level-l, after suppressing the approximation at level-k, where $0 \leq l < k \leq \log N$, $N * N$ being the resolution of the image. Given (l, k), we transform (DWT) all images from the training and validation sets to the subband face, $(A_l - A_k)$. Then we use any subspace method for feature extraction and generate a confusion matrix (see Eqn. 3.1) using nearest-neighbor policy followed by a normalization technique to convert score values in the range [0, 1]. From the confusion matrix the genuine and impostor score sets are

generated for obtaining goatishness and lambishness measures for each subject. Thus for each $(l,\ k)$ pair, we will obtain C goatishness and lambishness pairs corresponding to every subject, and therefore $(G_i^{(l,\ k)},\ L_i^{(l,\ k)})$ represents the corresponding pair for the i^{th} subject. Now a two step search is performed on $(G_i^{(l,\ k)},\ L_i^{(l,\ k)})$'s for each subject over all $(l,\ k)$ pairs $(l = 0, 1, .., L$ and $k = (l+1), ..., K)$. First the set, denoted as MG_i, of all $(l,\ k)$ pairs giving minimum G_i for i^{th} subject is obtained. MG_i can be written as,

$$MG_i = \{(l,\ k) \mid (l,\ k) = \arg\min_{(l,\ k)} G_i^{(l,\ k)}\}. \tag{3.9}$$

Among all the pairs in MG_i, a single pair denoted as $(l_i,\ k_i)$ is selected which gives minimum value for L_i. So, $(l_i,\ k_i)$ represents the optimal subband face for subject i and is obtained by,

$$(l_i,\ k_i) = \arg\min_{(l,\ k)} \hat{L}_i^{(l,\ k)} \mid (l,\ k) \in MG_i. \tag{3.10}$$

where $\hat{L}_i^{(l,\ k)}$'s are the lambishness values of the pairs selected in MG_i. This operation is repeated for all subjects. So final output of subband selection algorithm provides C optimal $(l,\ k)$ pairs, where $(l_i,\ k_i)$ or $(A_{l_i} - A_{k_i})$ represents optimal subband face for i^{th} subject in the database.

The two step searching procedure discussed above, can be considered equivalent to minimizing a weighted cost function with G_i and L_i, for all $(l,\ k)$ pairs, by assigning weights say, 0.8 and 0.2 , respectively. The reason behind using a two-step search, with G_i being minimized prior to L_i, instead of giving equal weightage to both of the measures is two-fold:

1. The minima of the sum, $S_i^{(l,\ k)} = G_i^{(l,\ k)} + L_i^{(l,\ k)}$, may not contain any of the minima from G_i or L_i. Thus it has a large scope of not detecting the local minima's from both set and can give a suboptimal result.

2. Searching with two-steps in a reverse way (L_i prior to G_i) does not work well due to the presence of a very few (single in almost all cases) minima's for L_i's, which provides no scope of minimizing G_i in any sense. Hence, an initial search for G_i gives a rich collection of local minima's from which searching for $min(L_i)$ yields good results.

The algorithm for obtaining the subject-specific subbands for best recognition performance is given below:

Algorithm for Optimal Subject-specific Subband Selection

1. *Compute Goatishness $G_i^{(l,\,k)}$ and Lambishness $L_i^{(l,\,k)}$ measures for all subjects and all pairs of (l, k)'s, as follows:*

   For l = 0,1,.., L and k = l, l+1,.., K do

   1.1 *If $l = k$, Generate l^{th} level approximation A_l for all images from training and validation set.*
   else, Generate subband face $(A_l - A_k)$ for all images from training and validation set.

   1.2 *Apply any subspace method (PCA, LDA, 2D-PCA, 2D-LDA or DCV) on the training data.*

   1.3 *Compute confusion matrix $CM^{(l,\,k)}$ on validation set after projecting them on the subspace obtained in step 1.2. Extract sets $GD_i^{(l,\,k)}$ and $ID_i^{(l,\,k)}$ for $i = 1, 2, ..., C$. (using Eqn. 3.2-3.3)*

   1.4 *Calculate $G_i^{(l,\,k)}$ and $L_i^{(l,\,k)}$ from $GD_i^{(l,\,k)}$ and $ID_i^{(l,\,k)}$, respectively for $i = 1, 2, ..., C$, using any of the criteria described in Section 3.3.1.*

2. *Selection of optimal subband $A_{l_i} - A_{k_i},\ i = 1, 2, .., C$:*

   For i = 1,2,..., C do

   2.1 *Select the (l, k) pairs corresponding to the minima's of G_i's to form MG_i (see Eqn. 3.9).*

   2.2 *Select a single pair $(l_i,\ k_i)$ from the set MG_i, which provides the minimum value of $\hat{L}_i$ (see Eqn. 3.10).*

3. *The optimal subband is selected as: $(l_i,\ k_i)\ for\ i = 1, 2, ..., C$.*

Once the optimal subband faces are selected, we construct C different subspaces where each one of them corresponds to one optimal subband face (i.e. one pair of *(l, k)* values) for a subject. Thus, i^{th} subspace denoted as LSA_i is constructed by generating the subband faces, $A_{l_i} - A_{k_i}$, for all the training samples and applying one

of the five subspace methods on them. During the identification phase, we create all subject-specific subband faces $T_{(l_i,\, k_i)},\ i = 1, 2, ..., C$ for the test image, T. Then the C different subband faces are projected onto the corresponding subspaces to obtain the distances from C different classes. More specifically, i^{th} test subband face $T_{(l_i,\, k_i)}$ is projected on i^{th} subspace LSA_i to obtain the (Euclidean) distance of the test sample from i^{th} class. A process of normalization (using min-max, with score value in range [0, 1]) is performed separately on each subspace to compare the C different distances computed in C different subspaces for obtaining the class label based on minimum or maximum membership rule. Let $D = [d_1, \ldots, d_i, \ldots, d_C]$ be the vector of distance values obtained from C different subspaces after normalization. Now the class label for test sample T, $\mathcal{L}(T)$, is obtained by using minimum membership rule on D.

$$\mathcal{L}(T) = \arg\min_c d_c. \tag{3.11}$$

The method for verification is simpler than the identification stage as we claim a person's identity and allow the system to verify it. Only one subband face $T_{(l_c,\, k_c)}$ is generated for the test sample T, where the claimed identity is class c. Then $T_{(l_c,\, k_c)}$ is projected on the c^{th} subspace LSA_c to obtain the distance measure from c^{th} class. The calculated distance measure is checked against a predefined threshold value to produce the output as acceptance or rejection. In the next section, we provide experimental results for our proposed method and also compare our method with the approach of Ekenel's multiresolution face recognition [36] (the closest and recent competitor of our approach).

3.4 Experimental Results

3.4.1 Databases Used

In this chapter, three face databases were used for experiments - Yale, PIE and ORL, where:

- The Yale face database has 15 subjects with 11 samples each, having variations in expression and illumination for each subject.
- A subset of the PIE (Pose Illumination and Expression) face database [112, 5] with 60 subjects, where only the frontal poses (camera: c27) were used. For

each subject, 42 samples (flashes from 21 different directions with and without the room lights on) were used for our study. Henceforth we use the following notations for each of the 21 flashes with and without room lights on:
(i) 21 flashes with room lights on: lights_flash02 - lights_flash22,
(ii) 21 flashes without room lights on: illum_flash02 - illum_flash22.
The face samples were extracted from the original gray level images for the PIE database, based on the specification of eye locations given in [5].

- The ORL face database has 40 subjects with 10 samples each. There is no change in illumination but significant (near frontal, no profile views) change in the face pose.

3.4.2 Performance Analysis on Three Standard Face Databases

To select subject-specific optimal subbands we split the image database into three disjoint sets, namely training, validation and testing set. Training set along with validation set are used for subband selection, and then the performance of selected subbands is observed on the testing set. The number of images used for a subject over the three sets, for all three databases are given in Table 3.1. Fig. 3.9 shows the images in training set for a single subject on Yale, PIE and ORL databases. For subband selection, the image size has been kept as 64*64 for all subspace methods except DCV where it is 25*25. The maximum values for l and k are 2 and 4 for DCV, while for other subspace methods they were chosen as 4 and 7 , respectively. For the PIE database, which has 42 frontal samples, only 4 face images corresponding to 2 frontal flashes (illum_flash08, lights_flash08, illum_flash11, lights_flash11) are used for training.

Table 3.2 shows the performances of subspace methods on original images for three databases. Table 3.3 shows the verification (in terms of Equal Error Rate or EER) and recognition (in terms of Peak Recognition Accuracy or PRA) performance of subband face representation integrated with subspace methods (PCA, LDA, 2D-PCA, 2D-LDA and DCV) on Yale database. The results in the column labeled "Subband Face(C_1)" corresponds to that obtained using optimal subband faces for subjects as selected by minimizing the first criterion discussed in Section 3.3.1. Same

Table 3.1: Sample distribution (per subject) in training, validation and testing sets for Yale, PIE and ORL databases.

Set	Yale	PIE	ORL
Training	4	4	3
Validation	3	12	4
Testing	4	26	3

Table 3.2: Peak Recognition Accuracy (PRA) and EER of original gray-level face image with PCA, LDA, 2D-PCA, 2D-LDA and DCV for Yale, PIE and ORL databases.

	Yale		PIE		ORL	
Methods	PRA	EER	PRA	EER	PRA	EER
PCA	81.67	7.78	58.85	17.05	80.83	3.33
LDA	85.00	6.67	88.14	2.39	79.17	4.21
2D-PCA	85.00	5.44	75.96	7.56	80.83	3.33
2D-LDA	81.67	5.00	83.59	4.10	86.67	2.31
DCV	81.67	5.71	85.90	3.29	82.50	5.00

is applicable for the other three columns of the table. For five subspace methods namely, PCA, LDA 2D-PCA, 2D-LDA and DCV, we obtained maximum accuracies of 95.00%, 100%, 98.33%, 93.33% and 95.00%, respectively. In case of EER, LDA provides a minimum value of 0.47% using the fourth criterion for selecting subbands. Among others, both PCA and DCV are observed to provide EER of 3.33% for second criterion. We gained a maximum accuracy of 15% over original face in case of LDA when tested with the fourth criterion.

The performance on PIE database is tabulated in Table 3.4. All criteria provide same PRA and EER of 85.38% and 4.51%, respectively for PCA. DCV produces the maximum PRA of 95.32% and minimum EER of 1.13% using the third criterion. Over all subspace methods, the maximum gain in PRA is obtained as 26.53% for PCA, for all the criteria.

Table 3.3: Peak Recognition Accuracy (PRA) and EER of *subband face representation* integrated with PCA, LDA, 2D-PCA, 2D-LDA and DCV with subject-specific subbands obtained using four criteria on Yale database.

	Subband Face(C_1)		Subband Face(C_2)		Subband Face(C_3)		Subband Face(C_4)	
Methods	PRA	EER	PRA	EER	PRA	EER	PRA	EER
PCA	91.67	6.67	95.00	3.33	91.67	6.67	93.33	6.07
LDA	96.67	2.26	96.67	2.26	96.67	2.26	**100**	**0.47**
2D-PCA	98.33	5.00	96.67	5.00	98.33	5.00	93.33	5.36
2D-LDA	91.67	5.00	93.33	5.00	93.33	5.00	91.67	6.67
DCV	95.00	3.33	95.00	3.33	95.00	3.33	81.67	5.71

Table 3.4: Peak Recognition Accuracy (PRA) and EER of *subband face representation* integrated with PCA, LDA, 2D-PCA, 2D-LDA and DCV with subject-specific subbands obtained using four criteria on PIE database.

	Subband Face(C_1)		Subband Face(C_2)		Subband Face(C_3)		Subband Face(C_4)	
Methods	PRA	EER	PRA	EER	PRA	EER	PRA	EER
PCA	85.38	4.51	85.38	4.51	85.38	4.51	85.38	4.51
LDA	93.27	2.55	92.76	2.76	92.63	2.88	91.86	3.09
2D-PCA	92.31	2.03	92.24	2.03	92.05	2.18	92.76	1.94
2D-LDA	87.88	4.29	86.15	4.76	86.92	4.55	87.56	4.33
DCV	95.13	1.15	94.94	**1.13**	**95.32**	**1.13**	92.63	1.98

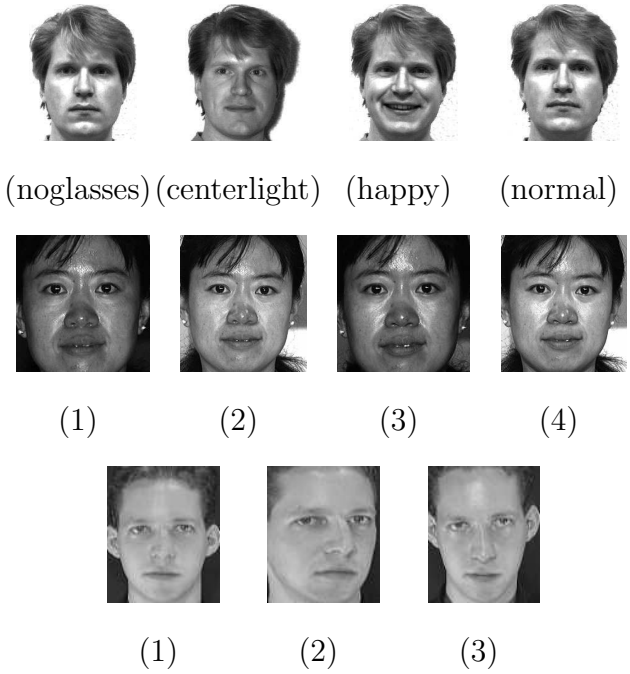


Figure 3.9: Training samples for three databases: first, second and third row shows the training set for Yale, PIE, and ORL databases, respectively.

The performance of subject-specific subband face on ORL database, which has a considerable amount of pose variation is given in Table 3.5. The maximum PRA obtained on this database is 91.67% for 2D-LDA using the first criterion. The same method also gives the minimum EER of 1.67%. LDA is observed to provide a maximum PRA gain of 10.00% with fourth criterion on this database.

3.4.3 Comparison with Ekenel's Multiresolution Face Recognition [36]

We have also compared the performance of our approach with a recent wavelet-based method denoted as Ekenel's multiresolution face recognition [36]. They used 2D DWT to decompose the original image upto level-III and obtained A_1, H_1, V_1, D_1 in the first level, A_2, H_2, V_2, D_2 in second level, and at the third level they generated 16 subband images by decomposing A_2, H_2, V_2 and D_2 further and thus obtained a total of 24 subsampled images. Daubechies 4-tap wavelet was used for wavelet tree decomposition. For each subband they extracted features using PCA and ICA, and nearest neighborhood criterion was used for classification. They used three different

Table 3.5: Peak Recognition Accuracy (PRA) and EER of *subband face representation* integrated with PCA, LDA, 2D-PCA, 2D-LDA and DCV with subject-specific subbands obtained using four criteria on ORL database.

	Subband Face(C_1)		Subband Face(C_2)		Subband Face(C_3)		Subband Face(C_4)	
Methods	PRA	EER	PRA	EER	PRA	EER	PRA	EER
PCA	88.33	4.83	88.33	4.64	88.33	4.64	81.67	7.50
LDA	88.33	4.16	87.50	4.12	87.50	4.12	89.17	4.16
2D-PCA	81.67	7.50	81.67	7.50	81.67	7.50	81.67	7.50
2D-LDA	**91.67**	**1.67**	91.67	1.81	91.67	2.50	90.00	1.98
DCV	85.83	3.44	86.67	3.50	87.50	3.33	88.33	3.59

distance metrics namely, the L_1 norm, the L_2 norm and the normalized correlation coefficient. In two separate sets of experiments they searched for the subbands insensitive to expression and illumination. They report that very less number of subbands (A_1, A_2, A_3) hold expression invariant information. The same comment does not hold for illumination invariation. In both cases, all such subbands performing equally good or better (compared to original face) were selected for further processing using data, feature and decision level fusion. In decision fusion they used sum, product and max rules.

We have experimented their technique using PCA, in two different ways. In the first case, we follow their implementation of selecting successful subbands based on the results obtained on testing set. This is however not technically sound in the sense that, selected successful subbands are over-tuned on the testing set and will obviously produce the best possible result only on the testing set. So the right way of selecting successful subbands is to evaluate them on a validation set (as given in Section 3.4.2) and then combine them to verify the performance on the testing set of a database. This is the second approach used for comparing the performance of our proposed approach with that of [36].

For the first case, Table 3.6 shows the best performing (successful) subbands selected using the testing set. Table 3.7 shows the performance of Ekenel's multireso-

lution face recognition [36] for Yale, PIE and ORL databases based on the successful subbands determined on testing set. From the results given in Table 3.7, we can observe that the process of decision fusion based on sum and product rules perform the best among all decision combination strategies proposed in [36]. However, in case of ORL, this strategy fails to improve upon that provided by the gray-level image using PCA.

In the second case of experimentation, the successful subbands are selected using the validation set, as used in our proposed technique (see Table 3.1). Results are given in Table 3.8-3.9, in a similar manner, as in Table 3.6-3.7. We can notice that the over-tuning done in the first case offers apparently better results, by comparing the performances given in Table 3.7 (first case) and the corresponding top seven rows of Table 3.9 (second case).

Table 3.6: Best performing and successful subbands for Ekenel's multiresolution face recognition [36] for Yale, PIE and ORL databases determined on testing set.

	Yale	PIE	ORL
Best Performing Subband	V_3	HA_2	A_2
Successful Subbands Selected for Data, Feature, Decision Fusion	A_1, V_2, V_3	A_1, H_1, V_1, D_1 H_2, V_2, D_2 H_3, V_3, D_3 HA_3, HH_3, HV_3, HD_3 VA_3, VH_3, VV_3, VD_3 DA_3, DH_3, DV_3, DD_3	A_1, A_2, A_3

Hence, to provide a meaningful comparative study, which is also practically useful, we use the second approach (based on validation set) to compare the performance of our underlying *subband face representation* with that suggested in [36], using data, feature and decision level fusion. We do so with identical feature extraction and distance metric, which is PCA and L_2 norm in both cases. Image size is also kept as same for both cases. We use the same validation set to select the subbands performing equally good and better than original image for multiresolution face recognition. The

Table 3.7: Performance of Ekenel's multiresolution face recognition [36] for Yale, PIE and ORL databases based on the successful subbands determined on testing set.

	Yale		PIE		ORL	
Methods	PRA	EER	PRA	EER	PRA	EER
Original Face	81.67	7.78	58.85	17.05	**80.83**	3.33
Best Performing Subband	88.33	5.95	90.64	7.50	70.83	8.49
Data Fusion	76.67	16.67	34.46	28.21	69.17	7.07
Feature Fusion	73.33	20.00	23.91	32.42	71.67	7.68
Decision Fusion-Sum Rule	**85.00**	8.38	88.65	3.24	70.00	7.38
Decision Fusion-Product Rule	**85.00**	8.67	**90.77**	8.82	**72.50**	6.82
Decision Fusion-Max Rule	81.67	10.00	66.09	49.67	67.50	7.5

feature (subband) selection criterion is however different. Then those selected subbands are combined at data, feature and decision level, as described in multiresolution approach [36] to obtain the results in Table 3.9.

For Yale database, original image with PCA and L_2 norm gives 81.67%. Experimentation on subband selection for fusion shows that A_1, A_2 and A_3 (see Table 3.8) perform same as A_0 (original image) on validation set. So A_1, A_2 and A_3 are combined at data, feature and decision level and the performance was observed on the testing set. Note that, any of the fusion strategies (Table 3.9) do not perform better than original image. It was also observed that A_1 is the best performing subband on testing set for Yale database. In our approach, four criteria used for subband selection provide accuracies in the range [91.67%-95.00%] (refer to first row of Table 3.3) while decision fusion using three (sum, product and max) rules in Ekenel's multiresolution face recognition provide a PRA of only 70% (Table 3.9). We are however able to obtain a 100% recognition in case of Yale database (see last row of Table 3.9), using fourth criterion C_4 of subband selection and LDA (see second row of Table 3.3).

For PIE database, we obtain comparable performance with Ekenel's multiresolution technique. PRAs for the best performing subband (HA_3), decision fusion for sum and product rules are 90.64%, 85.83% and 88.20% respectively, while our subject-

Table 3.8: Best performing and successful subbands for Ekenel's multiresolution face recognition [36] for Yale, PIE and ORL databases determined on validation set.

	Yale	PIE	ORL
Best Performing Subband	A_1	HA_3	A_2
Successful Subbands Selected for Data, Feature, Decision Fusion	A_1, A_2, A_3	A_1, H_1, V_1, D_1 A_2, H_2, V_2, D_2 H_3, V_3, D_3 HA_3, HH_3, HV_3, HD_3 VA_3, VH_3, VV_3, VD_3 DA_3, DH_3, DV_3, DD_3	A_1, A_2, A_3

specific subband face representation produces 85.38% (refer to first row of Table 3.4). As specified in the last row of Table 3.9, we obtain the best performance of 95.32%, using DCV and third criterion (C_3) for subband selection (see last row in Table 3.4).

In case of ORL database, Ekenel's multiresolution face recognition fails to work and all fusion techniques provide worse performance than the original face, while our technique is observed to better the PRA by 7.50% over original face in case of PCA (see Table 3.5). The best performance of our proposed technique in case of ORL was provided by C_1, C_2 and C_3 using 2D-LDA, which is 91.67%.

For a better understanding, we have added best overall performance provided by *subband face representation* on the three databases at the last row of Table 3.9. The EER values have also been provided which are far superior in our method than that provided by Ekenel's [36] method. In fact, the EER values of decision fusion strategies in [36] are generally inferior than the original face (Table 3.9) itself. The subband face performs appreciably well in presence of expression (Yale database), illumination (PIE and Yale database) and considerable pose (ORL database) variations. This gives us the insight that an optimally chosen subband face is a very powerful tool to conquer the curse of large illumination, expression and pose variations, as demonstrated in the results on Yale, PIE and ORL databases.

Table 3.9: Performance of Ekenel's multiresolution face recognition [36] for Yale, PIE and ORL databases based on the successful subbands determined on validation set.

	Yale		PIE		ORL	
Methods	PRA	EER	PRA	EER	PRA	EER
Original Face	81.67	7.78	58.85	17.05	80.83	3.33
Best Performing Subband	73.33	20.88	**90.64**	7.50	70.83	8.49
Data Fusion	66.67	23.60	31.15	29.59	69.17	7.07
Feature Fusion	65.00	24.90	27.12	30.96	71.67	7.68
Decision Fusion-Sum Rule	**70.00**	18.33	85.83	4.00	70.00	7.38
Decision Fusion-Product Rule	**70.00**	34.05	88.20	49.69	**72.50**	6.82
Decision Fusion-Max Rule	**70.00**	16.67	63.37	9.38	67.50	7.5
Proposed Method (with PCA)	95.00	3.33	85.38	4.51	88.33	4.64
Proposed Method (Overall)	**100.00**	**0.47**	**95.32**	**1.15**	**91.67**	**1.67**

3.5 Conclusion

In this chapter, we have proposed the use of subband face (reconstructed from selective wavelet subbands) representation for the task of face recognition. This representation is evaluated using PCA, 2D-PCA, LDA, 2D-LDA and DCV based subspace projections for Yale, ORL and PIE face databases. It is observed that the subband face performs significantly better than the wavelet decomposed subbands of the face [36], in the presence of variations in illumination, expression and pose. The novelty of this method is that the subbands containing more of discriminatory information are selected for face representation, whereas those with common features and redundant information are discarded. Therefore subband face is an effective representation for frontal face recognition achieving high recognition rate even under hard testing conditions. The performance of the subband face can be further improved by obtaining an optimal selection of discriminating subbands in case of full-tree decomposition of the image using wavelets, use of modular [67] subband faces and also by exploring more sophisticated criterion for subband selection.

CHAPTER 4

Dual Space Face Recognition using Feature and Decision Fusion

We propose a new face recognition technique by combining information from null space and range space of within-class scatter of a face space. In our technique, the combination of information is attempted in two different levels: (i) Feature level and (ii) Decision level. The combination of information at feature level poses a problem of optimally merging two eigenmodels obtained separately from null space and range space. We use two different methods: 1) Covariance Sum and 2) Gramm-Schmidt Orthonormalization to construct a new combined space, named as dual space, by merging two different set of discriminatory directions obtained separately from null space and range space. We employ forward and backward selection techniques to select the best set of discriminative features from dual space and use them for face recognition.

