LOCAL DISCRIMINANT EMBEDDING ALGORITHMS FOR PALM DORSA VEIN PATTERN BASED BIOMETRIC SYSTEMS USING THERMAL IMAGING

Thesis submitted in partial fulfilment of the requirements for the degree of

MASTER of ELECTRICAL ENGINEERING

By

DEBABRATA MAHIS

REGISTRATION NO. - 128907 of 2014 -15

ROLL NO. - M4ELE1615

Under the guidance of

Prof. AMITAVA CHATTERJEE

ELECTRICAL ENGINEERING DEPARTMENT

FACULTY COUNCIL OF ENGINEERING & TECHNOLOGY

JADAVPUR UNIVERSITY

KOLKATA-700032, INDIA

2016

JADAVPUR UNIVERSITY

FACULTY COUNCIL OF ENGINEERING & TECHNOLOGY

ELECTRICAL ENGINEERING DEPARTMENT

KOLKATA-700032, INDIA

CERTIFICATE of RECOMMENDATION

I hereby recommend that the thesis prepared under my supervision and guidance by **Debabrata Mahis** entitled "**LOCAL DISCRIMINANT EMBEDDING ALGORITHMS FOR PALM DORSA VEIN PATTERN BASED BIOMETRIC SYSTEMS USING THERMAL IMAGING**" be accepted in partial fulfilment of the requirements for award of the degree of "**MASTER of ELECTRICAL ENGINEERING**" at **JADAVPUR UNIVERSITY**.

The project, in my opinion, is worthy of acceptance.

SUPERVISOR

Dr. AMITAVA CHATTERJEE
(In-charge of Thesis)
Professor
Electrical Engineering Department
Jadavpur University
Kolkata – 700032, India

COUNTERSIGNED

Prof. SWAPAN KUMAR GOSWAMI
Head of the Department
Jadavpur University
Kolkata – 700032, India

Prof. SIVAJI BANDYOPADHYAY
Dean
Jadavpur University
Kolkata – 700032, India

JADAVPUR UNIVERSITY

FACULTY COUNCIL OF ENGINEERING & TECHNOLOGY

ELECTRICAL ENGINEERING DEPARTMENT

KOLKATA-700032, INDIA

CERTIFICATE of APPROVAL*

The foregoing thesis is hereby approved as a creditable study of an engineering subject carried out and presented in a satisfactory manner to warrant its acceptance as a pre-requisite to the degree for which it has been submitted. It is notified to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed and conclusion drawn therein but approve the thesis only for the purpose for which it has been submitted.

Final Examination for Evaluation of the Thesis BOARD OF EXAMINERS

_____

_____

(SIGNATURE OF EXAMINERS)

*Only in case the thesis is approved

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of his **Master of Electrical Engineering** studies.

All information in this document have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

NAME:	DEBABRATA MAHIS
REGISTRATION NO.:	128907 of 2014-15
ROLL NO.:	M4ELE1615
THESIS TITLE:	LOCAL DISCRIMINANT EMBEDDING ALGORITHMS FOR PALM DORSA VEIN PATTERN BASED BIOMETRIC SYSTEMS USING THERMAL IMAGING
DATE :	
PLACE: KOLKATA	_____
	(DEBABRATA MAHIS)

Acknowledgements

The privilege to pursue post graduate studies at the Electrical Engineering Department, Jadavpur University, has been of great pleasure for me. I feel contented to be a part of such a highly acclaimed and esteemed academic institution.

I hereby express my deepest and most profound gratitude towards my supervisor, Dr. Amitava Chatterjee, who has been not only a great mentor but also a lodestar to me in the field of technological research. The continuum of his thorough guidance and constructive criticism has exceedingly helped me throughout the tenure of the project work. I convey my sincere thanks to him for mentoring me about structured project planning and research methodologies. I feel immensely fortunate to have had the opportunity to work under his supervision on a state of the art research field. Finally, I thank him from the bottom of my heart not only for granting me the opportunity to work under him but also for reinstating, in my mind, the importance of liberty in work and freedom of thought for the successful realization of a true research idea.

I do express my heartfelt admiration and convey sincere thanks to the faculty members of the Electrical Engineering Department, Jadavpur University, especially, the faculty members of the Electrical Measurement and Instrumentation Section for teaching me elaborately and extensively the subjects in my course work and even addressing the smallest of my queries and doubts. The exceptional benevolence shown by Professor Sugata Munshi, Professor Mita Dutta, Professor Palash Kr. Kundu, Professor Biswajit Bhattacharyya, Professor Gautam Sarkar, and Professor Debangshu Dey for guiding and motivating me, throughout the tenure of my coursework, is praiseworthy.

I would like to express my sincere gratitude to Professor (Dr.) Swapan Kumar

Goswami, Head of the Department, Electrical Engineering Department, Jadavpur University, for his continued co-operation and allowing me to access freely and use the rich infrastructure of the department.

Also I thank the Laboratory-in-Charge and all other members of the Electrical Measurement and Instrumentation Laboratory, Electrical Engineering Department, Jadavpur University, for guiding me in the process of operating the various apparatuses present in the laboratory.

I am also grateful to my seniors Sandip Joardar, Sayan Banerjee, Saptarshi Chatterjee, Sayanti Chaudhuri, and friends Subhabrata Naskar, Pitar Mondal, Joy Banerjee, Shramana Guha Sarkar, Nabanita Chatterjee and juniors who, along with me, have collaboratively developed an ambient atmosphere of mutually helpful and progressive research environment in the Electrical Measurement and Instrumentation Laboratory, Electrical Engineering Department, Jadavpur University.

Finally, it is my utmost pleasure to have great family who have always supported, guided, and helped me in innumerable occasions even in the hours of need, in all my endeavours, for which I will always keep on respecting and adoring them.

DATE:

PLACE: Kolkata _____

(DEBABRATA MAHIS)

CONTENTS

LIST OF ACRONYMS

ACRONYM	FULL-FORM	PAGE
PDVP	Palm Dorsa Vein Pattern	2
FAR	False Acceptance Rate	3
FRR	False Rejection Rate	3
FNMR	False Non-Match Rate	3
ROC	Receiver Operating Characteristic	4
PCA	Principal Component Analysis	5
LDA	Linear Discriminant Analysis	5
FARO	Face Recognition against Occlusion	5
FACE	Face Analysis about Commercial Entities	5
DTCWT	Dual Tree Complex Wavelet Transform	7
CRC	Collaborative Representation Based Classification	8
FKP	Finger-Knuckle-Print	9
ICA	Independent Component Analysis	10
DTW	Dynamic Time Warping	11
IR	Infrered	12
EM	Electromagnetic	12
NIR	Near Infrared	12
MIR	Middle Infrared	12
FIR	Far Infrared	12
EIR	Extreme Infrared	12
LDE	Local Discriminant Embedding	15
KLDE	Kernel Local Discriminant Embedding	15
SOLDE	Stable Orthogonal Local Discriminant Embedding	15
FLDE	Fuzzy Local Discriminant Embedding	15
LPP	Locality Preserving Projection	17
NPE	Neighbourhood Preserving Embedding	17
MFA	Marginal Fisher Analysis	17
LLE	Locally Linear Embedding	21

PDSVP	Palm Dorsa Subcutaneous Vein Pattern	30
LE	Laplacian Eigenmap	30
MDS	Multi-Dimensional Scaling	31
LFDA	Local Fisher Discriminant Analysis	31
CLAHE	Contrast Limited Adaptive Histogram Equalization	46
AHE	Adaptive Histogram Equalization	46
GHE	Global Histogram Equalization	46
RMS	Root Mean Square	47
NLSE	Normalized Least Square Error	47
PSNR	Peak Signal to Noise Ratio	47
FKNN	Fuzzy K-Nearest Neighbourhood	57

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This chapter provides a general introduction about entire thesis. In this chapter, information about various biometrics are given.

CHAPTER 1

Introduction

1.1. Introduction:

The term "biometrics" is derived from the Greek words "bio" (means life) and "metrics" (means to measure) [80]. Biometrics refers to metrics related to human characteristics. Biometric authentication is used in computer science as a form of identification and process control. Biometric identifiers are distinctive, measurable characteristics used to label and describe individuals. The fear of cyber security give rise to biometric security systems. And now a days Palm Dorsa Vein Pattern (PDVP) has become a very useful biometric for biometric security system.

1.2. Biometric Security Systems:

Biometric security system is security process in which various human biometric traits are used to identify any individual's tangible appearance for a security purpose. Biometric security system is an enhanced, trust worthy, convenient security system for protecting someone's important documents. Military, Intelligence, Law Enforcement agencies, and that which are highly prone to identity theft have been implementing biometric verification and identification system for the last few decades for great security. In biometric security systems an individual's biometric characteristics are previously stored and that can be only accessed by only that authorized personnel. Security system based on anyone's knowledge, data can be very easy to replicate so can be easily spoofed. So security system based on biometric is very promising as it is tough to spoof. Biometric traits can be broadly classified to two categories.

1.2.1. Mode of Operation of a Biometric System [1, 80, 81]:

Biometric system operates in the following way:

1.2.1.1. Verification Mode [1, 81]:

In this mode of operation, biometric of an individual is compared with the pre-stored biometric of that individual's for confirmation. In this mode any individual first claim his identity to be authenticated before the system. This is a one-to-one comparison for confirmation of authentication about any individual's identity.

1.2.1.2. Identification Mode [1, 81]:

In this mode of operation, biometric of an individual is compared with the pre-stored biometric of all individual. So unlike the previous mode this is a one-to-many comparison for confirmation of authentication about any individual's identity.

1.2.2. Characteristics of an Ideal Biometric [1, 80, 81]:

An ideal biometric must have the following characteristics and properties:

i. **Universality [1]:** Universality of an ideal biometric refers that the features used by system, should be present universally in all concerned subjects.

ii. **Uniqueness [1]**: This refers to the particular biometric features must be unique to a person and by this feature one can distinguish two subjects.

iii. **Permanence [1]**: This property of a biometric feature refers to the property of the feature that it must be time invariant i.e. it must not change or decay over a long period of time.

iv. **Collectability [1]**: This refers to the ease of acquisition of biometric feature considered, can be acquired in a form suitable and usable by the biometric system.

v. **Performance [1, 80, 81]**: A biometric system must have satisfactory recognition accuracy and has different performance metrics which are described as follows:

1. **False Acceptance Rate (FAR) [80]**: It can be defined as the probability that the biometric system incorrectly matches the input pattern to a non-matching template in the database. In other words, it measures the per cent of invalid inputs which are incorrectly accepted. It is also known as False Match Rate (FMR).
2. **False Rejection Rate (FRR) [80]**: It can be represented as the probability that the biometric system incorrectly discard the input pattern, matching to a template in the database i.e. it incorrectly rejects the input data which match the template in database. It is also known as False Non-Match Rate (FNMR). As the sensitivity of the biometric device increases the FAR decreases but the FRR increase. In practical applications, the FAR should be very low to provide high enough confidence and the FRR should be sufficiently low. If the threshold set in the decision stage is reduced, it is expected that less False Non-Matches but more False Accepts would occur. On the contrary, setting a higher value for the

threshold corresponds to a smaller FAR but a larger FRR. Therefore, the threshold of the Biometric System should be carefully chosen.

3. **Receiver Operating Characteristic (ROC) Curve [80]**: The ROC curve is used for illustrating the relationship between FAR and FRR. It is a visual characterization of the trade-off between FAR and FRR.

vi. **Acceptability [1]:** The biometric system should be user-friendly i.e. it must be acceptable to wide range of classes.

vii. **Circumvention [1]:** This refers to that system must be robust to various imposter method i.e. the 'Proof of Life' must be present.

1.3. Classification of Biometrics:

Biometric traits are classified into two categories. And these Physiological biometric traits and Behavioural biometric traits are given in tabulated form as follows in table number 1.1.:

Physiological Biometric Traits	Behavioural Biometric Traits
i. Face	i. Voice
ii. Iris	ii. Signature
iii. Fingerprint	iii. Handwriting
iv. Retina	iv. Keystroke
v. Vein pattern	v. Gait
vi. Palm-print	
vii. Ear geometry	
viii. Hand shape geometry	
ix. DNA	
x. Knuckle	
xi. Heart sound	
xii. Body shape geometry	
xiii. Lips	

A brief description of the above mentioned biometric traits is given as follows:

1.3.1. Face Biometric [2-5]:

Face biometric is a very promising biometric for identification of any individual without bothering that individual for test sample, from the pre-stored image database unlike fingerprint, iris etc. biometric. A facial recognition system is capable of identifying a person from a digital

image or a video frame from a video source by comparing the test image with image database preserved previously. Among linear methods, namely principal component analysis (PCA) and linear discriminant analysis (LDA) are two classic method for face based biometric authentication. These two perform the identification problem in a reduced dimensional space rather in an image space. To increase robustness i.e. face recognition against occlusion and expression variations, FARO method is used for identification of individual based on face biometric [2]. Improved interval type-2 fuzzy logic Sugeno integral and modular neural networks method can also be used for face biometric based biometric authentication [3]. The face recognition by using hyperspectral image in the visible light bands acquire images with inherent spectral information of skin, a new proposed approach for biometric authentication [4]. Paper [5] describes about face analysis about commercial entities (FACE) which perform significantly when image acquisition system is not good i.e. for poor image quality. Though face biometric provide good recognition accuracy there are various problem also for face biometric like angle of face problem (Good recognition accuracy for frontal face and 20 degrees off and rest have problems), low illumination, occlusion, expression change, sunglass, hair etc. Some facial biometric traits are shown below:

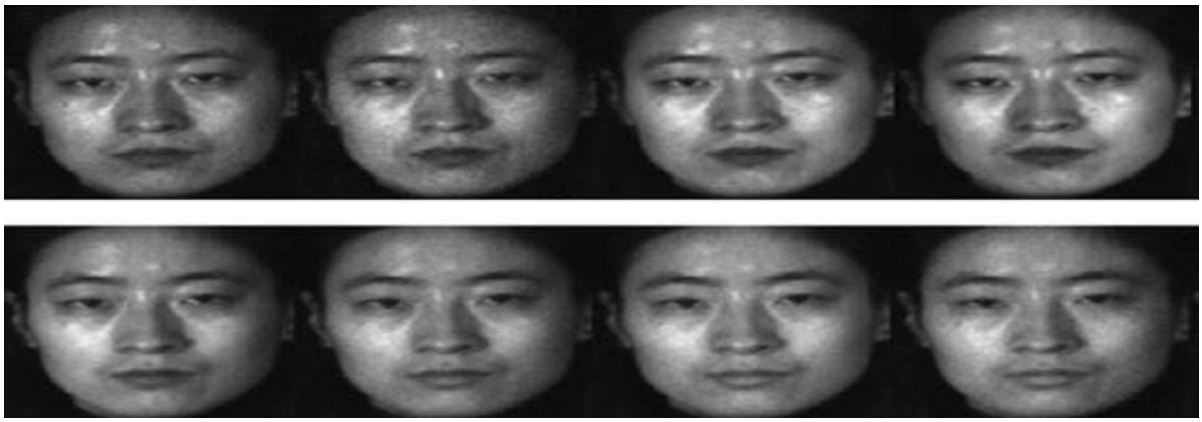


Figure 1.1. Various Face Biometric [4]

1.3.2. Iris Biometric [6-8]:

It is an internal part of body and very useful for identification or verification of an individual. Due to its uniqueness for every person (even for genetically identical persons have separate iris structure) and can be captured even from 10 centimetres - few metre distance away make it very useful biometric trait.

As the iris must be within focus volume of acquisition system and positioning of head is necessary i.e. cooperation of individual is very much required in data acquisition system.

Extended depth-of-field iris recognition system using unrestored wave front-coded imaginary is needed for image acquisition [6]. Poor image quality is great hindrance in iris based biometric authentication. Extended iris quality can be accessed by fusion of various quality factor such as defocus blur, illumination, iris resolution etc. [7] Some iris biometric are shown in figure1.2.

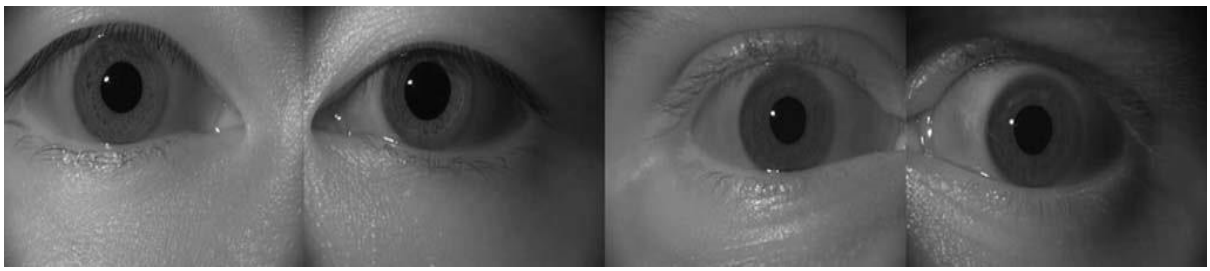


Figure 1.2. Various Iris Biometric [7]

In iris biometric security system adjustment of scanner, distance of scanner, non-co-operation of any individual, and poor image quality make the recognition rate poor.

1.3.3. Fingerprint Biometric [9-11]:

It is one of the most useful biometric for biometric authentication. As it is distinctive for every individual and easy getting of this biometric make it very convenient for biometric security system and even for forensic science also. Previously so called 'ink technique' is used for fingerprint but fingerprint scanner is now used for data acquisition. From local graph of pre-processed fingerprint global graph of finger print is formed and stored in a database and then the distance between the global graphs of the presented and the stored fingerprint images is calculated and based on a threshold and by this biometric authentication can be performed [10]. Fingerprint can be divided into various sub-region for the direction pattern for recognition accuracy measurement [11]. According to Henry system of classification three basic types of fingerprints are loop, whorl and arch. Classification can be more complex like arch may be plain or tented arches. Various types of fingerprint patterns are as follows:

Figure 1.3. Various Fingerprint Biometric [10]

1.3.4. Retina Biometric [12, 13]:

Now a days, retina biometric has become commercially popular. It is utilized by several government agencies including NASA, FBI etc. Even genetically same twins do not have identical retina biometric, and this uniqueness make it useful biometric trait. Retina based biometric authentication generally implicate by vessel based matching considering feature points. But Vessel segmentation and feature extraction is a time consuming process. So by extracting illuminance, contrast and structural features from a colour retina image and then by using an optimized function, combining the extracted feature, one can generate a similarity score between two images. And the identified person have highest similarity score [12]. Some of retina biometrics are as follows:

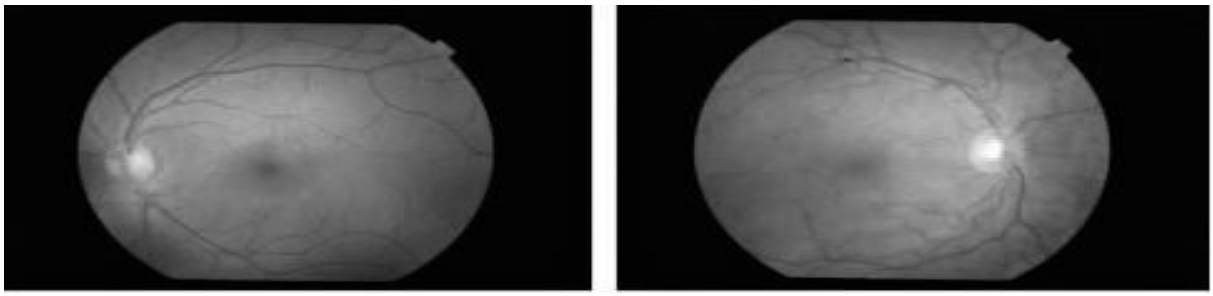


Figure 1.4. Various Retina Biometric [12]

1.3.5. Palm Print [16, 17]:

Palm print biometric is a popular and promising biometric because of its ease of acquisition, higher acceptance and reliability. The palm print of two different people is always separate and complexity of palm print make it very useful biometric trait. Palm print feature generally consists of three main features like palm length, palm width and palm area. Paper [16] give a description about the identification of an individual by using palm-print using dual tree complex wavelet transform (DTCWT). Three dimensional palm print is more robust and reliable than two dimensional palm print. Palm print is more robust to scars and dirt than fingerprint due to its long length [17]. Some palm print biometric is represented below:

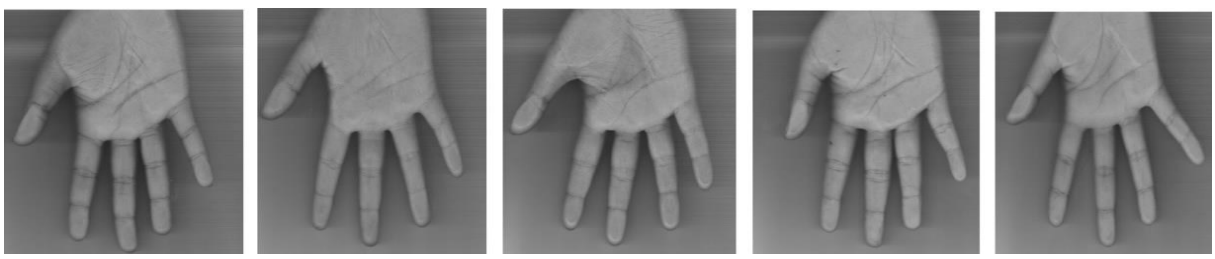


Figure 1.5. Various Palm-Print Biometric [16]

1.3.6. Ear geometry [18, 19]:

Ear biometric authentication is increasing biometric trait now a days. Unchanged shape with ageing, unchanged stable structure with change of facial expression, uniqueness of pattern (separate even for twins) make it a promising biometric trait for biometric authentication [19]. Paper [18] discuss about recognition based on ear biometric considering 2D images treating ear as planner surface and stored by using a homograph transform. Some ear biometrics are shown below:

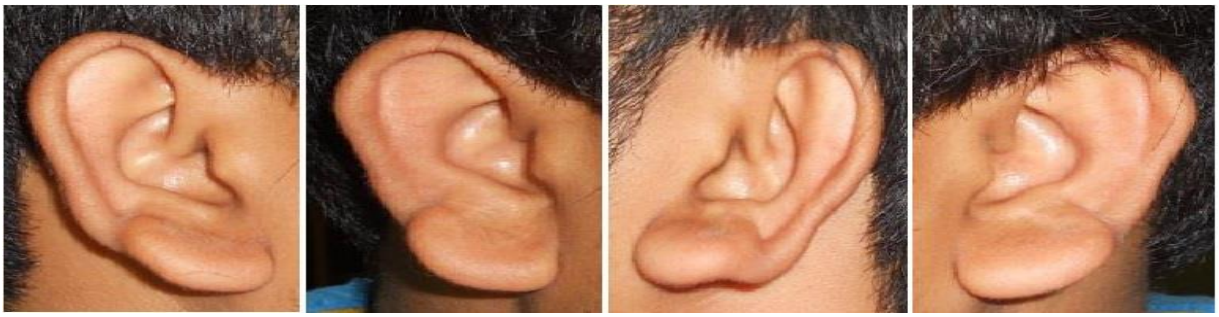


Figure 1.6. Various Ear Biometric [19]

1.3.7. Hand shape geometry [20]:

This biometric trait is less prone to theft. Unlike others it has non-invasive and low cost data acquisition, uniqueness between individual persons make it useful biometric for biometric security system. Biometric identification based on hand shape geometry by using collaborative representation based classification (CRC) is described in [20]. Some of hand shape geometry are as follows:

Figure 1.7. Various Hand Shape Biometric [20]

1.3.8. Knuckle [21, 22]:

Finger knuckle surface is considered as one of the emerging potential biometric traits for personnel authentication. This is due to its stable and unique inherent patterns present in the

outer surface of the finger back knuckle region. And finger knuckle has a high potentiality towards discriminating individuals with high accuracy. Finger knuckle print imaging technique can be used for personnel authentication [21]. Acquisition of knuckle in real time and based on this real time data biometric authentication can be done by using Finger–Knuckle-Print (FKP) algorithm [22]. Some of knuckle biometrics are shown below:

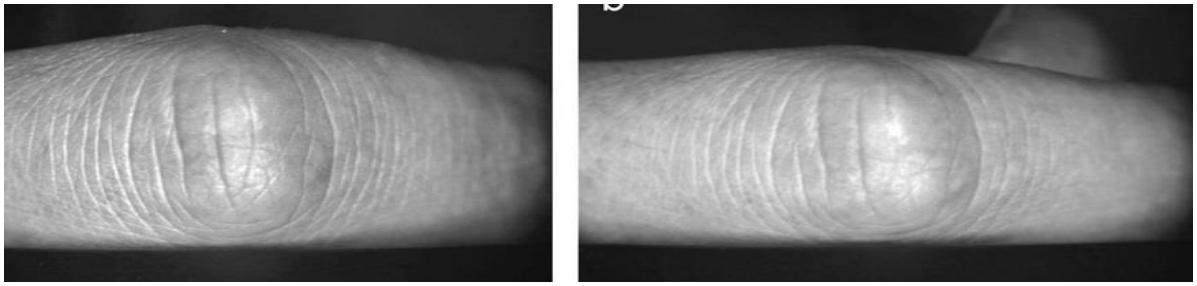


Figure 1.8. Various Knuckle Biometric [22]

1.3.9. DNA [23]:

Using DNA sequence encryption of any data makes it fully secured as it is very tough for spoofing. So DNA characteristics based biometric security system is very promising approach. By using the special characteristics of DNA a digital code of watermark for biometric security system is healthy approach, discussed in [23].