Combining information at decision level requires the construction of classifiers individually on null space and range space. Then these two classifiers are combined using three decision fusion strategies. Along with two classical decision fusion strategies sum rule and product rule, we employ our own decision fusion technique which exploits each classifier space separately to enhance combined performance. Our method of decision fusion uses Linear Discriminant Analysis (LDA) and nonparametric LDA on classifier's response to enhance class separability at classifier output space. Experimental results on three public databases, Yale, ORL and PIE will show the superiority of our method over a face recognition technique called Discriminative Common Vectors (DCV) [20], which is based only on the null space of within-class scatter.

Rest of the chapter is organized as follows. Section 4.1 provides a brief review on existing subspace methods for face recognition followed by a concise introduction

to our proposed method. Section 4.2 describes the method of obtaining two separate eigenmodels in range space and null space of within-class scatter matrix. It also explains the method of obtaining two separate classifiers on the basis of two disjoint sets of optimal discriminatory directions (eigenmodels as specified before) obtained from range space and null space. Section 4.3 explains the feature fusion strategy with the following details: The techniques used for merging two eigenmodels are studied in Section 4.3.1. Section 4.3.2 presents the search techniques and criterion used to reorder and select optimum set of discriminative directions. Section 4.3.3 provides the overall algorithm for our proposed feature fusion method. Section 4.4 provides an in-depth discussion on the decision fusion strategy. This section contains the following details: Section 4.4.1 elaborates two classical decision fusion strategies (sum rule and product rule) used for combining the classifiers obtained on null space and range space. Section 4.4.2 describes the decision fusion strategy proposed by us. Section 4.4.3 provides the overall algorithm for our proposed method of decision fusion. Finally, Section 4.5 discusses the experimental results of the proposed methods tested on three face databases and Section 4.6 concludes the chapter.

4.1 Subspace Methods

Most face-based biometric systems developed so far have attempted to either extract second-order features from the face image or project onto a lower dimensional subspace (PCA, LDA and their variants) in order to obtain discrimination. In PCA, the pixels of the images are considered as features and projected onto a lower dimensional space to obtain a larger scatter [35]. LDA maximizes the ratio of between-class scatter and within-class scatter [39]. A brief discussion of a few relevant subspace methods of face recognition is presented in the following.

In LDA, the number of training samples is usually less than the dimension of the sample space. This leads to the so-called small-sample-size (SSS) problem due to the singularity of the within-class scatter matrix. There are various attempts to solve the SSS problem of LDA. A two stage PCA+LDA approach has been used in [120] that uses discriminant eigenfeatures. Fisherface [12] and discriminant component analysis [143] also address this problem. Here the higher dimensional face data is projected to

a lower dimensional space using PCA and then LDA is applied to this PCA subspace. But the removed subspace may contain some useful discriminative information which is lost in this process.

Chen et al. [25] suggested that the null space of the within-class scatter S_w contains most discriminative information, and hence they proposed LDA in null space of S_w called N-LDA. However, when the number of training samples is large, the null space becomes small resulting in loss of discriminative information outside this null space. The performance of the null space depends on the dimension of the null space. Thus any preprocessing that reduces original sample space should be avoided. Another drawback of this approach is that it involves solving the eigenvalue problem for a very high dimensional matrix. Yu et al. [138] proposed an algorithm which basically uses simultaneous diagonalization method [39]. First, the null space of between-class scatter S_b is removed, assuming that the null space of S_b contains no discriminative information, and then the method seeks a projection to minimize S_w in the transformed subspace of S_b. As the rank of S_b is smaller than that of S_w, removing the null space of S_b may lose entire null space of S_w. So S_w is likely to be of full rank after this removal [14, 28, 45]. One more drawback is that the whitening of S_b is redundant in this method. Another novel method, proposed by Huang et al. [45] uses PCA+Null space approach. Here the core idea is that null space of S_w is useful for discrimination unlike that of S_b. Since the null space of total scatter matrix S_t is the intersection of the null spaces of S_b and S_w, this method applies PCA to remove the null space of S_t and then Null space method is used on reduced subspace of S_t. Hakan et al. [20] proposed a novel scheme to solve SSS problem called Discriminant common vectors (DCV) method. Among two algorithms to extract the discriminant common vectors for representing each person in the training set of face database, one algorithm uses within-class scatter matrix of the samples in the training set while the other uses the subspace methods and the Gramm-Schmidt orthogonalization procedure. These DCVs are used for classification of new faces.

Most of the null space based methods like DCV [20], PCA+Null space approach [45] explored null space of S_w while others like PCA+LDA [120] and Fisherface [12] eliminated a part of null space to obtain discriminative features. None of them ex-

plored both of the spaces to capture and then combine discriminative directions for enhancing discriminability across classes. Our approach efficiently exploit both the spaces by combining them in two different levels: (i) Feature level and (ii) Decision level.

For feature level fusion, we develop a dual space by combining the discriminative features from both range space and null space of within-class scatter matrix S_w. This allows us to utilize the whole set of discriminative directions present in an entire face space. As every face has a unique decomposition in null space and range space, we project all class means in both spaces to obtain two sets of projected means. Now each of these sets are used separately to search for the directions that discriminates them in that space. This step is equivalent to applying PCA on the set of projected means separately. These two eigenmodels are then combined using: 1) Covariance Sum method and 2) Gramm-Schmidt Orthonormalization [42]. These methods construct a new set of directions integrating the information from both spaces. Then we reorder and select the best combination among those directions to obtain the best discriminability across classes. The feature reordering and selection is performed using two techniques: 1) Forward Selection and 2) Backward Selection [39] on a validation set, based on a class separability criterion.

For decision level fusion, we extract two disjoint sets of optimal discriminatory basis separately from null space and range space to obtain two different classifiers. Then we combine the classifiers obtained on null space and range space using sum rule and product rule, two classical decision fusion techniques developed by Kittler [60],[59]. We also exploit each classifier separately using LDA and nonparametric LDA to enhance class separability at classifier response level and then combine them using sum rule. We denote the class scores provided by a classifier on a sample as response vector. Basically, we use response vectors as features vectors at decision level and employ LDA and nonparametric LDA to enhance class separability at classifier output space. Response vectors on a validation set (disjoint from training and testing sets of a database) is used as training data at decision level. Then the response vectors on testing set of the database are recalculated in the eigenmodel to improve combined classification accuracy.

4.2 Obtaining Eigenmodels in Range Space and Null Space of Within-class Scatter

In appearance based face recognition techniques, a face image of $w \times h$ pixels is represented by a vector in a $d\ (= wh)$-dimensional space. Therefore, each face image corresponds to a point in d-dimensional image space. Let the training set be defined as $X = [x_1^1, x_2^1, ..., x_N^1, x_1^2, ..., x_N^C]$, where C is the number of classes and N is the number of samples per class. An element x_j^i denotes j^{th} sample from class i and a vector in d-dimensional space $\mathcal{R}^d$. Total number of training samples is $M = NC$. Then within-class (S_w), between-class (S_b) and total (S_t) scatter matrices can be defined as,

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{N}(x_j^i - \mu_i)(x_j^i - \mu_i)^T, \tag{4.1}$$

$$S_b = \sum_{i=1}^{C} N(\mu_i - \mu)(\mu_i - \mu)^T, \tag{4.2}$$

$$S_t = \sum_{i=1}^{C}\sum_{j=1}^{N}(x_j^i - \mu)(x_j^i - \mu)^T = S_w + S_b. \tag{4.3}$$

where μ is the overall mean and μ_i is the mean of i^{th} class. Methods searching for discriminative directions only in null space of S_w try to maximize a criterion given as,

$$\begin{aligned} J(W_{opt}^{Null}) &= \arg \max_{|W^T S_w W|=0} |W^T S_b W| \\ &= \arg \max_{|W^T S_w W|=0} |W^T S_t W|. \end{aligned} \tag{4.4}$$

Similarly, the directions providing discrimination in range space can be searched by maximizing a criterion $J(W_{opt}^{Range})$ which is defined as,

$$J(W_{opt}^{Range}) = \arg \max_{|W^T S_w W|\neq 0} |W^T S_b W|. \tag{4.5}$$

To find the optimal projection vectors in null space, we project all samples in null space and then obtain W_{opt}^{Null} by applying PCA. As the basis vectors constituting null space do not contain intra-class variations, we obtain single common vector for each class. Any sample as well as mean of a class will provide the same common vector for that class after projection in null space. The same does not happen for range space as it contains the complete intra-class variations present in the face space. So, we

project only the class means in range space and apply PCA to obtain W_{opt}^{Range}. Let V and $\bar{V}$ be the range space and null space of S_w respectively and can be expressed as,

$$V = span\{\alpha_k \mid S_w\alpha_k \neq 0, \;\; k = 1, ..., r\}, \tag{4.6}$$

$$\bar{V} = span\{\alpha_k \mid S_w\alpha_k = 0, \;\; k = (r+1), ..., d\}. \tag{4.7}$$

where r ($< d$) is the rank of S_w, and $\{\alpha_1, ..., \alpha_d\}$ is the orthogonal eigenvector set of S_w. A typical example of null space ($\bar{V}$), range space (V) and eigenspectrum of a scatter matrix is shown in Figure 4.1. Now the matrices $Q = [\alpha_1, ..., \alpha_r]$ and

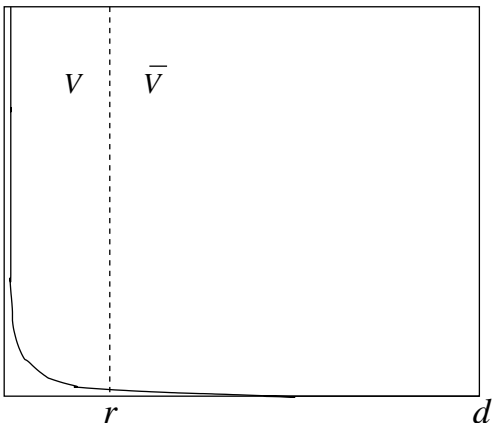


Figure 4.1: A typical eigenvalue spectrum with its division into null space and range space.

$\bar{Q} = [\alpha_{r+1}, ..., \alpha_d]$ correspond to the basis vectors for range space and null space respectively. As $\mathcal{R}^d = V \oplus \bar{V}$, every face $x_j^i \in \mathcal{R}^d$ has a unique decomposition of the form,

$$x_j^i = y_j^i + z_j^i \;\; where \tag{4.8}$$

$$y_j^i = Px_j^i = QQ^T x_j^i \in V, \tag{4.9}$$

$$z_j^i = \bar{P}x_j^i = \bar{Q}\bar{Q}^T x_j^i \in \bar{V}. \tag{4.10}$$

P and $\bar{P}$ are the orthogonal projection operators onto V and $\bar{V}$. As the range space is much smaller than null space we obtain z_j^i using y_j^i,

$$z_j^i = x_j^i - y_j^i = x_j^i - Px_j^i. \tag{4.11}$$

Now the vectors in Q can be constructed by selecting the first r eigenvectors of S_w obtained from eigen-analysis. Let x_{Null}^i and x_{Range}^i represent the vectors obtained from

the projection of μ_i (mean of i^{th} class) on null space and range space respectively and can be obtained by,

$$x^i_{Null} = \bar{Q}\bar{Q}^T\mu_i = \mu_i - QQ^T\mu_i, \tag{4.12}$$

$$x^i_{Range} = QQ^T\mu_i. \tag{4.13}$$

The scatter matrices S_{Null} and S_{Range} can be obtained from $X_{Null} = [x^1_{Null}, .., x^C_{Null}]$ and $X_{Range} = [x^1_{Range}, .., x^C_{Range}]$ respectively as,

$$S_{Null} = \sum_{i=1}^{C}(x^i_{Null} - \mu_{Null})(x^i_{Null} - \mu_{Null})^T, \tag{4.14}$$

$$S_{Range} = \sum_{i=1}^{C}(x^i_{Range} - \mu_{Range})(x^i_{Range} - \mu_{Range})^T. \tag{4.15}$$

where μ_{Null} and μ_{Range} are the mean of X_{Null} and X_{Range} respectively,

$$\mu_{Null} = \frac{1}{C}\sum_{i=1}^{C} x^i_{Null}, \tag{4.16}$$

$$\mu_{Range} = \frac{1}{C}\sum_{i=1}^{C} x^i_{Range}. \tag{4.17}$$

Now to obtain the discriminative directions in null space and range space, we need to maximize the following criteria,

$$J(W^{Null}_{opt}) = \arg\max_W |W^T S_{Null} W|, \tag{4.18}$$

$$J(W^{Range}_{opt}) = \arg\max_W |W^T S_{Range} W|. \tag{4.19}$$

Since, W^{Null}_{opt} and W^{Range}_{opt} are two independent sets of vectors we need to merge for obtaining maximum discriminability present in a face space. The merging techniques are described in the Section 4.3.1. The requirements for feature fusion strategy are two sets of discriminatory directions (i.e W^{Null}_{opt} and W^{Range}_{opt}), whereas for decision fusion we have to formulate the problem of designing two distinct classifiers, based on these two sets obtained separately from null space and range space.

For decision fusion, we project all the class means μ_i's on W^{Null}_{opt} to obtain $\Omega_{Null} = [\omega^1_{Null}, ..., \omega^C_{Null}]$. Similarly, the projection of μ_i's on W^{Range}_{opt}, gives $\Omega_{Range} = [\omega^1_{Range}, ..., \omega^C_{Range}]$, where:

$$\omega^i_{Null} = W^{Null}_{opt}\mu_i, \quad i = 1, ..., C \tag{4.20}$$

$$\omega^i_{Range} = W^{Range}_{opt}\mu_i, \quad i = 1, ..., C. \tag{4.21}$$

Ω_{Null} and Ω_{Range} represents the class templates in null space and range space, respectively. On arrival of a test sample, we project it onto W_{opt}^{Null} to compare with the templates in Ω_{Null}. Similarly, the projection of the test sample onto W_{opt}^{Range} will be compared with the templates in Ω_{Range}. So, on the presentation of a test sample x^{test}, the feature vectors are obtained as,

$$\omega_{Null}^{test} = W_{opt}^{Null} x^{test}, \tag{4.22}$$

$$\omega_{Range}^{test} = W_{opt}^{Range} x^{test}. \tag{4.23}$$

Then ω_{Null}^{test} is compared to the vectors in Ω_{Null} using Euclidean distance to obtain response vector $D_{Null}(x^{test}) = [d_{Null}^{1}(x^{test}), ..., d_{Null}^{C}(x^{test})]$. The response vector $D_{Range}(x^{test}) = [d_{Range}^{1}(x^{test}), ..., d_{Range}^{C}(x^{test})]$ in range space can be obtained by comparing ω_{Range}^{test} with the vectors in Ω_{Range}. The elements in $D_{Null}(x^{test})$ and $D_{Range}(x^{test})$ represent the similarity measures or score values for different classes. More specifically, $d_{Null}^{i}(x^{test})$ is the amount of evidence provided by null space classifier, D_{Null}, to the fact that x^{test} belongs to class i. Now the task of decision combination rule is to combine $D_{Null}(x^{test})$ and $D_{Range}(x^{test})$ for obtaining a combined decision vector, $\tilde{D}(x^{test})$. Detail discussion on the techniques for combining response vectors (also known as soft class labels and decision vectors) from two classifiers is provided in Section 4.4.

4.3 Feature Fusion

4.3.1 Techniques for Merging Eigenmodels

The problem of combining W_{opt}^{Null} and W_{opt}^{Range} is similar to the problem of merging two different eigenmodels obtained from two different sets of feature vectors X_{Null} and X_{Range}. Among two methods for combining them, first method uses X_{Null} and X_{Range} to obtain a set of combined discriminative directions W^{Dual}, while the other one forms a orthogonal basis W^{Dual} for a matrix $[W_{opt}^{Null}, W_{opt}^{Range}]$ using QR decomposition. Henceforth, we will refer to the first technique as Covariance Sum method [96] and the second technique as Gramm-Schmidt Orthonormalization, adopted from modified Gramm-Schmidt Orthonormalization technique [42] used for computing QR decomposition of a matrix. A brief description for Covariance Sum method and

Gramm-Schmidt Orthonormalization is given bellow.

4.3.1.1 Covariance Sum Method

Here the problem deals with the computation of an eigenmodel from two different sets of feature vectors X_{Null} and X_{Range} with the aim to combine the eigenmodels constructed separately on the two sets. As X_{Null} and X_{Range} contain the class means projected on null space and range space, this combination will merge the discriminative directions from both spaces. Let the combined feature vector set be $Z = [X_{Null}, X_{Range}]$ and total number of feature vectors in Z is $L = 2 * C$. The combined mean can be expressed as,

$$\mu_Z = \frac{1}{L}\left(C\mu_{Null} + C\mu_{Range}\right) = \frac{1}{2}\left(\mu_{Null} + \mu_{Range}\right) \tag{4.24}$$

Then the combined covariance matrix can be written as,

$$S_Z = \frac{1}{L}\left(\sum_{i=1}^{C} x^i_{Null}(x^i_{Null})^T + \sum_{i=1}^{C} x^i_{Range}(x^i_{Range})^T\right) - \mu_Z\mu_Z^T \tag{4.25}$$

Using Eqn. 4.14-4.15 in the above equation, we get,

$$\begin{aligned} S_Z &= \frac{1}{L}\left(CS_{Null} + C\mu_{Null}\mu_{Null}^T + CS_{Range} + C\mu_{Range}\mu_{Range}^T\right) - \mu_Z\mu_Z^T \\ &= \frac{1}{2}S_{Null} + \frac{1}{2}S_{Range} + \frac{1}{4}(\mu_{Null} - \mu_{Range})(\mu_{Null} - \mu_{Range})^T. \end{aligned} \tag{4.26}$$

Last term in the Eqn. 4.25 allows for a change of mean. Now the eigenvectors $\Phi = [\beta_1, ..., \beta_d]$ of S_Z can be computed from,

$$S_Z\Phi = \Phi\Theta \;\; and \;\; \Phi^T\Phi = I. \tag{4.27}$$

The range space of S_Z will give $W^{Dual}(= [\beta_1, ..., \beta_k])$, where k is the rank of S_Z.

4.3.1.2 Gramm-Schmidt Orthonormalization

Gramm-Schmidt Orthonormalization technique [42] is used to compute the QR factorization. Given A, a m-by-n $(m > n)$ matrix, it's QR factorization can be written as $A = QR$, where $Q \in \mathcal{R}^{m\times m}$ is orthogonal and $R \in \mathcal{R}^{m\times n}$ is upper triangular. If A has full column rank then the first n columns of Q form an orthogonal basis for range(A).

Thus, calculation of the QR factorization is one way to compute an orthogonal basis for a set of vectors.

In our case $A = [W_{opt}^{Null}, W_{opt}^{Range}] \in \mathcal{R}^{d \times 2(C-1)}$, $m = d$ and $n = 2(C-1)$, where $d >> 2(C-1)$. The following algorithm computes the factorization $A = QR$ where the columns of Q constitute the orthogonal basis for A and form W^{Dual}.

Algorithm for Gramm-Schmidt Orthonormalization

for $k = 1 : n$

$R(k,k) = \|A(1:m,k)\|_2$

$Q(1:m,k) = A(1:m,k)/R(k,k)$

for $j = k+1 : n$

$R(k,j) = Q(1:m,k)^T A(1:m,j)$

$A(1:m,j) = A(1:m,j) - Q(1:m,k)R(k,j)$

end

end

Finally, Q is assigned to W^{Dual}.

4.3.2 Search Criterion and Techniques for Optimal Feature Selection

After obtaining W^{Dual}, we select an optimal feature set from W^{Dual} to form W_{opt}^{Dual} by maximizing a class separability criterion. A set of features providing maximum value of separability criterion on a validation set (disjoint from training and testing set) is selected for evaluating on the testing subset of the database. The class separability criterion used in our method is given by,

$$J = \frac{tr(S_b)}{tr(S_w)}. \tag{4.28}$$

where S_b and S_w are between-class and within-class scatter given in Eqn. 4.1-4.2. We use backward and forward selection techniques which avoid exhaustive enumeration [39] as used in branch and bound selection technique. Even though these methods do not guarantee the selection of best possible feature subset, we use them for their simplicity.

4.3.2.1 Backward Selection

The backward selection procedure starts from the full set of n features. Then, eliminating one feature from n, all possible subsets of *n-1* features are obtained and criterion values are evaluated for each of them. Then the subset corresponding to maximum value of the criterion is selected as the best subset containing *n-1* features. This step repeats until the desired number of features are obtained. We evaluate all possible number of optimal features on a validation set and select one providing maximum separability. A toy example demonstrating backward selection with four features is shown in Figure 4.2(a).

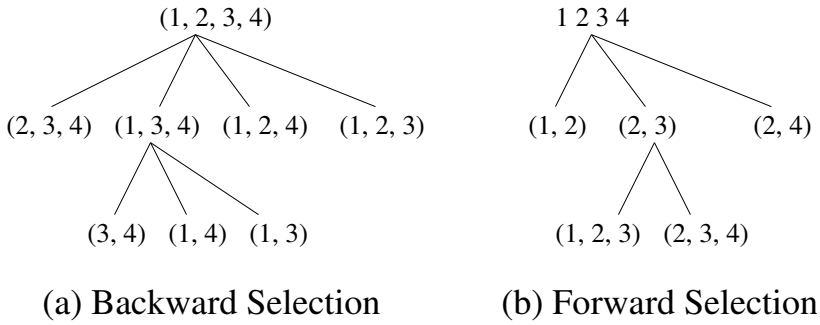


Figure 4.2: Stepwise feature subset selection

4.3.2.2 Forward Selection

Forward selection on the other hand starts from the evaluation of individual features. For n number of features, forward selection evaluates the value of criterion on every individual feature and select one which provides maximum value. Now one more feature is added to this selected feature to form a subset of two. All possible such subsets are evaluated against the criterion. The subset providing the maximum value of the criterion is selected as the best subset of two features. This continues until the algorithm finds a subset of desired number of features. We compute all possible number of optimal features on a validation set and select the feature set providing maximum separability to evaluate on testing set. An example describing the operation of forward selection is shown in Figure 4.2(b).

4.3.3 Algorithm for Feature Fusion

In this section, we describe the overall algorithm of our proposed method of feature fusion from both range space and null space of within-class scatter. All null space based methods try to extract discriminative directions present in the null space and ignore range space with the view that it holds the complete intra-class variations. The performance of null space based methods is highly sensitive to the size of null space which in term depends on the original image size and also number of training samples. For large databases, huge number of training samples creates a negative effect on the performance of null space based methods. So, in the cases where 1) image size is small and 2) total number of training samples is large, most of the discriminative information goes to range space of within-class scatter. Our observation says that there is a nice performance balance between range space and null space. While image size is constant, total number of training samples controls the performance of both spaces. When number of training samples is less, null space performs far better than range space whereas increment in number of training samples enhances the performance of range space noticeably. So one way of capturing and utilizing the discriminative information present in entire face space is to merge the discriminative information present in both spaces. We have attempted to do the merging with the help of feature fusion. The steps of our algorithm is given below:

1. *Compute S_w from training set $X = [x_1^1, x_2^1, ..., x_N^1, x_1^2, ..., x_N^C]$, using Eqn. 4.1.*
2. *Perform eigen-analysis on S_w and select first r eigenvectors to form the basis vector set Q for range space V.*
3. *Project all class means onto null space and range space using the basis vectors for range space only (see Eqn. 4.12-4.13). The sets of class means, projected on null space and range space, are denoted by X_{Null} and X_{Range} respectively.*
4. *Compute the scatter matrices S_{Null} and S_{Range} of X_{Null} and X_{Range}.*
5. *Perform eigen-analysis of S_{Null} and S_{Range} to obtain discriminative directions W_{opt}^{Null} and W_{opt}^{Range} in null space and range space separately.*
6. *Perform feature fusion by applying either*

(a) *Covariance Sum method on X_{Null} and X_{Range}, to obtain W^{Dual}.*

or

(b) *Gramm-Schmidt Orthonormalization method on [W_{opt}^{Null},W_{opt}^{Range}], to obtain orthonormal basis W^{Dual}.*

7. *Select optimal feature set by using either*

   (a) *Forward Selection procedure to select optimal features W_{opt}^{Dual} from W^{Dual} based on class separability criterion given in Eqn. 4.28.*

   or

   (b) *Backward Selection in the same manner to form W_{opt}^{Dual}.*

8. *Project all training samples on W_{opt}^{Dual}.*

9. *Recognition stage:*

   (a) *Project probe sample, presented for recognition, on W_{opt}^{Dual}.*

   (b) *Find distance of the projected probe to nearest training sample from each class.*

   (c) *Label the probe with the class corresponding to the minimum distance value.*

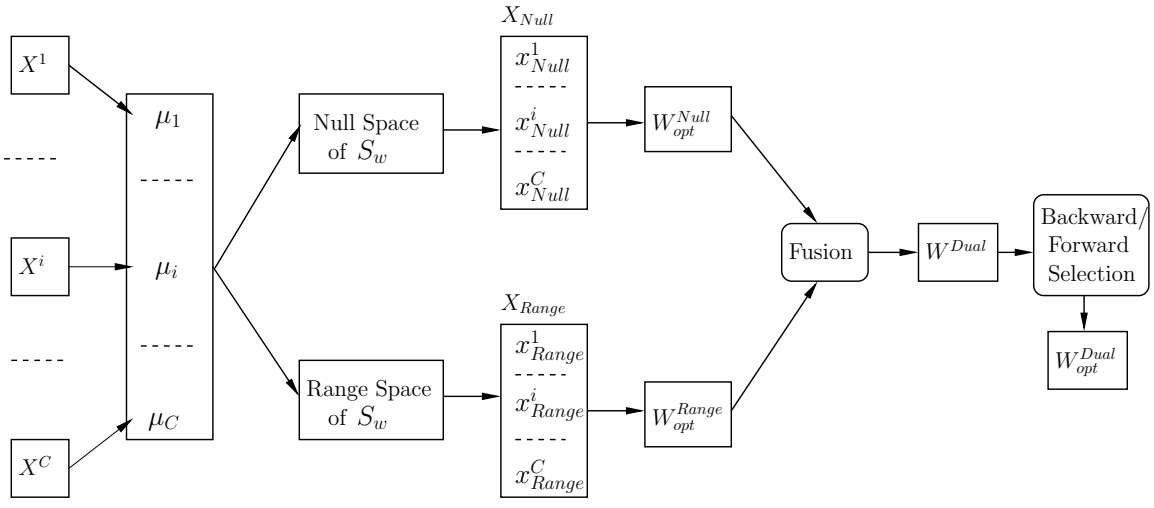


Figure 4.3: Architecture of our proposed method of feature fusion.