1.3.10. Heart sound [24-26]:

Heart sound is generally used for determining the heart condition. But now a days, heart sound is used as a biometric for identification of any individual. The signal produced by heart is natural signal, produced by sudden closure of artio-ventricular and semilunar valve, and impossible to recreate artificially, so difficult to deceit. Generally heart sound consists of three major sounds, first heart sound (S1), second heart sound (S2) and Murmur.

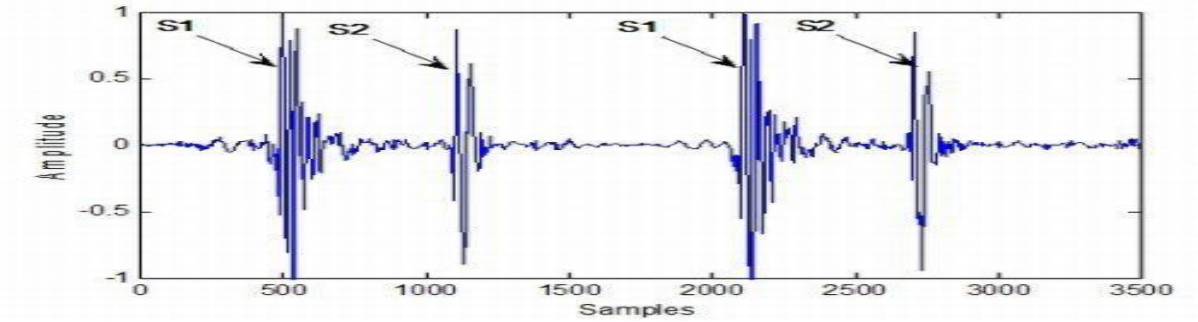


Figure 1.9. Typical PCG Signal with the Waveforms of S1 and S2 Sounds [26]

The feature of heart sound can be extracted by using Wavelet Packet Decomposition [25].

1.3.11. Body shape geometry [27]:

Body shape of several person are not identical i.e. different person have different body shape information. Its uniqueness makes it a promising biometric. Body shape can be divided into seventy three different points and several distance vector (generally four) like wrist to elbow, elbow to shoulder etc. are made from different poses and by these characteristics human identification is done [27].

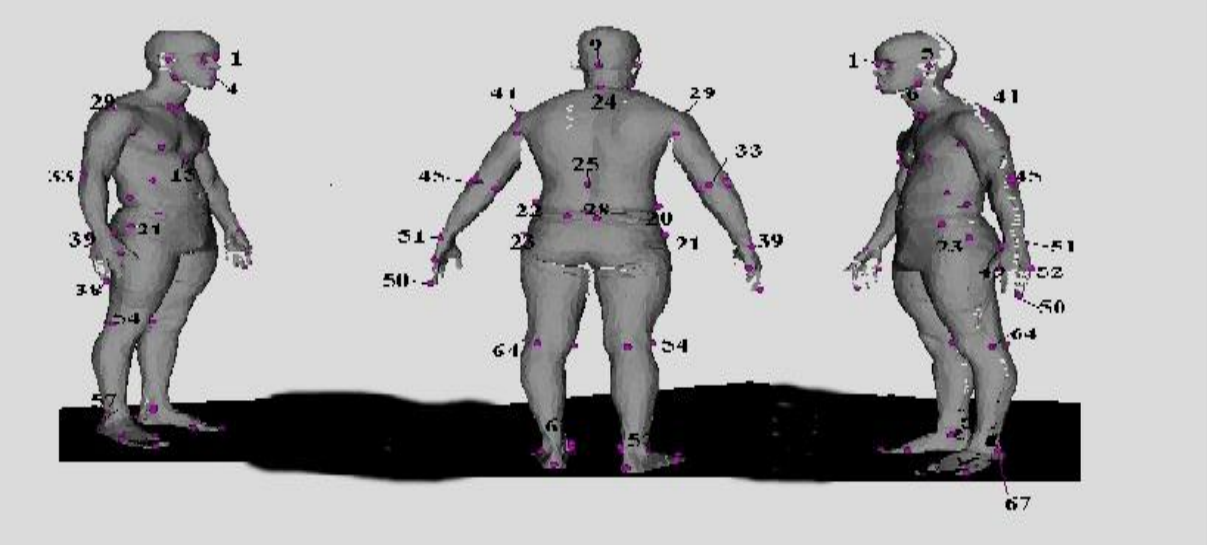


Figure 1.10. Body Shape Biometric [27]

1.3.12. Lips [28-30]:

Lip texture of different individual is exclusively different. Variety of using lip of speaker used for biometric identification. Since the original lip features are usually of high-dimension, the independent component analysis (ICA) used for dimension reduction and biometric authentication is proposed in [27]. Speaker identification using lip feature by PCA is proposed in [28].

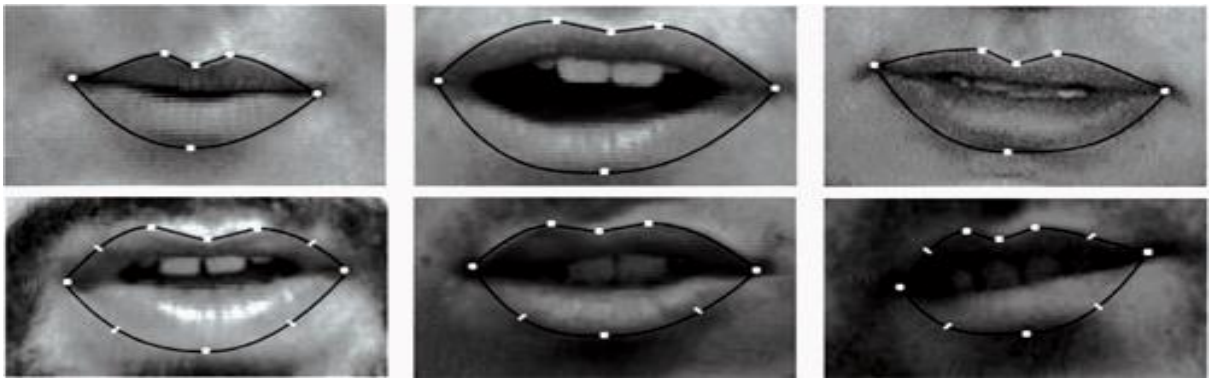


Figure 1.11. Various Lip Biometric [30]

1.3.13. Voice [31]: Voice biometrics or speaker authentication is a method to verify the person's identification using that person's voice. It involves checking the uttered speech, whether it belongs to the claimed speaker or not. An I-vector based speaker authentication can be done by minimizing the intra-speaker variability and maximizing inter-class variability. This can be done by mapping the I-vector of intra-speaker to same point and I-vector of inter-speaker to a large distance [31].

1.3.14. Signature [32, 33]:

Signature is a potential biometric for biometric authentication. A protected on-line signature based biometric authentication system where the original signature template are protected by transforming them in a non-invertible way and the transformed templates are compared employing a Dynamic Time Warping (DTW) matching strategy [32]. Signature can also be acquired online and can be used for identification [33].

1.3.15. Handwriting [34]:

Handwriting is useful biometric trait and can be successfully used for authentication of person. The uniqueness and ease of data acquisition make it a promising biometric. The main problem in handwriting based identification is intra-class variability and inter-class similarity. So a stable hash value can be generated by using fuzzy data of biometric characteristics [34].

1.3.16. Keystroke [35]:

Keystroke is not so widely used biometric till some years before due to its low accuracy compare to other biometric trait. As data acquisition is cheap and non-invasive, keystroke biometric is now in use for biometric authentication.

1.3.17. Gait [36, 37]:

Gait means manner of walking of a person. Generally gait of any person should be different from other person. It is a very auspicious biometric for security surveillance or authentication because it can be acquired from a distance and without contact and without the knowledge of subject. It is very tough to spoofing this biometric. These all advantages make gait as a good biometric for authentication as well as security system. For identification purpose gait biometric can be acquired also by wireless accelerometer sensor in three orthogonal directions like up–down, forward–backward, and sideways [36].

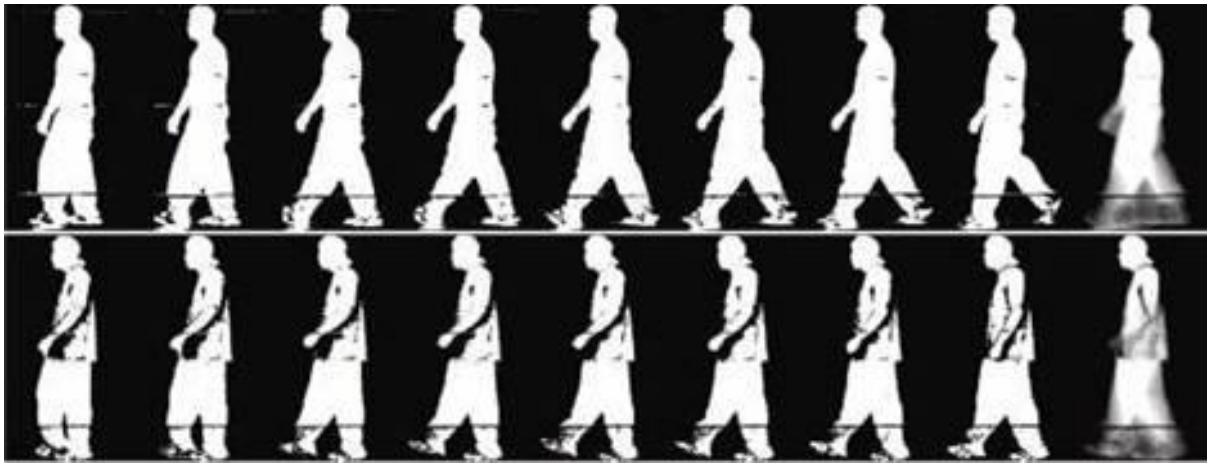


Figure 1.12. Various Gait Biometric [45]

1.4. Thermal Imaging Based Biometric Trait [38-44]:

Thermal imaging based biometric security system is a dramatically increased new trend in biometric system and commercial and industrial applications. Thermal imaging of any subject is called thermography. Thermography is generally detection of radiation as a function of temperature (in long infra-red range) from any subject and creation an image from that radiation. According to black body radiation law infrared (IR) radiation by any object is above absolute zero temperature. And the radiation is directly proportional to the temperature of an object. So thermography camera does this detection and makes thermograms from that detected radiation. Thermal camera detects the difference of temperature even if it is very little. Thermal images are normally grayscale in nature: black objects are cold, white objects are hot and the depth of grey indicates variations between the two. Some thermal cameras, add colour to images to help users to identify objects at different temperatures. In the entire Electromagnetic (EM) Spectrum, infra-red refers to a specific region with a wavelength range [0.75, 1000] μm, which can be further subdivided into four sub-bands: Near-IR (NIR) [0.75, 2] μm, Middle-IR (MIR) [2, 6] μm, Far-IR (FIR) [6, 14] μm, and Extreme-IR (EIR) [14, 1000] μm [20]. Thermal imaging based biometric system has some significant advantages like [21] –

i. It does not depend on any external light source i.e. it does not depend upon the illumination condition.

ii. It is a non-contact and a non-destructive test method, used from a safe distance.

iii. The results are available in real-time and there is little or no processing needed.

iv. Image acquisition procedure is easy.

Though there are many advantages it has some shortcomings like ambient temperature of surroundings and humidity has an effect on thermal image even perspiration of human hand. One can face difficulty in locating the original position of required biometric. While acquiring thermal image these factors affect the quality of acquired image using IR technology. Despite this shortcomings thermal image based biometric security system is very promising, popular and useful biometric authentication. Some of thermal image based biometric [38] are discussed in brief as follows:

1.4.1. Thermal imaging of face biometric [39-42]:

For biometric security application thermal image based on face is promising growing interest due to its robustness in nature. Recognition by thermal image of face is not affected by illumination condition of surroundings and this makes thermal image based biometric authentication a promising approach. A crucial step for biometric authentication based on thermal image of face is segmentation as it does not pinpoint only face like simple face detection rather the shape of face. Reference [27] present a method for face segmentation in IR images based on the Sobel Edge detector and morphological operations while [28] describes a method for face segmentation using a Bayesian Approach. Some thermal images of face biometric are as follows:

Figure 1.13. Thermal Image of Face Biometric [42]

1.4.2. Thermal imaging of Vein Pattern Biometric [43, 44]:

Uniqueness and robustness of palm dorsa vein pattern make it a convenient biometric for biometric security application. The thermal radiation emitted from the Dorsum of the human hand is used for FIR imaging the Palm Dorsum [43]. From extensive medical research it is known that veins have higher temperature than surrounding tissues [44].The acquisition of

thermal image is contactless and non-destructive type. The universality and toughness to replicate for spoofing, make vein pattern a suitable biometric for identification. Vein can't be acquired by general digital camera. So by thermal camera it can be acquired. Interested region can be extricate by Gabor filter and identification can be done using LDA, box and branch point based method.

1.4.2.1. FIR Imaging of Vein Pattern Biometric [15, 80]: FIR imaging uses the IR radiation emitted by an object for image acquisition of the object [80]. Therefore, FIR imaging of the Palm Dorsum can be obtained by using the thermal radiation emitted from the dorsum of human hand. In case of FIR imaging external source of light is not required, so variation of illumination has no effect in FIR imaging. FIR imaging technique is useful for capturing the large vein pattern on the back side of hand i.e. on dorsum of hand. FIR imaging of vein pattern is quite good approach for biometric authentication though there are some factors which affect the FIR imaging of vein pattern like ambient temperature, low contrast of image, and exact location of vein etc. [80]. Some FIR image of vein pattern are represented as follows:

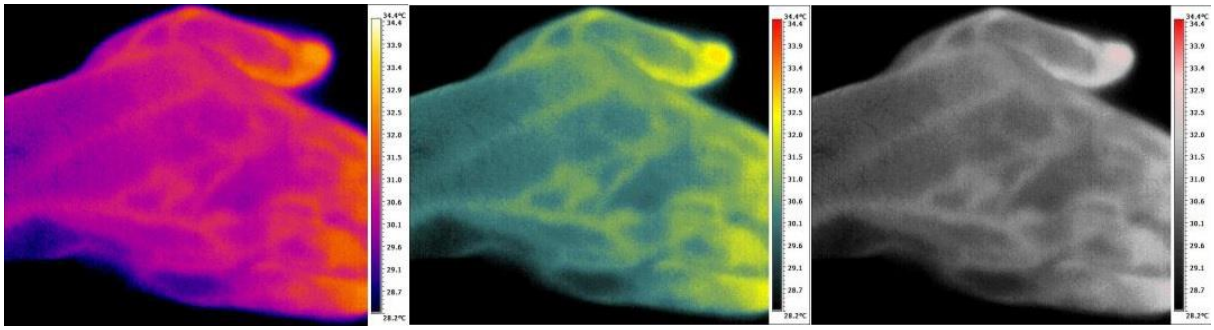


Figure 1.14. Some FIR Image of Vein Biometric [80]

1.4.2.2. NIR Imaging of Vein Pattern Biometric [15, 80, 82]: When the NIR light is incident on the Palm Dorsum the reduced haemoglobin in the blood flowing through the veins absorb more incident radiation energy than that by the surrounding tissues [80, 82]**.** The NIR light penetrates to a depth of about 3mm into the subcutaneous tissue [80, 82]. NIR imaging of vein pattern is robust to external environment and medical condition of object. NIR imaging can capture very small vein pattern even in palm and wrist also. The skin colour is not a factor affecting the NIR imaging of vein pattern. In spite of these above mentioned advantages of NIR imaging of vein pattern, it has some shortcomings like any visible marks on skin can be mistaken as vein, and it requires an external source of infrared light beam on desired body part. Some NIR image of vein pattern are represented as follows:

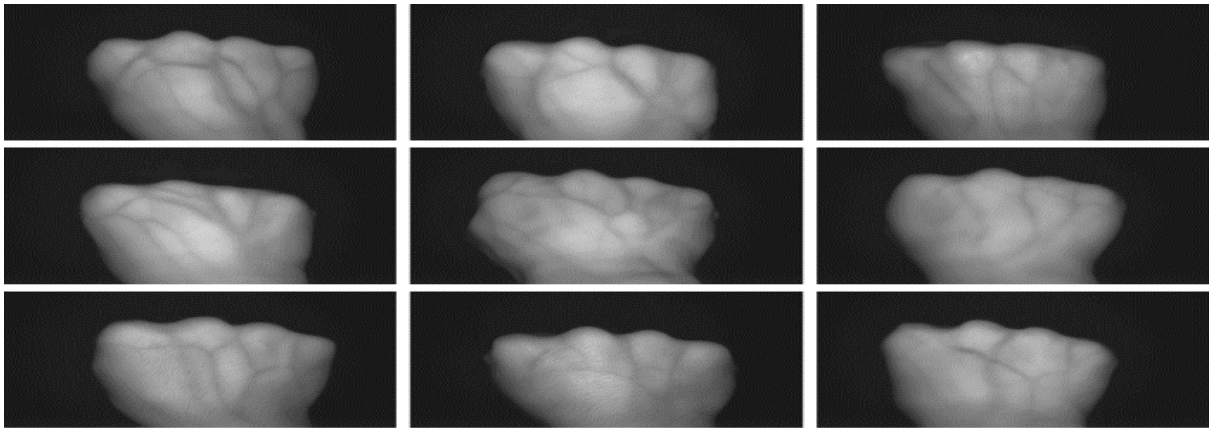


Figure 1.15. NIR Image of Vein Biometric [46]

1.5. Brief of the Thesis:

This thesis has effectively demonstrated how several state-of-the-art variants of local discriminant embedding (LDE) algorithms can be effectively utilized for dimensionality reduction and subsequent biometric authentication and recognition systems, developed using thermal imaging of palm dorsa vein patterns (PDVP). To be specific, the thesis has concentrated in developing such useful solutions utilizing two contemporary improved variants of LDE called Stable Orthogonal Local Discriminant Embedding (SOLDE) algorithm and Fuzzy Local Discriminant Embedding (FLDE) algorithm.

1.6. Summary:

Chapter 1 has described about various biometric (physical and behavioural) and their corresponding features, advantages and disadvantages. And also a brief review about the thermal image based biometric for biometric security system is given. Feature of an ideal biometric and mode of operation of a biometric security system is also discussed.

Chapter 2 will describe about various dimension reduction method like Linear Discriminant Analysis (LDA) and Local Discriminant Embedding (LDE) and variants of LDE called Kernel Local Discriminant Embedding (KLDE), Stable Orthogonal Local Discriminant Embedding (SOLDE) algorithm and Fuzzy Local Discriminant Embedding (FLDE) algorithm.

CHAPTER 2

Linear Discriminant Analysis and Local Discriminant Embedding

This chapter provides a brief introduction to various dimension reduction method like Linear Discriminant Analysis and Local Discriminant Embedding. In this chapter, variants of Local Discriminant Embedding is also described.

CHAPTER 2

Linear Discriminant Analysis AND Local Discriminant Embedding

2.1. Introduction:

In machine learning scenario, while dealing with high dimensional data, problems like requirement of high storage space and involvement of high computational burden arise often. To overcome these problems, one has to often employ methodologies that will result in reduction in the dimension of the data i.e. one must map the data from high dimensional data space to low dimensional data space. These reduction methods are broadly classified into two categories: supervised methods and unsupervised methods [70]. There are many supervised methods like Linear Discriminant Analysis (LDA) [70], Locality Preserving Projection (LPP) [67], Neighbourhood Preserving Embedding (NPE), Marginal Fisher Analysis (MFA) [57] etc. Among these methods, linear discriminant analysis is a popular method that is extensively used for dimension reduction. In LDA [70-72] method, generally the distance between the data pairs belonging to different classes is maximized so that the data pairs from different classes are mapped far apart in the resultant low dimension data space. At the same time, the distance between the data pairs belonging to the same class is minimized so that these data pairs are mapped close to each other in the resultant low dimensional space.

Mapping the input data from high dimensional space to low dimensional space is an almost mandatory, unavoidable step for biometric pattern recognition. It often happen that the Euclidean distances between data points are not capable of capturing the inherent similarities between data points. As a result traditional subspace method is no longer a trustworthy method for data distribution in low dimensional data space. One can overcome this above mentioned problem by using Local Discriminant Embedding (LDE), used for manifold learning and pattern classification. The neighbourhood relations and class relations of data are used to construct the embedding for classification problems in LDE [73]. In LDE [73-75] algorithm, the embedding for the sub manifold of each class is achieved by solving an optimization problem. After being embedded into a low-dimensional subspace, data points from the same class maintain their inherent neighbourhood relations and neighbouring points of different

classes are no longer close to each other. So with the help of local discriminant embedding, new test data can thus be more reliably classified, by employing the nearest neighbour rule, considering the local inherent structure of data. So linear discriminant embedding gives good dimension reduction approach as it consider the inherent geometry of data along with class relations.

2.2. Linear Discriminant Analysis [70-72]:

Linear discriminant analysis (LDA), a generalization of Fisher's linear discriminant [59] used in pattern recognition, finds the linear combination of features that separates objects belonging to different classes. LDA [71] is closely related to analysis of variance and it also expresses any feature as a linear combination of other features or measurements. In modern image processing techniques, most often it is required to work with images acquired from state –of-art cameras, which capture high dimensional data. In order to work with these acquired images the dimension of image must be reduced for subsequent reduction of computational burden and computational time, associated with the execution of the algorithms. LDA [71] essentially computes the difference between the data of different classes. LDA is almost similar to PCA. But there are some differences between PCA and LDA.

2.2.1. Differences between PCA and LDA [71]:

Both PCA and LDA perform dimensionality reduction for biometric recognition. The differences between PCA and LDA are as follows:

- **i.** The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification [71].
- **ii.** In PCA, the shape and location of the original data sets changes when data are mapped from high dimensional space to low dimensional space, whereas, LDA does not change the location but only tries to provide more class separability and draw a decision region between the given classes [71].

2.2.2. Different Approaches for LDA [71]:

Data sets are transformed from high dimensional data space to low dimensional data space and test vectors are classified in transformed low dimensional data space. The classification of test

vectors in transformed space can be achieved using two different approaches and those are as follows:

i. **Class-Dependent Transformation [71]:** In class-dependent transformation approach for classification, the ratio of interclass variance to intra-class variance is maximized. The main objective of this approach is to maximize this ratio so that proper separability of classes can be secured. The class-specific type approach involves using two optimizing criteria for transforming the data sets independently from input data space to transformed data space.

ii. **Class-Independent Transformation [71]:** In class-independent transformation approach for classification the ratio of overall variance to intra-class variance is maximized. This type of transformation approach uses only one optimizing criteria for transforming the input data sets to transformed data sets. So in this approach all data are transformed by using only one optimizing criteria irrespective of their classes. This type of LDA considers that each and every class is different from other classes.

2.2.3. Computation of Linear Discriminant Analysis [70, 71]:

Computational operation of classical linear discriminant analysis is carried out as follows:

Let us assume that there are n number of training data $\{x_1, x_2, \dots, x_n\} \in R^P$ in a data space where $\{x_i\}_{i=1}^{n}$ represent the labelled data. X can be partitioned into C disjoint classes. LDA computes a linear projection that maps the input data x_i to a low dimensional data y_i. The linear transform performed is as follows:

$$y_i = W^T x_i \tag{2.1}$$

In LDA for transformation matrix computation, three scatter matrices which are within class scatter matrix, between class scatter matrix, and total scatter matrix have to be computed. Within class scatter matrix can be computed in the following way [70, 71]:

$$S_w = \frac{1}{n}\sum_{i=1}^{C}\sum_{x_i \in C_i}(x_i - m_i)(x_i - m_i)^T = \sum_{i=1}^{C} S_{w_i} \tag{2.2}$$

Where S_{w_i} is the covariance matrix of class C_i and m_i is the mean of vectors. This m_i can be computed as follows [70, 71]:

$$m_i = \frac{1}{n_i}\sum_{i=1}^{n_i} x_i \tag{2.3}$$

Where n_i denotes the number of samples in class i. Between classes scatter matrix can be computed in the following way [70, 71]:

$$S_b = \sum_{i=1}^{C} \frac{n_i}{n}(m_i - m)(m_i - m)^T = \sum_{i=1}^{C} S_{w_i} \tag{2.4}$$

Where m represents mean of all vectors and n_i denotes the number of samples in class i.

Total scatter matrix can be calculated as follows [70, 71]:

$$S_t = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)(x_i - m)^T \tag{2.5}$$

From above definition it can be seen that $S_t = S_w + S_b$. If data are centred by subtracting their mean, between class scatter matrix and total scatter matrix can be constructed as follows [70, 71]:

$$S_b = \sum_{i=1}^{c} \frac{n_i}{n} m_i {m_i}^T \tag{2.6}$$

$$S_t = \frac{1}{n}\sum_{i=1}^{n} x_i {x_i}^T \tag{2.7}$$

Then by using the above mentioned equations from (2.2) to (2.7) the transformation matrix can be computed. The transformation matrix can be computed from the equation given below [70]:

$$W = \arg\max_{w} \frac{tr\{W^T S_b W\}}{tr\{W^T S_t W\}} \tag{2.8}$$

Then by using this transformation matrix W, equation (2.1) transforms the input data from high dimensional data space to low dimensional data space. Transforming the entire data set from

high dimensional space to low dimension data space, proper discriminant boundaries classify the data.

2.3. Local Discriminant Embedding (LDE) [56, 69]:

LDE uses relations between data pairs of local neighbourhood for embedding the input data from high dimensional data space to low dimensional data space. Various dimension reduction methods like LLE [54, 55], ISOMAP [50, 51] mainly consider local or global property of training data for data preservation. Because it remains a difficult issue to map new test data to the low dimensional space, their algorithms cannot be easily extended for classification problems. Also these methods do not promise good discriminating capability. LDE has good discriminating ability and it is directly applicable for use of new test data. After the mapping of input data by using LDE, data belonging the to same class in neighbourhood are mapped very close to each other in transformed data space, whereas, data belonging to different classes are mapped distant from each other.

2.3.1. Local discriminant Embedding Algorithm [56, 69]:

In this approach, neighbour and class relations of data are used for constructing the embedding for classification problems. LDE [56] does the mapping from high dimensional space to low dimensional space. LDE [56] uses locality to get intra-class compactness. Suppose there are n number of training samples $\{x_1, x_2, ..., x_n\} \in \Re^p$ from C classes. These training samples are used to determine a linear transformation matrix A such that

$$y_i = A^T x_i, i = 1,2,..., n \tag{2.9}$$

The LDE algorithm can be detailed in the following steps:

2.3.1.1. Construct the Neighbourhood Graphs [56, 69]: Suppose G and G' are two adjacency graphs to be constructed. Adjacency graphs G and G' are constructed with the help of training data matrix X to find the local inherent geometric structure of training data, where $X = \{x_1, x_2, ..., x_n\}$ is a vertex set and it has n number of data points. To construct the adjacency graph G, put an edge between nodes i and j, and the data points must belong to the same class i.e. x_i is among the k nearest neighbours of x_j and both x_i and x_j belong to the same class.

To construct adjacency graph $G^{'}$ put an edge between nodes i and j, and x_j is among the k nearest neighbours of x_i and these two belong to different classes.

2.3.1.2. Construct the Weight Matrices [56, 69]: Suppose there are the training data matrix $X = \{x_1, x_2, ..., x_n\}$ with n number of data points. These data belong to C classes. The weight matrix W of G, where each element W_{ij} refers to the weight of the edge between x_i and x_j can be computed in the following way:

$$W_{ij} = \begin{cases} K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i = c_j \\ 0, & \text{otherwise} \end{cases} \quad (2.10)$$

Where $K(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / p\right)$, p is a suitable parameter, and $\|x_i - x_j\|$ denotes the l_2-norm of the difference between x_i and x_j, and $N_k(x_i)$ stands for the k nearest neighbours of x_i and c_i is the class labels of x_i. In a similar manner, one can compute the weight matrix $W^{'}$ of $G^{'}$, where each element $W_{ij}^{'}$ refers to the weight of the edge between x_i and x_j, belong to different classes. $W_{ij}^{'}$ Can be computed in the following way:

$$W_{ij}^{'} = \begin{cases} 1 - K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i \neq c_j \\ 0, & \text{otherwise} \end{cases} \quad (2.11)$$

2.3.1.3. Embedding [56, 69]: From generalized eigen equation, the eigenvectors corresponding to the largest eigenvalues are computed. These eigenvectors make the transformation matrix $A = [a_1, a_2, ..., a_n]$ for embedding. The generalized eigen equation will be as following:

$$X(D^{'} - W^{'})X^T a = \lambda X(D - W)X^T a \quad (2.12)$$

Where D and $D^{'}$, the diagonal matrices, with diagonal elements are as follows:

$$D = \sum_j W_{ij} \quad (2.13)$$

$$D^{'} = \sum_j W_{ij}^{'} \quad (2.14)$$

Then transformation of input data from high dimensional data space to low dimensional data space is accomplished by using equation (2.9). Considering the weight matrices W and $W^{'}$ the optimized objective function can be represented as follows:

$$\text{Maximize } J(V) = \max \sum_{ij} \left\| A^T x_i - A^T x_j \right\|^2 W_{ij}^{'} \tag{2.15}$$

$$\text{Subject to } \sum_{ij} \left\| A^T x_i - A^T x_j \right\|^2 W_{ij} = 1 \tag{2.16}$$

2.3.2 Variants of Local Discriminant Embedding [56, 69, 77]:

In recent times, several variants of the basic LDE method have been proposed by researchers all over the world e.g. Kernel Linear Discriminant Embedding (KLDE) [56], Stable Orthogonal Linear Discriminant Embedding (SOLDE) [69], and Fuzzy Linear Discriminant Embedding (FLDE) [76-77] etc. These are described in brief below.

2.3.2.1. Kernel Local Discriminant Embedding (KLDE) [56]: The classification power of linear learnings algorithm is generally limited and often found inadequate for real data based pattern recognition problems. One possible way to overcome these problems is to transform the input data to a high dimensional data space. KLDE is the best possible method for this purpose. Kernel representation of LDE is represented in the following way.

Suppose there are n number of training samples $\{x_1, x_2, ..., x_n\} \in \Re^p$ from C classes, mapped to a high dimensional data space F via a nonlinear mapping $\phi : \Re^p \to F$. For simplification of the discussion, assume that only one projection direction a in F for local discriminant embedding is computed, where a can be expressed in terms of a combination of mapped data i.e. $a = \sum_{i=1}^{n} \alpha_i \phi(x_i)$ and is therefore determined by the expansion coefficients α_i. Optimization problem (2.15) and (2.16) can be kernelized as following:

$$\max U(a) = a^T K(D^{'} - W^{'}) K a \tag{2.17}$$

$$\text{Subject to } a^T K(D - W) K a = 1 \tag{2.18}$$

Where K is a kernel matrix with $K_{ij} = K(x_i, x_j)$. $A = [a_1, a_2, ..., a_n]$ consists of the expansion coefficients. It follows that the corresponding generalized eigenvalue problem to obtain a of (2.17) and (2.18) is formulated as:

$$K(D' - W')Ka = \lambda K(D - W)Ka \tag{2.19}$$

Where the generalized eigenvector associated with the largest eigenvalue.

2.3.2.2. Stable Orthogonal Local Discriminant Embedding (SOLDE) [56, 69]: Stable orthogonal linear discriminant embedding (SOLDE) is an advanced version of basic LDE that uses both similarity and diversity of same class data pairs for construction of discriminant objective function. The SOLDE algorithm [69] works as follows:

SOLDE tries to find a set of orthogonal basis vectors $a_1, a_2, ..., a_n$ i.e. $\langle a_i, a_j \rangle = 0$ if $i \neq j$ otherwise $\langle a_i, a_j \rangle = 1$, where $\langle , \rangle$ denotes the inner product of two vectors. So the objective function of orthogonal algorithm can be written as follows:

$$a_1 = \arg\max_a \frac{a^T X L_m X^T a}{a^T X L_d X^T a} \tag{2.20}$$

And

$$a_k = \arg\max_a \frac{a^T X L_m X^T a}{a^T X L_d X^T a} \tag{2.21}$$

So that $a_k^T a_1 = a_k^T a_2 = ... = a_k^T a_{k-1} = 0$.

Here $L_d = \alpha L_s - (1-\alpha)L_v$, where α is a suitable constant and the value of α is assign to $0.5 \leq \alpha \leq 1$, $L_s = D^s - S$ and $L_v = D^v - V$, and $D_{ii}^s = \sum_j S_{ij}$, and $D_{ii}^v = \sum_j V_{ij}$.

The solutions of the above objective functions [68] i.e. optimal orthogonal basis vectors of (2.20) and (2.21) can be easily solved. So the algorithm for obtaining the orthogonal basis vectors and mapping of data from high dimensional data space to low dimensional space is as follows:

i. PCA Projection [47]: The input data x_i is mapped into a PCA subspace such that the denominator matrix in equation (2.21) is non-singular. The conversion matrix to PCA subspace is denoted as W_{PCA}.

ii. Adjacency Graph construction [56, 69]: Three adjacency graphs $G_{g-s} = \{X, S\}$, $G_{g-v} = \{X, V\}$ and $G_{g-d} = \{X, F\}$ are constructed with a vertex set X i.e. the training matrix. In this training matrix there are n number of nodes and three weight matrices namely similarity matrix S, diversity matrix V, and discriminant matrix F which are constructed respectively. The computation of Similarity, diversity, and discriminant weight matrices are described in detail in chapter 3.

iii. Orthogonal Projection Vectors Calculation [69]: Let $[a_1, a_2, ..., a_k]$ be the orthogonal projection vectors. One can denote $A^{k-1} = [a_1, a_2, ..., a_{k-1}]$ and $B^{k-1} = \left(A^{k-1}\right)^T \left(XL_d X^T\right)^{-1} A^{k-1}$. Then the orthogonal basis vectors can be calculated in the following way [69]:

1. At first, the eigenvector a_1 i.e. eigenvector corresponding to maximum eigenvalue of $\left(XL_d X^T\right)^{-1} XL_m X^T$ is computed.
2. Follow the same procedure of computing eigenvectors corresponding to maximum eigenvalue for calculating k number of orthogonal basis vectors of M^k. M^k can be represented as follows:

$$M^k = \left[I - \left(XL_d X^T\right)^{-1} A^{k-1} \left\{B^{(k-1)}\right\}^{-1} \left\{A^{(k-1)}\right\}^T\right] \left(XL_d X^T\right)^{-1} XL_m X^T \qquad (2.22)$$

3. Orthogonal Embedding [69]: The projection matrix is denoted as $A_{o/p} = [a_1, a_2, ..., a_p]$ and mapping of data points from high dimensional data space to low dimensional data space is as follows:

$$y = W^T x, \qquad \text{Where } W = W_{PCA} A_{o/p}$$

Where y is the data after mapping and W is the transformation matrix.

2.3.2.3. Fuzzy Local Discriminant Embedding (FLDE) [76, 77]: The fuzzy local discriminant embedding (FLDE) algorithm has been recently proposed, in which the fuzzy k-nearest neighbour (FKNN) [76] is implemented to reduce the detrimental effects (like variation in illumination, expression etc.) to obtain good performance. A membership degree matrix is firstly calculated using FKNN, then the membership degree is incorporated into the definition of the laplacian scatter matrix to obtain the fuzzy laplacian scatter matrix. A brief description

about FLDE is given here. FLDE is described elaborately in chapter 4. Fuzzy LDE algorithm is as follows:

Let $G = \{X, W\}$ be an undirected graph with vertex set $\{x_1, x_2, ..., x_n\} \in \Re^p$ that belongs to C classes and W gives a similarity matrix. The low dimensional data points are $Y = \{y_1, y_2, ..., y_n\}$. A fuzzy objective function for good discriminant approach is formulated as follows:

$$\min \frac{1}{2} \sum_{ij} \left(y_i - y_j \right)^2 U_{ij} W_{ij} \tag{2.23}$$

One can denote the projection matrix as $A = [a_1, a_2, ..., a_k]$ and mapping of data points from high dimensional data space to low dimensional data space is as follows:

$$Y = A^T X \tag{2.24}$$

The objective function will be as follows:

$$\begin{aligned}
&\frac{1}{2} \sum_{ij} (y_i - y_j)^2 U_{ij} W_{ij} \\
&= \frac{1}{2} \sum_{ij} \left(a^T x_i - a^T x_j \right)^2 U_{ij} W_{ij} \\
&= \sum_{ij} a^T x_i U_{ij} W_{ij} x_i^T a - \sum_{ij} a^T x_i U_{ij} W_{ij} x_j^T a \\
&= \left(a^T X D^F X^T a - a^T X W^F X^T a \right) \\
&= a^T X L_F X^T a
\end{aligned} \tag{2.25}$$

The fuzzy diagonal matrix D^F and the fuzzy laplacian matrix L_F can be expressed as follows:

$$L_F = D_F - W_F \text{ And } D_{ii}^F = \sum_{j \neq i} U_{ij} W_{ij}, \forall i \tag{2.26}$$

i. Fuzzy Neighbourhood Graph Construction [76, 77]: Let G_{fuzzy} and G'_{fuzzy} be two graphs constructed on training data. They are constructed on same training data matrix X . For construction of G_{fuzzy} neighbourhood data belonging to the same class are considered. Similarly, for construction of G'_{fuzzy} neighbourhood data belonging to different class are considered.

ii. Intra-class Compactness Fuzzy Graph and Separability Fuzzy Graph Construction [77]: From fuzzy neighbourhood graph G_{fuzzy} intra-class compactness fuzzy graph is computed as follows:

$$U_{ij}^{G} = \begin{cases} 0.51 + 0.49(n_{ij}/k), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i = c_j \\ 0.49(n_{ij}/k), & \text{otherwise} \end{cases} \quad (2.27)$$

Here n_{ij} stands for the number of the neighbours of the j[th] data (pattern) that belong to the i[th] class and k is the size of the neighbourhood to be considered. In a similar way, seperability fuzzy graph is also computed from G'_{fuzzy} as follows:

$$U_{ij}^{G'} = \begin{cases} 0.51 + 0.49(n_{ij}/k), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i \neq c_j \\ 0.49(n_{ij}/k), & \text{otherwise} \end{cases} \quad (2.28)$$

iii. Fuzzy Laplacian Scatter Matrices Computation [77]: Fuzzy affinity matrix W_{fuzzy}^{G}, the fuzzy diagonal matrix D_{fuzzy}^{G} and the fuzzy laplacian matrix L_{fuzzy}^{G} of the graph G_{fuzzy} are computed as follows:

$$W_{fuzzy}^{G} = U_{ij}^{G} .* W_{ij} \quad (2.29)$$

$$L_{fuzzy}^{G} = D_{fuzzy}^{G} - W_{fuzzy}^{G} \quad (2.30)$$

$$D_{fuzzy}^{G} = \sum_{j \neq i} U_{ij}^{G} W_{ij} \quad (2.31)$$

In a similar way, the fuzzy affinity matrix $W_{fuzzy}^{G'}$, the fuzzy diagonal matrix $D_{fuzzy}^{G'}$ and the fuzzy laplacian matrix $L_{fuzzy}^{G'}$ of the graph G'_{fuzzy} are computed.

iv. Embedding [77]: Solving the following general eigen equation, the eigenvectors corresponding to the largest eigenvalues of the transformation matrix are obtained.

$$X\left(D_{fuzzy}^{G'} - G_{fuzzy}^{G'}\right)X^{T}a = \lambda X\left(D_{fuzzy}^{G} - G_{fuzzy}^{G}\right)X^{T}a \quad (2.32)$$

Then the desired embedding is accomplished by equation (2.24).

In addition to these variants of LDE, other variants like Two-dimensional LDE can also be used as a dimension reduction technique.

2.4 Summery:

In this chapter a brief discussion about linear and nonlinear dimensional reduction methods as well as dimension enlargement method (KLDE) is discussed. Linear discriminant analysis and local discriminant embedding are illustrated in brief in this chapter.

Various local discriminant embedding like kernel linear discriminant embedding, stable orthogonal linear discriminant embedding, and fuzzy linear discriminant embedding is discussed and their computational procedures are also presented in this chapter.

Chapter 3 will describe about SOLDE algorithm and vein pattern based biometric authentication using SOLDE.

CHAPTER 3

Stable Orthogonal Local Discriminant Embedding for Palm Dorsa Based Vein Pattern Biometric System

The chapter presents a variant of LDE algorithm called Stable Orthogonal Local Discriminant Embedding algorithm. In this chapter Stable Orthogonal Local Discriminant Embedding algorithm is used for palm dorsa vein pattern based biometric system.

CHAPTER 3

Stable Orthogonal Local Discriminant Embedding for Palm Dorsa Vein Pattern Based Biometric System

3.1 Introduction:

PALM dorsa subcutaneous vein pattern (PDSVP) is comparatively new approached physiological feature as compared to other physiological features that have been used immensely for biometric applications, such as face biometric [2-5], ear biometric [18, 19], finger-print biometric [9-11], palm-print biometric [16,17], iris biometric [6-8], hand shape biometric [20], and gait [36, 37] biometric. PDSVP [14, 15, 43, 44] potentially has all the characteristics that can potentially be useful to build an effective recognition system. Data protection and security is a major concern in today's digital era. Because of presence of large digital data and cheap hardware methods based on keys or password can be easily defraud or stolen. Depending upon the user known password or password token is not good as it can be easily spoofed. As PDSVP sprawl beneath the human skin, and can only be transfigure through surgical intervention, PDSVP has large immunity factor to counterfeit of imposter. PDSVP is the sign of life. Therefore these remarkable advantageous facet of PDSVP makes it a prospective feature for any biometric system.