The architecture of our proposed method is given in Fig. 4.3 where X^i denote the training set for class i. X_{Null} and X_{Range} are the two sets of means projected on null space and range space separately. W_{Opt}^{Null} and W_{Opt}^{Range} are the sets of discriminative directions available in null space and range space and W_{Opt}^{Dual} represents the optimal combined discriminative directions obtained from both spaces. The decision fusion strategy is described in the following section.

4.4 Decision Fusion

The response vectors obtained from null and range classifiers are combined using three decision fusion techniques. Among them sum rule and product rule [59], [60] are adopted from already existing literature and third one is proposed by us. The response vectors on a sample x obtained from null and range classifiers can be expressed as a matrix of size $2 \times C$ and is denoted by $DP(x)$ [68],

$$DP(x) = \begin{bmatrix} d_{Null}^1(x) & \dots & d_{Null}^i(x) & \dots & d_{Null}^C(x) \\ d_{Range}^1(x) & \dots & d_{Range}^i(x) & \dots & d_{Range}^C(x) \end{bmatrix} = \begin{bmatrix} D_{Null}(x) \\ D_{Range}(x) \end{bmatrix} \tag{4.29}$$

The task of a combination rule, $\mathcal{F}$, is to construct $\tilde{D}(x) = [\tilde{d}^1(x), \dots, \tilde{d}^i(x), \dots \tilde{d}^C(x)]$, the fused soft class label provided by two classifiers as:

$$\tilde{D}(x) = \mathcal{F}(D_{Null}(x), D_{Range}(x)) \tag{4.30}$$

The fused soft class label can be hardened to a crisp class label c, by *maximum/ minimum membership rule*, based on the fact that measurements of a response vector represent similarity/dissimilarity:

$$\mathcal{L}(x) = \arg\max_c \tilde{d}^c(x). \tag{4.31}$$

where $\mathcal{L}(x)$ represents the class label assigned to x.

4.4.1 Existing Techniques for Decision Fusion

Given $DP(x)$, sum and product rules operate class-wise on each column of $DP(x)$ to produce $\tilde{D}(x)$ [60, 68].

- **Sum Rule:** Sum Rule computes the soft class label vectors using:

$$\tilde{d}^j(x) = d_{Null}^j(x) + d_{Range}^j(x), \ j = 1, ..., C \tag{4.32}$$

- **Product Rule:** Product Rule computes the soft class label vectors as:

$$\tilde{d^j}(x) = d^j_{Null}(x) * d^j_{Range}(x),\ j = 1, ..., C \tag{4.33}$$

4.4.2 Proposed Technique for Decision Fusion

Unlike sum rule and product rule, our technique of decision fusion learns each of the classifier's behavior by using training data at decision level. Learning a classifier involves training with the response of that classifier. So the response vectors on a validation set of a database which is disjoint from training and testing set can be used as training information at decision level for learning the behavior of a classifier. We employ LDA and nonparametric LDA on training information at decision level to build up discriminative eigenmodel. Then response vectors of the input test images are projected in that eigenmodel and for each test response vector, the similarity measures for the respective classes in the eigenmodel provide a new response vector which replaces the old test response vector.

To accomplish the whole task, we divide our database into three disjoint subsets, training, validation and testing sets. Now the training set along with validation set and testing set respectively constructs the "training response vector set" and "testing response vector set", respectively at classifier output space. Then "training response vector set" is used as training data to construct a LDA or nonparametric LDA-based eigenmodel at classifier response level.

Here, we formally define validation and testing set which are independent of any classifier. Let, the validation and testing set of a database are denoted by X_{VA} and X_{TS} and can be defined as,

$$X_{VA} = [vx^1_1, vx^1_2, \ldots, vx^1_V, vx^2_1, \ldots, vx^i_j, \ldots, vx^C_V], \tag{4.34}$$

$$X_{TS} = [tx^1_1, tx^1_2, \ldots, tx^1_S, tx^2_1, \ldots, tx^i_j, \ldots, tx^C_S]. \tag{4.35}$$

where vx^i_j represents the j^{th} validation sample from class i and V is the number of validation samples per class. Similarly, j^{th} testing sample from class i is denoted by tx^i_j and number of testing samples per class is represented as S. The "training response vector set" and "testing response vector set" for null space classifier are

denoted by RV_{Null}^{TR} and RV_{Null}^{TS} and are represented as,

$$\begin{aligned} RV_{Null}^{TR} &\equiv \{D_{Null}(vx_1^1), D_{Null}(vx_2^1), \ldots, D_{Null}(vx_V^1), D_{Null}(vx_1^2), \\ &\ldots \ , D_{Null}(vx_V^2), \ldots, D_{Null}(vx_j^i), \ldots, D_{Null}(vx_V^C)\} \qquad (4.36) \\ RV_{Null}^{TS} &\equiv \{D_{Null}(tx_1^1), D_{Null}(tx_2^1), \ldots, D_{Null}(tx_S^1), D_{Null}(tx_1^2), \\ &\ldots \ , D_{Null}(tx_S^2), \ldots, D_{Null}(tx_j^i), \ldots, D_{Null}(tx_S^C)\} \qquad (4.37) \end{aligned}$$

Considering "training response vector set", RV_{Null}^{TR}, as training as well as testing set, we calculate an accuracy denoted by $A_{Null}^{Original}$, with the condition that same samples are never compared. So, the calculation of $A_{Null}^{Original}$ involves the following steps:

1. For $i = 1$ to $V * C$

   (i) Select a response vector $D_{Null}(vx_j^c)$ where $i = (c-1) * V + j$ from the set RV_{Null}^{TR}.

   (ii) Calculate the Euclidean distance of $D_{Null}(vx_j^c)$ from all other response vectors in the set RV_{Null}^{TR}.

   (iii) Find the class label of the response vector providing minimum distance from $D_{Null}(vx_j^c)$ and assign it to $D_{Null}(vx_j^c)$.

   (iv) Cross check the obtained class label with the original label.

2. Calculate $A_{Null}^{Original}$ as $\frac{Number\ of\ correctly\ labeled\ response\ vectors}{Total\ number\ of\ response\ vectors} * 100\%$. *Total Number of response vectors* is $V * C$ in this case.

Then RV_{Null}^{TR} is used to construct an LDA or nonparametric LDA based eigenmodel. As the length of a response vector is equal to the number of classes (C) present in a database, we obtain C eigenvectors in the eigenmodel. Let, the C eigenvectors in the order of decreasing eigenvalues be represented by $EV_{Null} = [e_{Null}^1, ...e_{Null}^i, ..., e_{Null}^C]$. We obtain C eigensubspaces ESV_{Null}^i's for $i = 1, 2..., C$ where i^{th} eigensubspace ESV_{Null}^i contains first i eigenvectors from the set EV_{Null}. Thus C^{th} eigensubspace, ESV_{Null}^C, is same as EV_{Null}. The performance of i^{th} eigensubspace ESV_{Null}^i on "training response vector set", which is denoted by $A_{Null}^{ESV_{Null}^i}$, is evaluated by projecting RV_{Null}^{TR} onto ESV_{Null}^i and calculating accuracy in the same manner as described for the calculation of $A_{Null}^{Original}$. For each eigensubspace, we compare its performance with

$A_{Null}^{Original}$ and if $A_{Null}^{ESV_{Null}^{i}} > A_{Null}^{Original}$, i^{th} eigensubspace ESV_{Null}^{i} will be termed as *qualified eigensubspace*. Then, we project all "testing response vector set", RV_{Null}^{TS}, onto ESV_{Null}^{i} and calculate new response vectors by normalizing the measured distances from respective classes in the eigensubspace. Thus, each qualified eigensubspace gives a recalculated "testing response vector set". The "testing response vector set" recalculated in i^{th} eigensubspace will be denoted as $RV_{Null}^{TS}(ESV_{Null}^{i})$. If any of the eigensubspace is unable to provide better accuracy than $A_{Null}^{Original}$, then the "testing response vector set" will remain unchanged. As a result, we may obtain original "testing response vector set", RV_{Null}^{TS}, or a number of newly calculated "testing response vector set"s corresponding to each of the qualified eigensubspace.

The output space of the range classifier is also explored in the same way as described above. For range space, "training response vector set" and "testing response vector set" are denoted by RV_{Range}^{TR} and RV_{Range}^{TS} and are expressed as,

$$\begin{aligned} RV_{Range}^{TR} &\equiv \{D_{Range}(vx_1^1), D_{Range}(vx_2^1), \ldots, D_{Range}(vx_V^1), D_{Range}(vx_1^2), \\ &\ldots \ , D_{Range}(vx_V^2), \ldots, D_{Range}(vx_j^i), \ldots, D_{Range}(vx_V^C)\} \qquad (4.38) \\ RV_{Range}^{TS} &\equiv \{D_{Range}(tx_1^1), D_{Range}(tx_2^1), \ldots, D_{Range}(tx_S^1), D_{Range}(tx_1^2), \\ &\ldots \ , D_{Range}(tx_S^2), \ldots, D_{Range}(tx_j^i), \ldots, D_{Range}(tx_S^C)\} \qquad (4.39) \end{aligned}$$

Now one test response vector set from each of the classifier space are selected to form the DP's for test samples. After the formation of DP's, sum rule is deployed to calculate final soft class label for test samples. For multiple number of "testing response vector set"s from both classifiers, we evaluate performance for all possible pair of "testing response vector set"s and the final accuracy on test set is assigned as the maximum of them. In the worst case, if there is no qualified eigensubspace, the performance will be same as the performance given by sum rule.

Our proposed method of decision fusion can be stated as a preprocessing stage before final combination. The preprocessing stage uses training information at decision level to learn classifier's behavior and employ LDA and nonparametric LDA to enhance combination performance. Formally, given a test sample x, its DP (as

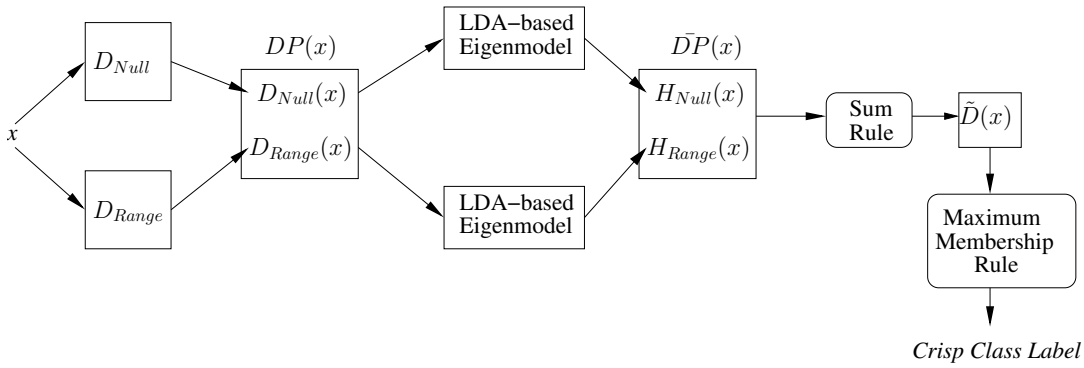


Figure 4.4: The fusion architecture of the proposed method of decision fusion.

defined in Eqn. 4.29) is replaced by $\bar{DP}(x)$, as:

$$\bar{DP}(x) = \begin{bmatrix} H_{Null}(x) \\ H_{Range}(x) \end{bmatrix} \tag{4.40}$$

where, H can be D or $\bar{D}$ based upon the success of the eigensubspaces constructed from LDA or nonparametric LDA-based eigenmodel. Figure 4.4 gives the diagram of the proposed decision fusion technique.

The general discussion on LDA and nonparametric LDA are given in the following:

4.4.2.1 Linear Discriminant Analysis

The objective in LDA [39] [35] (which is also known as Fisher's Linear Discriminant) is to find the optimal projection such that the between-class scatter is maximized and within-class scatter is minimized. It is thus required to maximize the Fisher's criterion (J) as shown below:

$$J = tr(S_w^{-1} S_b) \tag{4.41}$$

where $tr(.)$ is the trace of the matrix and S_b and S_w are the between-class and within-class scatter matrices, respectively. A within-class scatter matrix shows the scatter of samples around their respective class expected vectors, and is expressed by:

$$S_w = \sum_{i=1}^{C} P_i E\{(x - M_i)(x - M_i)^T | i\} = \sum_{i=1}^{C} P_i \Sigma_i. \tag{4.42}$$

M_i and Σ_i are mean and covariance matrix of i^{th} class, respectively and C is total number of classes. P_i is the number of training sample for i^{th} class and is equal to N

in our case. On the other hand, a between-class scatter matrix is the scatter of the expected vectors around the mixture mean as:

$$S_b = \sum_{i=1}^{C} P_i (M_i - M_0)(M_i - M_0)^T. \tag{4.43}$$

where M_0 represents the expected vector of the mixture distribution and is given by:

$$M_0 = E\{x\} = \sum_{i=1}^{C} P_i M_i. \tag{4.44}$$

A linear transformation from an n-dimensional x to an m-dimensional y ($m < n$) is expressed by:

$$y = A^T x. \tag{4.45}$$

where A is an $n \times m$ rectangular matrix and the column vectors are linearly independent. The problem of feature extraction for classification is to find the A which optimizes J. It can be easily proved that J is optimized if A constitutes first *C-1* eigenvectors of $S_w^{-1} S_b$ [39].

4.4.2.2 Nonparametric Linear Discriminant Analysis

The number of features, *C-1*, selected in LDA [39] is suboptimal in Bayes sense. Therefore, if the estimate of Bayes error in feature space (**y** space) is much larger than the one in the original space (**x** space), the feature extraction process should be augmented. If $tr(S_w^{-1} S_b)$ is used as a criterion, LDA selects the first (*C-1*)-dimensional subspace containing classification information using the scatter of mean vectors, while the second (*n-C+1*)-dimensional space containing information due to the covariance-differences is neglected, where n is the dimensionality of **x** space. Fig. 4.5 shows some of the cases where traditional LDA does not work. Therefore, we need to select additional features from the (*n-C+1*)-dimensional subspace. The basis of nonparametric LDA [39] is the nonparametric formulation of scatter matrices, using *k*-nearest neighbor (*kNN*) techniques, which measure between-class and within-class scatter on a local basis, both being generally of full rank. In addition, the nonparametric nature of the scatter matrix inherently leads to extracted features that preserve class structure important for accurate classification.

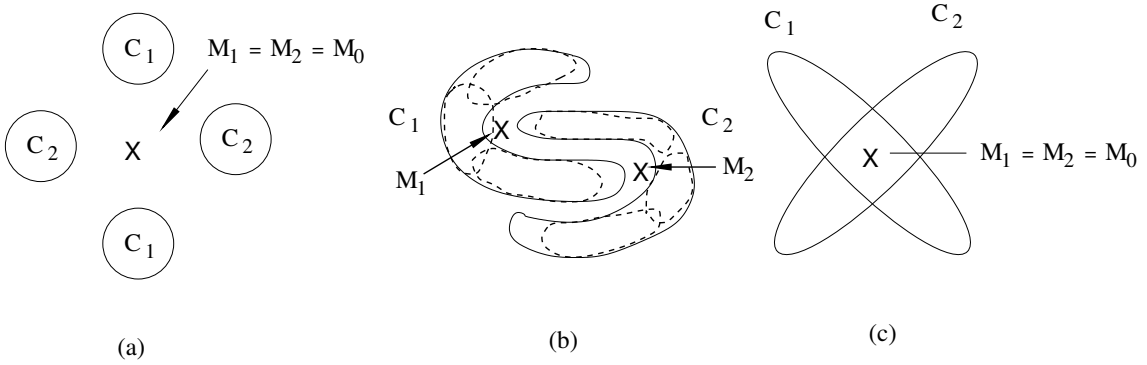


Figure 4.5: Examples where LDA based on the criteria $\mathbf{tr}(\mathbf{S_w^{-1}S_b})$ fails to discriminate between classes.

Let C_1 and C_2 be two classes, and $x^{(i)}$ refers to samples from C_i. The C_i-local mean of a sample $x^{(l)}$ from class C_l is denoted by $m_i(x^{(l)})$ and is computed as:

$$m_i(x^{(l)}) = \frac{1}{k}\sum_{j=1}^{k} x_{jNN}^{(i)}, \tag{4.46}$$

where $x_{jNN}^{(i)}$ is the j^{th} NN from C_i for the sample $x^{(l)}$, a sample of class C_l.

Nonparametric between-class scatter matrix for two classes, denoted by $\mathcal{S}_b$ can be defined as:

$$\begin{aligned} \mathcal{S}_b &= P_1 E\{(x^{(1)} - m_2(x^{(1)}))(x^{(1)} - m_2(x^{(1)}))^T | C_1\} \\ &+ P_2 E\{(x^{(2)} - m_1(x^{(2)}))(x^{(2)} - m_1(x^{(2)}))^T | C_2\} \end{aligned} \tag{4.47}$$

Similarly, nonparametric within-class scatter matrix for two-class is denoted by $\mathcal{S}_w$ and can be defined as:

$$\begin{aligned} \mathcal{S}_w &= P_1 E\{(x^{(1)} - m_1(x^{(1)}))(x^{(1)} - m_1(x^{(1)}))^T | C_1\} \\ &+ P_2 E\{(x^{(2)} - m_2(x^{(2)}))(x^{(2)} - m_2(x^{(2)}))^T | C_2\} \end{aligned} \tag{4.48}$$

We calculate $\mathcal{S}_b$ and $\mathcal{S}_w$ for C classes using the above defined scatters for two classes (equations 4.47 and 4.48). For a set of C classes we thus get a total of C^2 scatter

matrices, which can be organized in a matrix form as:

$$SM = \begin{bmatrix} \mathcal{S}_w^{1,1} & \mathcal{S}_b^{1,2} & \dots & \dots & \mathcal{S}_b^{1,C} \\ \mathcal{S}_b^{2,1} & \mathcal{S}_w^{2,2} & \dots & \dots & \mathcal{S}_b^{2,C} \\ \dots & \dots & \dots & \dots & \dots \\ \mathcal{S}_b^{i,1} & \dots & \mathcal{S}_w^{i,i} & \dots & \mathcal{S}_b^{i,C} \\ \dots & \dots & \dots & \dots & \dots \\ \mathcal{S}_b^{C,1} & \mathcal{S}_b^{C,2} & \dots & \dots & \mathcal{S}_w^{C,C} \end{bmatrix}$$

The diagonal elements of SM, $\mathcal{S}_w^{i,i}$ $i = 1, 2, ..., C$, represent the within-class scatter matrices for each of the C classes. Similarly, the non-diagonal elements of SM, $\mathcal{S}_b^{i,j}$ *where* $i, j = 1, 2, ..., C$ *and* $i \neq j$, are the between-class scatter matrices for two different classes i and j. $\mathcal{S}_b$ and $\mathcal{S}_w$ for C classes, denoted by $\mathcal{S}_{bC}$ and $\mathcal{S}_{wC}$, can be obtained by summing up all non-diagonal and diagonal entries of SM respectively. So

$$\mathcal{S}_{wC} = \sum_{i=1}^{C} \mathcal{S}_w^{i,i}, \tag{4.49}$$

$$\mathcal{S}_{bC} = \sum_{i=1}^{C} \sum_{\substack{i \neq j \\ j=1}}^{C} \mathcal{S}_b^{i,j} \tag{4.50}$$

Once $\mathcal{S}_{bC}$ and $\mathcal{S}_{wC}$ are obtained, optimum linear features can be selected by maximizing $tr(\mathcal{S}_{wC}^{-1}\mathcal{S}_{bC})$.

4.4.3 Algorithm for Decision Fusion

In this section, we describe the overall algorithm for our proposed method of decision fusion of two classifiers constructed on range space and null space of within-class scatter. We attempt to combine the complementary information present in null space and range space by decision fusion. The steps of our algorithm is given below:

1. **Construction of two classifiers (D_{Null} and D_{Range}) based on null space and range space.**

   i) Compute S_w from training set $X = [x_1^1, x_2^1, ..., x_N^1, x_1^2, ..., x_N^C]$ using Eqn. 4.1.

   ii) Perform eigen-analysis of S_w and select first r eigenvectors to form the basis vector set Q for range space V.

iii) Project all class means onto null space and range space using the basis vectors for range space only (see Eqn. 4.12-4.13). The sets of class means projected on null space and range space are denoted by X_{Null} and X_{Range}, respectively.

iv) Compute the scatter matrices, S_{Null} and S_{Range} of X_{Null} and X_{Range}, respectively.

v) Perform eigen-analysis of S_{Null} and S_{Range} to obtain discriminatory directions W_{opt}^{Null} and W_{opt}^{Null} in null space and range space separately.

Let, D_{Null} and D_{Range} be two classifiers defined on the null space and range space, respectively.

2. **Decision Fusion of D_{Null} and D_{Range}.**

i) Construct $DP(x)$ for a test sample, x.

ii) Obtain $\tilde{D}$ (using any one among (a) or (b) or (c))

(a) Apply sum rule to obtain $\tilde{D}$.

(b) Apply product rule to obtain $\tilde{D}$.

(c) Apply proposed method of decision fusion (see Fig. 4.4 given in section 4.4.2) to obtain $\tilde{D}$.

iii) Use maximum membership rule to obtain crisp class labels for a test sample x.

Experimental results for our proposed method is given in the following section.

4.5 Experimental Results and Discussion

We observe the performance of our proposed method on three public face databases: Yale, ORL and PIE. Yale has 15 subjects with 11 samples per class. ORL and PIE consist of 40 and 60 subjects with 10 and 42 samples per class respectively. More details about these databases are provided in Section 3.4.1.

4.5.1 Effect of Number of Training Samples on the Performance of Null Space and Range Space

To demonstrate the effect of increasing number of training sample on the performance of null space and range space, we compute the recognition accuracies of null space and range space for Yale, ORL and PIE databases with different number of training samples. For Yale and ORL databases, the accuracies are calculated for the number of training samples per class ranging from two to nine. As the number of training samples per class for PIE database is higher (42), the number of training samples is varied in the range 2 to 18 (e.g. 2, 4, 6,..., 18). The image size has been kept as 25*25 for all databases. The null space technique used in our experimentation is adopted from the method called Discriminative Common Vectors (DCV) [20].

The performance of null space and range space for Yale database is shown in Table 4.1. The performance of null space is best when the number of training samples per class is only two and three. Performance of null space decreases with increasing number of samples and provides the minimum accuracy of 50% when number of training samples is maximum (nine). This result validates the claim of negative effect of increasing number of training samples on the performance of null space. Initial increase in the number of training samples enhances performance of null space, because the captured common features for each class hold more robust commonalities present across samples containing variations in expression, illumination and pose. In case of range space, the performance increases with increasing number of training samples upto a certain point and again deteriorates. This performance trend can be easily explained by stating that initial increase in number of samples drives the discriminative information to the range space. However too many number of training samples leads to a huge amount of intra-class variations which in turn effects the overall accuracy.

The performances of null space and range space on ORL database with training samples varying from two to nine are shown in Table 4.2. As the ORL database has only pose variation, the increasing number of training samples helps the null space to capture the common features which are robust and efficient for classification. So, the performance of null space increases with increasing number of training samples and

Table 4.1: Effect of increasing number of training samples on the performance of null space and range space for Yale database.

No. of Training Samples	**Null Space**	**Range Space**
2	**93.33**	64.44
3	**93.33**	76.67
4	90.48	**82.86**
5	88.89	81.11
6	85.33	76.67
7	81.67	76.67
8	75.56	75.56
9	50.00	56.67

attains a maximum value 95.00% for eight training samples but again decreases down to 92.00% for nine training samples due to the small size of null space. The performance of range space exhibits an interesting behavior. The accuracy of range space increases with increasing number of training samples for the following two reasons: (i) the subspace learns more and more about the pose variations across the database and (ii) discriminative informations go to range space due to the increase in the number of training samples. Thus maximum accuracy for range space is obtained as 97.50% for eight training samples. But the further inclusion of more training samples reduces the performance due to the following reasons: (i) the new training samples does not provide any extra discriminatory information with respect to the information already learned by the classifier from previous training samples, (ii) moreover, they add confusing information to the discriminatory features.

For PIE database the performance of null space and range space is evaluated and shown (see Table 4.3) for even number of training samples ranging from two (2) to eighteen (18). The performances of null space and range space with increasing number of training samples can be explained by a similar logic as described in case of Yale and ORL databases. PIE has only illumination variation.

Table 4.2: Effect of increasing number of training samples on the performance of null space and range space for ORL database.

No. of Training Samples	**Null Space**	**Range Space**
2	82.19	55.00
3	85.00	72.86
4	88.75	82.50
5	88.50	85.00
6	91.25	90.63
7	93.33	93.33
8	**95.00**	**97.50**
9	92.00	90.00

Now to summarize the above discussion on the effect of increasing number of training samples on the performance of null space and range space, we state the following points:

- For Yale database, the optimal performances for null space and range space occur for relatively lesser number of training samples (2/3 samples for null space and 4 samples for range space).
- For ORL and PIE database, the maximum accuracies for null space and range space are obtained for relatively larger number of training samples (8 and 10 samples for ORL and PIE, respectively).

However the reasons behind the above observations are explained in the following:

- Yale database consists of diverse variations in expressions as well as illumination, and even has face with spectacles. Even though Yale database is the smallest among the three databases used for experimentation, it provides a rich combination of the challenging factors (expression, illumination and camouflage) for automatic face recognition, known in practice. Due to the high diversity among the training samples, the captured common features are not

Table 4.3: Effect of increasing number of training samples on the performance of null space and range space for PIE database.

No. of Training Samples	**Null Space**	**Range Space**
2	80.21	40.17
4	82.58	72.36
6	88.33	75.28
8	90.44	89.75
10	**93.49**	**97.55**
12	55.33	94.56
14	53.87	94.44
16	50.58	95.90
18	39.86	96.04

robust and efficient for classification for higher number of training samples. The common features extracted from a number of faces having expression variation does not help in classifying the samples from the same class showing illumination variation. So, the performance of null space on Yale database is high with very less number of training samples. On the other hand, the variability among samples contributes in a different way in case of range space. Higher number of training samples contributes to a large amount of intra-class variations in range space. Thus, the information learned by range space from larger number of training samples does not help in discrimination and adds to the degree of confusion.

- On the contrary, ORL and PIE databases contain only a specific type of variability (pose for ORL and illumination for PIE). So, higher number of training samples provides robust common features for null space and offers discriminative information (instead of adding confusing information) to range space. Hence, the performance increases with higher number of training samples in the case of these two databases, unlike that in Yale database.

From Tables 4.1-4.3, we conclude the following: *A large quantity of training samples drive discriminative information to the range space of within-class scatter.* For a database with a large number of subjects, the number of training samples will not allow null space to perform well. The main aim of our work is to capture the discriminative informations available in range space and combine them efficiently with the discriminative informations obtained from null space. Thus, we exploit the whole gamut of discriminative informations present in the entire face space and utilize them for enhancing classification performance.

4.5.2 Performance of Dual Space Face Recognition Approach

We split a image database into three disjoint sets called as training, validation and testing set. Feature fusion technique requires validation set for selecting optimal feature, W_{opt}^{Dual} and the performance of selected features is observed on the testing set. For decision fusion, training set along with validation set is required for generating "training response vector set" that is used for learning each classifier and to construct LDA or nonparametric LDA based eigenmodel at classifier output space. The sample distribution for a single subject (which is same for feature fusion and decision fusion) over the three sets for all three databases are given in Table 4.4.

Table 4.4: Sample distribution (per subject) in training, validation and testing sets for Yale, ORL and PIE databases.

Set	**Yale**	**ORL**	**PIE**
Training	4	3	4
Validation	3	4	12
Testing	4	3	26

The unique decomposition of a face into null space and range space is displayed in Fig 4.6, with three faces drawn from Yale, ORL and PIE databases respectively. Fig. 4.7 shows the images used for training for a single subject on Yale, ORL and PIE databases. Table 4.5 shows the performance of null and range spaces for Yale, ORL and PIE databases. The performances of our dual space based methods using

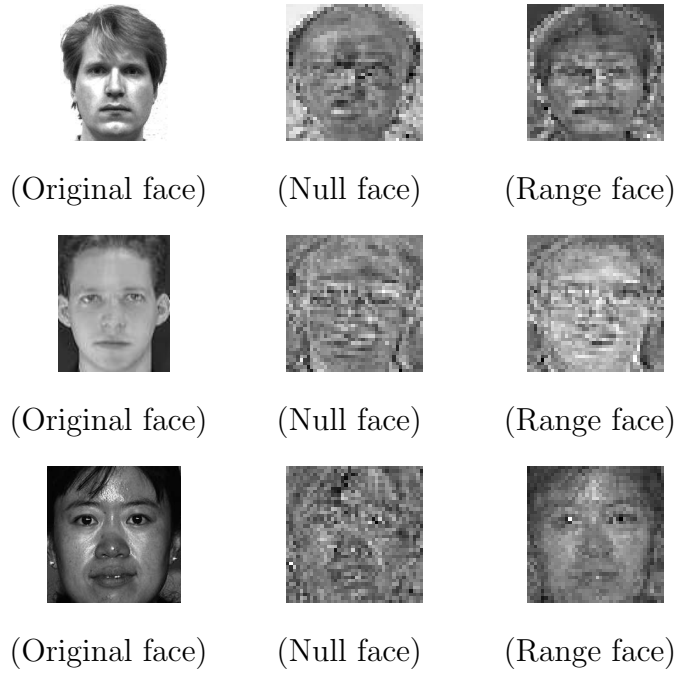


Figure 4.6: Unique decomposition of faces into null space and range space of within-class scatter drawn from Yale (first row), ORL (second row) and PIE (third row) databases.

feature (first four columns) and decision fusion (last four columns) are tabulated in Table 4.6. BS and FS are the abbreviations for Backward Selection and Forward Selection respectively.

For Yale database, dual space based feature fusion and decision fusion methods provide 85.00% and 86.67% (see Table 4.6), respectively. We obtain 85.99% accuracy using feature fusion (Gramm-Schmidt Orthonormalization and forward selection) and 97.50% using decision fusion (P_{NLDA}) for ORL database. Feature fusion using covariance sum and decision fusion using both P_{LDA} and P_{NLDA} provide 88.40% and 100% respectively on PIE database. Decision fusion using sum rule and product rule fails to work better than the null space itself. As we can notice that backward and forward search works equally well for all databases. Results on three databases show that there exist a good scope of combining information from null and range space.

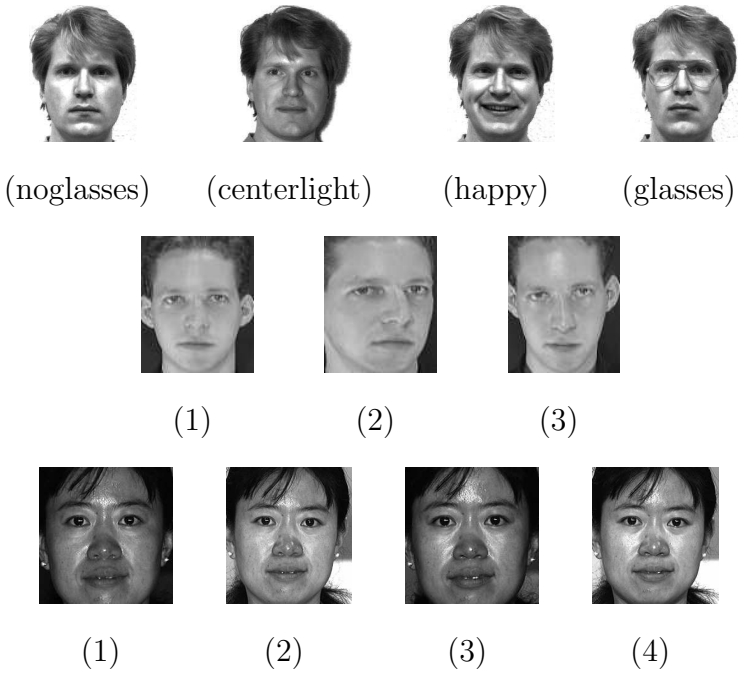


Figure 4.7: Training samples for three databases: First, second and third row shows the training set for Yale, ORL, PIE databases, respectively.

Table 4.5: Performance of null space and range space on Yale, ORL and PIE Databases.

Methods	Null Space	Range Space
Yale	81.67	75.00
ORL	79.17	71.67
PIE	86.35	73.14

Table 4.6: Performance of Dual Space Face Recognition on Yale, ORL and PIE databases. P_{LDA} and P_{NLDA} are our proposed decision fusion techniques, using LDA and nonparametric LDA respectively. BS: Backward Selection; FS: Forward Selection.

	Feature Fusion				**Decision Fusion**			
	Covariance Sum		**Gramm-Schmidt**		**Existing Techniques**		**Proposed Methods**	
Databases	BS	FS	BS	FS	Sum	Product	P_{LDA}	P_{NLDA}
Yale	**85.00**	**85.00**	**85.00**	**85.00**	**86.67**	**86.67**	**86.67**	**86.67**
ORL	85.83	85.93	85.83	**85.99**	81.67	81.67	96.67	**97.50**
PIE	**88.40**	**88.40**	87.28	87.35	83.72	83.85	**100.00**	**100.00**

4.6 Conclusion

In this chapter, a dual space based face recognition approach using feature and decision fusion from null space and range space of within-class scatter is proposed. The discriminative directions spread over the total face space. Null space and range space of with-class scatter constitutes the complete face space. So finding discriminative direction only in null space ignores the utilization of discriminative directions present in range space. We tried to merge the discriminative directions available in both spaces using covariance sum and Gramm-Schmidt Orthonormalization method. We also combine the complementary discriminatory information present in null space and range space by decision fusion. Along with sum and product rule for decision fusion we propose a new technique which learns both classifier's (built on null space and range space) behavior to enhance class separability (using LDA and nonparametric LDA) at decision level using classifier's response as training information. Experimental results on three standard databases show that, feature and decision fusion indeed combines information from both spaces to improve classification performance.

CHAPTER 5

Enhancing Decision Combination of Face and Fingerprint by Exploitation of Individual Classifier Space: An approach to Multimodal Biometry

This chapter presents a new approach to combine decisions from face and fingerprint classifiers by exploiting the individual classifier space on the basis of availability of class specific information present in the classifier's output space. We use the response vectors on a validation set for enhancing class separability (using linear discriminant analysis and nonparametric linear discriminant analysis) in the classifier output space and thereby improving performance of the face classifier. Fingerprint classifier often does not provide this information due to high sensitivity to minutiae points, producing partial matches across subjects. The enhanced face and fingerprint classifiers are combined using sum rule. We also propose a generalized algorithm for Multiple Classifier Combination (MCC) based on our approach. Experimental results show superiority of the proposed method over other already existing fusion techniques like sum, product, max, min rules, decision template and Dempster-Shafer theory. The rest of the chapter is organized as follows: Section 5.1 provides brief description on a few relevant works attempted for decision combination and a succinct introduction to our proposed method of decision fusion. Section 5.2 gives the theoretical formulation of our proposed classifier combination technique; Section 5.3 describes our method of exploitation of base classifier using prior information and then combine them; Section 5.4 presents the experimental results to support our methodology and Section 5.5 concludes this chapter. The description of other decision combination strategies used in our experimentation for comparison are provided in Section 2.3.3.3.

5.1 Introduction

In recent years, the concept of combining multiple experts [60],[68] in a unified framework to generate a robust decision based on the individual decisions delivered by multiple co-operating experts has been examined in the context of biometry [115],[22]. Unimodal biometrics have certain limitations like nonuniversality, spoof attacks, sensitivity to noisy data. This causes unimodal (e.g. face-based, fingerprint-based systems) recognition techniques to reach a near saturation stage in terms of performance in future. A potential way of overcoming such limitations consists in combining multiple modalities. This solution usually called as multimodal biometrics has already shown good promise. We have chosen face and fingerprint for their strong universality and uniqueness property respectively.

Generally, decision fusion in multimodal biometrics can be treated as a Multiple Classifier Combination (MCC) problem. The success and failure of a decision combination strategy largely depends on the extent to which the various possible sources of information are exploited from each classifier space in designing the decision combination framework. Kittler et al. [60] proposed a set of decision fusion strategies namely, sum, product, min, max, median and majority vote rule. The paper states that the ensemble of classifiers providing decisions in the form of crisp class labels can use majority voting, whereas if classifier outputs are in the form of posterior probabilities they can be combined using sum, product, max, min rules. Snelick et al. [115] has used classifier specific weight (MW: Matcher Weight) and class specific weight (UW: User Weight) with sum rule to combine face and fingerprint at a large scale. Besides these techniques, researchers have suggested decision-level data fusion. Several modalities like still image and speech are combined by using fuzzy k-means (FKM), fuzzy vector quantization (FVQ) algorithms and median radial basis function (MRBF) network [22]. Kuncheva et al. [68] has proposed decision templates (DT) for multiple classifier fusion. Per class DT's are estimated on a validation set and are then matched to the decision profiles (DP) of new incoming test samples by some similarity measure. Dempster-Shafer (DS) combination [103] finds out belief degrees for every class as well as for every classifier (for a test sample), and multiply (as per product rule) them to give a combined soft class label vector.

We try to enhance the performance of decision fusion by exploitation of individual classifier space on the basis of availability of contextual information. Contextual information refers to the fact that, when a sample of a particular subject is presented the similarity it shows for different classes is subject specific. This means that if the outputs of a classifier output for the samples of a specific class are well clustered and distinct from other clusters, there exists a scope of using this information to enhance the class separability at classifier output space in order to improve the individual classifier's performance. This will in turn enhance the performance of fusion. Face and fingerprint are the two modalities to be combined in our case. Prior knowledge (training data at decision level) about each classifier space helps us to know that subject-wise contextual information is present in case of face but not for fingerprint. Fingerprint classifier is unable to show subject-wise contextual information because of it's sensitivity to the number, occurrence and distribution of matched minutiae between two fingerprints.

Hence in our approach, we try to improve the face space as much as possible and then combine with fingerprint. For enhancing the performance of face classifier, we use the classifier output (known as response vectors) on a validation set as the training data at fusion level for building up an LDA or nonparametric LDA-based eigenmodel and thereby enhancing class-separability.

5.2 Theoretical Basis

Let $x \in \mathcal{R}^n$ be a feature vector and $\{1, 2, ..., C\}$ be the set of C classes. A classifier is called a mapping:

$$D : \mathcal{R}^n \rightarrow [0, 1]^C$$

The output of a classifier D will be called as response vector and denoted by $R_D(x)$, a C dimensional vector, where:

$$R_D(x) = (r_D^1(x), r_D^2(x), ..., r_D^C(x)) \ , \quad r_D^i(x) \in [0, 1] \tag{5.1}$$

The components $\{r_D^i(x)\}$ can be regarded as estimates of the posterior probabilities of (dis)similarity or (dis)belief provided by classifier D, for all classes, for a sample x.

This means for a classifier D:

$$r_D^i(x) = P(i|x), \quad i = 1, 2, ..., C. \tag{5.2}$$

Response vector is also known as the soft class label provided by a classifier. The decision of a classifier can be hardened to a crisp class label c, where $c \in \{1, 2, ..., C\}$, by *maximum (minimum) membership rule*, based on the fact that elements of response vector represent similarity (dissimilarity):

$$D(x) = \arg\max_c r_D^c(x). \tag{5.3}$$

In a traditional multiple classifier system, a feature vector x is classified into one of the C classes using L classifiers $\{D_1, D_2, ..., D_L\}$, using the feature vectors x_l, $l = 1, 2, ..., L$ respectively. Measurement level (also called response vector level) combination strategies give final decision by fusing the response vectors from multiple classifiers. Formally, for a feature vector x, response vectors from multiple classifiers can be organized as a matrix called decision profile (DP):

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \dots & d_{1,j}(x) & \dots & d_{1,C}(x) \\ \dots & \dots & \dots & \dots & \dots \\ d_{i,1}(x) & \dots & d_{i,j}(x) & \dots & d_{i,C}(x) \\ \dots & \dots & \dots & \dots & \dots \\ d_{L,1}(x) & \dots & d_{L,j}(x) & \dots & d_{L,C}(x) \end{bmatrix} = \begin{bmatrix} D_1(x) \\ \dots \\ D_i(x) \\ \dots \\ D_L(x) \end{bmatrix}$$

We denote i^{th} row of the above matrix as $D_i(x) = [d_{i,1}(x), ..., d_{i,C}(x)]$, where $d_{i,j}(x)$ is the degree of support given by classifier D_i to the hypothesis that x belongs to class j. $D_i(x)$ is the response vector of classifier D_i for the sample x. The task of any combination rule is to construct $\tilde{D}(x)$, the fused output of L classifiers as:

$$\tilde{D}(x) = \mathcal{F}(D_1(x), \dots, D_i(x), \dots, D_L(x)). \tag{5.4}$$

Some fusion techniques known as class-conscious [68], do column-wise class-by-class operation on *DP(x)* matrix to obtain $\tilde{D}(x)$. Example of this type of fusion techniques are: sum, product, min, max, etc [60]. Another fusion approach known as class-indifferent [68], use entire *DP(x)* to calculate $\tilde{D}(x)$. The later needs training at fusion level in the sense that it has to create one or more than one templates per class from

the response vectors on a validation set to compare with *DP(x)* (decision profile for a test sample x). It can be stated that, these methods use prior knowledge about the classifier's behavior to improve combined performance. These methods will perform well when classifier output preserves class specific contextual information. In case if any classifier is highly sensitive to the features used and exhibits large intra-class variance in the classifier output space, these methods do not perform well, and even deteriorate beyond the performance of individual base classifiers.

In that case, the only way to overcome this problem is to conditionally improve each row $D_i(x)$ (response vector of classifier D_i) using response vectors of corresponding classifier on a validation set as training data at fusion level and then use class-conscious techniques to enhance the performance of classifier combination. Conditional row-wise operation here implies the use of training data (generated on a validation set) as prior knowledge about classifier's (D_i) behavior to replace it's response vector $D_i(x)$ with $\bar{D}_i(x)$ in *DP(x)* and thereby improving combined performance. If any classifier's output does not provide consistency in class-specific information, training data does not help to improve the final decision. In such a case, it is wiser to keep $D_i(x)$ unchanged for that classifier. This observation is the essence of our approach to influence the final classification performance.

In our approach, we have accomplished the task of using training data (prior knowledge about a classifier) by using the response vectors on a validation set as training data at the fusion level for building a LDA or nonparametric LDA-based eigenmodel. Now for a particular classifier, the ability of the eigenmodel in enhancing class-separability in response vector eigenspace dictates the usefulness of training data at fusion level. The improvement of class-separability by eigenmodel strengthens the performance of it's corresponding base classifier. This is evident from the fact that LDA performs well when the training data uniformly samples the underlying class distribution, or in other words training data represents class-specific information [86]. Empirically, we can validate it by observing that the use of eigenmodel on classifier's output space provides better performance than the direct classifier output on a validation set of response vectors.

To accomplish this task, we divide a database in four disjoint sets, namely train,

validation$_1$, *validation*$_2$ and test. Train set along with *validation*$_1$, *validation*$_2$ and test set will create train, validation and test response vector sets respectively in classifier output space. "Train response vector set" is used to build up an LDA or nonparametric LDA-based eigenmodel. "Validation response vector set" is used to check the usefulness of eigenmodel for improving class-separability and thereby checking the availability of class-specific contextual information present in the output of a classifier. If the LDA-based eigenmodel is observed to improve the performance of a classifier for the "validation response vector set", we recalculate class score values (by replacing $D_i(x)$ with $\bar{D}_i(x)$ in *DP(x))* for "test response vector set" for that classifier. Otherwise, all $D_i(x)$'s from "test response vector set" remain unchanged for that classifier.

This means that our method operates on *DP(x)* row-wise (classifier-wise) to produce a improved version of *DP(x)*, denoted by $\bar{DP}(x)$. We can easily visualize our approach as a preprocessing step before final fusion, which takes each row of *DP(x)* (i.e. $D_i(x)$) as input and replaces with $H_i(x)$, where $H_i(x)$ can be $\bar{D}_i(x)$ or $D_i(x)$ depending on the presence of class-specific information in the output of the corresponding classifier. In the next section, we will describe how we exploit availability of class-specific information present in classifier output space to strengthen base classifier and thereby improve the combined performance.

5.3 Improving Combined Performance by Strengthening Base Classifiers: Proposed Method

Based on the concepts described at the end of the previous section, the strength of a base classifier is enhanced as follows:

Use the train response vectors as intermediate feature vectors to build a eigenmodel based on LDA and nonparametric LDA at classifier output space. Then calculate improved response vectors $\bar{D}$ for "test response vector set" in that eigenspace. Theoretically, improvement of $\bar{D}$ over D is evident from the fact that LDA builds an eigenmodel where class response vectors will be well-clustered as well as well-separated. Empirically, it is tested by performing a performance evaluation of this eigenmodel on a "validation response vector set".

We exploit classifier space in two ways:

1. LDA and
2. Nonparametric LDA.

LDA and nonparametric LDA are discussed in Section 4.4.2.1 and Section 4.4.2.2, respectively.

We combine face and fingerprint (two modalities for person authentication) using our proposed algorithm, described in this section. As already discussed in Section 5.2, our aim is to improve combined performance by conditionally replacing response vector $D_i(x)$ by our proposed measure, $\bar{D}_i(x)$, in *DP(x)* to obtain $\bar{DP}(x)$, classifier-wise by using train and validation response vector sets (obtained on two disjoint validation sets, $validation_1$ and $validation_2$).

Now for each of the L classifier, we construct train, validation and test response vector sets, denoted by RV_i^{TR}, RV_i^{VA} and RV_i^{TS}, where $i = 1, 2, ..., L$. Precisely, RV_i^{TR}, RV_i^{VA} and RV_i^{TS} form the sets of the i^{th} rows of the DP's constructed from $validation_1$ $validation_2$ and test set feature vectors respectively. Now let $validation_1$ $validation_2$ and test set feature vectors for i^{th} classifier are denoted by:
(i) $validation_1$ feature vector set: $\{ux^{11}, ux^{21}, \ldots, ux^{M_i^{TR}1}, ux^{12}, \ldots, ux^{mc}, \ldots, ux^{M_i^{TR}C}\}$,
(ii) $validation_2$ feature vector set: $\{vx^{11}, vx^{21}, \ldots, vx^{M_i^{VA}1}, vx^{12}, \ldots, vx^{mc}, \ldots, vx^{M_i^{VA}C}\}$,
and (iii) test feature vector set: $\{tx^{11}, tx^{21}, \ldots, tx^{M_i^{TS}1}, tx^{12}, \ldots, tx^{mc}, \ldots, tx^{M_i^{TS}C}\}$,
where for i^{th} classifier the number of response vectors per class for the three disjoint sets are $M_i^{TR}, M_i^{VA}, M_i^{TS}$ respectively. The superscript "mc" on a feature vector (e.g. ux^{mc}) denotes the m^{th} sample from c^{th} class of the corresponding set. So the response vector sets can be visualized as:

$$
\begin{aligned}
RV_i^{TR} &\equiv \{D_i(ux^{11}), \ldots, D_i(ux^{M_i^{TR}1}), D_i(ux^{12}), \ldots, D_i(ux^{mc}), \ldots, D_i(ux^{M_i^{TR}C})\},\\
RV_i^{VA} &\equiv \{D_i(vx^{11}), \ldots, D_i(vx^{M_i^{VA}1}), D_i(vx^{12}), \ldots, D_i(vx^{mc}), \ldots, D_i(vx^{M_i^{VA}C})\},\\
RV_i^{TS} &\equiv \{D_i(tx^{11}), \ldots, D_i(tx^{M_i^{TS}1}), D_i(tx^{12}), \ldots, D_i(tx^{mc}), \ldots, D_i(tx^{M_i^{TS}C})\}.
\end{aligned}
$$

The steps of our algorithm is given below for L classifiers, even though we are dealing with the case where $L = 2$.