Now a days, reduction of dimension technique [49, 55] on the ground of manifold learning [60] is trending interest in pattern recognition technique. In classical linear dimensionality reduction techniques, like principal component analysis (PCA) [47, 48] and linear discriminant analysis (LDA) [49] only see the global Euclidean structure of data whereas manifold learning process focus on finding a projected data, that well preserve the inherent pattern of data. Some of manifold learning based algorithms are ISOMAP [50, 51], Laplacian Eigenmaps (LE) [52, 53], and Locally Linear Embedding (LLE) [54, 55].

ISOMAP [50, 51], a nonlinear dimensionality reduction method is the combination of the Floyd-Warshall algorithm with classic Multidimensional Scaling. By using straight line Euclidean distance measurement method it measures the intrinsic geometry of the local neighbourhoods and geodesic distance for the inherent geometry of nonlocal

neighbourhoods. It defines the geodesic distance to be the sum of edge weights along the shortest path between two nodes and then it stores all the pairwise distances in a low dimensional space. Laplacian Eigenmaps [52, 53] also does dimensionality reduction using spectral techniques. LE creates an undirected graph from input data set using each data as a node and computes the connectivity by considering the proximity of node i.e. k-nearest neighbourhood. In low dimensional data space it indicates the pairwise data of local neighbourhood. LLE [54, 55] was presented almost at the same time as ISOMAP but LLE is generally believed to be a better dimensionality reduction method than ISOMAP including faster optimization when implemented to take advantage of sparse matrix algorithms. LLE seeks a lower-dimensional projection of the data which preserves the distances within local neighbourhoods. It can be thought of as a series of local Principal component analyses which are globally compared to find the best nonlinear embedding.

There are various manifold learning based discriminant methods. And by those methods one can store the inherent geometry of neighbourhoods and also can get the discriminant structures of data. Some of those discriminant approaches are Local Discriminant Embedding (LDE) [56], Marginal Fisher Analysis (MFA) [57], Multi-dimensional Scaling (MDS) [58], and Local Fisher Discriminant Analysis (LFDA) [59]. All these methods strive to achieve very similar objectives. These approaches essentially store the local geometric structures, which portray intra-class compactness. This is done by minimizing the sum of the distances of local neighbourhoods of same class. By this approach neighbourhoods' data from the same class can be mapped to a single data point in a reduced space. So these discriminant approaches consider only the similarity in same class but not the diversity in the same class of local neighbourhoods. Figure 3.1. [69] shows the methodology of conserving the data of same class.

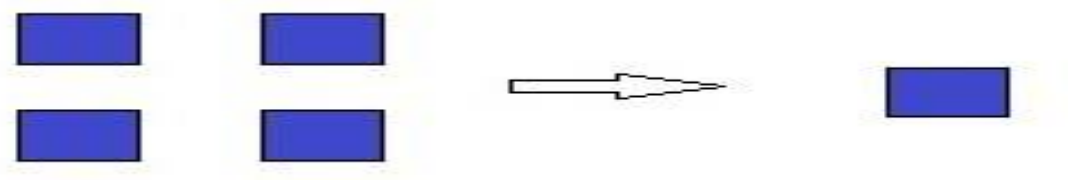


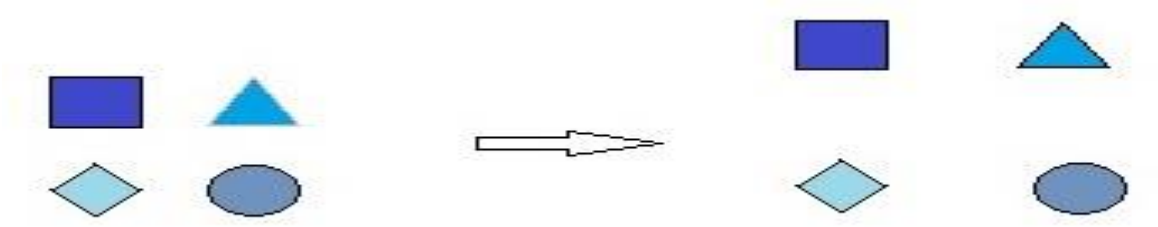
Figure – 3.1. Same Class Data Conservation [69]

So here one can see that the intra-class variation of local neighbourhoods is totally discarded.

The diversity of data i.e. non-identical feature of data which is also significant information can be used to obtain intra-class diversity. These diversity of data can be stored by maximizing the

variance [47] of data. In practice inherent structures of data are unknown and complex and testing data and training data are different from each other. So only the similarity of data or the diversity of data is not adequate to determine the discriminant feature and, hence, the resultant recognition accuracy will be also poor.

The discriminant approaches also take into account the interclass diversity of data by maximizing the Euclidean distance between the data of different classes of local neighbourhoods. Figure 3.2. [69] shows the methodology of conserving the data of different classes.



3.2. Different Class Data Conservation [69]

In the literature, the optimal projection directions are usually the eigenvectors of generalized Eigen-equations and do not satisfy mutual orthogonality [61, 62], though orthogonal projection directions are more efficient in storing the data from high dimensional space to low dimensional space and to make better the discriminant feature [61-64]. However any of these discriminant methodologies do not use the similarity of intra-class data and diversity of intra-class data together to get the projection directions i.e. to get the total recognition accuracy [69].

To make a more comprehensive consideration of this situation, recently a state-of-the-art improved variant of LDE has been proposed for reduction of dimension called stable orthogonal linear discriminant embedding (SOLDE) algorithm [69]. The speciality of SOLDE is that it takes into account the intra-class similarity and intra-class diversity of local neighbourhoods. In this methodology two adjacency graphs to display the inherent geometry are constructed and using those one can obtain the intra-class similarity and diversity and also one can obtain the discriminating feature. By utilizing this state-of-art machine learning algorithm one can incorporate similarity, diversity, and discriminating feature in the objective function of dimension reduction problem, and get the recognition rate to verify the efficacy of SOLDE algorithm. The original SOLDE algorithm was proposed for face recognition purposes. The present thesis shows how this modified version of the LDE algorithm can be

effectively employed for real life thermal imaging of PDVP based biometric authentication/recognition problems

The rest of the chapter is constructed in the following manner.

I. In section 3.2., LDE is introduced and it discusses the issues in discriminant methodology based on manifold learning.
II. In section 3.3., the objective functions are presented and the orthogonal algorithm is detailed to acquire the projection directions.
III. In section 3.4., a theoretical study in support of the SOLDE algorithm presented in [69] is detailed.
IV. In section 3.5., experimental results reporting the recognition accuracy performance is given.
V. In section 3.6., concluding remarks about SOLDE methodology are presented.

3.2. Brief Review of LDE and Related Problem Statements:

3.2.1. Local Discriminant Embedding [56, 69]:

In this approach neighbour and class relations of data are used for constructing the embedding for classification problems. As mentioned before in chapter 2, LDE [56] does the mapping from high dimensional space to low dimensional space. LDE [56] uses locality to obtain intra-class compactness. Let us assume that there are n number of training samples $\{x_1, x_2, ..., x_n\} \in \Re^p$ from C classes. These training samples are used to determine a linear transformation matrix A such that

$$y_i = A^T x_i, i = 1,2,..., n \tag{3.1}$$

Where y_i denotes the low dimensional representation of training sample x_i. The aim of LDE is to keep close the samples from same classes of neighbourhood, whereas prevent to keep close the data sample from different classes of neighbourhood. The intra-class compactness of training samples can be obtained by optimizing the following objective function

$$\min \sum_{ij} (y_i - y_j)^2 W_{ij} \tag{3.2}$$

The elements W_{ij} in weight matrix W can be defined as follows:

$$W_{ij} = \begin{cases} K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i = c_j \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

Where, as given in chapter 2, $K(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / p\right)$, and p is a suitable parameter, and $\|x_i - x_j\|$ denotes the l_2-norm of a vector, difference between x_i and x_j, and $N_k(x_i)$ and $N_k(x_j)$ stand for the k nearest neighbours of x_i and x_j respectively and c_i is the class labels of x_i. Here W_{ij} are the elements of W.

From equation 3.1, it can be easily seen that:

i. It impairs the local topology of data [65, 69]: By using equation (3.2) the inherent geometry of local data is preserved i.e. if the distance between two data is smaller, those data are stored together when they are mapped from high dimensional space to low dimensional space. In equation (3.2), if the distance between the data is higher it gets large weightage and those data can be stored very close to each other when they are mapped from high dimensional space to low dimensional space. So it can be seen that the main objective of projection i.e. storing intra-class data of local neighbourhoods very close to each other is violated [65, 69].

ii. It also degrades the generalization ability [65, 66]: In objective function equation (3.2) it can be seen that it maps the close points from same class in high dimensional space to low dimensional space irrespective of the similarity or diversity of data from same class. So only considering the similarity among the data, and ignoring the diversity among the data, data from same class are mapped from high dimensional space to low dimensional space [69]. But diversity among the data from same class has significant weightage during mapping. Generally the inherent structure of data are unknown and complex in real world application and training data and testing data are different from each other. So considering similarity only gives poor recognition accuracy. So SOLDE proposes to take in account both similarity and diversity of data [66, 69].

iii. It does not give proper inherent structure of data [65, 66, 69]: As similarity between data are taken only and diversity between the data are ignored, intrinsic structure of data are not properly unfold.

3.3. Stable Orthogonal Linear Discriminant Embedding [69]:

Although a brief introduction of the SOLDE algorithm was presented in chapter 2, in this section an elaborate description about SOLDE algorithm is given. By this algorithm one can detect the discriminant structure and considering both the similarity and diversity of data, data is stored in low dimensional space from high dimensional space.

3.3.1. Similarity [69]: Suppose there are n number of training data $\{x_1, x_2, ..., x_n\} \in \Re^p$ in a data space. Now an adjacency graph $G_{g-s} = \{X, S\}$ is constructed with the help of training data matrix X to find the local inherent geometric structure of training data. Here $X = \{x_1, x_2, ..., x_n\}$ is a vertex set. Similarity weight matrix S characterize the similarity among the data points of same class of local neighbourhoods.

With the help of LPP [67], the elements S_{ij} of S matrix can be defined as follows:

$$S_{ij} = \begin{cases} K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i = c_j \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

3.3.2. Diversity [69]: With the help of training data $X = \{x_1, x_2, ..., x_n\} \in \Re^p$ in a data space, an adjacency graph $G_{g-v} = \{X, V\}$ is constructed to find the local inherent geometric structure of training data where $X = \{x_1, x_2, ..., x_n\}$ is a vertex set and it has n number of data points. Diversity weight matrix V characterizes the diversity among the data points of same class of local neighbourhoods.

The elements V_{ij} of V matrix can be defined as follows:

$$V_{ij} = \begin{cases} 1 - K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i = c_j \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

So from equations (3.4) and (3.5) it can be seen that if some local points lie on a compact region in low data space region then these data points have high similarity and negligible diversity. If some local points lie on a sparse region in low data space region then these data points have high diversity and negligible similarity.

Suppose y_i is the mapped data of x_i from high dimensional space to low dimensional space. Now considering the similarity and diversity of data points, the two objective functions as follows have to be optimized [69]:

$$\min \sum_{ij} (y_i - y_j)^2 S_{ij} \tag{3.6}$$

$$\max \sum_{ij} (y_i - y_j)^2 V_{ij} \tag{3.7}$$

The objective function (3.6) shows that smaller the distance between the nearby points, data pairs are placed close to each other after mapping [69]. But this objective function gives more weightage to the data pairs whose distance between them is large. So mapping based on this objective function maps the data pairs close to each other which have large distance between them. So this violates the preservation of the local data pairs of neighbourhoods.

By using objective function (3.7) this shortcoming can be overcome [69]. This is because the adjacency graph G_{g-v} puts a large penalty if nearby points and with a large distance from same class are mapped very close to each other. So maximizing the objective function secures the data points from same class with large distance to be mapped with a high distance in low dimensional space.

So from this analyses, it can be see that, objective function (3.7) assures that data pairs with large distances from same class are mapped with high distance and nearby data points from same class are mapped very close to each other in low dimensional data space with inherent geometric structure.

3.3.3. Discriminant Objective Function [69]: With the help of training data points a discriminant adjacency graph G_{g-d} is constructed and F is the weight matrix. Now the elements F_{ij} of weight matrix F can be given as follows:

$$F_{ij} = \begin{cases} K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i \neq c_j \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

Mapping the data from high dimensional space to a line in a low dimensional space so that the data points with large distance, remain as distant as possible. So the objective function will be [69]

$$\max \sum_{ij} (y_i - y_j)^2 F_{ij} \tag{3.9}$$

This objective function guarantees that the nearby points from different classes will be mapped far apart in low dimensional data space. So one can see that objective function maximizes the distance of nearby points from different classes. Thus one can now easily define local discriminating structure of data.

The detailed algorithm of the aforementioned method is given in Algorithm 3.1. [69].

Algorithm 3.1.: Similarity and Diversity and Discriminant Matrix

---

INPUT: Training Matrix $X \in \Re^{mxn}$

OUTPUT: Similarity Matrix $S \in \Re^{nxn}$ and Diversity Matrix $V \in \Re^{nxn}$ and Discriminant Matrix $F \in \Re^{nxn}$

BEGIN:

Initialize $B = 0$

FOR $cntr1 = 1:n$

FOR $cntr2 = 1:n$

$$B(cntr1, cntr2) = \exp\left(-\left\| X(:, cntr1) - X(:, cntr2) \right\|^2\right)$$

END FOR

END FOR

Initialize $S = B$

Initialize $F = B$

FOR $cntr1 = 1:n$

FOR $cntr2 = 1:n$

IF $(cntr1 \neq cntr2)$

IF class $(cntr1)$ = class $(cntr2)$

$S(cntr1, cntr2) = 0$

ELSE

$F(cntr1, cntr2) = 0$

END IF

END IF

END FOR

END FOR

$V = I^{nxn} - S$

END

3.3.4. Main Objective Function [69]: Considering objective functions (3.6), (3.7) and (3.9) main objective function of SOLDE can be computed. Here, Rayleigh quotients form that is formally similar to MFA is used to build the objective function of SLODE. Suppose a is a projection vector. Substituting $y_i = a^T x_i$ into the objective function (3.6) can be represented as follows [69]:

$$\begin{aligned} &\sum_{ij} (y_i - y_j)^2 S_{ij} \\ &= \sum_{ij} (y_i - y_j)(y_i - y_j)^T S_{ij} \\ &= \sum_{ij} \left(a^T x_i - a^T x_j\right)\left(a^T x_i - a^T x_j\right)^T S_{ij} \\ &= \left(a^T X D^s X^T a - a^T X S X^T a\right) \\ &= a^T X L_s X^T a \end{aligned} \tag{3.10}$$

Where D^s is a diagonal matrix whose elements are the column or row sum of S as S is a symmetric matrix. So D^s can be expressed as follows:

$$D_{ii}^s = \sum_j S_{ij}$$

Here L_s is the laplacian matrix of G_{g-s} and can be expressed as follows:

$$L_s = D^s - S$$

Similarly substituting $y_i = a^T x_i$ into the objective function (3.7) can be expressed in the following way [69]:

$$\begin{aligned} &\sum_{ij} (y_i - y_j)^2 V_{ij} \\ &= \sum_{ij} (y_i - y_j)(y_i - y_j)^T V_{ij} \\ &= \left(a^T X D^v X^T a - a^T X V X^T a\right) \\ &= a^T X L_v X^T a \end{aligned} \tag{3.11}$$

Where D^v is a diagonal matrix whose elements are the column or row sum of V as V is a symmetric matrix. So D^v can be expressed as follows:

$$D_{ii}^v = \sum_j V_{ij}$$

Here L_v is the laplacian matrix of G_{g-v} and can be expressed as follows:

$$L_v = D^v - V .$$

Similarly substituting $y_i = a^T x_i$ in objective function (3.9) can be expressed in the following way [69]:

$$\begin{aligned} &\sum_{ij} (y_i - y_j)^2 F_{ij} \\ &= \sum_{ij} (y_i - y_j)(y_i - y_j)^T F_{ij} \\ &= \left(a^T X D^m X^T a - a^T X F X^T a\right) \\ &= a^T X L_m X^T a \end{aligned} \tag{3.12}$$

Where D^m is a diagonal matrix whose elements are the column or row sum of F as F is a symmetric matrix. So D^m can be expressed as following:

$$D_{ii}^m = \sum_j F_{ij}$$

Here L_m is the laplacian matrix of G_{g-d} and can be expressed as follows:

$$L_m = D^m - F .$$

Substituting (3.10), (3.11) and (3.12) into objective functions (3.6), (3.7) and (3.9) respectively, the optimal objective function can be written as

$$a = \arg\max \frac{a^T X L_m X^T a}{a^T X L_s X^T a - a^T X L_v X^T a} \tag{3.13}$$

This optimal objective function considers that the similarity and diversity of nearby data of same class are similarly important. The main role of $a^T X L_v X^T a$ is to avoid the problem caused by $a^T X L_s X^T a$ that characterize the intra-class compactness. In order to get a compact intra-class representation, a large weight to $a^T X L_s X^T a$ and a small weight to $a^T X L_v X^T a$ are assigned. So the optimal objective function becomes [69]:

$$a = \arg\max \frac{a^T X L_m X^T a}{a^T X (\alpha L_s - (1-\alpha) L_v) X^T a} \tag{3.14}$$

Where $L_d = \alpha L_s - (1-\alpha) L_v$ and α is a suitable constant and value of α is assign to $0.5 \le \alpha \le 1$. As utilized in the original SOLDE algorithm, in our experiments too we set the value of α as 0.9. When the denominator of equation (3.14) is non-singular, the optimal projection vector a can be obtained by the maximum eigenvalue solution of the following generalized eigenvalue equation [69].

$$X L_m X^T a = \lambda X L_d X^T a \tag{3.15}$$

But, if the denominator of (3.14) is singular the optimal projection can not be directly obtained by the generalized eigenvalue equation. Then at first PCA is applied to project the data into a subspace so that matrix $X L_d X^T$ becomes non-singular and then the generalized value equation is solved. It is sure that the projection vectors do not satisfy the mutual orthogonality and it can be shown than that orthogonal projection vector is more effective to preserve the inherent geometry and discriminant factor is also improved [69]. So for maximizing the objective function, a set of orthogonal projection vectors in SOLDE is computed.

The detailed algorithm of the aforementioned method is given in Algorithm 3.2 [69].

Algorithm 2: Objective Function Matrices Computation

INPUT: Similarity Matrix $S \in \Re^{nxn}$ and Diversity Matrix $V \in \Re^{nxn}$ and Discriminant Matrix $F \in \Re^{nxn}$

OUTPUT: Laplacian Matrices L_s, L_v and L_m respectively and Diagonal Matrices D_s, D_v, and D_m respectively.

Initialize $D_s = 0$

Initialize $D_v = 0$

Initialize $D_m = 0$

FOR $cntr = 1:n$

$$D_s(cntr, cntr) = \sum_{cntr} S_{cntr,cntr}$$

$$L_s = D_s - S$$

END FOR

FOR $cntr = 1:n$

$$D_v(cntr, cntr) = \sum_{cntr} V_{cntr,cntr}$$

$$L_v = D_v - V$$

END FOR

FOR $cntr = 1:n$

$$D_m(cntr, cntr) = \sum_{cntr} F_{cntr,cntr}$$

$$L_m = D_m - H$$

END FOR

END

3.3.5. Orthogonal Algorithm [68, 69]: In this orthogonal algorithm, for maximizing the objective function (3.14), find a set of orthogonal basis vectors $a_1, a_2, ..., a_n$ i.e. $\langle a_i, a_j \rangle = 0$ If $i \neq j$ otherwise $\langle a_i, a_j \rangle = 1$ where $\langle , \rangle$ denotes the inner product of two vectors. So the objective function of the orthogonal algorithm can be written as [69] follows:

$$a_1 = \arg\max_a \frac{a^T X L_m X^T a}{a^T X L_d X^T a} \tag{3.15}$$

And

$$a_k = \arg\max_a \frac{a^T X L_m X^T a}{a^T X L_d X^T a} \tag{3.16}$$

So that $a_k^T a_1 = a_k^T a_2 = ... = a_k^T a_{k-1} = 0$.

The solutions of the above objective functions [68, 69] i.e. optimal orthogonal basis vectors of (3.15) and (3.16) can be easily obtained. So the algorithm for obtaining the orthogonal basis vectors and mapping of data from high dimensional data space to low dimensional space is as follows:

a. PCA Projection [47, 68]: The input data x_i is mapped into a PCA subspace such that the denominator matrix in objective function is non-singular. Let denote the conversion matrix to PCA subspace as W_{PCA}. By this PCA projection the transformed matrix will be non-singular i.e. the rank of the matrix and the number of features are equal.

b. Adjacency Graph Construction [69]: Three adjacency graphs $G_{g-s} = \{X, S\}$, $G_{g-v} = \{X, V\}$ and $G_{g-d} = \{X, F\}$ are constructed with a vertex set X i.e. the training matrix. In this training matrix there are n number of nodes and three weight matrix S, V and F are constructed respectively. To construct the adjacency graphs G_{g-s} and G_{g-v} put an edge between nodes i and j, and the data point must belongs to same class i.e. x_i is among the k nearest neighbourhood of x_j and both x_i and x_j belongs to same class. To construct adjacency graph G_{g-d} put an edge between nodes i and j. And x_j is among the k nearest neighbourhood of x_i and these two belongs to different class. It can be seen that adjacency graphs G_{g-s} and G_{g-v} inform about inherent geometry and adjacency graph informs about discriminant structure.

c. Weight Matrix Calculation [69]: Similarity, diversity, and discriminant weight matrices S, V and F respectively calculated by using equation (3.4), (3.5) and (3.8) respectively.

d. Orthogonal Projection Vectors Calculation [69]: Let $[a_1, a_2, ..., a_k]$ be the orthogonal projection vectors and $A^{k-1} = [a_1, a_2, ..., a_{k-1}]$, $B^{k-1} = \left(A^{k-1}\right)^T \left(XL_d X^T\right)^{-1} A^{k-1}$ are denoted then the orthogonal basis vectors can be calculated in following way [69]:

   1. At first a_1 is computed as the eigenvectors corresponds to maximum eigenvalue of $\left(XL_d X^T\right)^{-1} XL_m X^T$.
   2. Follow the same procedure of computing eigenvectors corresponds to maximum eigenvalue for calculating k number of orthogonal basis vectors of M^k. And M^k can represented as follows:

$$M^k = \left[I - \left(XL_d X^T\right)^{-1} A^{k-1} \left\{B^{(k-1)}\right\}^{-1} \left\{A^{(k-1)}\right\}^T\right] \left(XL_d X^T\right)^{-1} XL_m X^T \qquad (3.17)$$

e. Orthogonal Embedding [69]: The projection matrix is denoted as $A_{o/p} = [a_1, a_2, ..., a_p]$ and mapping of data points from high dimensional data space to low dimensional data space is as follows:

$$Y = W^T X, \qquad \text{Where } W = W_{PCA} A_{o/p}$$

Where Y is data after mapping and W is the transformation matrix.