5.3.1 The Algorithm

The steps of our algorithm are as follows:

1. *For i= 1,2,..., L do*

   1.1 Evaluate performance of classifier D_i on validation response vector set RV_i^{VA} and denote it by $P_{D_i}^{DIR}(RV_i^{VA})$.

   1.2 Use train response vector set RV_i^{TR} of classifier D_i to construct eigenmodel based on LDA or nonparametric LDA.

   1.3 Evaluate the performance of classifier D_i on RV_i^{VA} in the eigenmodel constructed in previous step. For $m = 1, 2, ..., M_i^{VA}$ and $c = 1, 2, ..., C$, select each $D_i(x^{mc})$ from set RV_i^{VA} and do the following:

   (a) Project $D_i(x^{mc})$ on LDA or nonparametric LDA-based eigenmodel.

   (b) Calculate the class scores by measuring the Euclidean distance from $D_i(x^{mc})$ to the nearest sample from every class in the eigenmodel to generate $\bar{D}_i(x^{mc})$.

   (c) Use minimum membership rule on $\bar{D}_i(x^{mc})$ to produce crisp class label for x^{mc}.

   (d) Cross check the generated class label with actual class label of x^{mc}.

   Calculate the performance accuracy on RV_i^{VA} and denote it by $P_{D_i}^{EM}(RV_i^{VA})$. If $P_{D_i}^{EM}(RV_i^{VA}) > P_{D_i}^{DIR}(RV_i^{VA})$ then continue, else go back to step 1.1 to continue for next classifier.

   1.4 For an unknown test sample x, select $D_i(x)$ from set RV_i^{TS} and calculate $\bar{D}_i(x)$ using the following steps.

   (a) Project $D_i(x)$ on LDA or nonparametric LDA-based eigenmodel generated in step 1.2.

   (b) Calculate the new class scores for x as $\bar{D}_i(x)$ by measuring the Euclidean distance from $D_i(x)$ to the nearest response vector from every class in the eigenspace.

   The elements of RV_i^{TS} are now $\bar{D}_i(x)$ instead of $D_i(x)$.

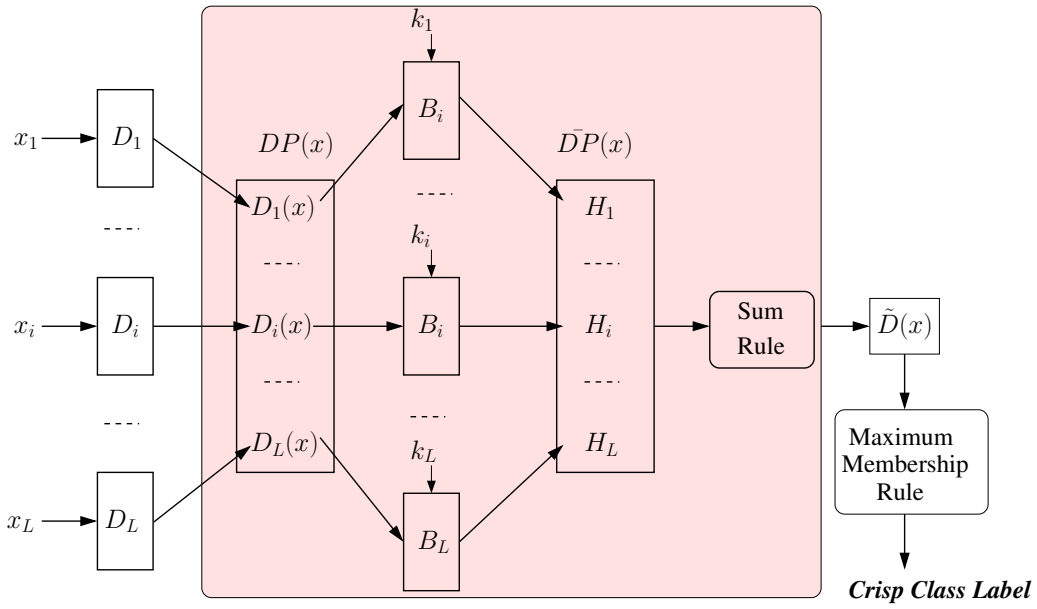


Figure 5.1: Operation of proposed method for classifier fusion.

2. *Use all of RV_i^{TS}, where $i = 1, 2, .., L$, to construct set of DP's for test samples. Now, for a test sample x, $\bar{DP}(x) = [H_1(x), H_2(x), ..., H_L(x)]^T$, where H can be D or $\bar{D}$ based on the performance of the classifier observed in 1.3.*

3. *Use sum rule on $\bar{DP}(x)$ to get soft class label vector $\tilde{D}(x)$.*

4. *Maximum or minimum (based on the implementation of min-max normalization technique) membership rule on $\tilde{D}(x)$ will provide crisp class label for x.*

The architecture illustrating our proposed methodology of decision fusion is given in Fig 5.1. The blocks labeled B_i's *i=1,2,..,L*, use a binary variable k_i, $i = 1, 2, ..., L$ for decision-making purpose. When $k_i = 1$ (the LDA-based eigenmodel is beneficial), the output of classifier $D_i(x)$ on test sample x is replaced by $\bar{D}(x)$ in *DP(x)*, while $k_i = 0$ ensures that output $D_i(x)$ is passed on as the output unaltered. In the diagram $\bar{DP}(x)$ represents modified $DP(x)$ after replacing it's rows $D_i(x)$ by $\bar{D}_i(x)$ conditionally (see step 2 of the algorithm).

A comparative study of the performance of the proposed technique of decision fusion using class-specific information for a classifier, with that of those in existing literature, is given in the following section. The decision fusion techniques used for comparison in our experimentation are described in section 2.3.3.3.

5.4 Experimental Results

In this chapter, we have combined face and fingerprint features at decision level. We have used PIE [5] and ORL [3] databases for face. We have labeled FP_A and FP_B as fingerprint databases acquired from the FVC (Fingerprint Verification Competition) 2002 [1] and 2004 website [2]. Both the fingerprint databases have 10 subjects with 8 images per subject. So we have selected 10 persons from both PIE and ORL databases with 42 and 10 images per subject respectively. Assuming that face and fingerprint are statistically independent for an individual, we have associated an individual from face database with an individual from fingerprint database to create a virtual subject. Now each database is divided into three disjoint sets namely train set, validation set and test set. Table 5.1 and Table 5.2 show the distribution of images per subject for the train, validation and test sets for face and fingerprint respectively, as used in our experimentation.

Table 5.1: Sample distribution (per subject) in train, validation and test sets for face databases.

	PIE		ORL	
Set	$Test_1$	$Test_2$	$Test_1$	$Test_2$
Train	4	4	1	1
Validation	4	4	3	3
Test	34	34	6	6

Table 5.2: Sample distribution (per subject) in train, validation and test sets for fingerprint databases.

	FP_A *or* FP_B			
Set	$Test_1$	$Test_2$	$Test_3$	$Test_4$
Train	1	1	1	1
Validation	2	3	3	4
Test	5	4	4	3

We have associated each face database with a fingerprint database and thus obtained 4 possible combinations: (i) PIE & FP_A, (ii) PIE & FP_B, (iii) ORL & FP_A and (iv) ORL & FP_B. For each combination of a face and fingerprint database, total number of virtual test samples for a subject are obtained by all possible combination of test samples provided for the corresponding subject from face and fingerprint databases. For example, for the combination of PIE and FP_A databases the number of virtual test samples per person is 170 (=34*5). We have created two and four different test sets (also corresponding train and validation) for each face and fingerprint databases respectively. $Test_i$ denotes i^{th} test set for any database. For a particular face and fingerprint database combination, $Test_{ij}$ represents the combination of i^{th} test set from face and j^{th} test set from fingerprint database. PCA [123] is used for extracting features from face. We have used PCA to keep the performance of base classifier for face reasonably weak and thus allow some space for fusion techniques to perform better than that of the base classifiers. The algorithm for fingerprint matching have been adopted from existing literature [55] [100] [49], using elastic graph matching technique over a minutiae feature set.

As described in Section 5.2 and in Section 5.3, the distribution of a database should be in four disjoint sets as train, $validation_1$, $validation_2$ and test sets. Three set pairs {Train, $validation_1$}, {Train, $validation_2$} and {Train, Test} in feature space, create train, validation and test response vector sets in classifier output space. Validation response vector set is needed to check the availability of class specific information and applicability of LDA or nonparametric LDA in classifier output space. If the application of LDA or nonparametric LDA-based eigenmodel produces better performance than that of direct classifier output on validation response vector set, then we can utilize the scope of strengthening our base classifiers using LDA and nonparametric LDA. Intuitively, a stronger base classifier always helps to enhance combination performance, as it helps to enhance mean accuracy of a classifier ensemble it belongs to. This is evident from our experimental results provided in this section. In our experimentation, we assume the test set to be same as $validation_2$ for the following reasons:

- The number of samples per class for fingerprint is too small to divide in four

sets.

- We put as much as possible samples per class in test set to make it far bigger than other sets and also to maintain reasonably low accuracy on test set for base classifiers.
- It can be easily argued that the cross validation test on test set will give us the same decision as the case when cross validation check is done on $validation_2$ set. This is because $validation_2$ set can be treated as a subset of test set.

The relatively weak performance of base classifiers (as shown in Table 5.4) allow the fusion to demonstrate its utility. The state of the art classifiers for face and fingerprint recognition are not used in this work as their performance leaves a narrow scope for fusion to improve upon them. We have performed experiments with different cases, where: (i) Both base classifiers are comparable (both either bad or good) in performance and (ii) One of the classifiers is less/more efficient in comparison to the other. A few examples of such combinations are tabulated in Table 5.3. Table 5.4 shows the performance as per Percentage Accuracy of face and fingerprint base classifiers on all test sets for all the databases, as specified before.

Table 5.3: Some test cases to simulate different type (performance-wise) of classifier combination.

Face		Fingerprint		Description
Set	Performance	Set	Performance	
$Test_1$ of ORL	83.33	$Test_1$ of FP_A	86.00	Both good
$Test_2$ of ORL	76.67	$Test_3$ of FP_B	60.00	Both bad
$Test_1$ of PIE	81.47	$Test_4$ of FP_B	63.33	Face good and Fingerprint bad
$Test_2$ of ORL	76.67	$Test_1$ of FP_A	86.00	Face bad and Fingerprint good

Table 5.5 and Table 5.6 show the performance of face and fingerprint classifiers after applying LDA and nonparametric LDA on classifier output space for the same

Table 5.4: Base classifier's performance (in Percentage Accuracy) for four face and fingerprint databases (two each).

Test Set	PIE	ORL	FP_A	FP_B
$Test_1$	81.47	83.33	86.00	78.00
$Test_2$	78.53	76.67	82.50	67.50
$Test_3$	-	-	80.00	60.00
$Test_4$	-	-	76.67	63.33

Table 5.5: Base classifier's performance (in Percentage Accuracy) using LDA for the face and fingerprint databases (two each).

Test Set	PIE	ORL	FP_A	FP_B
$Test_1$	86.18	86.67	64.00	34.00
$Test_2$	91.76	95.00	57.50	57.50
$Test_3$	-	-	63.33	55.33
$Test_4$	-	-	66.67	60.00

combinations as in Table 5.4. We witness the effectiveness of LDA and nonparametric LDA on face classifier output space by comparing the performance of face classifier using LDA, (see 1^{st} *and* 2^{nd} column of Table 5.5) and nonparametric LDA (see 1^{st} *and* 2^{nd} column of Table 5.6) with the one not using any of the two (see 1^{st} *and* 2^{nd} column of Table 5.4). The same conclusion cannot be drawn in case of fingerprint classifier, because of the deterioration of the performance of fingerprint classifier on the application of LDA (see 3^{rd} *and* 4^{th} column of Table 5.5) and nonparametric LDA (see 3^{rd} *and* 4^{th} column of Table 5.6) as compared to the performance of base classifier without using the two (see 3^{rd} *and* 4^{th} column of Table 5.4). Henceforth (as described in the algorithm in section 5.3.1), we use LDA and nonparametric LDA only on face classifier space while keeping fingerprint classifier space unperturbed, and then use the sum rule for combination. This can be considered to be a preprocessing step before fusion on classifier space to improve their performance. One can argue that we can use a better method for each classifier itself to make it strong. But that

argument is invalid in the context of the issue addressed in this chapter. We need completely different techniques to improve individual classifiers in feature space. For example, we cannot use the same new technique to improve both face and fingerprint classification. But here we are applying the same framework as a preprocessing step to all classifiers output responses, which are homogeneous.

Table 5.6: Base classifier's performance (in Percentage Accuracy) using nonparametric LDA for the face and fingerprint databases (two each).

Test Set	PIE	ORL	FP_A	FP_B
$Test_1$	85.00	86.67	66.00	30.00
$Test_2$	83.52	96.67	67.50	62.50
$Test_3$	-	-	60.00	56.67
$Test_4$	-	-	70.00	56.67

The combined results for each combination of face and fingerprint databases are tabulated in separate tables. Table 5.7-5.10 show the combined performance for PIE & FP_A, PIE & FP_B, ORL & FP_A and ORL & FP_B database combinations. P_{LDA} and P_{NLDA} are our pair of proposed methods using LDA and nonparametric LDA respectively. Results on all database combinations have shown the superiority of our method over existing fusion techniques. The superiority of our method over it's best competitor is explained by choosing four test cases (see in Table 5.3), which exhibit situations (performance) where: (1) both face and fingerprint classifiers are good, (2) both are bad, (3) face classifier is good while fingerprint classifier is bad and (4) the reverse of case (3). Four test cases have been used, based on the performance of individual base classifiers, which are explained below:

1. **Both classifiers are good:** The base classifiers performance for $Test_1$ of ORL and $Test_1$ of FP_A are 83.33% and 86% respectively. Both P_{LDA} and P_{NLDA} provide maximum accuracy of 98.00% as compared to 95.67% as given by product rule (see 1^{st} row of Table 5.9).

2. **Both classifiers are bad:** The performance of base classifiers for $Test_2$ of ORL and $Test_3$ of FP_B are 76.67% and 60.00% respectively. Nonparametric

Table 5.7: Combined performance (in Percentage Accuracy) with PIE and FP_A databases for different decision fusion strategies.

$TestSet$	Sum	Max	Min	Product	DT	DS	P_{LDA}	P_{NLDA}
$Test_{11}$	93.82	83.71	87.41	93.71	76.53	75.29	97.76	**98.29**
$Test_{12}$	91.32	83.23	82.65	88.01	78.89	76.59	94.25	**95.74**
$Test_{13}$	90.98	82.55	80.39	86.86	79.71	78.14	95.29	**96.27**
$Test_{14}$	90.29	82.25	76.67	86.08	83.92	84.51	96.76	**97.84**
$Test_{21}$	96.94	81.88	88.23	97.41	76.35	77.53	**98.53**	97.82
$Test_{22}$	93.90	81.25	82.79	91.47	80.51	78.09	**96.10**	94.71
$Test_{23}$	93.63	80.69	80.59	91.37	80.88	79.31	**96.57**	94.31
$Test_{24}$	93.23	80.98	76.67	89.31	84.61	84.90	**98.04**	95.39

Table 5.8: Combined performance (in Percentage Accuracy) with PIE and FP_B databases for different decision fusion strategies.

$TestSet$	Sum	Max	Min	Product	DT	DS	P_{LDA}	P_{NLDA}
$Test_{11}$	90.59	88.94	79.76	89.00	78.18	78.29	93.00	**93.47**
$Test_{12}$	84.41	85.66	68.38	77.72	74.12	75.07	**86.32**	85.22
$Test_{13}$	81.17	**85.29**	62.35	73.14	69.02	68.73	80.68	80.29
$Test_{14}$	87.65	83.73	62.45	73.82	75.49	76.67	**88.92**	**88.92**
$Test_{21}$	90.76	87.29	79.88	88.41	79.23	78.82	**94.71**	93.41
$Test_{22}$	87.21	85.37	69.12	80.96	75.44	76.18	**89.85**	86.84
$Test_{23}$	84.41	84.90	62.25	76.86	70.20	70.29	**86.27**	81.86
$Test_{24}$	89.31	84.02	62.45	76.57	77.35	78.43	**92.65**	89.80

LDA and LDA produce the same accuracy of 96.67%, whereas sum rule, the next best method in this case, gives 85.56% (see second last row of Table 5.10).

3. **Face classifier is good while fingerprint classifier is bad:** The performance of face and fingerprint classifier on $Test_1$ of PIE and $Test_4$ of FP_B are 81.47% and 63.33% respectively. The combined performaces for this pair, using P_{LDA}, P_{NLDA} and sum rule are 88.92%, 88.92% and 87.65% (see 4^{th} row of Table 5.8).

4. **Face classifier is bad while fingerprint classifier is good:** The performance of face and fingerprint classifiers on $Test_2$ of ORL and $Test_1$ of FP_A are 76.67% and 86.00% respectively. We obtained a maximum of 98.00% accuracy for both P_{LDA} and P_{NLDA}, as compared to 95.00% accuracy provided by sum rule (see 5^{th} row of Table 5.9).

Table 5.9: Combined performance (in Percentage Accuracy) with ORL and FP_A databases for different decision fusion strategies.

$TestSet$	Sum	Max	Min	Product	DT	DS	P_{LDA}	P_{NLDA}
$Test_{11}$	95.33	84.67	90.00	95.67	88.00	87.00	**98.00**	**98.00**
$Test_{12}$	92.50	85.00	82.52	89.58	87.08	88.75	96.67	**97.08**
$Test_{13}$	91.67	85.00	80.00	87.22	89.44	90.56	**97.22**	96.67
$Test_{14}$	91.11	84.44	76.67	86.67	92.77	92.77	**97.22**	96.67
$Test_{21}$	95.00	76.67	87.00	93.00	85.67	88.33	**98.00**	**98.00**
$Test_{22}$	90.42	76.67	82.50	88.33	84.17	85.83	**96.25**	95.83
$Test_{23}$	89.44	76.67	80.00	86.67	87.22	88.33	**95.56**	95.00
$Test_{24}$	87.78	76.67	76.67	84.44	90.56	90.00	95.00	**95.56**

In each of the cases described above, our proposed algorithm performs better than other methods. We have gained a percentage accuracy of 11.11% using both LDA and nonparametric LDA over their best competitor sum rule in test case 2, where we have combined two bad performing classifiers. In addition, one can observe (in general) that the performance of our proposed methods in the last two columns in Tables 5.7-5.10 are always better than any of the other approaches which do not use

Table 5.10: Combined performance (in Percentage Accuracy) with ORL and FP_B databases for different decision fusion strategies.

$TestSet$	Sum	Max	Min	Product	DT	DS	P_{LDA}	P_{NLDA}
$Test_{11}$	93.00	92.67	88.00	92.33	83.00	82.33	96.00	**97.67**
$Test_{12}$	88.75	88.33	70.83	81.67	78.33	80.83	92.50	**92.92**
$Test_{13}$	87.22	85.56	65.56	78.33	77.22	82.22	91.11	**92.78**
$Test_{14}$	91.11	84.44	62.78	78.89	80.00	85.55	**94.44**	92.78
$Test_{21}$	92.67	86.00	81.67	91.33	85.00	84.67	98.33	**98.67**
$Test_{22}$	85.83	80.83	69.17	77.08	83.75	83.75	96.25	**97.08**
$Test_{23}$	85.56	82.22	64.44	74.44	79.44	79.44	**96.67**	**96.67**
$Test_{24}$	88.89	78.89	65.00	76.67	80.56	87.22	**97.22**	**97.22**

class-specific information. We have an exception in the 3^{rd} row of Table 5.8, where max rule outperforms the rest. This validates the efficiency of our approach of decision combination over already existing fusion techniques. As evident from Tables 5.7-5.10, sum rule is the closest competitor of our method in most of the cases. Product rule is the next best. Decision Template (DT) and Dempster-Shafer (DS) combination fail to work well because of their blind belief on all classifiers to provide class-specific information in their outputs.

5.5 Conclusion

As unimodal biometry is going to saturate in near feature in terms of performance and is having several limitations like nonuniversality, spoof attacks and noisy data, multimodal biometry has become a hot field for research. Due to the heterogeneous nature of the features extracted from different modalities, it is very difficult to combine them at feature level. So decision level fusion has become an obvious choice for combining different modalities due to the homogeneity of multiple classifiers at decision level.

Here we have tried to strengthen each classifier by using training data at fusion level, depending upon the presence of class-specific information in classifier output

space. If the training data uniformly sample the underlying class distribution and represent the class properly, LDA and nonparametric LDA can enhance the class separability and hence improve base classifier performance when applied on classifier output space. This may not be applicable for a classifier which gives response vectors not representing class distribution and nonuniformly sampling the distribution. In this case, application of LDA and nonparametric LDA even deteriorate the performance of base classifier. We have verified the performance enhancement on a validation response vector set, and then used it as a basis for the application of LDA and nonparametric LDA on response vectors for test samples. After this preprocessing step, we have combined classifiers using sum rule. The experimental results show that our proposed method performs better than sum, product, max, min rule, Decision Template (DT) and Dempster-Shafer (DS) combination.

CHAPTER 6

Conclusion

In this thesis, we have proposed three novel methods to solve the face recognition problem and a multiple classifier combination strategy for combining multiple modalities. Several difficulties in an automatic face recognition problem (such as illumination, pose, expression, camouflage etc.) are addressed in our proposed face recognition techniques. The decision combination of face and fingerprint classifiers is attempted to obtain a robust multimodal biometric system. Following is the summary and contributions of the work presented in this thesis.

6.1 Contribution of the Thesis

- In Chapter 3, we propose the *subband face* as a new representation for the face recognition task. Only the discriminatory information of a face is retained or captured in a subband face. Discrete wavelet transform is used to decompose the original face image into approximation and detail subbands. We perform multi-level dyadic decomposition of a face image using the Daubechies filters [30]. The subband face may be reconstructed from selected subbands by suppressing the approximation at a suitable higher level and retaining only the details. An inherent information fusion is being performed by the reconstruction process which retains only the inter-class discriminatory informations and discards the inter-class common informations. The information of a face in the details at lower levels of decomposition, usually contains noise and redundant pixels which do not constitute any discriminatory information for a face. It is thus often necessary to eliminate these details for the representation of a face, and in such cases the subband face is obtained by partial reconstruction. We present four different criteria as cost functions to obtain an optimal subband

face for each subject, and compare their performances. The performance of the subband face representation on several linear subspace techniques: PCA, LDA, 2D-PCA, 2D-LDA and Discriminative Common Vectors (DCV) with Yale, ORL and PIE face databases show that the subband face based representation performs significantly better than the proposed multiresolution face recognition by Ekenel [36], for frontal face recognition in the presence of varying illumination, expression and pose. Peak Recognition Accuracy (PRA) for all three databases are: 100% for Yale (using LDA), 95.32% for PIE (using DCV) and 91.67% for ORL (using 2D-LDA).

- In chapter 4, we try to enhance classification performance by combining the discriminative features from both range space and null space of within-class scatter matrix S_w at (i) feature level and (ii) decision level.

  (i) Feature Fusion Strategy: Feature level information fusion allows us to utilize the whole set of discriminative directions present in an entire face space. As every face has a unique decomposition in null space and range space, we project all class means in both spaces to obtain two sets of projected means. Now each of these sets are used separately to search for the directions that discriminates them in that space. This step is equivalent to applying PCA separately on the set of projected means. These two eigenmodels are then combined using: 1) Covariance Sum and 2) Gramm-Schmidt Orthonormalization [42]. These methods construct a new set of directions integrating the information from both spaces. We then reorder and select the best combination among those directions using two techniques: 1) Forward Selection and 2) Backward Selection [39] on a validation set based on a class separability criterion to obtain the best discriminability across classes.

  (ii) Decision Fusion Strategy: For decision fusion, we extract two disjoint sets of optimal discriminatory basis separately from null space and range space of within-class scatter matrix S_w, to obtain two different classifiers. Then we combine the classifiers obtained on null space and range space using sum rule and product rule, the two classical decision fusion techniques developed by Kittler et al. [60],[59]. We also exploit each classifier space separately using

LDA and nonparametric LDA on them to enhance class separability at classifier response level and then combine them using sum rule. We refer to the set of class scores provided by a classifier on a sample, as *response vector*. Response vectors are used as features vectors at decision level and we then employ LDA and nonparametric LDA to enhance class separability at classifier output space. Response vectors on a validation set (disjoint from training and testing sets of a database) are used as training data at decision level. Then the response vector for a test case is recalculated in the newly formed eigenmodel (LDA or nonparametric LDA-based) to improve the combined classification accuracy. Decision fusion using both P_{LDA} and P_{NLDA} provide **100%** on PIE database.

- In chapter 5, we describe a new approach to combine decisions from face and fingerprint classifiers for multimodal biometry by exploiting the individual classifier space on the basis of availability of class specific information present in the classifier space. We train the face classifier using response vectors on a validation set for enhancing class separability using LDA and nonparametric LDA in the classifier output space and thereby improving performance of the face classifier. Fingerprint classifier does not often provide this information due to high sensitivity of available minutiae points, producing partial matches across subjects. The enhanced face and fingerprint classifiers are combined using sum rule. We also propose a generalized algorithm for Multiple Classifier Combination (MCC) based on our approach. Experimental results on two face and two fingerprint databases exhibit the superiority of the proposed method over other already existing fusion techniques.

6.2 Scope of Future Work

Here we provide some of the open problems that can be attempted:

1. In chapter 3, we proposed four different criteria for selecting optimum subject-specific subband face. We believe a better criterion can indeed improve the performance of the technique.
2. In the same context, the introduction of subject-specific subspace method can

also be introduced. But class specific linear subspace method requires sufficient number of training samples per class which can be attained by generating multiple face samples (image rendering using 3D model of a face) from a single or a few face images.

3. The subband face introduced in this thesis is obtained by suppressing approximation and in some case, eliminating lower level details from dyadic decomposed subbands. The search of optimal subband face can be extended by using a full tree decomposition with 2D-DWT and designing a more efficient algorithm and criterion to accomplish the search procedure.

4. In chapter 4, we combined null space and range space of within-class scatter. Different combination strategies at feature and decision level can be explored to improve the classification accuracy.

5. One can investigate the scope of combining subband face representation with dual space method to implement a superior face recognition system.

6. Effect of weighted modular PCA [67] on subband face representation can also form a nice extension of subband face based recognition system.

7. In chapter 5, face and fingerprint are combined using an efficient decision combination technique. Here, one can explore superior techniques for decision fusion of face and fingerprint classifiers. A robust multimodal biometrics can be developed based on three or more cues (face, fingerprint, iris, voice, palm-print, etc.) using our proposed method of multiple classifier combination.

Appendix A

Fingerprint Recognition System

Fingerprint is an important biometric signature to identify an individual. Fingerprint impression is time invariant after adolescence, unless the finger is damaged or bruised due to accidents or injuries. So, it finds wide applications for all type of secured access, including entry to buildings, missile launch control, nuclear reactor, bank vaults, confidential files and most recently, for e-Commerce accesses, such as, ATM counters, web-enabled transactions etc.

A fingerprint is formed from an impression of the pattern of ridges on a finger. A ridge is defined as a single curved segment, and a valley is the region between two adjacent ridges. The minutiae, which are the local discontinuities in the ridge flow pattern, provide the features that are used for identification. Details such as the type, orientation, and location of minutiae are taken into account when performing minutiae extraction. The different stages of pre-processing for recognition of fingerprints are:

1. Image Enhancement
    (a) Normalization
    (b) Orientation Image Estimation
    (c) Frequency Image Estimation
    (d) Filtering
2. Image Segmentation
3. Binarization
4. Thinning
5. Minutiae Extraction
6. Minutiae Matching

These stages are discussed in the following sections.

A.1 Image Enhancement

The quality of the ridge structures in a fingerprint image is an important characteristic, as the ridges carry the information of characteristic features required for minutiae extraction. Ideally, in a well-defined fingerprint image, the ridges and valleys should alternate and flow in a locally constant direction. This regularity facilitates the detection of ridges and consequently, allows minutiae to be precisely extracted from thinned ridges. However, in practice, a fingerprint image may not be well defined due to the noise that corrupt the clarity of the ridge structures. Thus image enhancement are often employed to reduce the noise and enhance the definition of ridges against valleys.

The fingerprint image enhancement algorithm receives an input gray-level fingerprint image and outputs the enhanced image after applying a set of intermediate steps which are: (a) Normalization, (b) Orientation Image Estimation, (c) Frequency Image Estimation and (d) Filtering using Gabor filter. This algorithm is mostly adopted from the technique proposed by Hong, Wan and Jain [49].

A.1.1 Notation

- **A gray-level Fingerprint image (I):** A gray-level fingerprint image, I, is defined as an $N \times N$ matrix, where $I(i,j)$ represents the intensity of the pixel at the i^{th} row and j^{th} column. The mean and variance of a gray-level fingerprint image, I, are defined as,

$$M(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I(i,j) \tag{A.1}$$

$$VAR(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(I(i,j) - M(I)\right)^2 \tag{A.2}$$

- **An orientation image (O):** An orientation image, O, is defined as an $N \times N$ image, where $O(i,j)$ represents the local ridge orientation at pixel (i,j). Local ridge orientation is usually specified for a block rather that at every pixel. An

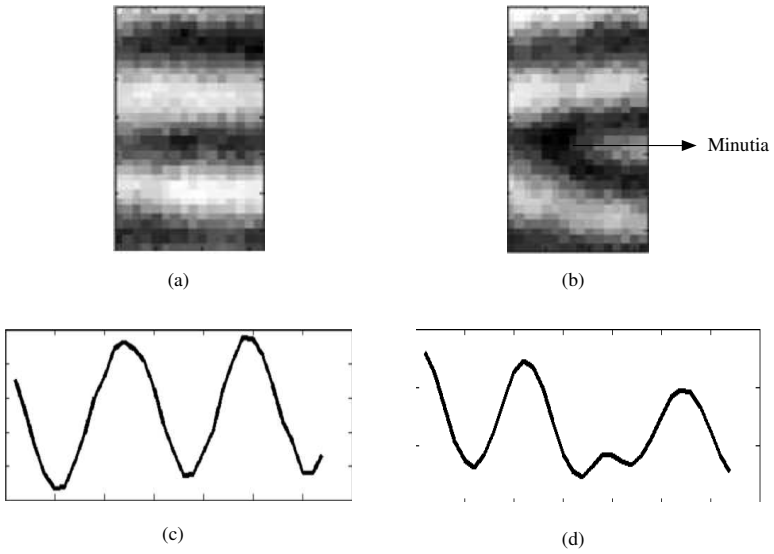


Figure A.1: (a) and (b) A local window containing ridge valley structures; (c) and (d) Curve (i.e. x-signature) obtained along the direction normal to the ridge orientation.

image is divided into a set of $W \times W$ nonoverlapping blocks and a single local ridge orientation is defined for each block.

- **A frequency image (F):** A frequency image, F, is an $N \times N$ image, where $F(i,j)$ represents the local ridge frequency of the ridge and valley structures in a local neighborhood along a direction normal to the local ridge orientation. The ridge and valley structures in a local neighborhood where minutiae or singular points appear do not form a well-defined sinusoidal-shaped wave (see Fig. A.1). In such situations, the frequency is defined as the average frequency in the neighborhood of block (i,j). Like orientation image, frequency image is specified block-wise.

A.1.2 Normalization

Let $I(i,j)$ denote the gray-level value at pixel (i,j), M and VAR denote the estimated mean and variance of I, respectively, and $G(i,j)$ denote the normalized gray-level value at pixel (i,j). The normalized image is defined as follows:

$$G(i,j) = \begin{cases} M_0 + \sqrt{\frac{VAR_0(I(i,j)-M)^2}{VAR}} & \text{if } I(i,j) > M \\ M_0 - \sqrt{\frac{VAR_0(I(i,j)-M)^2}{VAR}} & \text{Otherwise} \end{cases} \quad \text{(A.3)}$$

(a) (b)

Figure A.2: Input fingerprint images.

(a) (b)

Figure A.3: Normalized images of two input fingerprints shown in Fig. A.2.

where M_0 and VAR_0 are the desired mean and variance values, respectively. Normalization is a pixel-wise operation. It does not change the clarity of the ridge and valley structures. It reduces the variations in gray-level values along ridges and valleys, which facilitates the subsequent processing steps. Fig A.3 shows the result of normalization on two different fingerprint images shown in Fig A.2.

A.1.3 Orientation Image

The orientation image represents an intrinsic property of the fingerprint images and defines invariant coordinates for ridges and valleys in a local neighborhood. By viewing a fingerprint image as an orientated texture, a least mean square orientation estimation algorithm is given below:

1. Divide G into blocks of size $W \times W$ (16×16).

2. Compute the gradients $\varrho_x(i,j)$ and $\varrho_y(i,j)$ at each pixel (i,j). The gradient operation in our case is done by sobel operator.

3. Estimate the local orientation of each block centered at pixel (i,j) using the following equation,

$$\gamma_x(i,j) = \sum_{u=i-\frac{W}{2}}^{i+\frac{W}{2}} \sum_{v=j-\frac{W}{2}}^{j+\frac{W}{2}} 2\varrho_x(u,v)\varrho_y(u,v), \tag{A.4}$$

$$\gamma_y(i,j) = \sum_{u=i-\frac{W}{2}}^{i-\frac{W}{2}} \sum_{v=j-\frac{W}{2}}^{j-\frac{W}{2}} \left(\varrho_x^2(u,v)\varrho_y^2(u,v)\right), \tag{A.5}$$

$$\theta(i,j) = \frac{1}{2}tan^{-1}\left(\frac{\gamma_y(i,j)}{\gamma_x(i,j)}\right). \tag{A.6}$$

where $\theta(i,j)$ is the least square estimate of the local ridge orientation at the block centered at pixel (i,j) and represents the direction which is orthogonal to the dominant direction of the Fourier spectrum of the $W \times W$ window.

4. Noise, corrupted ridge and valley structures may result incorrect estimation of local ridge orientation. As local ridge orientation varies slowly in a local neighborhood low-pass filtering can be used to correct $\theta(i,j)$. For low-pass filtering the orientation image is converted to a continuous vector field and is defined as,

$$\Phi_x(i,j) = cos(2\theta(i,j)), \tag{A.7}$$

$$\Phi_y(i,j) = sin(2\theta(i,j)). \tag{A.8}$$

where $\Phi_x(i,j)$ and $\Phi_y(i,j)$ are the x and y components of the vector field, respectively. The low-pass filtering can be done as follows,

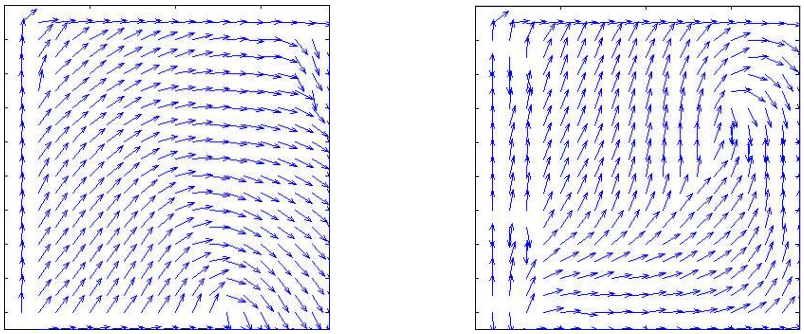
$$\Phi'_x(i,j) = \sum_{u=-\frac{D_\Phi}{2}}^{\frac{D_\Phi}{2}} \sum_{v=-\frac{D_\Phi}{2}}^{\frac{D_\Phi}{2}} D(u,v)\Phi_x(i-uW, j-vW), \tag{A.9}$$

$$\Phi'_y(i,j) = \sum_{u=-\frac{D_\Phi}{2}}^{\frac{D_\Phi}{2}} \sum_{v=-\frac{D_\Phi}{2}}^{\frac{D_\Phi}{2}} D(u,v)\Phi_y(i-uW, j-vW). \tag{A.10}$$

where D is a two-dimensional low-pass filter with unit integral and $D_\Phi \times D_\Phi$ specifies the size of the filter.

5. Compute the local ridge orientation at (i,j) using,

$$O(i,j) = \frac{1}{2}tan^{-1}\left(\frac{\Phi'_y(i,j)}{\Phi'_x(i,j)}\right) \tag{A.11}$$



(a) (b)

Figure A.4: Orientation images of two input fingerprints shown in Fig. A.2.

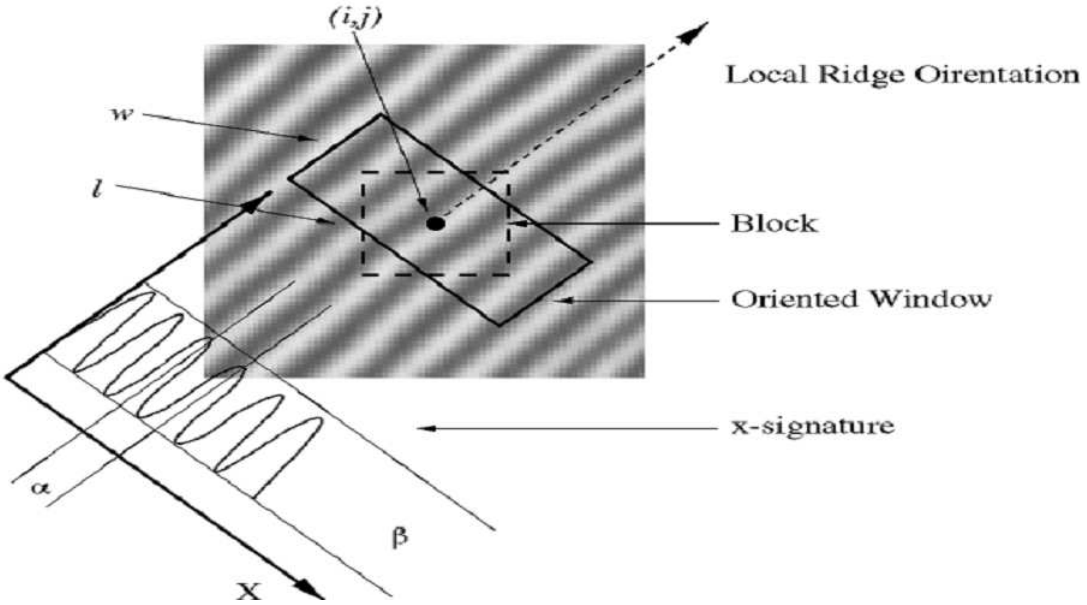


Figure A.5: A X-signature with $w = W$ and $l = H$ in our case.

The orientation image of the input fingerprint images shown in Fig A.2 are given in Fig A.4.

A.1.4 Ridge Frequency Image

In a local neighborhood, where no minutiae and singular points appear, the gray levels along ridges and valleys can be modeled as a sinusoidal shaped wave along a direction normal to the local ridge orientation (Fig. A.5). So, local ridge frequency is

another intrinsic property of a fingerprint image. The steps for obtaining frequency image, F, is as follows:

1. Divide G into blocks of size $W \times W$ (16×16).
2. For each block centered at pixel (i, j), compute an orientation window of size $H \times W$ (32×16) that is defined in the ridge coordinate system (see Fig. A.5).
3. For each block centered at pixel (i, j), compute the x-signature, $X[0], X[1], \ldots, X(H-1)$, of the ridges and valleys within the orientated window, where,

$$X[k] = \frac{1}{W} \sum_{d=0}^{W-1} G(u, v), \quad k = 0, 1, ..., H-1, \tag{A.12}$$

$$u = i + \left(d - \frac{W}{2}\right) cos\, O(i, j) + \left(k - \frac{H}{2}\right) sin\, O(i, j), \tag{A.13}$$

$$v = j + \left(d - \frac{W}{2}\right) sin\, O(i, j) + \left(\frac{H}{2} - k\right) cos\, O(i, j). \tag{A.14}$$

   If no minutiae and singular points appear in the orientated window, the x-signature forms a discrete sinusoidal-shape wave, which has the same frequency as that of the ridges and valleys. Let $T(i, j)$ be the average number of pixels between two consecutive peaks in x-signature, then the frequency $\Omega(i, j)$ is computed as: $\Omega(i, j) = 1/T(i, j)$. If no consecutive peaks can be detected from the x-signature, then the frequency is assigned a value of -1 to differentiate it from valid frequency values.
4. For a fingerprint image scanned at a fixed resolution like 500dpi, the value of the frequency lies in certain range ([1/3, 1/25] for 500dpi). So any estimated frequency out of this range will be assigned a value of -1 to indicate invalid frequency.
5. The blocks containing minutiae and/or singular points will provide a well-defined sinusoidal-shaped wave. So, the frequency values for these blocks need to be interpolated from the frequency of the neighboring blocks which have a well-defined frequency. The interpolation is performed as follows:

(i) For each block centered at (i, j),

$$\Omega'(i,j) = \left\{ \begin{array}{cc} \Omega(i,j) & \text{if } \Omega(i,j) \neq -1 \\ \frac{\sum_{u=-K_\Omega/2}^{K_\Omega/2} \sum_{v=-K_\Omega/2}^{K_\Omega/2} K_g(u,v)\mu(\Omega(i-uW,j-vW))}{\sum_{u=-K_\Omega/2}^{K_\Omega/2} \sum_{v=-K_\Omega/2}^{K_\Omega/2} K_g(u,v)\nu(\Omega(i-uW,j-vW)+1)} & \text{Otherwise} \end{array} \right\} \tag{A.15}$$

where

$$\mu(x) = \left\{ \begin{array}{cc} 0 & \text{if } x \leq 0 \\ x & \text{Otherwise} \end{array} \right\} \tag{A.16}$$

$$\nu(x) = \left\{ \begin{array}{cc} 0 & \text{if } x \leq 0 \\ 1 & \text{Otherwise} \end{array} \right\} \tag{A.17}$$

K_g is a discrete Gaussian kernel with mean and variance of zero and nine, respectively, and $K_\Omega = 7$ is the size of the kernel.

(ii) If there exists at least one block with the frequency value -1, then swap Ω and Ω' and go to step (i).

6. Finally, a low-pass filter is used to remove the outliers,

$$F(i,j) = \sum_{u=-L_\Omega/2}^{L_\Omega/2} \sum_{v=-L_\Omega/2}^{L_\Omega/2} L(u,v)\Omega'(i-uW, j-vW), \tag{A.18}$$

where L is a two-dimensional low-pass filter with unit integral and $L_\Omega = 7$ is the size of the filter.

The frequency image after low-pass filtering is shown in Fig. A.6.

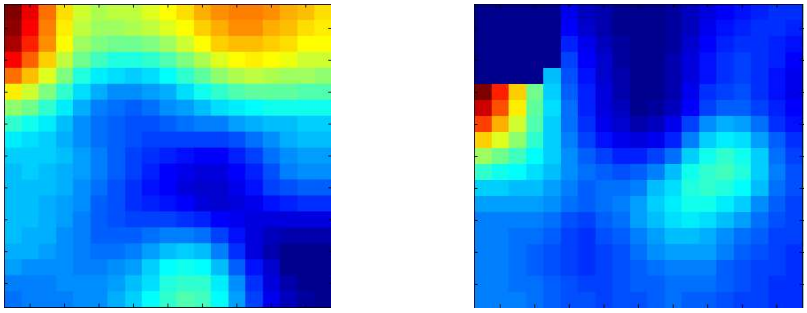
A.1.5 Filtering

The derived orientation and frequency images are used for noise removal. A bandpass filter tuned to the corresponding frequency and orientation can remove undesired noise. Gabor filters have both frequency-selective and orientation-selective properties and have optimal joint resolution in both spatial and frequency domains [32, 53]. The even-symmetric Gabor filter has the general form [53],

$$h(x, y : \phi, f) = exp\left\{ -\frac{1}{2}\left[\frac{x_\phi^2}{\delta_x^2} + \frac{y_\phi^2}{\delta_y^2} \right] \right\} cos\left(2\pi f x_\phi\right), \tag{A.19}$$

$$x_\phi = x\, cos\, \phi + x\, sin\, \phi, \tag{A.20}$$

$$y_\phi = -x\, sin\, \phi + y\, cos\, \phi. \tag{A.21}$$



(a) (b)

Figure A.6: Frequency images (after filtering) of two input fingerprints shown in Fig. A.2.

where ϕ is the orientation of the Gabor filter, f is the frequency of a sinusoidal plane wave, and δ_x and δ_y are the space constants of the Gaussian envelope along x and y axes, respectively. The modulation transfer function (MTF) of the Gabor filter can be represented as,

$$H(u, v : \phi, f) = 2\pi\delta_x\delta_y exp\left\{-\frac{1}{2}\left[\frac{(u_\phi - u_0)^2}{\delta_u^2} + \frac{(v_\phi - v_0)^2}{\delta_v^2}\right]\right\} \quad (A.22)$$

$$+ \; 2\pi\delta_x\delta_y exp\left\{-\frac{1}{2}\left[\frac{(u_\phi + u_0)^2}{\delta_u^2} + \frac{(v_\phi + v_0)^2}{\delta_v^2}\right]\right\}, \quad (A.23)$$

$$u_\phi = u\,cos\,\phi + v\,sin\,\phi, \quad (A.24)$$

$$v_\phi = -u\,sin\,\phi + v\,cos\,\phi, \quad (A.25)$$

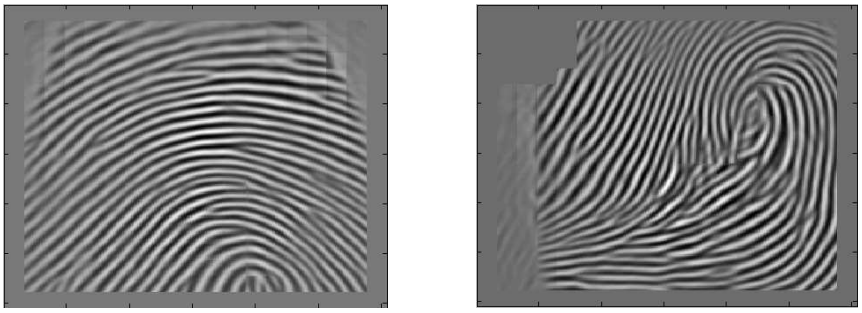
$$u_0 = \frac{2\pi\,cos\,\phi}{f}, \quad (A.26)$$

$$v_0 = \frac{2\pi\,sin\,\phi}{f}. \quad (A.27)$$

where $\delta_u = 1/2\pi\delta_x$ and $\delta_v = 1/2\pi\delta_y$. To apply Gabor filters to an image, three parameters must be specified:

- The frequency, f,
- The filter orientation ϕ, and
- The standard deviations of the Gaussian envelope, δ_x and δ_y.

In this case, the frequency of the filter is given by local ridge frequency and the orientation by local ridge orientation. The selection of the values of δ_x and δ_y involves



(a) (b)

Figure A.7: Enhanced images of two input fingerprints shown in Fig. A.2.

a trade-off. The larger the values, the more robust to noise the filters are, but the more likely the filter will create spurious ridges and valleys. On the other hand, the smaller the values of δ_x and δ_y, the less likely the filters will create spurious ridges and valleys, and they will be less effective in removing noise. The enhanced image E is expressed as follows,

$$E(i,j) = \sum_{u=-W_g/2}^{W_g/2} \sum_{v=-W_g/2}^{W_g/2} h\left(u, v : O(i,j), F(i,j)\right) G(i-u, j-v). \tag{A.28}$$

where $W_g = 11$ specifies the size of the Gabor filters. Fig. A.8 shows a number of 16×16 local blocks and corresponding Gabor filters tuned to the frequency and orientation of the corresponding blocks. Fig. A.7 demonstrates the enhanced images of the input fingerprints shown in Fig. A.2.

A.2 Image Segmentation

Segmentation is the process of separating the foreground regions from the background regions. The background corresponds to the regions containing no valid fingerprint information. When minutiae extraction algorithms are applied to the background regions of an image, it results in the extraction of noisy and false minutiae. Thus, segmentation is employed to discard these background regions to facilitate reliable minutiae extraction. In a fingerprint, the background regions generally exhibit a low gray-scale variance, whereas the foreground regions have a very high variance. Hence, a method based on variance thresholding [87] can be used to perform the

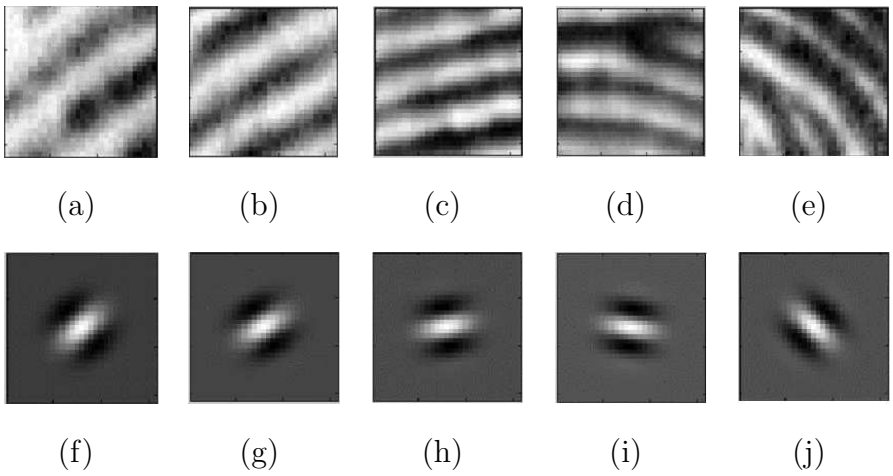


Figure A.8: (a), (b), (c), (d) and (e) shows blocks of size 16×16; (f), (g), (h), (i) and (j) shows the Gabor filters tuned to the orientation and frequency of the corresponding blocks.