3.4. Theoretical Analysis of SOLDE Algorithm [69]:

A theoretical analysis of SOLDE algorithm was presented in [69] which is also presented here now. SOLDE is quite similar to MFA but, in a way, superior to MFA. As previously discussed, objective function (3.1) considers intra-class similarity only, but SOLDE considers both intra-class similarity and diversity of data. So combining the objective function (3.6) and (3.7) one objective function can be computed as follows [69]:

$$\begin{aligned} &\min \sum_{ij} (y_i - y_j)^2 S_{ij} - \sum_{ij} (y_i - y_j)^2 V_{ij} \\ &= \sum_{ij} (y_i - y_j)^2 \left(S_{ij} - V_{ij}\right) \\ &= \sum_{ij} (y_i - y_j)^2 \left(2S_{ij} - 1\right) \end{aligned} \qquad (3.18)$$

Though objective function (3.18) is similar to equation (3.1) but inherent structure is well established in SOLDE. If the distance is changed continuously and the parameter p is fixed then it can be seen from figure 3.3. that S_{ij} is always non-negative and $(2S_{ij}-1)$ may be negative or positive.

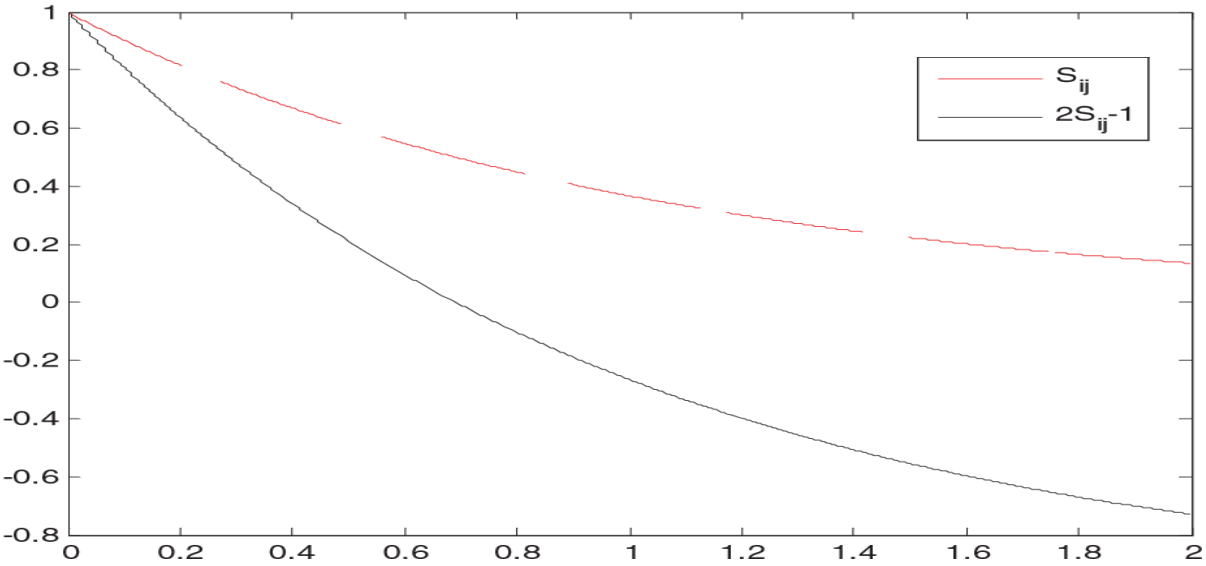


Figure 3.3. S_{ij} Versus $(2S_{ij}-1)$ When p Is Fixed [69].

From Fig. 3.3. [69], it can be seen that if two data points are very close to each other, i.e. distance is very small, both $(2S_{ij}-1)$ and S_{ij} are positive and large. In this case, both (3.1) and (3.18) map data points very close to each other into the low dimensional data space and mainly stores the similarity of data. But in real applications, the data distribution is uneven and the k nearest neighbours of some data may lie on a sparse region, thus the distance between these points may be large. In this case, $(2S_{ij}-1)$ may be negative, and is equivalent to maximizing the variation among these nearby data, preserves the diversity of data. While S_{ij} is still positive, and still preserves the similarity of data.

SOLDE considers both the similarity and diversity of data belonging to local neighbourhood. As mentioned earlier, MFA violates the local preserving topology while preserving data pairs with small distance as it does not consider the diversity of intra-class data pairs. Whereas SOLDE gives better recognition accuracy as it considers both similarity and diversity between the intra-class data.

3.5. Experimental Setup and Performance Evaluation:

In this section, we first present the experimental setup employed in Electrical Measurement and Instrumentation Laboratory of Electrical Engineering Department, Jadavpur University, to acquire the PDVP thermal images by employing NIR imaging philosophy. Then we experiment these algorithms for this image database to determine how effective they can be for the desired biometric recognition/authentication purposes.

3.5.1. The Experimental Setup [38, 83 - 87]: Palm Dorsa Subcutaneous Vein Pattern (PDSVP) is a promising and reliable biometric in biometric authentication [38]. The automatic data acquisition and extracting PDSVP is quite challenging in real world application. An experimental setup is employed in Electrical Measurement and Instrumentation Laboratory of Electrical Engineering Department, Jadavpur University, to acquire the PDVP thermal images by employing NIR imaging philosophy automatically [83]. A Near-Infrared (NIR) Pan-Tilt camera (FOSCAM® FI8918W [84]) is used to automatically locate the palm dorsum of the subject and acquiring the NIR images of PDSVP.

Seek-&-Freeze algorithm is used for data acquiring purpose because of its robustness. At first any acquired image i.e. the test image is compared with an existing reference image by extracting the SIFT keypoint descriptors [85] of both images using SIFT transform and then descriptors are compared with another descriptors using Brute Force matching Algorithm. So that the test image is compared with reference image based on their keypoint descriptors which are invariant to image scaling and rotation [85] and also partially invariant to illumination and 3D displacement of camera viewpoint [85]. Then the position is locked and data acquisition by the NIR camera is activated [85].The position of the palm dorsum and the camera after Seek-&-Freeze algorithm locks position is shown below in figure 3.4. [83].

Figure 3.4. Position of the Palm Dorsum and the Camera after Position Lock [83]

Hardware setup used in this data acquiring process is shown below in figure 3.5. [83] A low-cost Raspberry® Pi 2 Model B (comprising ARMv7 processor and 1 GB RAM which greatly enhances the computational ability even at such small size of the module) [86], Netgear® JNR1010 N150 (used for wireless communication between the data acquisition device and the Raspberry® Pi 2 Model B that greatly enhances the range of area that the data acquisition sensor can cover without having the overhead constraint of being close to the CPU) [87], and NIR camera FOSCAM® 8918W (for near-Infrared image acquisition of the PDSVP) [84] forms the most significant components of the integrated hardware setup.

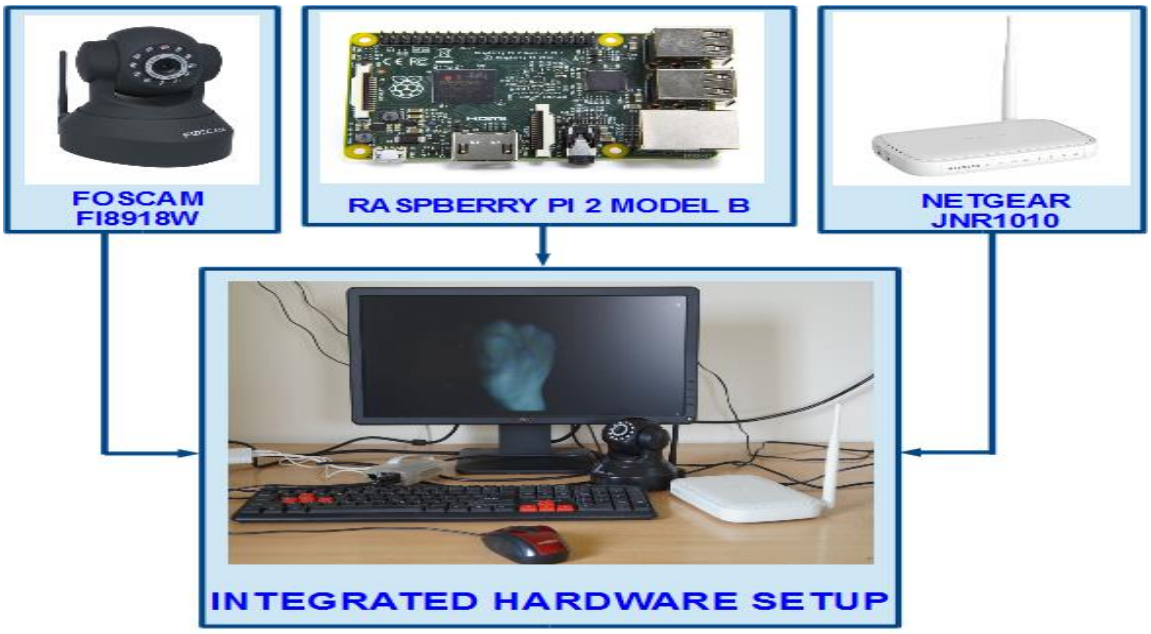


Figure 3.5. Integrated Hardware Setup Developed for PDSVP Based Biometric System [83]

Contrast Limited Adaptive Histogram Equalization (CLAHE) is used for enhancing the PDSVP [83].One significant advantage of Adaptive Histogram Equalization (AHE) is that, rather than considering the whole image matrix at a time, like Global Histogram Equalization (GHE), AHE concentrates on small neighbourhoods called "tiles or blocks" (8×8 in this case) [83]. Then only the PDSVP is extracted by using Local Adaptive thresholding technique. Then final image is achieved by performing morphological closing followed by morphological opening to fuse the narrow breaks between long stretches of the veins and to eliminate small dot-noise [83]. Then the final image is stored into the JU-NIR-V2: NIR Vein Database [83].

3.5.2. Performance Evaluation: Extensive experimentations have been carried out to demonstrate the efficiency and effectiveness of the SOLDE algorithm for our proposed biometric classification problems. As mentioned before, we have done this experiments on JU-

NIR-V2: NIR Vein Database. As mentioned before, we set the parameter α as 0.9 for this algorithm SOLDE.

The JU-NIR-V2: NIR Vein Database contains 570 images of 57 individuals where each person has 10 images. In this experiments all the images are resized to 32 X 25 pixels manually. Here we take 2, 3, 4 and 5 images per person as training data respectively and rest of images are taken for testing. Also we vary the number of features per sample 200, 400, 600 and 800 respectively.

Firstly we have optimized the above parameter (training samples per class and features per sample image) and then we determine the recognition rate with quantitative measures like Norm, Root Mean Square (RMS), Normalized Least Square Error (NLSE), Correlation, and Peak Signal to Noise Ratio (PSNR). The recognition accuracy by using various measuring method is given below in tabulation form, in tables 3.1 to 3.5.

Table 3.1.: Recognition rate calculated by Norm method:

Image Size	Number of Training Samples Per Class	Recognition Rate Mean ± Standard Deviation (%)			
		Number of Features Considered Per Sample Image			
		200	400	600	800
Image size = (32 x 25)	2	88.87 ± 0.02630	89.68 ± 0.0201	92.11 ± 0.02000	93.74 ± 0.02070
	3	89.67 ± 0.03390	90.89 ± 0.02000	93.37 ± 0.01860	95.07 ± 0.01550
	4	94 .02 ± 0.03375	95.16 ± 0.01660	96 .00 ± 0.01770	96.37 ± 0.01550
	5	95.79 ± 0.02100	96.28 ± 0.01890	96.41 ± 0.01390	97.07 ± 0.01170

Table 3.2.: Recognition rate calculated by Root Mean Square (RMS) method:

Image Size	Number of Training Samples Per Class	Recognition Rate Mean ± Standard Deviation (%)			
		Number of Features Considered Per Sample Image			
		200	400	600	800
Image size = (32 x 25)	2	89.53 ± 0.01800	90.03 ± 0.01800	90.91 ± 0.03682	92.63 ± 0.02137
	3	90.07 ± 0.02700	91.65 ± 0.02706	93.65 ± 0.02083	95.11 ± 0.02147
	4	93.92 ± 0.02700	95.13 ± 0.01404	96.12 ± 0.01989	97.13 ± 0.01781
	5	96.17 ± 0.02380	96.95 ± 0.01435	97.42 ± 0.01793	97.92 ± 0.00874

Table 3.3.: Recognition rate calculated by Normalised Least Square Error (NLSE) method:

Image Size	Number of Training Samples Per Class	Recognition Rate Mean ± Standard Deviation (%)			
		Number of Features Considered Per Sample Image			
		200	400	600	800
Image size = (32 x 25)	2	88.74 ± 0.02587	89.96 ± 0.01680	92.09 ± 0.03363	92.88 ± 0.02187
	3	91.97 ± 0.02518	93.62 ± 0.03042	94.21 ± 0.02701	95.02 ± 0.01402
	4	94.67 ± 0.02098	95.11 ± 0.00938	96.02 ± 0.01242	96.72 ± 0.01702
	5	95.68 ± 0.01606	96.25 ± 0.01389	97.00 ± 0.01971	98.20 ± 0.01573

Table 3.4.: Recognition rate calculated by Correlation method.

Image Size	Number of Training Samples Per Class	Recognition Rate Mean ± Standard Deviation (%)			
		Number of Features Considered Per Sample Image			
		200	400	600	800
Image size = (32 x 25)	2	89.46 ± 0.01636	90.03 ± 0.01104	91.11 ± 0.01077	92.52 ± 0.02139
	3	92.10 ± 0.01898	93.15 ± 0.02014	94.26 ± 0.01603	95.03 ± 0.01933
	4	93.31 ± 0.02012	94.02 ± 0.01313	95.56 ± 0.01402	96.26 ± 0.01375
	5	95.26 ± 0.01852	95.82 ± 0.01261	96.62 ± 0.01830	97.11 ± 0.01627

Table3.5.: Recognition rate calculated by Peak Signal to Noise Ratio (PSNR) method.

Image Size	Number of Training Samples Per Class	Recognition Rate Mean ± Standard Deviation (%)			
		Number of Features Considered Per Sample Image			
		200	400	600	800
Image size = (32 x 25)	2	88.86 ± 0.02530	90.91 ± 0.02046	91.89 ± 0.01632	93.03 ± 0.01853
	3	90.26 ± 0.00955	92.16 ± 0.01669	93.78 ± 0.01529	94.28 ± 0.02235
	4	93.03 ± 0.01424	94.56 ± 0.01662	95.82 ± 0.01513	96.21 ± 0.02200
	5	95.47 ± 0.01674	96.23 ± 0.01193	96.96 ± 0.01754	97.88 ± 0.01972

Graphical representation of variation of recognition accuracy with the variation of number of features in a sample and also with the number of training samples per class is presented as given below in figures 3.6. to 3.10.

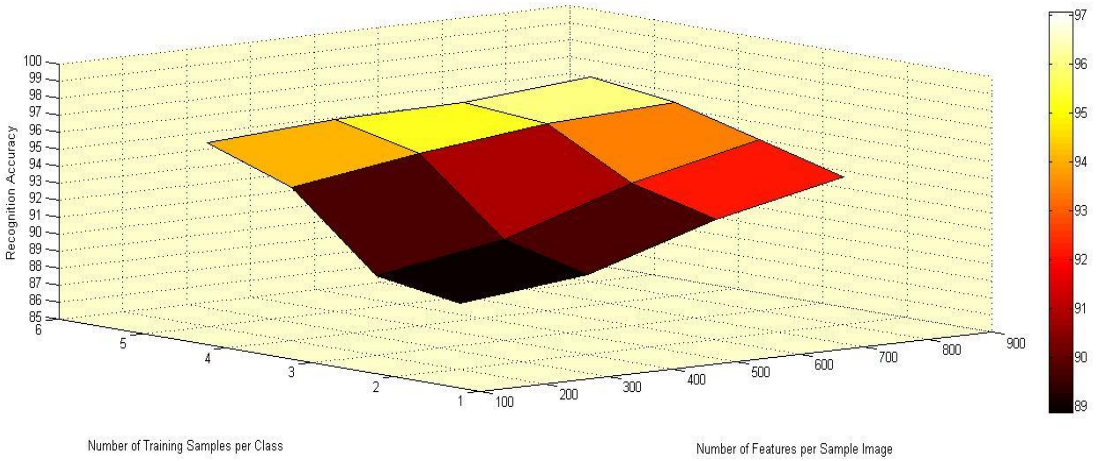


Figure 3.6. Graphical Representation of Recognition Accuracy Measured by Using NORM

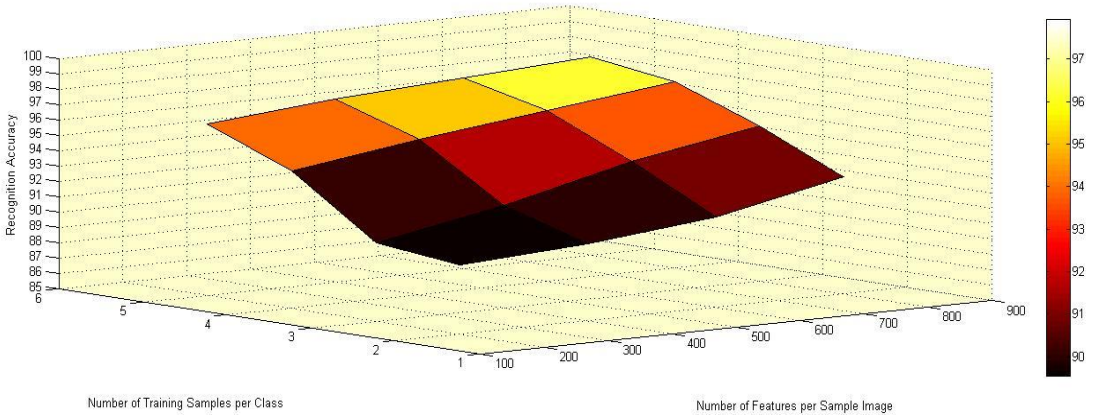


Figure 3.7. Graphical Representation of Recognition Accuracy Measured by Using RMS

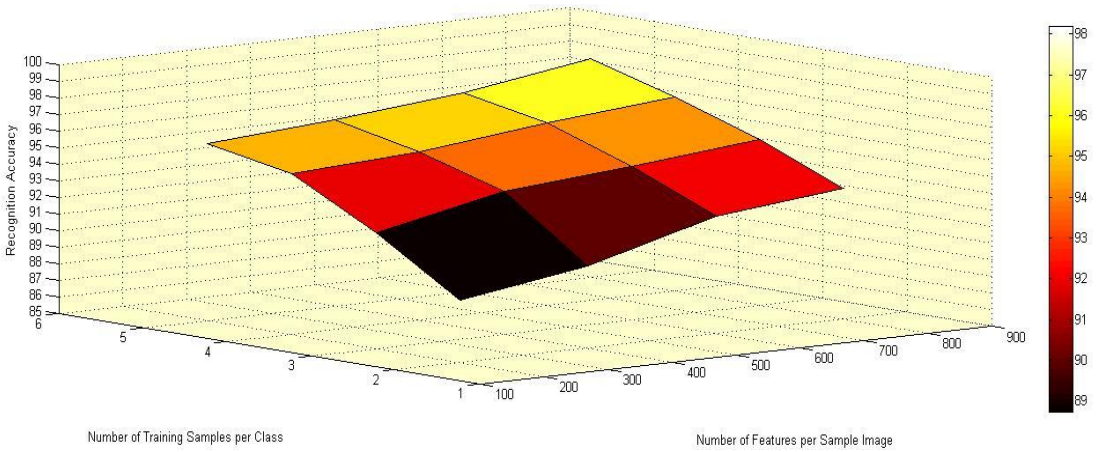


Figure 3.8. Graphical Representation of Recognition Accuracy Measured by Using NLSE

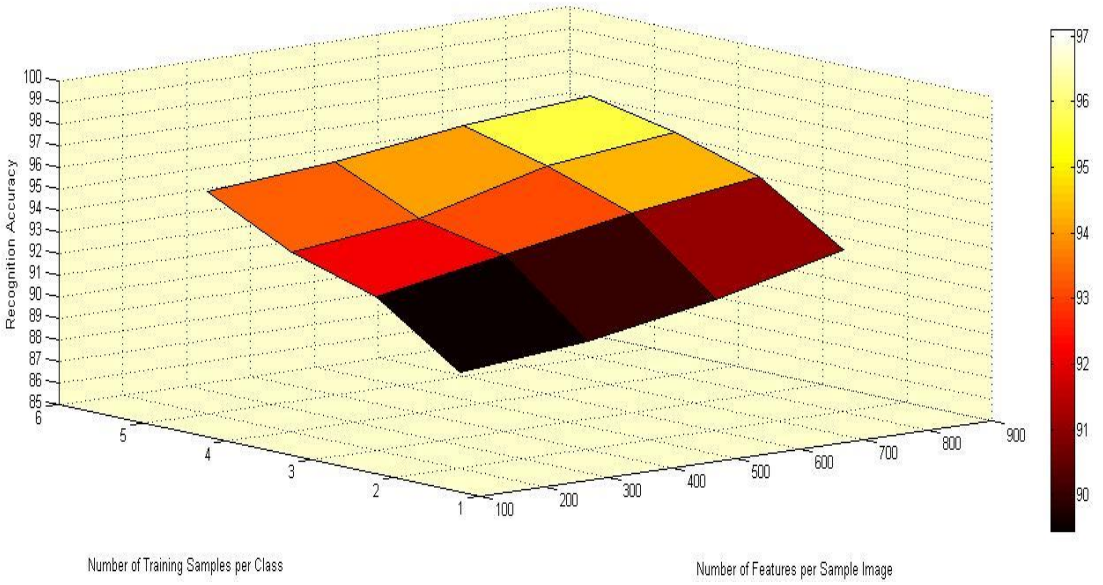


Figure 3.9. Graphical Representation of Recognition Accuracy Measured by Using Correlation

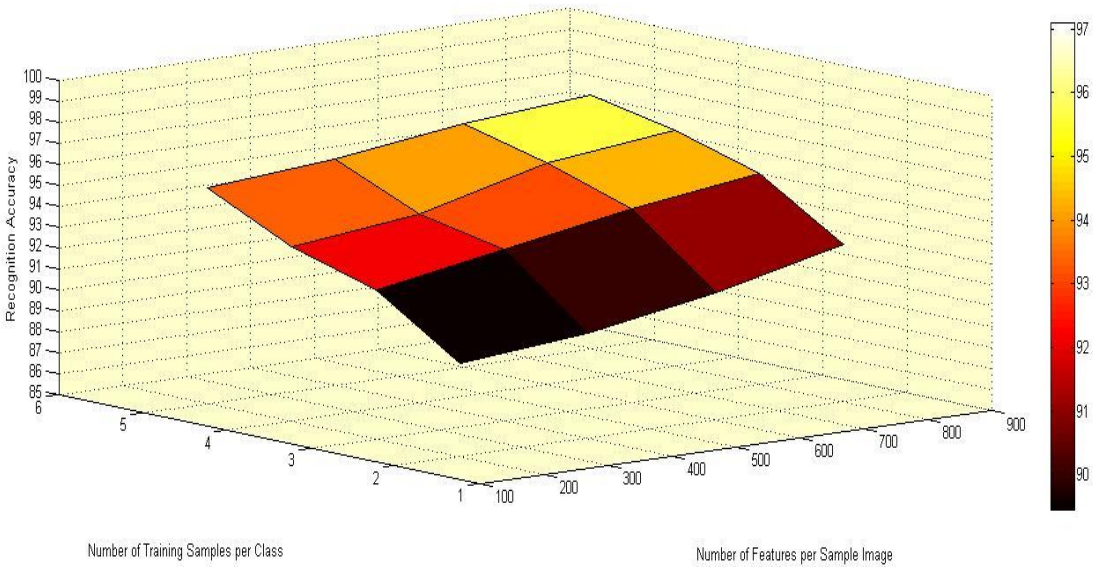


Figure 3.10. Graphical Representation of Recognition Accuracy Measured by Using PSNR

The graphical representation of performance comparison for various measuring methods with variation in number of training samples per class and the comparison among various measuring methods with variation in the number of features per sample are also given below in figures 3.11. to 3.18.