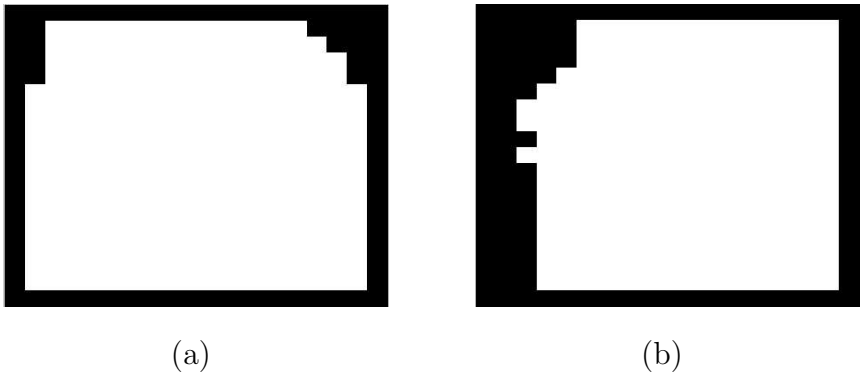
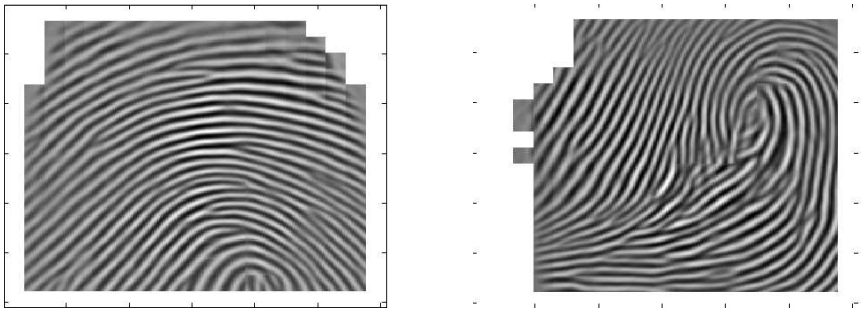


Figure A.9: (a) and (b) shows masks of two input fingerprints shown in Fig. A.2.



(a) (b)

Figure A.10: (a) and (b) shows the enhanced fingerprint images after masking.

segmentation. Firstly, the image is divided into blocks and the gray-level variance is calculated for each block in the image. If the variance is less than the global threshold, then the block is assigned to be a background region (i.e. assign $R(i,j) = 0$ for the block centered at (i,j)); otherwise, it is assigned to be the part of the foreground (i.e. assign $R(i,j) = 1$), where R is called the mask. The gray-level variance for a block of size $W \times W$ is defined as:

$$V(k) = \frac{1}{W^2} \sum_{i=0}^{W-1} \sum_{j=0}^{W-1} \left(E(i,j) - M(k)\right)^2 . \tag{A.29}$$

where $M(k)$ is the mean gray-level value for block k and $E(i,j)$ is the enhanced image. Then, the enhanced image after masking can be represented by,

$$H(i,j) = \left\{ \begin{array}{cc} E(i,j) & \text{if } R(i,j) = 1 \\ 255 & \text{Otherwise} \end{array} \right\} \tag{A.30}$$

Fig. A.9 and Fig. A.10 shows the masks and enhanced images after masking for input fingerprints, respectively.

A.3 Image Binarization

Binarization is the process of converting a gray-level image into a binary image where the black pixels represent ridges and white pixels represent valleys. This process improve the contrast between the ridges and valleys and consequently facilitates the extraction of minutiae.

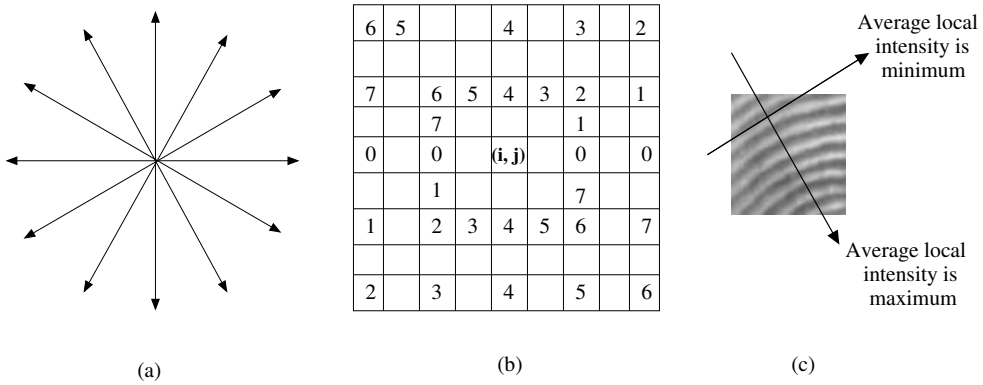
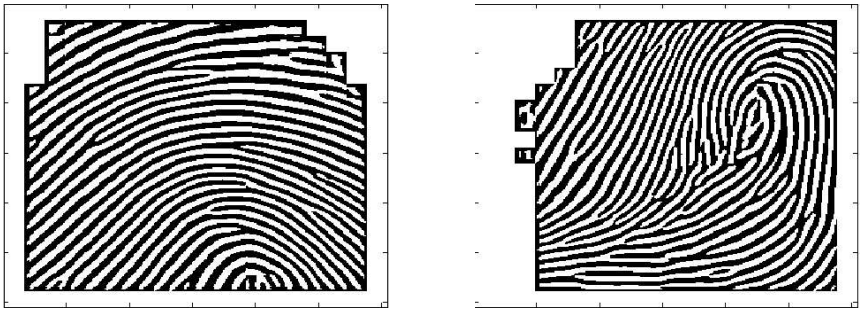


Figure A.11: (a) The eight directions; (b) The 9×9 mask for computing slit-sums; (c) For ridge pixels, average local intensity or slit-sum is minimum along the direction of the ridge and maximum along the normal direction of ridge.

A slit-sum method with local threshold proposed by Stock and Swonger [118] is used to binarize a image. This method uses pixel alignment along eight (8) discrete directions, $0, \pi/8, 2\pi/8, ..., 7\pi/8$ (see Fig. A.11(a)) and a 9×9 mask (see Fig. A.11(b)) to center at the pixel of interest. The basic idea here is that for each pixel that belongs to ridge line, there exists an orientation whose average local intensity is lower than those of remaining orientations (see Fig. A.11(c)). The gray-level values along eight directions are added respectively to obtain each slit-sum by using the following equations,

$$
\begin{aligned}
S_0 &= H(i, j+4) + H(i, j+2) + H(i, j-2) + H(i, j-4), \\
S_1 &= H(i-2, j+4) + H(i-1, j+2) + H(i+1, j-2) + H(i+2, j-4), \\
S_2 &= H(i-4, j+4) + H(i-2, j+2) + H(i+2, j-2) + H(i+4, j-4), \\
S_3 &= H(i-4, j+2) + H(i-2, j+1) + H(i+2, j-1) + H(i+4, j-2), \\
S_4 &= H(i-4, j) + H(i-2, j) + H(i+2, j) + H(i+4, j), \\
S_5 &= H(i-4, j-2) + H(i-2, j-1) + H(i+2, j+1) + H(i+4, j+2), \\
S_6 &= H(i-4, j-4) + H(i-2, j-2) + H(i+2, j+2) + H(i+4, j+4), \\
S_7 &= H(i-2, j-4) + H(i-1, j-2) + H(i+1, j+2) + H(i+2, j+4).
\end{aligned}
$$

where $S_0, S_1, ..., S_7$ represent the sum of gray-level values for eight discrete direction (slit). Let S_{max}, S_{min} and S_{sum} be the maximum, minimum and sum of the



(a) (b)

Figure A.12: Binarized images of two input fingerprints shown in Fig. A.2.

slit-sums calculated in eight different directions. That is,

$$\begin{aligned} S_{max} &= \max_{i=0,..,7} S_i, \\ S_{min} &= \min_{i=0,..,7} S_i, \\ S_{sum} &= \sum_{i=0}^{7} S_i \end{aligned}$$

Finally, the binarized image can be obtained by applying one of the following equations.

$$B(i,j) = \left\{ \begin{array}{ll} 1 & \text{if } H(i,j) \geq S_{sum}/12 \\ 0 & \text{Otherwise} \end{array} \right\} \tag{A.31}$$

$$B(i,j) = \left\{ \begin{array}{ll} 1 & \text{if } (S_{max} + S_{min}) \geq S_{sum}/4 \\ 0 & \text{Otherwise} \end{array} \right\} \tag{A.32}$$

$$B(i,j) = \left\{ \begin{array}{ll} 1 & \text{if } (4H(i,j)S_{max} + S_{min}) \geq 3S_{sum}/8 \\ 0 & \text{Otherwise} \end{array} \right\} \tag{A.33}$$

We used the first equation for obtaining binarized image. Fig. A.12 shows the binarized output for the input fingerprints shown in Fig. A.2.

A.4 Image Thinning

For thinning, we used the method proposed by Zhang and Suen [142]. The input for this method is a binary fingerprint image with pixels of ridge and valley having value 1 (black) and value 0 (white), respectively. A 3×3 window as shown in Fig. A.13

P_9 (i−1, j−1)	P_2 (i, j−1)	P_3 (i+1, j−1)
P_8 (i−1, j)	P_1 (i, j)	P_4 (i+1, j)
P_7 (i−1, j+1)	P_6 (i, j+1)	P_5 (i+1, j+1)

Figure A.13: The locations of the nine pixels in a window.

is used in the thinning process, where P_1 is the center pixel and $P_2, ..., P_9$ are its neighbors. Let P'_i be the pixel value of P_i. The method consists of two iterations applied to any pixels with a pixel value 1, and determines whether a center point must be to preserved or deleted. In the first iteration, the center point is deleted from the image, if the following conditions are satisfied:

(a) $2 \leq N(P_1) \leq 6$;

(b) $S(P_1) = 1$;

(c) $P'_2 * P'_4 * P'_6 = 0$;

(d) $P'_4 * P'_6 * P'_8 = 0$;

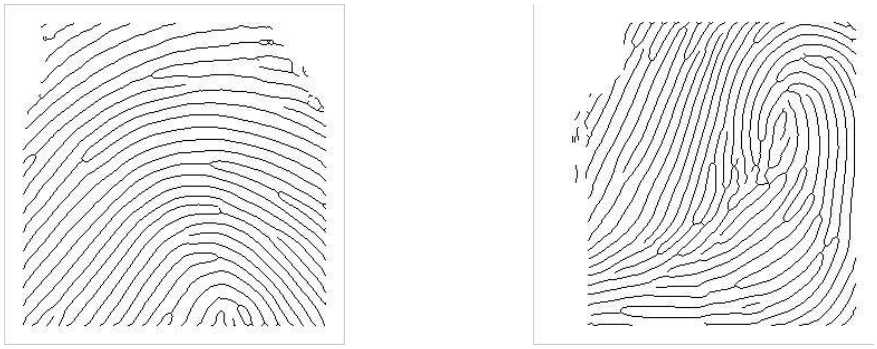
where $N(P_1)$ is the number of nonzero neighbors of P_1, that is, $N(P_1) = P'_2 + P'_3 + ... + P'_9$ and $S(P_1)$ is the number of transitions from 0 to 1 for the order set $\{P'_2, P'_3, ..., P'_9\}$. $N(P_1) = 0$, $N(P_1) = 1$, and $N(P_1) > 6$ are corresponding to the center pixel P_1 is an isolation point, end point and intermediate point, respectively. If $S(P_1) > 1$, the center point P_1 is a bridge that preserves the connectivity of a skeleton. For these four cases, the center pixel P_1 will not be deleted to break the skeleton.

In the second iteration, the conditions (a) and (b) remain the same, conditions (c) and (d) are modified as follows:

(c') $P'_2 * P'_4 * P'_8 = 0$;

(d') $P'_2 * P'_6 * P'_8 = 0$;

Conditions (c) and (d) for the first iteration are used to remove south-east boundary points and north-west corner points which are not skeletal pixels. Similarly, con-



(a) (b)

Figure A.14: Thinned images of two input fingerprints shown in Fig. A.2.

ditions (c') and (d') are used to remove north-west boundary points and south-east corner points. The iteration process is repeated until no more pixels can be removed. The skeleton images of the input fingerprint images are shown in Fig. A.14.

A.5 Minutiae Extraction

Once a binary skeleton has been obtained, a simple image scan allows the pixel corresponding to minutiae to be detected. This stage consists of two stages [100],

- Detection of ridge endings and bifurcations from thinned images, by using neighborhood constraints and
- Post-processing to remove undesired features.

For minutiae detection the concept of *Crossing number* is used. Pixels corresponding to minutiae are characterized by a crossing number different from 2. The crossing number [6] $cn(p)$ of a pixel p in a binary image is defined as half the sum of the differences between pairs of adjacent pixels in the 8-neighborhood of p:

$$cn(p) = \frac{1}{2} \sum_{i=1,..,8} |val(p_{i \bmod 8}) - val(p_{i-1})|, \tag{A.34}$$

where $p_0, p_1, ..., p_7$ are the pixels belonging to an ordered sequence of pixels defining the 8-neighborhood of p and $val(p) \in \{0, 1\}$ is the pixel value. It is simple to note (Fig. A.15) that a pixel p with $val(p) = 1$:

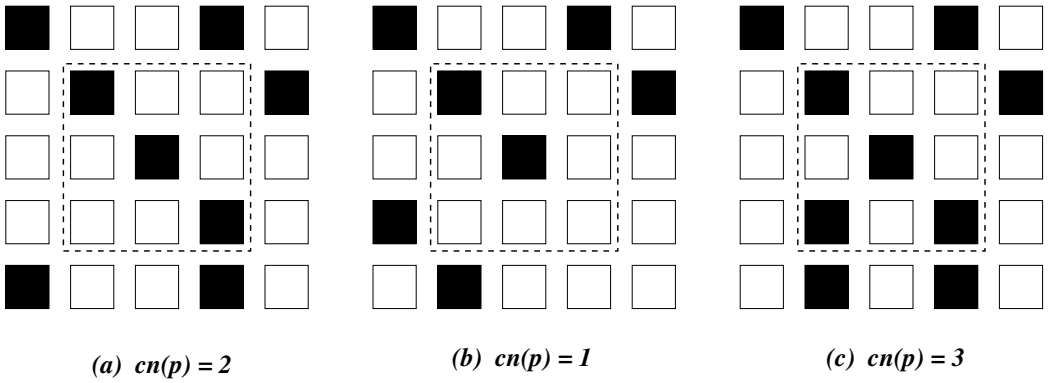


Figure A.15: (a)Intra-ridge pixel; (b) Termination minutia; (c) Bifurcation minutia.

- is an intermediate ridge point, if $cn(p) = 2$;
- corresponds to a termination minutia, if $cn(p) = 3$;
- defines a more complex minutia (bifurcation, crossover, etc.), if $cn(p) \geq 3$

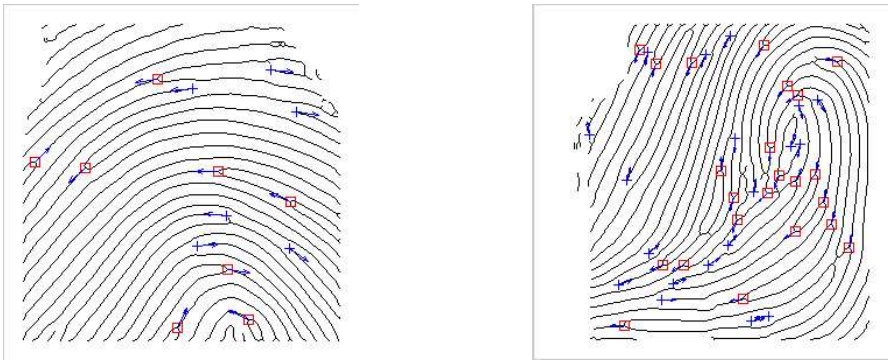
For each extracted minutiae point, the following information is recorded:

- x and y coordinates,
- orientation, α, of the associated ridge segment, and
- type of the minutiae (ridge ending or bifurcation).

In the second stage, there are 3-4 levels of elimination: nearby features, L-shaped false bifurcations, boundary effects and spike elimination.

- **Ridge break elimination:** Two end points with the same orientation and within a distance threshold T_1 are eliminated.
- **Spike elimination:** An end point which is connected to a bifurcation point and is also within a distance threshold T_2 is eliminated.
- **Boundary effects:** The minutiae detected within a specified border of the boundary of the foreground areas are deleted.

The post-processing stage eliminates spurious feature points based on the structural and spatial relationships of the minutiae. In Fig. A.16, the minutiae features after post-processing, for two input fingerprint images are shown. Bifurcations are marked



(a) (b)

Figure A.16: Thinned fingerprint images marked with minutiae.

by square box and ridge endings by a "+" sign. Minutiae directions are marked with an arrow (→) sign.

A.6 Minutiae Matching

The matching is done using Elastic graph matching as proposed by Ratha et al. [100]. The output of this stage is a matching score which says the goodness of the matching. There are three main steps involved in this process:

1. Registration,
2. Minutiae Pairing, and
3. Matching Score Computation.

A.6.1 Registration

In order to match two point sets with unknown orientation, scale, and translation, the two point sets must be registered with respect to each other. The orientation, scale, and translation parameters are estimated using a generalized Hough Transform [8].

The input to the registration algorithm consists of two sets of minutiae points $\mathcal{P}$ and $\mathcal{Q}$ extracted from two fingerprint images. They are denoted by,

$$\mathcal{P} = \left\{(p_x^1, p_y^1, \alpha^1), ..., (p_x^P, p_y^P, \alpha^P)\right\},$$

$$\mathcal{Q} \;=\; \left\{(q_x^1, q_y^1, \beta^1), ..., (q_x^Q, q_y^Q, \beta^Q)\right\}. \tag{A.35}$$

where $|\mathcal{P}| = P$, $|\mathcal{Q}| = Q$, and (p_x^i, p_y^i, α^i) are the three features associated with the i^{th} minutiae in set $\mathcal{P}$. The basic assumption made here is that the second point set $\mathcal{Q}$ is a rotated, scaled, and translated version of the first set $\mathcal{P}$, where points may be shifted by random noise, some points may be added and some points deleted. The task of fingerprint registration is to recover this unknown transformation. Since we do not know whether the two fingerprints are the same or not (i.e., images of the same finger), we attempt to find the 'best' transformation in the sense that when applying the transformation to the minutiae points of the set $\mathcal{P}$, as many of these points as possible overlap with the minutiae points from the set $\mathcal{Q}$. Two overlapping points are considered as a match only if they have the same direction. There may be minutiae points in either set that do not match with any point in the other set.

The matching score is computed for each transformation after discretizing the set of all allowed transformations. Consider a transformation $F_{s,\theta,\Delta x,\Delta y} : R^2 \rightarrow R^2$, given by,

$$F_{s,\theta,\Delta x,\Delta y}\begin{pmatrix} x \\ y \end{pmatrix} = s\begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}, \tag{A.36}$$

where s, θ, and $(\Delta x, \Delta y)$ are the scale, rotation, and shift parameters, respectively. The space of transformations consists of quadruples $(s, \theta, \Delta x, \Delta y)$, where each parameter is discretized into a finite set of values:

$$s \in s_1, .., s_K \quad , \quad \theta \in \theta_1, ..., \theta_L,$$
$$\Delta x \in \Delta x_1, ..., \Delta x_M \quad \text{and} \quad \Delta y \in \Delta y_1, ..., \Delta y_N$$

Matching scores for the transformations are collected in the accumulator array A, where the entry $A(k, l, m, n)$ counts the evidence for the transformation $F_{s_k,\theta_l,\Delta x_m,\Delta y_n}$. The array A is filled as follows: For each pair (p, q), where $p = (p_x^i, p_y^i)$ is a point in the set $\mathcal{P}$ and $q = (q_x^j, q_y^j)$ is a point in the set $\mathcal{Q}$, find all possible transformations that map p to q and increment the evidence for these evidence for these transformations in array A. For every pair of values, (s_k, θ_l), there is exactly one shift vector $(\Delta x, \Delta y)^t$ such that $F_{s_k,\theta_l,\Delta x,\Delta y}(p) = q$, and it can be obtained as,

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = q - s_k\begin{pmatrix} cos\theta_l & sin\theta_l \\ -sin\theta_l & cos\theta_l \end{pmatrix} p. \tag{A.37}$$

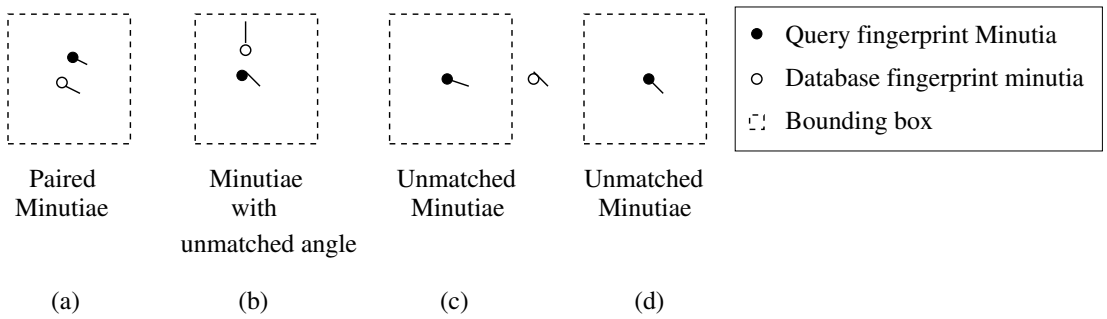


Figure A.17: Various cases in minutiae matching. (a) Matched Minutiae; (b) Minutiae with unmatched angle; (c) and (d) Unmatched minutia.

The values Δx and Δy need to be quantized to the nearest bins corresponding to Δx_m and Δy_n.

A.6.2 Minutia Pairing

The minutiae need to be paired or matched if their component or features (x, y, α) are equal within some tolerance after registration. For minutia pairing, we use a bounding box (also called tolerance box) of specific width and angle tolerance in order to accommodate the shift in the minutia angle attribute. One bounding box is drawn for every minutiae of query fingerprint image to search for it's match. We can obtain four possible cases during matching: (i) Minutiae from both fingerprints match in all components, hence denoted by paired minutiae (Fig. A.17(a)), (ii) The minutia from query fingerprint matches with minutia from database in x and y coordinate, but direction remain unmatched. These minutiae are called *minutiae with unmatched angle* (Fig. A.17(b)), (iii) No minutia from database fingerprint falls within bounding box of the minutia from query image (Fig. A.17(c)) and (iv) No minutia from database fingerprint matches with the minutia from query image (Fig. A.17(d)). In last two cases, the minutia from query image is denoted by *unmatched minutia*. In our case, bounding box size is set to 50. Angle tolerance is set to 0.1 radians.

A.6.3 Matching Score Computation

Once the count of matched minutiae, N_{MM}, is obtained, a matching score (represents the degree of match) is computed using one of the following formulas:

$$MS = \frac{\sqrt{N_{MM}}}{P * Q}, \tag{A.38}$$

$$MS = \frac{2N_{MM}}{P + Q}. \tag{A.39}$$

where P and Q are the number of minutiae features in fingerprints $\mathcal{P}$ and $\mathcal{Q}$, used for matching. We also used a normalized score, which takes into account the closeness (in the Euclidean sense) of paired minutiae features. This is computed as:

$$NMS = \frac{MS}{\frac{D}{N_{MM}+1}}. \tag{A.40}$$

where, D is the sum of all the Euclidean distances of the paired minutiae obtained after registration.

A.7 Experimental Results

The following databases are used for evaluating our fingerprint recognition system,

- DB_A consists of 10 subjects each having 8 samples (FVC 2002 Database [1]).
- DB_B consists of 10 subjects each having 8 samples (FVC 2004 Database [2]).

Table A.1: Accuracy (in Percentage Accuracy) for DB_A and DB_B.