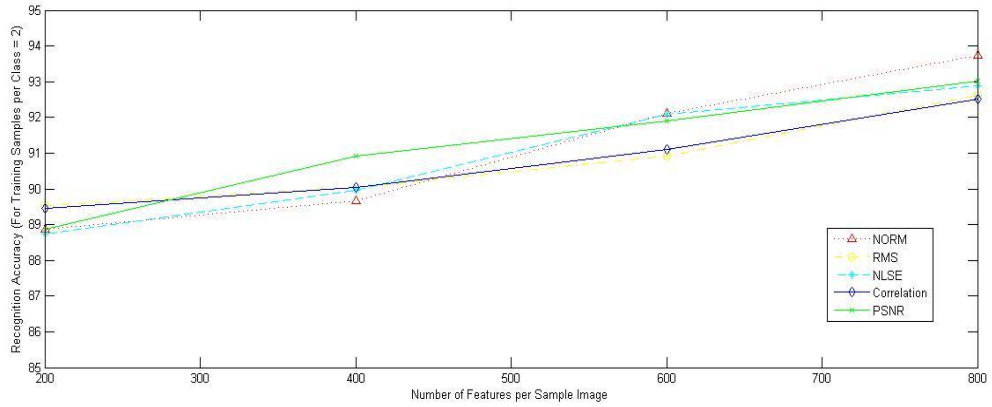


Figure 3.11. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 2)

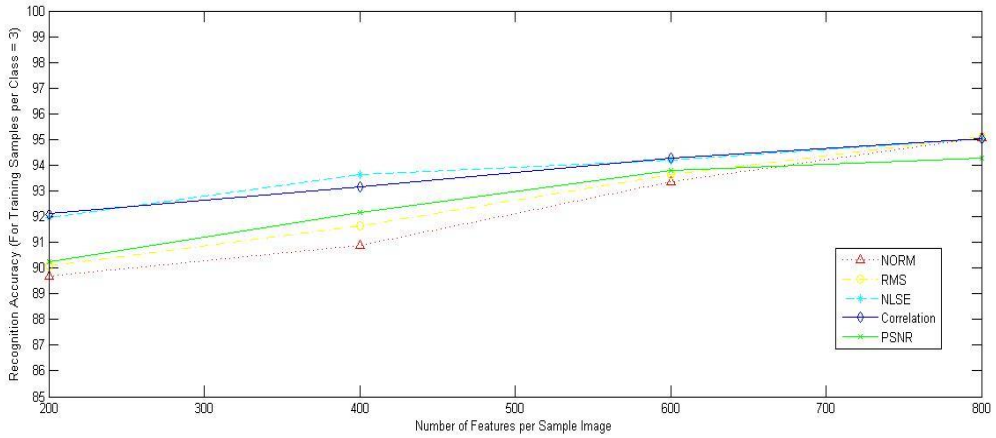


Figure 3.12. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 3)

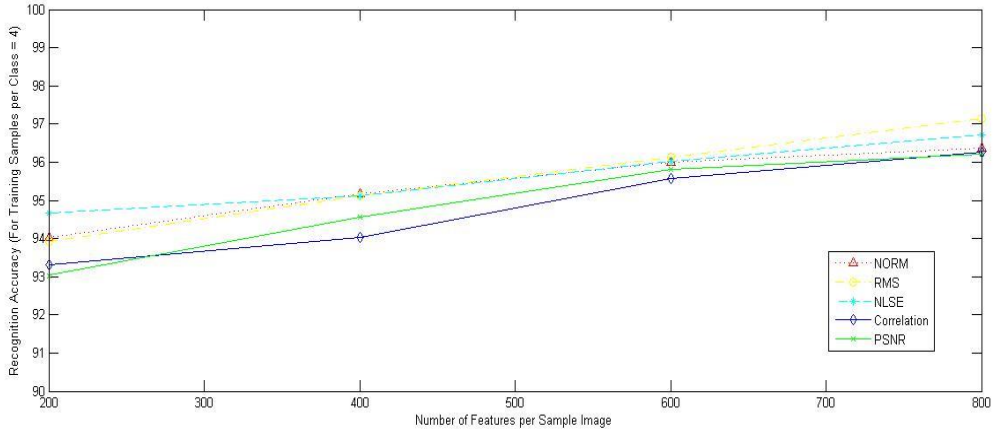


Figure 3.13. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 4)

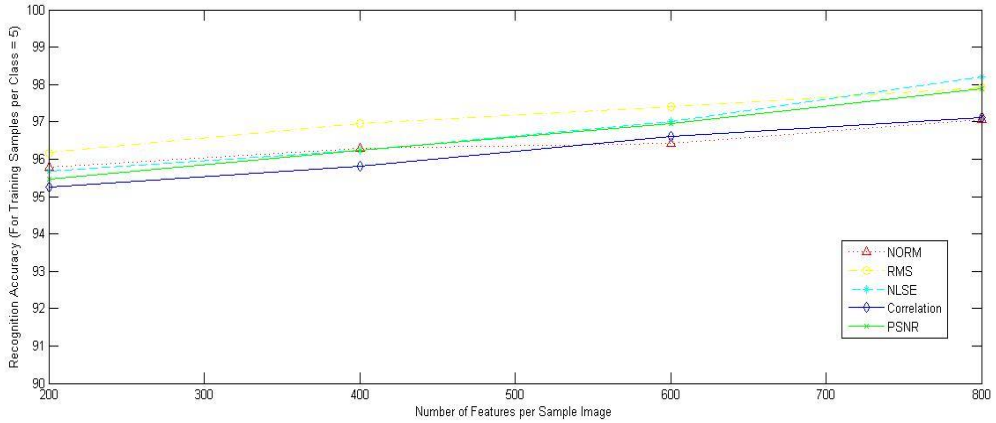


Figure 3.14. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 5)

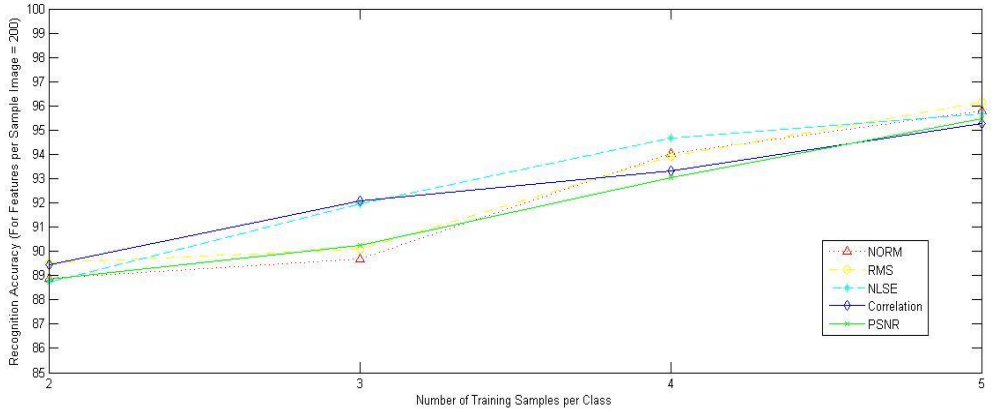


Figure 3.15. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 200)

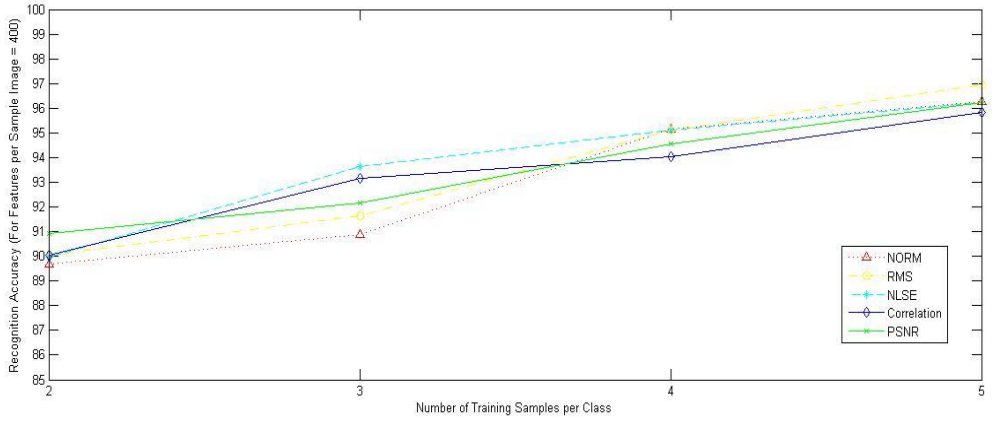


Figure 3.16. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 400)

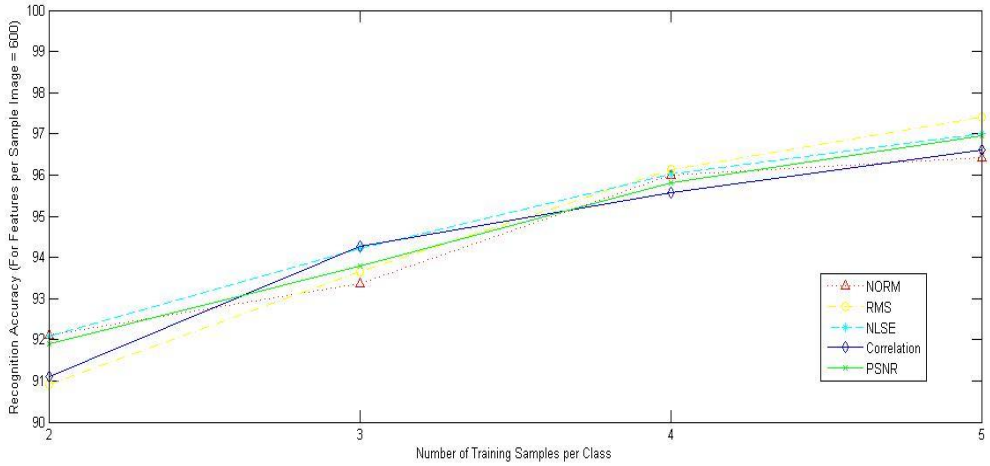


Figure 3.17. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 600)

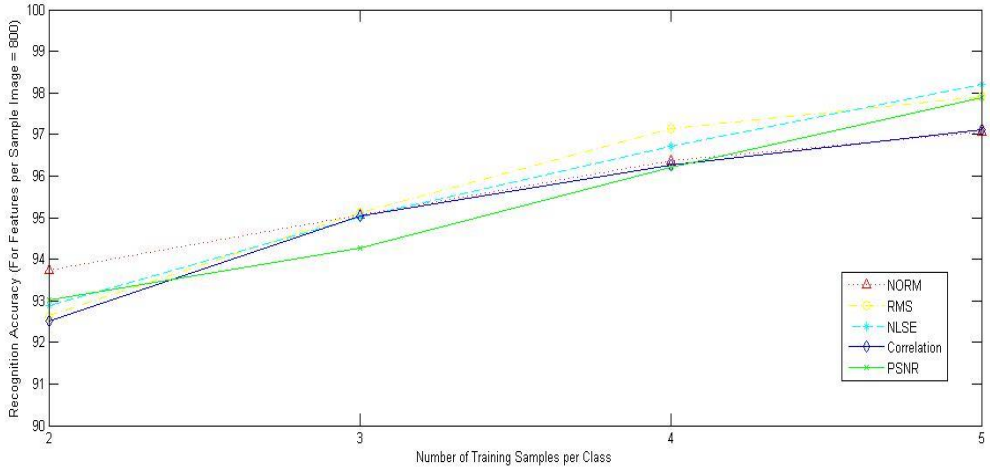


Figure 3.18. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 800)

From above experiments it can be easily seen that SOLDE give a good recognition accuracy. And this good recognition accuracy is due to consideration of diversity in data points in same class with intra-class similarity among data points.

It can also observe that for all performance measuring algorithms i.e. for Norm, RMS method, NLSE method, Correlation method, and PSNR method as the training samples per class is increased, the recognition accuracy is increased. Recognition accuracy is also increased with the increasing number of features per sample image.

From above experiments it is clear that among all the above measuring methods, by using Normalized Least Square Error (NLSE) we get the best performance i.e. best recognition accuracy. NLSE has also been demonstrated to be more stable than other measuring methods.

3.6. Summery:

The present chapter has shown how a state-of-the-art variant of LDE algorithm called SOLDE algorithm can be effectively employed for thermal imaging of PDVP based biometric systems. Various performance measuring methods like Norm, RMS, Correlation, NLSE, and PSNR are used for measuring the recognition accuracies and their performance comparisons are also carried out.

Here intra-class variation among data along with the intra-class similarity between data pairs are considered and this consideration of both similarity and diversity gives a better recognition accuracy than the recognition accuracy computed considering only intra-class similarity between data pairs. And among all the measuring methods, using NLSE we get the best performance i.e. best recognition accuracy and least variations in performance.

Chapter 4 will describe about another variants of LDE i.e. FLDE algorithm and vein pattern based biometric authentication using FLDE.

CHAPTER 4

Fuzzy Local Discriminant Embedding for Palm Dorsa Vein Pattern Based Biometric System

The chapter presents another variant of LDE algorithm called Fuzzy Local Discriminant Embedding algorithm. In this chapter, Fuzzy Local Discriminant Embedding algorithm is used for palm dorsa vein pattern based biometric system.

CHAPTER 4

Fuzzy Local Discriminant Embedding for Palm Dorsa Vein Pattern Based Biometric System

4.1. Introduction:

As mentioned in previous chapters, in machine learning scenario high dimensional data increases the required data space and also increases the computational burden and time. To overcome this problem one has to reduce the dimension of data i.e. one must map the data from high dimensional data space to low dimensional data space. Linear dimensionality reduction methods map the high dimensional input data to low dimensional subspace and this subspace can provide a compact representation of the input data. As mentioned before, there are mainly two fundamental linear dimensionality reduction methods and those are Principal Component Analysis (PCA) [47] and Linear Discriminant Analysis (LDA) [70, 71]. PCA attempts to find a linear mapping, which preserves the total variance by maximising the trace of feature covariance matrix. The optimal projections of PCA are corresponding to the first k number of largest eigenvalues of the total covariance matrix of data [47]. While considering k number of largest eigenvalues, PCA may probably discard much useful information and the recognition accuracy may be poor due to loss of information. However, in case of LDA, the optimal set of projection vectors is found such that it maximizes the determinant of the between-class scatter matrix and at the same time minimise the determinant of the within-class scatter matrix. But, since the dimension of vectors is high and the number of observations can be small, an intrinsic limitation of traditional LDA is that it fails to work when the within-class scatter matrix becomes singular, which is known as the small sample size (SSS) problem [70, 71]. It has already mentioned that Local Discriminant Embedding (LDE) uses the information of neighbours and class relations of data pairs for pattern recognition. However, in real world applications, data may not be properly acquired and variation of illumination may also be present, which will pose significant challenges. Fuzzy Local Discriminant Embedding (FLDE) algorithm is another state-of-the-art variant of the classical LDE algorithm that attempts to reduce this external problem and it has been shown that, in case of face recognition problems, it could achieve recognition accuracy much higher than LDE [77]. In this thesis this FLDE algorithm has been successfully implemented as another candidate variant of LDE algorithm,

for thermal imaging of PDVP based biometric systems, and its performance has been extensively studied for the benchmark database created and acquired in our laboratory. In FLDE, a membership degree matrix is firstly calculated using fuzzy k-nearest neighbourhood (FKNN) algorithm [76], then this membership degree information is incorporated into the definition of the laplacian scatter matrix to obtain the fuzzy lplacian scatter matrix.

Manifold learning based methods are quite adept in finding the intrinsic non-linear structure hidden in the observed space. Among many manifold learning based algorithms or locality preserving techniques, ISOMAP [50, 51], Laplacian Eigenmap (LE) [52, 53], and Locally Linear Embedding (LLE) [54, 55] are widely used. Recently, He and Niyogi [25] and He et al. [67] proposed locality preserving projections (LPP), which is a linear subspace learning method derived from LE. LPP preserves the local information of data and detect the essential pattern [67]. The optimal projection axes best preserve the local structure of the underlying distribution in the Euclidean space. From analysis it can be found that LPP is closely connected with PCA and LDA and LPP can find an embedding space that preserves local information of data. However, it is an unsupervised method. Local discriminant embedding (LDE) was developed by Chen et al. [56], but the underlying ideas of LDE are almost the same as LPP except for using the labelled information [56]. LDE [56] achieves good discriminating performance by considering the information of neighbour and class relations between data points. LDE incorporates the class information into the construction of embedding and derives the embedding for nearest-neighbour classification in a low-dimensional space which learns the embedding for the sub-manifold of each class by solving an optimisation problem. For eliminating the external problem like illumination variation etc. the advantages of the technology of fuzzy sets is used [78]. A number of studies have already been carried out for fuzzy image filtering, fuzzy image segmentation and fuzzy edge detection with an ultimate objective to overcome the factor of uncertainty being inherently present in many problems of image processing and pattern recognition [76]. To achieve the local distribution information of original samples Fuzzy k-nearest neighbour (FKNN) is implemented. As already mentioned, here a membership degree matrix using FKNN is computed first and then this membership degree is incorporated into the definition of the laplacian scatter matrix to obtain the fuzzy laplacian scatter matrices. In FLDE two fuzzy neighbour graphs are constructed differing from the existing graph based algorithms. In fuzzy neighbourhood graphs neighbour relations of nearest data points of same class are maintained and data from different classes are kept distant after FLDE [77].

The rest of the chapter is constructed as follows:

I. In Section 4.2., a brief description about LDE and introduce FKNN is given.
II. In Section 4.3., the idea of LDE and description of FLDE is introduced elaborately.
III. In section 4.4., a brief description about theoretical justification of FLDE is given.
IV. In section 4.5., the experimental results reporting the performance evaluation i.e., in essence, the recognition accuracy is presented.
V. In section 4.6., the observations about FLDE methodology for our system is concluded.

4.2. Brief Review of LDE and Fuzzy K-Nearest Neighbour Algorithm:

4.2.1. Local Discriminant Embedding [56, 77]:

Although LDE has already been presented in detail, here again the algorithm is introduced in very brief which will facilitate to understand further developments made on its basic premises to propose the FLDE algorithm. In this approach neighbour and class relations of data are used for constructing the embedding for classification problems. LDE [56] does the mapping from high dimensional space to low dimensional space. LDE [56] uses locality information to get intra-class compactness. Suppose there are n number of training samples $\{x_1, x_2, ..., x_n\} \in \Re^p$ from C classes. These training samples are used to determine a linear transformation matrix $A = [a_1, a_2, ..., a_n]$ such that

$$y_i = A^T x_i, i = 1,2,...,n \tag{4.1}$$

Where y_i denotes the low dimensional representation of training data sample x_i. It has already been mentioned that the aim of LDE is to keep close the sample from same classes of neighbourhood, whereas prevent to keep close the sample from different classes of neighbourhood. Then, with the introduction of G and G' as the two undirected graphs over data points, and W and W' as the weights in G and G′ respectively, one has to optimize the following objective function

$$\max \sum_{ij} (y_i - y_j)^2 W'_{ij} \tag{4.2}$$

Subjected to

$$\sum_{ij} (y_i - y_j)^2 W_{ij} = 1 \tag{4.3}$$

The elements W_{ij} in weight matrix W can be defined as follows:

$$W_{ij} = \begin{cases} K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i = c_j \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

And the elements $W_{ij}^{'}$ in weight matrix $W^{'}$ can be defined as follows:

$$W_{ij}^{'} = \begin{cases} K(x_i, x_j), & \text{If } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \text{ and } c_i \neq c_j \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

Where, as before, $K(x_i, x_j) = \exp\left(-\left\|x_i - x_j\right\|^2 / p\right)$, and p is a suitable parameter, and $\left\|x_i - x_j\right\|$ denotes the L2-norm of the difference between x_i and x_j, and $N_k(x_i)$ stands for the k nearest neighbourhoods of x_i and c_i is the class label of x_i.