Case	DB_A	DB_B
Case A	88.75	69.00
Case B	90.00	70.00

The recognition performance for two databases are tabulated in Table A.1. The accuracies for each database is calculated for two different cases, where the matching score is calculated using : Case A Eqn. (A.40), and Case B Eqn. (A.38). Case A provides a stricter check than case B by taking Euclidean Distance (D) under consideration. In all cases, only one sample per subject is used for training and rest of the images are used for recognition.

Bibliography

[**1**] FVC 2002 website. http://bias.csr.unibo.it/fvc2002.

[**2**] FVC 2004 website. http://bias.csr.unibo.it/fvc2004.

[**3**] ORL database. http://www.uk.research.att.com/facedatabase.html.

[**4**] Oxford English Dictionary. Oxford Edition, 2004.

[**5**] PIE feature point labels. http://www.hid.ri.cmu.edu/Hid/databases_pieLabels.html.

[**6**] C. Arcelli and G. S. D. Baja. A Width Independent Fast Thinning Algorithm. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 4(7):463–474, 1984.

[**7**] D. H. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, 3(2):110–122, 1981.

[**8**] D. H. Ballard and C. M. Brown. *Computer Vision.* Englewood Cliffs, NJ: Prentice-Hall, 1982.

[**9**] Robert J. Baron. Mechanisms of Human Facial Recognition. *International Journal of Man-Machine Studies*, 15(2):137–178, 1981.

[**10**] A.M. Bazen and S. H. Gerez. Segmentation of Fingerprint Images. In *Workshop on Circuits Systems and Signal Processing (ProRISC 2001)*, pages 276–280, 2001.

[**11**] A.M. Bazen and S. H. Gerez. Systematic Methods for the Computation of the Directional Fields and Singular Points of Fingerprint. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 24(7):905–919, 2002.

[**12**] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 19(7):711–720, Jul. 1997.

[**13**] J. Berry and D. A. Stoney. *The History and Development of Fingerprinting (H. C. Lee and R. Gaensslen eds.).* CRC Press, 2nd ed., pp. 1-40, Florida:, 2001.

[**14**] Y. Bing, J. Lianfu, and C. Ping. A new LDA-based Method for Face Recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 1, pages 168–171, Aug. 2002.

[**15**] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press Inc., New York, 1995.

[**16**] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, October 1993.

[**17**] Federal Bureau of Investigation. *The Science of Fingerprints: Classification and Uses.* U.S. Government Printing Office, Washington, D. C., 1984.

[**18**] B.Yegnanarayana. *Artificial Neural Networks.* Prentice-Hall of India, New Delhi, 1999.

[**19**] S. Cagnoni and A. Poggi. A Modified Modular Eigenspace Approach to Face Recognition. In *10th International Conference on Image Analysis and Processing (ICIAP99)*, pages 490–495, Venice, Italy, September 1999.

[**20**] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. Discriminative Common Vectors for Face Recognition. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 27(1):4–13, Jan. 2005.

[**21**] V. Chatzis, A. G. Bors, and I. Pitas. Multimodal Decision Level Fusion for Person Authentication. *IEEE Transactions on System and Man and Cybernetics*, 29(6):674–681, November 1999.

[**22**] Vassilios Chatzis, Adrian G. Bors, and Ioannis Pitas. Multimodal Decision-Level Fusion for Person Authentication. *IEEE Transactions on System and Man and Cybernetics*, 29(6):674–680, November 1999.

[**23**] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and Machine Recognition of Faces: A Survey. *Proc. of the IEEE*, 83(5):705–740, May 1995.

[**24**] K. Chen, L. Wang, and H. Chi. Methods of Combining Multiple Classifiers with Different Features and Their Application to Test-Independent Speaker Identification. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(3):417–445, 1997.

[**25**] Li-Fen Chen, Hong-Yuan Mark Liao, Ming-Tak Ko, Ja-Chen Lin, and Gwo-Jong Yu. A New LDA-based Face Recognition System which can solve the Small Sample Size Problem. *Pattern Recognition*, 33:1713–1726, 2000.

[**26**] Jen-Tzung Chien and Chai-Chen Wu. Discriminant Waveletfaces and Nearest Feature Classifier for Face Recognition. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 24(12):1644–1649, Dec. 2002.

[**27**] G. W. Cottrell and M. K. Fleming. Face Recognition using Unsupervised Feature Extraction. In *Int. J. Conf. on Neural Networks*, pages 322–325, Paris, 1990.

[**28**] D. Q. Dai and P. C. Yuen. Regularized Discriminant Analysis and Its Application to Face recognition. *Pattern Recognition*, 36:845–847, 2003.

[**29**] Y. Dai and Y. Nakano. Recognition of Facial Images with Low Resolution using a Hopfield Memory Model. *Pattern Recognition*, 31(2):159–167, 1998.

[**30**] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, CBMS series, April 1992.

[**31**] J. Daugman. Combining Multiple Biometrics. http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html.

[**32**] J. Daugman. Uncertainty Relation for Resolution in Space, Spatial-Frequency, and Orientation Optimized by Two-dimensional Visual Cortical Filters. 2:1160–1169, 1985.

[**33**] D. Demers and G. W. Cottrell. Nonlinear Dimensionality Reduction. In *Advances in Neural Information Processing Systems*, pages 580–587, 1993.

[**34**] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation. In *IC-SLP*, pages 1351–1354, Sydney, Australia, November 1998.

[**35**] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley and Sons, Singapore, 2000.

[**36**] Hazim Kemal Ekenel and Bulent Sankur. Multiresolution Face Recognition. *Image and Vision Computing*, 23:469–477, 2005.

[**37**] M. J. Er, S. Wu, and J. Lu. Face Recognition using Radial Basis Function (RBF) Neural Networks. In *38th Conference on Decision & Control*, pages 2162–2167, Phoenix, Arizona USA, 1999.

[**38**] N. J. Fliege. *Multirate Digital Signal Processing*. John Wiley and Sons, Ltd., 2004.

[**39**] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition*. Morgan Kaufmann, San Francisco, 2004.

[**40**] C. Garcia, G. Zikos, and G. Tziritas. Wavelet Packet Analysis for Face Recognition. *Image and Vision Computing*, 18:289–297, 2000.

[**41**] Nathalie George, Raymond J. Dolan, Gereon R. Fink, Gordon C. Baylis, Charlotte Russell, and Jon Driver. Contrast Polarity and Face Recognition in the Human Fusiform Gyrus. *Nature neuroscience*, 2(6):574–580, Jun. 1999.

[**42**] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore and London.

[**43**] R. C. Gonzalez and R. E. Woods. Digital image processing, 2nd edition. *Addisson Wesley*, page 229, 1999.

[**44**] S. Gutta, J. R. J. Huang, P. Jonathan, and H. Wechsler. Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces. *IEEE Trans. on Neural Networks*, 11:948–959, July 2000.

[**45**] R. Haung, Q. Liu, H. Lu, and S. Ma. Solving the Small Size Problem of LDA. In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 29–32, Aug. 2002.

[**46**] T. K. Ho, J. J. Hull, and S. N. Srihari. Decision Combination in Multiple Classifier Systems. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 16(1):66–75, Jan. 1994.

[**47**] J. Hong and X. Tan. A New Approach to Point Pattern Matching. In *Int. Conf. on Pattern Recognition (9th)*, pages 82–84, 1988.

[**48**] Lin Hong and Anil Jain. Integrating Faces and Fingerprints for Personal Identification. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 20(12):1295–1307, December 1998.

[**49**] Lin Hong, Yifei Wan, and Anil Jain. Fingerprint Image Enhancement: Algorithm and Performance Evaluation. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 20(8):777–789, August 1998.

[**50**] A. Jonathan Howell and Hilary Buxton. Invariance in Radial Basis Function Neural Networks in Human Face Classification. *Neural Processing Letters*, 2(3):26–30, 1995.

[**51**] J. Huang. *Detection Strategies for Face Recognition Using Learning and Evolution*. PhD thesis, George Mason University, May 1998.

[**52**] S.S. Iyengar, L. Prasad, and H. Min. *Advances in Distributed Sensor Technology*. Prentice Hall, 1995.

[**53**] A. K. Jain and F. Farrokhnia. Unsupervised Texture segmentation Using Gabor Filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[**54**] A. K. Jain, S. Prabhakar, and S. Pankanti. On the Similarity of Identical Twin Fingerprints. *Pattern Recognition*, 35(8):2653–2663, 2002.

[**55**] Anil Jain, Lin Hong, and Ruud Bolle. On-Line Fingerprint Verification. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 19(4):302–314, Arpil 1997.

[**56**] Nancy Kanwisher. Domain Specificity in Face Perception. *Nature neuroscience*, 3(8):759–776, Aug. 2000.

[**57**] K. Karu and A.K. Jain. Fingerprint Classification. *Pattern Recognition*, 29(3):389–404, 1996.

[**58**] M. Kawagoe and A. Tojo. Fingerprint Pattern Classification. *Pattern Recognition*, 17:295–303, 1984.

[**59**] J. Kittler, Y. P. Li, J. Mantas, and M. U. Ramos Sanchez. Combining Evidence in Multimodal Person Identity Recognition Systems. In *Proc. of Intl. Conf. on Audio and Video based Biometric Person Authentication*, pages 327–334, Crans-Montana, Switzerland, Mar. 1997.

[**60**] Josef Kittler, Mohamad Hatef, Robert P. W. Duin, and Jiri Matas. On Combining Classifiers. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[**61**] T. Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag, New York, 1988.

[**62**] Hui Kong, Eam Khwang Teoh, Jian Gang Wang, and R Venkateswarlu. Two Dimensional Fisher Discriminant Analysis: Forget About Small Sample Size Problem. In *International Conference on Acoustics Speech and Signal Processing*, volume 2, pages 761–764, March 2005.

[**63**] W. M. Koo and A. Kot. Curvature-Based Singular point Detection. In *Int. Conf. on Audio- and Video-based Biometric Person authentication (3rd)*, pages 229–234, 2001.

[**64**] Z. M. Kovacs-Vajna. A Fingerprint Verification System Based on Triangular Matching and Dynamic Time Warping. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 22:1266–1276, 2000.

[**65**] J. J. Kulikowski, S. Marcelja, and P. Bishop. Theory of Spatial Position and Spatial Frequency Relations in the Receptive Fields of Simple Cells in Visual Cortex. *Biological Cybernetics*, 43:187–198, 1982.

[**66**] A. Kumar, D. C. M. Wong, H. C. Shen, and A. K. Jain. Personal Verification using Palmprint and Hand Geometry Biometrics. In *Fourth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 669–678, Guildford, U.K., June 2003.

[**67**] A. Pavan Kumar, V. Kamakoti, and Sukhendu Das. System-on-Programmable-Chip Implementation for Online Face Recognition. *Accepted for Publication in the special issue in Pattern Recognition Letters, Jan 2006.*

[**68**] Ludmila I. Kuncheva, James C. Bezdek, and Robert P. W. Duin. Decision Templates for Multiple Classifier Fusion: An Experimental Comparison. *Pattern Recognition*, 34:299–314, 2001.

[**69**] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Von ser Malburg, R. P. Wurtz, and W. Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers*, 42:300–310, March 1997.

[**70**] L. Lam and C. Y. Suen. Optimal Combination of Pattern Classifiers. *Pattern Recognition Letters*, 16:945–954, 1995.

[**71**] 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*, 27(5):553–568, 1997.

[**72**] Andreas Lanitis, Christopher J. Taylor, and Timothy Francis Cootes. Automatic Interpretation and Coding of Face Images using Flexible Models. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[**73**] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face Recognition: A Convolutional Neural Network Approach. *IEEE Trans. on Neural Networks*, 8(1):98–112, January 1997.

[**74**] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. Face Recognition: A Convectional Neural Network Approach. *IEEE Trans. on Neural Networks*, 8(1):98–113, 1998.

[**75**] Kuang-Chih Lee, Jeffrey Ho, and David J. Kriegman. Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 27(5):684–698, May 2005.

[**76**] W. F. Leung, S. H. Leung, W. H. Lau, and A. Luk. Fingerprint Recognition using Neural Network. In *Workshop Neural Network for Signal Processing*, pages 226–235, 1991.

[**77**] Ming Li and Baozong Yuan. 2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix. *Pattern Recognition Letters*, 26:527–532, 2005.

[**78**] Shang-Hung Lin, Sun-Yuan Kung, and Long ji Lin. Face Recognition/Detection by Probabilistic Decision-based Neural Network. *IEEE Trans. on Neural Networks*, 8(1):114–131, January 1997.

[**79**] Z. Liposcak and S. Loncaric. Face Recognition from Profiles using Morphological Operations. In *International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 47–52, 1999.

[**80**] K. Liu, Y.Q. Cheng, and J.Y. Yang. Algebraic Feature Extraction for Image Recognition Based on an Optimal Discriminant Criterion. *Pattern Recognition*, 26(6):903–911, June 1993.

[**81**] D. Maio and D. Maltoni. An Efficient Approach to On-Line Fingerprint Verification. In *Int. Symp. on Artificial Intelligence (8th)*, pages 132–138, 1995.

[**82**] D. Maio and D. Maltoni. Direct Gray-scale Minutiae Detection in Fingerprints. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 19(1), 1997.

[**83**] Stephane G. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 11(7):674–693, Jul. 1989.

[**84**] R. J. Mammone. *Artificial Neural Networks for Speech and Vision*. Chapman and Hall, Cambridge, 1993.

[**85**] S. Marcelja. Mathematical Description of the Responses of Simple Cortical Cells. *Journal of Optical Society of America*, 70(11):1297–1300, 1980.

[**86**] Alex M. Martinez and Avinash C. Kak. PCA versus LDA. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 23(2):228–233, February 2001.

[**87**] B. M. Mehre. Fingerprint Image Analysis for Automatic Identification. *Machine Vision and Applications*, 6(2):124–139, 1993.

[**88**] B.M. Mehre and B. Chatterjee. Segmentation of Fingerprints Images - A Composite Approach. *Pattern Recognition*, 22(4):381–385, 1989.

[**89**] B.M. Mehre, N. N. Murthy, S. Kapoor, and B. Chatterjee. Segmentation of Fingerprints Images using the Directional Image. *Pattern Recognition*, 20(4):429–435, 1987.

[**90**] L. O'Gorman and J. V. Nickerson. Matched Filter Design for Fingerprint Image Enhancement. In *Int. Conf. on Acoustic Speech and Signal Processing*, pages 916–919, 1988.

[**91**] L. O'Gorman and J. V. Nickerson. An Approach to Fingerprint Filter Design. *Pattern Recognition*, 22(1):29–38, 1989.

[**92**] E. Osuna, R. Freund, and F. Girosi. Training Support Vector Machines: An Application to Face Detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–199, 1997.

[**93**] S. Pankanti, S. Prabhakar, and A. K. Jain. On the Individuality of Fingerprints.

IEEE Tran. on Pattern Analysis and Machine Intelligence, 24(8):1010–1025, 2002.

[**94**] Vinod Pathangay and Sukhendu Das. Exploring the use of Selective Wavelet Subbands for PCA based Face Recognition. In *Proc. National Conference on Image Processing (NCIP'05)*, pages 182–185, Bangalore, India, Mar. 2005.

[**95**] P. Penev and J. Atick. Local Feature Analysis: A General Statistical Theory for Object Representation. *Network: Computation in Neural Systems*, 7:477–500, 1996.

[**96**] Peter Hall and David Marshall, and Ralph Martin. Merging and Splitting Eigenspace Models. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 22(9):1042–1049, September 2000.

[**97**] T. Poggio and K. K. Sung. Example-based Learning for View-based Human Face Detection. *ARPA Image Understanding Workshop*, 2, November 1994.

[**98**] D. A. Pollen and S. F. Ronner. Spatial Computation Performed by Simple and Complex Cells in the Visual Cortex of the Cat. *Vision Research*, 22:101–118, 1982.

[**99**] A. Ranade and A. Rosenfeld. Point Pattern Matching by Relaxation. *Pattern Recognition*, 12(2):269–275, 1993.

[**100**] Nalini K. Ratha, Kalle Karu, Shaoyun Chen, and Anil Jain. A Real Time Matching System for Large Fingerprint Databases. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 18(8):799–813, August 1996.

[**101**] N.K. Ratha, S. Y. Chen, and A.K. Jain. Adaptive Flow orientation-based Feature Extraction in Fingerprint Images. *Pattern Recognition*, 28(11):1657–1672, 1995.

[**102**] N.K. Ratha, K. Karu, S. Chen, and A.K. Jain. A Real-Time Matching System for Large Fingerprint Databases. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 18(8):799–813, 1996.

[**103**] G. Rogova. Combining the Results of Several Neural Network Classifiers. *Neural Networks*, 7:777–781, 1994.

[**104**] A. Rosenfeld and A. Kak. *Digital Picture processing*. Academic, New York, 1976.

[**105**] A. Ross and R. Govindarajan. Feature Level Fusion Using Hand and Face Biometrics. In *SPIE Conference on Biometric Technology for Human Identification*, pages 196–204, Florida, U.S.A., March 2005.

[**106**] A. Ross and A. K. Jain. Information Fusion in Biometrics. *Pattern Recognition Letters, Special Issue on Multimodal Biometrics*, 24(13):2115–2125, 2003.

[**107**] Pankanti S., N.K. Ratha, and R.M. Bolle. Structure in Errors: A Case Study in Fingerprint Verification. In *ICPR*, pages III: 440–443, 2002.

[**108**] C. Sanderson and K. K. Paliwal. Information Fusion and Person Verification using Speech and Face Information. In *Research Paper IDIAP-PR 02-33*, September 2002.

[**109**] Bernhard Scholkopf, Alex J. Smola, and Andre Bernhardt. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319, 1998.

[**110**] B. G. Sherlock, D. M. Monro, and K. Miller. Algorithm for Enhancing Fingerprint Images. *Electronics Letters*, 28(18):1720, 1992.

[**111**] B. G. Sherlock, D. M. Monro, and K. Miller. Fingerprint Enhancement by Directional Fourier Filtering. *IEE Proceedings Vision Image and Signal Processing*, 141(2):87–94, 1994.

[**112**] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression Database. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, Dec. 2003.

[**113**] Simon Haykin. *Neural networks: A Comprehensive Foundation*. Prentice-Hall International, New Jersey, 1999.

[**114**] L. Sirovich and M. Kirby. Low-dimensional Procedure for the Characterization of Human Faces. *Journal of Optical Society of America*, 4(3):519–524, March 1987.

[**115**] Robert Snelick, Umut Uludag, Alan Mink, Michael Indovina, and Anil Jain. Large-Scale Evaluation of Multimodal Biometric Authentication Using State-

of-the-Art Systems. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 20(3):450–455, March 2005.

[**116**] J. Sprinzak and M. Werman. Affine Point Matching. *Pattern Recognition Letters*, 15:337–339, 1994.

[**117**] V. S. Srinivasan and N. N. Murthy. Detection of Singular Points in Fingerprint Images. *Pattern Recognition*, 25(2):139–153, 1992.

[**118**] R. M. Stock and C. W. Swonger. Development and Evaluation of a Reader of Fingerprint Minutiae. Technical Report CAL NO. XM-2478-X-1, Cornell Aeronautical Laboratory, 1969.

[**119**] T. J. Stonham. Practical Face Recognition and Verification with WISARD. In *Aspects of Face Processing*, pages 426–441, 1984.

[**120**] Daniel L. Swets and John (Juyang) Weng. Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 18(8):831–836, Aug. 1996.

[**121**] E. Szekély and V. Szekély. Image Recognition Problems of Fingerprint Identification. *Microprocessors and Microsystems*, 17(4):215–218, 1993.

[**122**] C. E. Thomaz, R. Q. Feitosa, and A. Veiga. Design of Radial Basis Function Network as Classifier in face Recognition using Eigenfaces. In *Vth Brazilian Symposium on Neural Networks*, pages 118–123, 1998.

[**123**] M. A. Turk and A. Pentland. Face Recognition Using Eigenfaces. In *Proc. 11th Int. Conf. Pattern Recognition*, pages 586–591, 1991.

[**124**] Matthew Turk and Alex Paul Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[**125**] D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell. Connectionist Models of Face Processing: A survey. *Pattern Recognition*, 27(9):1209–1230, 1994.

[**126**] P. Verlinde and G. Cholet. Combining Decision Fusion Paradigms using k-NN based Classifiers, Decision Trees and Logistic Regression in a Multi-modal Identity Verification Application. In *Proc. of Int. Conf. on Audio and Video based Biometric Person Authentication*, pages 188–193, Washington D. C., U. S. A., March 1999.

[**127**] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory.* Springer Verlog, Heidelberg, DE, 1995.

[**128**] Y. Wang, T. Tan, and A. K. Jain. Combining Face and Iris Biometrics for Identity Verification. In *Proc. of Int. Conf. on Audio and Video based Biometric Person Authentication*, pages 805–813, Guildford, June 2003.

[**129**] D. M. Weber. A Cost Effective Fingerprint Verification Algorithm for Commercial Applications. In *South African Symp. on Communication and Signal Processing*, 1992.

[**130**] J. Weng, N. Ahuja, and T. S. Huang. Learning Recognition Segmentation of 3-D Objects from 2-D Images. In *Int. Workshop Face Gesture Recognition*, Zurich, Switzerland, 1995.

[**131**] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.

[**132**] K. Woods, K. Bowyer, and W. P. Kegelmeyer. Combination of Multiple Classifiers using Local Accuracy Estimates. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 19(4):405–410, April 1997.

[**133**] J. D. Woodward, C. Horn, J. Gatune, and A. Thomas. *Biometrics : A look at Facial Recognition.* RAND Documented Briefing, Tech. Rep., 2003.

[**134**] Huilin Xiong, M. N. S. Swamy, and M. O. Ahmad. Two-Dimensional FLD for Face Recognition. *Pattern Recognition*, 38:1121–1124, 2005.

[**135**] L. Xu, A. Krzyzak, and C. Y. Suen. Methods for Combining Multiple Classifiers and their Applications to Handwriting Recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 22(3):418–435, 1992.

[**136**] M. H. Yang. Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition using Kernel Methods. In *IEEE International Conference on Face and Gesture Recognition*, pages 215–220, Washington, May 2002.

[**137**] Y. Yoshitomi, T. Miyaura, S. Tomito, and S. Kimura. Face Identification using Thermal Image Processing. In *IEEE International Workshop on Robot*

and Human Communication, pages 374–379, 1997.

[**138**] Hua Yu and Jie Yang. A Direct LDA Algorithm for High-Dimensional Data – with Application to Face Recognition. *Pattern Recognition*, 34:2067–2070, 2001.

[**139**] Alan L. Yuille. Deformable Templates for Face Recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.

[**140**] Bai-Ling Zhang, Haihong Zhang, and Shuzhi Sam Ge. Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory. *IEEE Tran. on Neural Networks*, 15(1):166–177, Jan. 2004.

[**141**] Jian Yang Zhang, Alejandro F. Frangi, and Jing yu Yang. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 26(1):131–137, Jan. 2004.

[**142**] T. Y. Zhang and C. Y. Suen. A Fast Parallel Algorithm for Thinning Digital Pattern. *Communications of the ACM*, 27(3):236–239, 1984.

[**143**] W. Zhao. Discriminant Component Analysis for Face Recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 2, pages 818–821, Sep. 2000.

List of Publications

- **Related to Thesis**

1. Arpita Patra and Sukhendu Das, Dual Space based Face Recognition using Feature Fusion, Accepted for Publication in *International Conference on Visual Information Engineering (VIE)*, September 2006.
2. Arpita Patra and Sukhendu Das, Enhancing Decision Combination of Face and Fingerprint by Exploitation of Individual Classifier Space: An approach to Multi-modal Biometry, Communicated to *Pattern Recognition (PR) Journal* in March 2006.
3. Arpita Patra, Vinod Pathangay and Sukhendu Das, An Efficient Method of Face Recognition using Subject-Specific Subband Faces, Communicated to *Image and Vision Computing (IVC) Journal* in April 2006.
4. Arpita Patra and Sukhendu Das, Dual Space based Face Recognition using Decision Fusion, Communicated to *Information Fusion (IF) Journal* in May 2006.
5. Arpita Patra and Sukhendu Das, A Dual Space Approach to Face Recognition using Feature Fusion, Communicated to *Computer Vision and Image Understanding (CVIU) Journal* in May 2006.

- **Miscellaneous**

1. Lalit Gupta, Vinod Pathangay, Arpita Patra, A. Dyana and Sukhendu Das, Indoor vs. Outdoor Scene Classification using Modified Probabilistic Neural Network, Accepted for Publication in *EURASIP Journal on Applied Signal Processing*, 2006.