As mentioned before, the optimisation can be reduced to the following generalised eigenvalue problem. From generalized eigen equation one can find the eigenvectors corresponds to largest eigenvalues. These eigenvectors make the transformation matrix $A = [a_1, a_2, ..., a_n]$ for embedding. The generalized eigen equation is given as:

$$X(D^{'} - W^{'})X^T a = \lambda X(D - W)X^T a \tag{4.6}$$

Where D and $D^{'}$, the diagonal matrices, with diagonal elements are given as:

$$D = \sum_j W_{ij} \tag{4.7}$$

$$D^{'} = \sum_j W_{ij}^{'} \tag{4.8}$$

4.2.2. Fuzzy K-Nearest Neighbour Algorithm [76, 77]:

The filtration of classification affects the within-class and between-class scatter matrices and, in the process, can augment the performance of the classifier. In this context, fuzzy partition can be utilized as an important notion under consideration. Suppose there are n number of transformed (transformed by PCA) training data $\{x_1, x_2, ..., x_n\} \in \Re^p$ in a data space and a fuzzy C class partition of these data specifies the degrees of membership of each data to the classes. The partition matrix is denoted by $U = [U_{ij}]$ for $i = 1,2,...,C$ and $j = 1,2,...,n$ and this partition matrix satisfies two following properties:

$$\sum_{i=1}^{C} U_{ij} = 1 \tag{4.9}$$

And $$0 < \sum_{j=1}^{n} U_{ij} < 1 \tag{4.10}$$

By using FKNN algorithm, the membership degree computation can be realised through the following steps [76, 77]:

i. Firstly the Euclidean distance matrix between pairs of data in training set is computed.

ii. Then set diagonal elements of this Euclidean distance matrix to infinity (practically these diagonal elements are set to a very large value).

iii. Then the distance matrix is sorted (considering each of its columns separately) in ascending order. Then corresponding class labels of the patterns located in the closest neighbourhood of the pattern under consideration are collected (as we consider k number of neighbours, this returns a list of k integers).

iv. Then the membership degree to class i for j^{th} pattern is computed using the expression proposed in the literature [79]:

$$U_{ij}^{G} = \begin{cases} 0.51 + 0.49(n_{ij}/k), & \text{If } i \in N_k(j) \\ 0.49(n_{ij}/k), & i \notin N_k(j) \end{cases} \tag{4.11}$$

In the above expression, n_{ij} stands for the number of the neighbours of the j^{th} data (pattern) that belongs to the i^{th} class.

4.3. Fuzzy Local Discriminant Embedding (FLDE) [77]:

An elaborate description about fuzzy local discriminant embedding is now presented here. This algorithm essentially incorporates the fuzzy membership degree function for the purpose of pattern recognition.

4.3.1. Basic Idea [77]:

Suppose there are n number training data $X = \{x_1, x_2, ..., x_n\} \in \Re^p$ in a data space from C number of classes. Let $G = \{X, W\}$ is an undirected graph with vertex set X and similarity matrix is given as W. The transformed low-dimensional data points are $Y = \{y_1, y_2, ..., y_n\}$. A

Fuzzy objective function that can be considered for good discriminant approach is given as follows [77]:

$$\min \frac{1}{2}\sum_{ij}\left(y_i - y_j\right)^2 U_{ij} W_{ij} \tag{4.12}$$

The projection matrix is denoted as $A = [a_1, a_2, ..., a_k]$ and mapping of data points from high dimensional data space to low dimensional data space can be carried out as follows:

$$Y = A^T X \tag{4.13}$$

By simple algebraic computation the objective function can be reduced as follows [77]:

$$\begin{aligned} &\frac{1}{2}\sum_{ij}(y_i - y_j)^2 U_{ij} W_{ij} \\ &= \left(a^T X D^F X^T a - a^T X W^F X^T a\right) \\ &= a^T X L_F X^T a \end{aligned} \tag{4.14}$$

Here the fuzzy diagonal matrix D^F and the fuzzy laplacian scatter matrix L_F can be computed by using following expressions:

$$D_{ii}^F = \sum_{j \neq i} U_{ij} W_{ij}, \forall i \tag{4.15}$$

$$L_F = D_F - W_F \tag{4.16}$$

4.3.2. The FLDE Algorithm [77]:

As mentioned before, suppose there are n number of training data $X = \{x_1, x_2, ..., x_n\} \in \Re^p$ in a data space from C number of classes and let c_i be the class labels of x_i. Then the complete FLDE algorithm can be given as follows:

i. Fuzzy Neighbourhood Graphs Construction: Fuzzy neighbourhood graphs are required for computation of the fuzzy membership degree functions. Let G_{fuzzy} denote a fuzzy neighbourhood graph and to construct G_{fuzzy} with the help of training data matrix X ,

data belong to the same class in local neighbourhood are considered. Fuzzy neighbourhood graph G_{fuzzy} is constructed as follows [77]:

$$G_{fuzzy} : if \begin{cases} C_i = C_j \\ x_i \in N_k(x_j) \; or \; x_j \in N_k(x_i) \end{cases} \tag{4.17}$$

ii. Let G'_{fuzzy} denote another fuzzy neighbourhood graph and to construct G'_{fuzzy} with the help of training data matrix X , data belong to different classes in local neighbourhood are considered.

Fuzzy neighbourhood graph G'_{fuzzy} is constructed as follows [77]:

$$G'_{fuzzy} : if \begin{cases} C_i \neq C_j \\ x_i \in N_k(x_j) \; or \; x_j \in N_k(x_i) \end{cases} \tag{4.18}$$

Here $N_k(x_i)$ and $N_k(x_j)$ stand for the k nearest neighbourhoods of x_i and x_j respectively.

iii. Construction of the Intra-class Compactness Fuzzy Graph and Separability Fuzzy Graph [77]: Intra-class Compactness Fuzzy Graph is computed from neighbourhood graph G_{fuzzy} using the following expression:

$$G_{fuzzy} : U_{ij}^{G} = \begin{cases} 0.51 + 0.49(n_{ij}/k), & \text{If } x_i \in N_k(x_j) \; or \; x_j \in N_k(x_i), \; and \; c_i = c_j \\ 0.49(n_{ij}/k), & otherwise \end{cases} \tag{4.19}$$

Here n_{ij} stands for the number of the neighbours of the j[th] data (pattern) that belong to the i[th] class and k is the number of neighbourhoods to be considered. In a similar way seperability fuzzy graph is also computed from G'_{fuzzy} as follows [77]:

$$G'_{fuzzy} : U_{ij}^{G'} = \begin{cases} 0.51 + 0.49(n_{ij}/k), & \text{If } x_i \in N_k(x_j) \; or \; x_j \in N_k(x_i), \; and \; c_i \neq c_j \\ 0.49(n_{ij}/k), & otherwise \end{cases} \tag{4.20}$$

Where fuzzy affinity matrix W_{fuzzy}^{G}, the fuzzy diagonal matrix D_{fuzzy}^{G} and the fuzzy laplacian matrix L_{fuzzy}^{G} of the graph G_{fuzzy} can be defined as follows [77]:

$$W_{fuzzy}^{G} = U_{ij}^{G} .* W_{ij} \tag{4.21}$$

$$L_{fuzzy}^{G} = D_{fuzzy}^{G} - W_{fuzzy}^{G} \tag{4.22}$$

$$D_{fuzzy}^{G} = \sum_{j \neq i} U_{ij}^{G} W_{ij}, \forall i \tag{4.23}$$

In a similar way the other fuzzy affinity matrix $W_{fuzzy}^{G'}$, the fuzzy diagonal matrix $D_{fuzzy}^{G'}$ and the fuzzy Laplacian matrix $L_{fuzzy}^{G'}$ of $G_{fuzzy}^{'}$ can be computed in the same manner, as given below [77]:

$$W_{fuzzy}^{G'} = U_{ij}^{G'} .* W_{ij}^{'} \tag{4.24}$$

$$L_{fuzzy}^{G'} = D_{fuzzy}^{G'} - W_{fuzzy}^{G'} \tag{4.25}$$

$$D_{fuzzy}^{G'} = \sum_{j \neq i} U_{ij}^{G'} W_{ij}^{'}, \forall i \tag{4.26}$$

iv. Transformation [77]: From generalized eigen equation one can find the eigenvectors corresponding to the largest eigenvalues. These eigenvectors make the transformation matrix $A = [a_1, a_2, ..., a_n]$ for embedding. The generalized eigen equation can be represented as follows:

$$X(D_{fuzzy}^{G'} - W_{fuzzy}^{G'})X^T a = \lambda_i X(D_{fuzzy}^{G} - W_{fuzzy}^{G})X^T a \tag{4.27}$$

Then embedding of data from high dimensional data to low dimensional data is done using following equation:

$$y_i = A^T x_i, \quad i = 1,2,..., n \tag{4.28}$$

The detailed algorithm of the aforementioned method is given in Algorithm 4.1.

Algorithm 4.1: FLDE Algorithm

INPUT: Training Matrix $X \in \Re^{mxn}$, k-Nearest Neighbour,

OUTPUT: Weight Matrices W and $W^{'}$,

BEGIN:

Initialize $B = 0$

FOR $cntr1 = 1:n$

FOR $cntr2 = 1:n$

$$B(cntr1, cntr2) = \exp\left(-\left\| X(:, cntr1) - X(:, cntr2) \right\|^2\right)$$

END FOR

END FOR

Initialize $W = B$

Initialize $W = B$

FOR $cntr1 = 1:n$

FOR $cntr2 = 1:n$

IF $(cntr1 \neq cntr2)$

IF class $(cntr1)$ = class $(cntr2)$

$W(cntr1, cntr2) = 0$

ELSE

$W'(cntr1, cntr2) = 0$

END IF

END IF

END FOR

END FOR

$B_{sort} = sort(B)$

$class\_mat =$ Class of elements in B_{sort}

Initialize $U = 0$

Initialize $U' = 0$

FOR $i = 1:n$

FOR $j = 1:n$

$val1 =$ Class of i

$val2$ = Class of j

$n = find(val2 == val1)$

IF $(val1 = val2)$

$u = 0.51 + 0.49 * (n/k)$

$u' = 0.49 * (n/k)$

ELSE

$u = 0.49 * (n/k)$

$u' = 0.51 + 0.49 * (n/k)$

END IF

$U(i, j) = u$

$U'(i, j) = u'$

END FOR

END FOR

END

$W_{fuzzy} = U * W$

$W'_{fuzzy} = U' * W'$

Initialize $D_{fuzzy} = 0$

Initialize $D'_{fuzzy} = 0$

Initialize $L_{fuzzy} = 0$

Initialize $L'_{fuzzy} = 0$

FOR $cntr = 1:n$

$D_{fuzzy}(cntr, cntr) = \sum_{cntr} W_{cntr,cntr}$

$L_{fuzzy} = D_{fuzzy} - W$

END FOR

FOR $cntr = 1:n$

$$D'_{fuzzy}(cntr, cntr) = \sum_{cntr} W'_{cntr,cntr}$$

$$L'_{fuzzy} = D'_{fuzzy} - W'$$

END FOR

END

4.4 Theoretical Justification of FLDE [77]:

A theoretical justification of FLDE has already been presented in [77] and this has been detailed now. Equation (4.27) can be solved by Lagrangian multiplier method as follows [77]:

$$\frac{\partial}{\partial A_i} tr\left\{A_i^T\left(D_{fuzzy}^{G'} - W_{fuzzy}^{G'}\right)A_i - \lambda_i\left(\left(D_{fuzzy}^{G} - W_{fuzzy}^{G}\right) - 1\right)\right\} = 0 \qquad (4.29)$$

Where λ_i is the Lagrangian multiplier. Thus one can obtain (4.27), where A_i the generalised eigenvector corresponds to generalised eigenvalue λ_i.

4.5. Experimental Setup and Performance Evaluation:

4.5.1. The Experimental Setup [80, 83]: In this thesis, the FLDE algorithm has been successfully employed for both FIR and NIR thermal imaging of PDVP based biometrics. The experimental setup employed in our laboratory to acquire the PDVP NIR thermal images and the associated image pre-processing steps employed before the processed images are stored in a database have already been explained in detail in Chapter 3. In addition to that, the FIR thermal images of PDVP patterns are obtained from several subjects in our laboratory using KT-384 thermal imager which is manufactured by Sonel®, Poland). This is a fully radiometric camera and it has non-cooled micro-bolometric matrix type detector with 384×288 pixel thermal resolution and thermal sensitivity of less than 0.08°C [80]. Complete integrated system developed for real-time human recognition using PDSVP is shown in figure below.

Figure 4.1. Complete Integrated System Developed for Real-Time Human Recognition Using PDSVP [80]

4.5.2. Performance Evaluation: We first present our extensive results obtained for the FIR thermal imaging, based on the database created in our laboratory which is named as JU-FIR-V1 database.

The JU-FIR-V1 database contains 435 images of 29 individuals with 15 images for each person. In these experiments all the images are resized to 32 X 32 pixels. For extensive experimental investigation, we consider different situations of 2, 3, 4 and 5 images per person as training data respectively and, in each case, rest of images from the database are utilized for the testing purpose. Also we vary the number of features per sample image as 200, 400, 600 and 800 respectively. For each experiment we consider 25 nearest neighbours.

Firstly we have optimized the above parameter (training samples per class and features per sample image) and then we find out the recognition rate with other measures like Norm, Root Mean Square (RMS), Normalized Least Square Error (NLSE), Correlation, and Peak Signal to Noise Ratio (PSNR).

The recognition accuracies obtained by using various measuring method is given below in tabular form, in tables 4.1. to 4.5.

Table 4.1.: Recognition rate calculated by NORM method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	400	600	800
		2	90.01±0.0179	90.81±0.0108	91.38±0.0138	92.04±0.0111
		3	93.19±0.0083	93.85±0.0126	94.25±0.0157	94.89±0.0121
		4	95.17±0.0126	95.58±0.0175	95.84±0.0125	96.26±0.0063
		5	97.24±0.0093	97.37±0.0067	97.56±0.0094	97.89±0.0153

Table 4.2.: Recognition rate calculated by Root Mean Square (RMS) method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	400	600	800
		2	90.26±0.0184	90.79±0.0179	91.28±0.0145	91.57±0.0137
		3	93.78±0.0181	94.08±0.0190	94.59±0.0088	94.93±0.0088
		4	95.79±0.0117	96.11±0.0137	96.27±0.0158	96.77±0.0050
		5	96.72±0.0078	97.48±0.01921	97.62±0.0095	97.82±0.0137

Table 4.3.: Recognition rate calculated by Normalized Least Square Error (NLSE) method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	400	600	800
		2	90.39±0.0139	90.56±0.0096	91.19±0.0146	92.12±0.0136
		3	94.40±0.0090	94.54±0.0133	94.72±0.0146	95.28±0.0158
		4	95.70±0.0125	96.48±0.0093	96.86±0.0128	97.17±0.0175
		5	97.17±0.0117	97.32±0.0078	97.47±0.0117	98.42±0.0192

Table 4.4.: Recognition rate calculated by Correlation method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	400	600	800
		2	89.92±0.0176	90.43±0.0073	90.97±0.0105	91.29±0.0076
		3	93.86±0.0095	94.28±0.0154	94.93±0.0165	95.68±0.0145
		4	96.05±0.0582	96.54±0.0125	97.12±0.0162	97.48±0.0175
		5	97.31±0.0212	97.41±0.0071	97.62±0.0128	98.02±0.0154

Table 4.5.: Recognition rate calculated by Peak Signal to Noise Ratio (PSNR) method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	400	600	800
		2	90.03±0.0135	91.11±0.0126	91.82±0.0016	92.52±0.0042
		3	94.51±0.0176	95.02±0.0113	95.97±0.01241	96.90±0.0012
		4	95.82±0.0182	96.23±0.0151	97.02±0.0151	97.57±0.0156
		5	97.41±0.0052	97.82±0.0012	98.07±0.0128	98.79±0.0128

Graphical representation of variation of recognition accuracy with the variation of number of features in a sample and also with the number of training samples per class is presented as given below in figures 4.2. to 4.6.

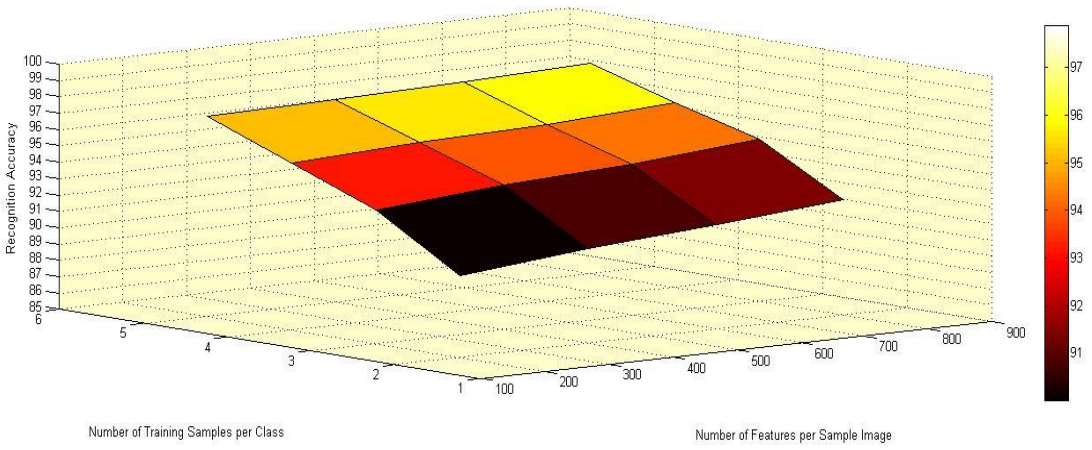


Figure 4.2. Graphical Representation of Recognition Accuracy Measured by Using NORM

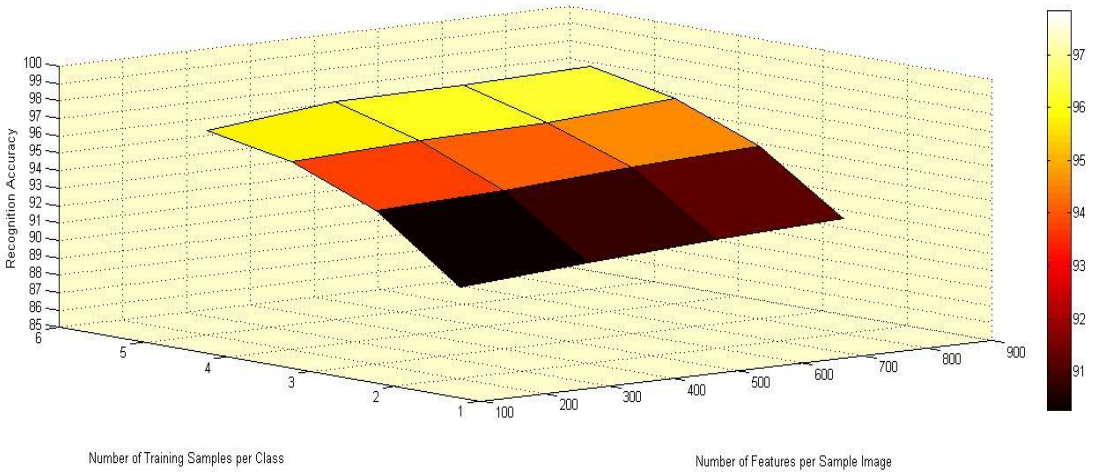


Figure 4.3. Graphical Representation of Recognition Accuracy Measured by Using RMS

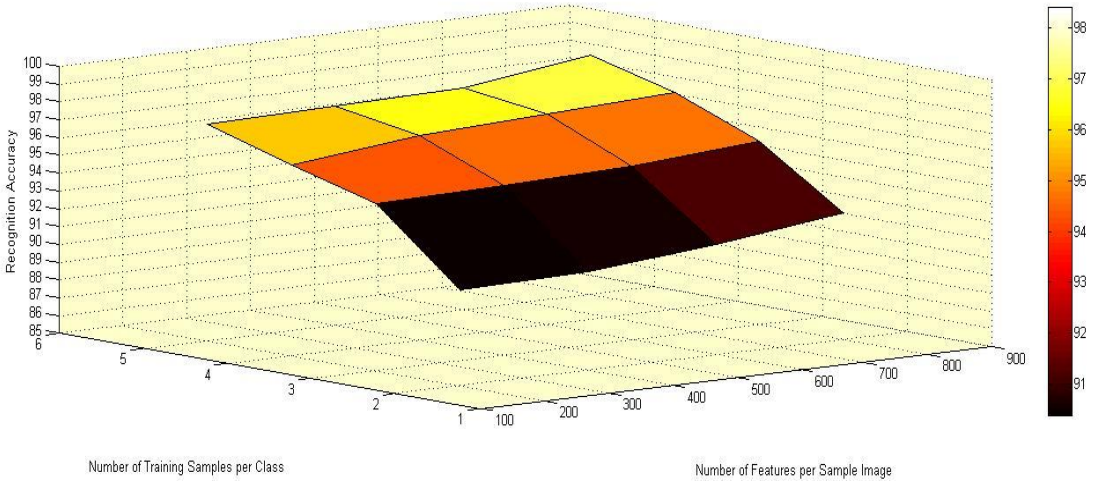


Figure 4.4. Graphical Representation of Recognition Accuracy Measured by Using NLSE

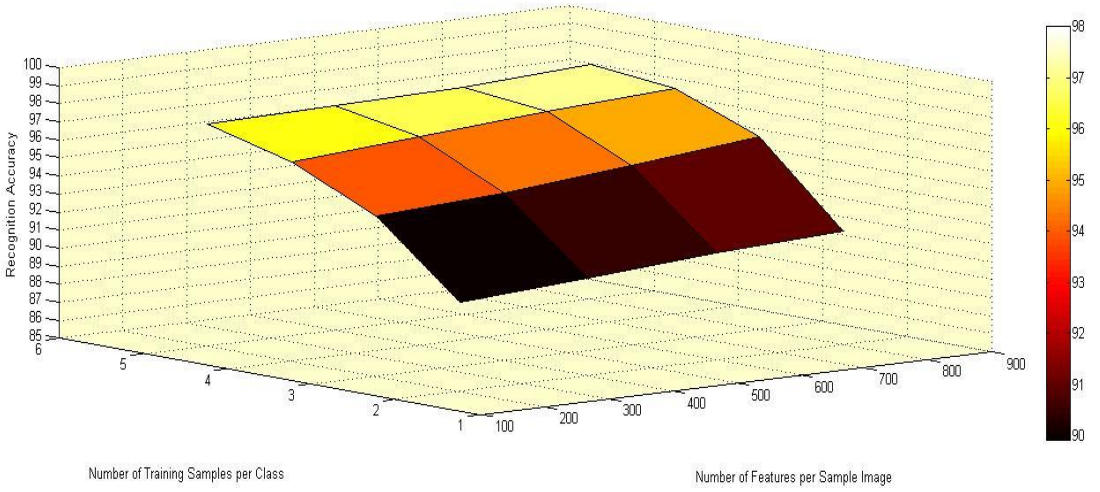


Figure 4.5. Graphical Representation of Recognition Accuracy Measured by Using Correlation

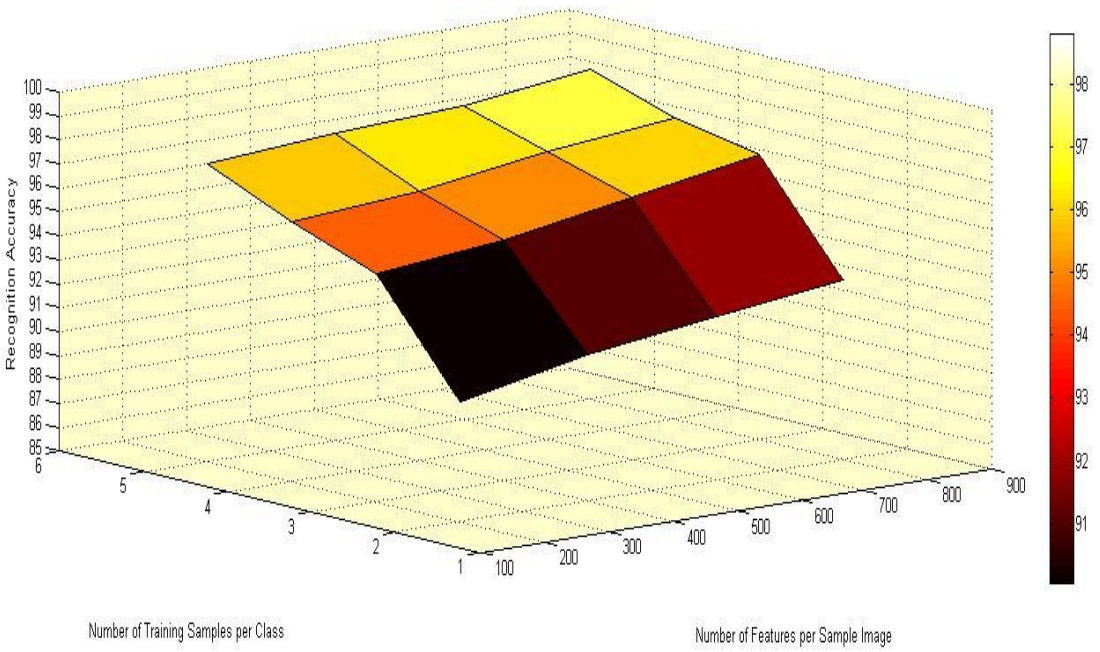


Figure 4.6. Graphical Representation of Recognition Accuracy Measured by Using PSNR

The graphical representation of performance comparison for various measuring methods with variation in number of training samples per class and the comparison among various measuring methods with variation in the number of features per sample image are also given below in figures 4.7. to 4.14.

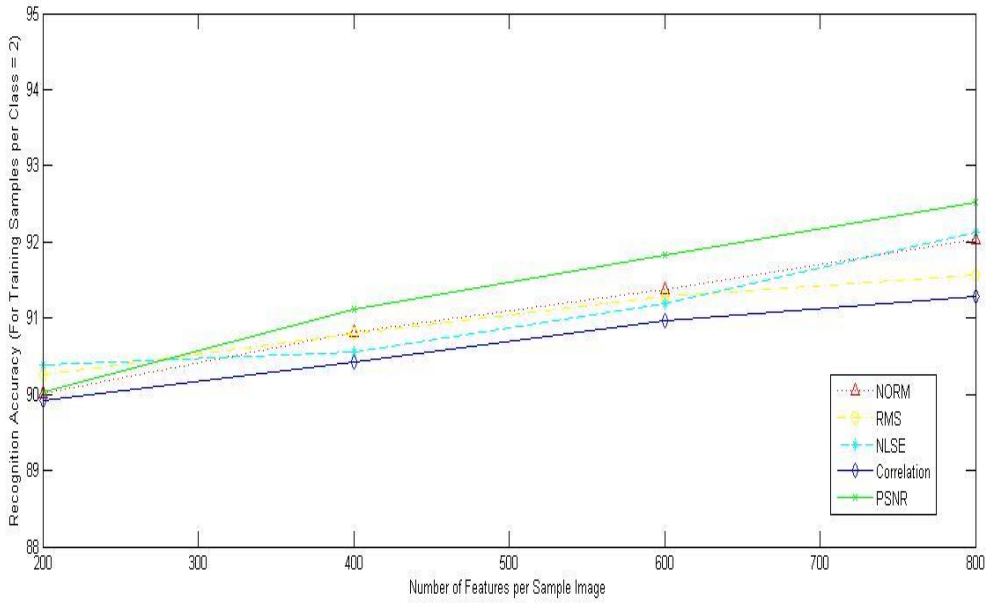


Figure 4.7. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 2)

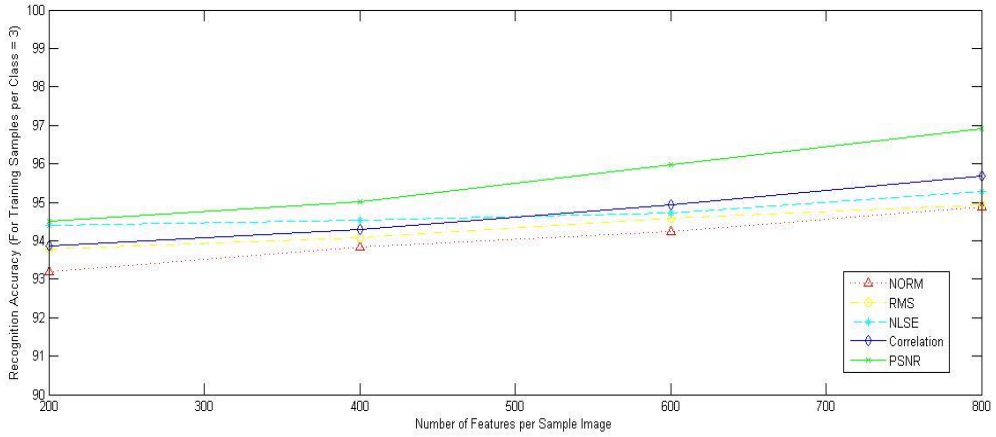


Figure 4.8. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 3)

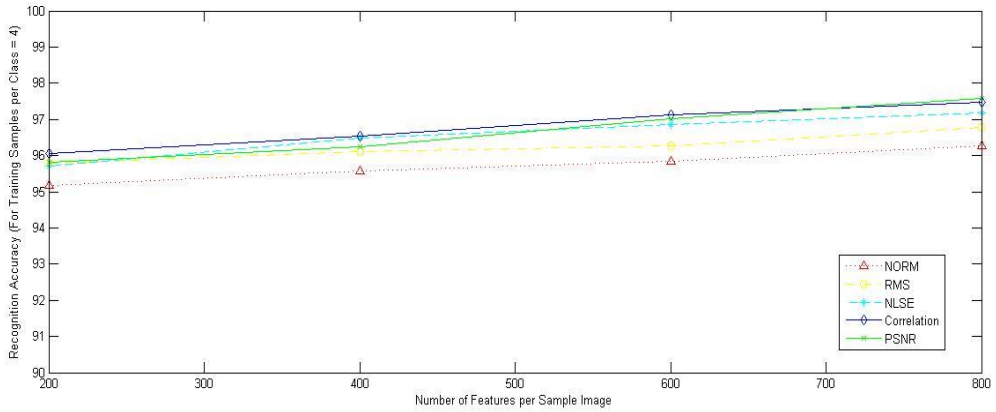


Figure 4.9. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 4)

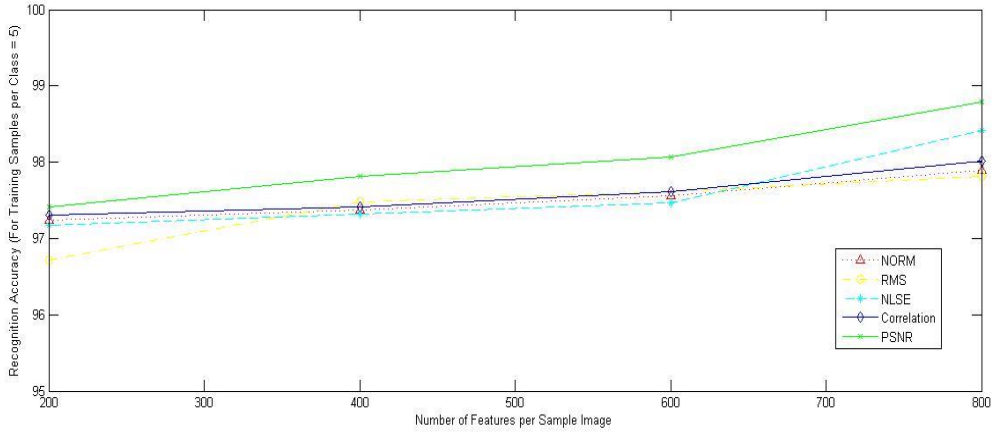


Figure 4.10. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 5)

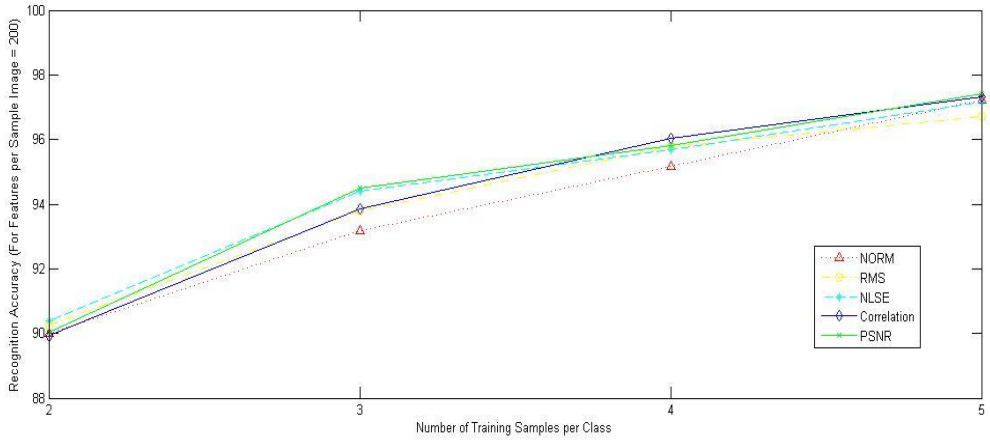


Figure 4.11. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 200)

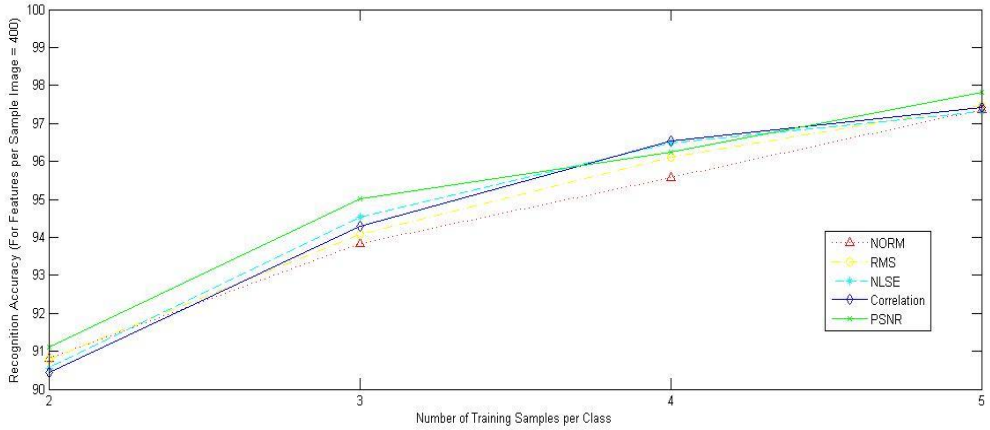


Figure 4.12. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 400)

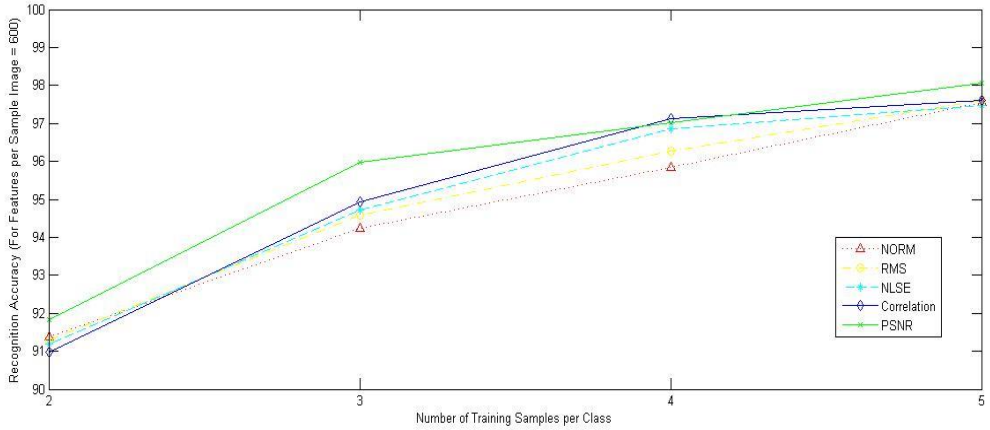


Figure 4.13. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 600)

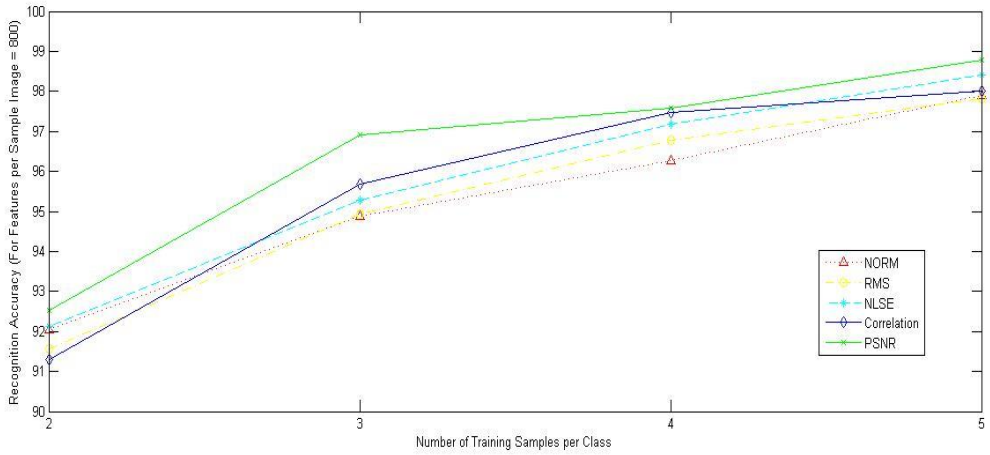


Figure 4.14. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 800)

From above experiments it can be easily seen that FLDE give a good recognition accuracy for FIR image of palm dorsa vein pattern. And this good recognition accuracy is because of addition of fuzzy membership to each data by using k-nearest neighbourhood.

It can also observe that for all measuring algorithm i.e. for Norm, RMS method, NLSE method, Correlation method, and PSNR method as the training samples per class is increased, the recognition accuracy is increased. Recognition accuracy is also increased with the increasing number of features per sample image. We can also observe that increasing the no of training sample increases recognition accuracy more.

From above experiments it is clear that from all the above measuring method by using Peak Signal to Noise Ratio (PSNR) for measuring we get the best performance i.e. best recognition accuracy.

We have also done this experiments on JU-NIR-V2: NIR Vein Database. The JU-NIR-V2: NIR Vein Database contains 570 images of 57 individuals where each person has 10 images. In this experiments all the images are resized to 32 X 25 pixels manually. Here we take 5, 4, 3, and 2 images per person as training data respectively and rest of images are taken for testing. Also we vary the number of features per sample 200, 150, 100 and 50 respectively. For each experiment we consider 25 nearest neighbours.

Firstly we have optimized the above parameter (training samples per class and features per sample image) and then we determine the recognition rate with quantitative measures like Norm, Root Mean Square (RMS), Normalized Least Square Error (NLSE), Correlation, and

Peak Signal to Noise Ratio (PSNR). The recognition accuracy by using various measuring method is given below in tabulation form, in tables 4.6. to 4.10.

Table 4.6.: Recognition rate calculated by NORM method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
			200	150	100	50
Image size = (32 x 32)	Nearest neighbours considered is = 25	5	100.0±0.0000	100.0±0.0000	100.0±0.0000	100.0±0.0000
		4	100.0±0.0000	100.0±0.0000	99.88±0.0037	99.65±0.0016
		3	100.0±0.0000	99.85±0.0033	99.72±0.0058	99.17±0.0080
		2	99.95±0.0009	99.73±0.0045	99.49±0.0072	98.70±0.0077

Table 4.7.: Recognition rate calculated by Root Mean Square (RMS) method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
			200	150	100	50
Image size = (32 x 32)	Nearest neighbours considered is = 25	5	100.0±0.0000	100.0±0.0000	100.0±0.0000	100.0±0.0000
		4	100.0±0.0000	99.99±0.0003	99.86±0.0024	99.67±0.0055
		3	99.95±0.0016	99.79±0.0041	99.64±0.0048	99.50±0.0079
		2	99.90±0.0018	99.59±0.0031	99.50±0.0047	99.34±0.0041

Table 4.8.: Recognition rate calculated by Normalized Least Square Error (NLSE) method:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
			200	150	100	50
Image size = (32 x 32)	Nearest neighbours considered is = 25	5	100.0±0.0000	100.0±0.0000	100.0±0.0000	99.81±0.0059
		4	100.0±0.0000	99.94±0.0018	99.86±0.0018	99.60±0.0036
		3	99.97±0.0018	99.83±0.0023	99.56±0.0036	99.16±0.0069
		2	99.82±0.0034	99.70±0.3730	99.32±0.0036	99.07±0.0065

Table 4.9.: Recognition rate calculated by Correlation method is given below:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	150	100	50
		5	100.0±0.0000	100.0±0.0000	100.0±0.0000	99.86±0.0043
		4	99.86±0.0044	99.73±0.0055	99.63±0.0059	99.45±0.0072
		3	99.64±0.0058	99.34±0.0063	99.21±0.0070	98.96±0.0084
		2	99.51±0.0048	99.14±0.0088	99.07±0.0063	98.83±0.0058

Table 4.10.: Recognition rate calculated by PSNR method is given below:

Image Size	Number of Nearest Neighbours to Be considered	Number of Training Samples per Class	Recognition Rate Mean ± Standard Deviation (%)			
			Number of Features Per Sample Image			
Image size = (32 x 32)	Nearest neighbours considered is = 25		200	150	100	50
		5	100.0±0.0000	99.86±0.0044	99.81±0.0037	99.72±0.0044
		4	99.85±0.0046	99.72±0.005	99.58±0.0072	99.47±0.0097
		3	99.64±0.0057	99.59±0.0055	99.34±0.0087	99.25±0.0099
		2	99.36±0.0063	99.14±0.0066	98.93±0.0077	98.89±0.0087

Graphical representation of variation of recognition accuracy with the variation of number of features in a sample and also with the number of training samples per class is presented as given below in figures 4.15. to 4.19.:

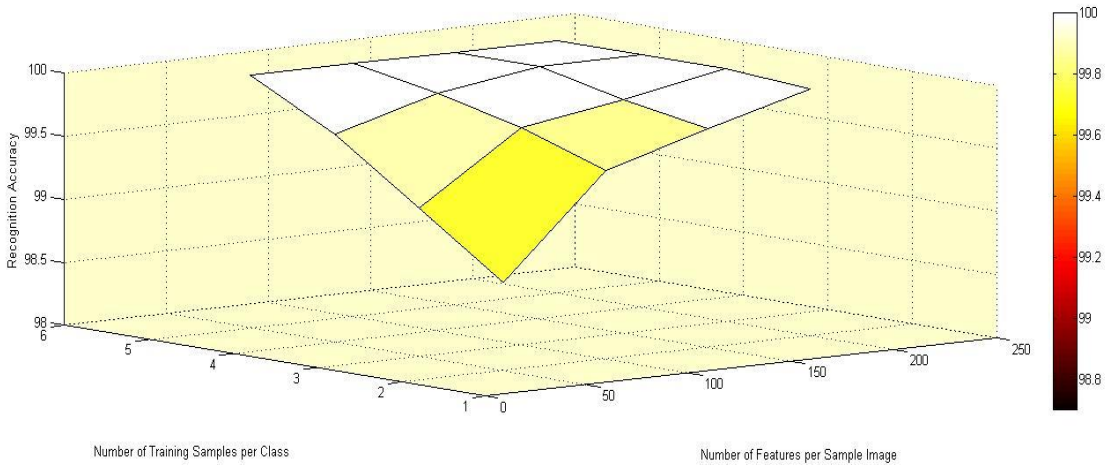


Figure 4.15. Graphical Representation of Recognition Accuracy Measured by Using NORM

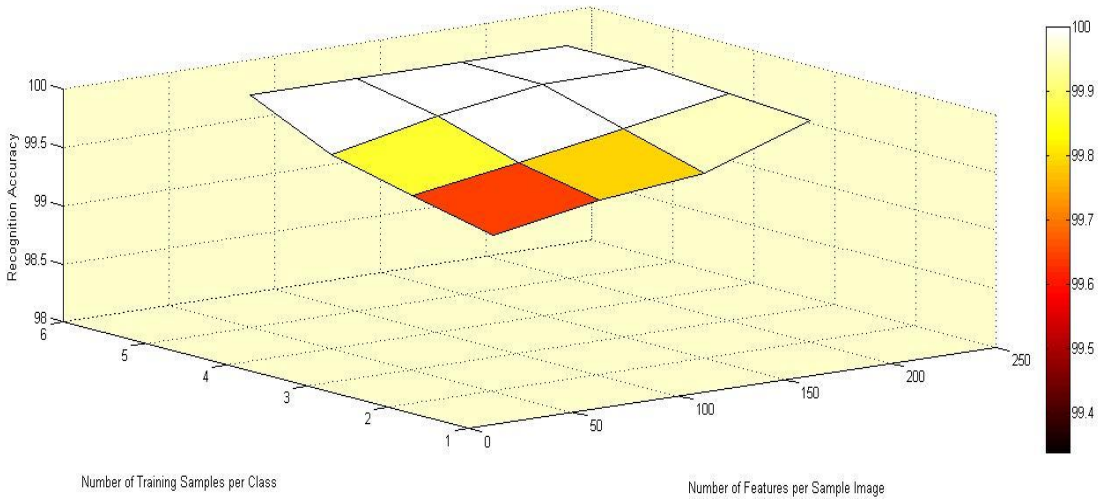


Figure 4.16. Graphical Representation of Recognition Accuracy Measured by Using RMS

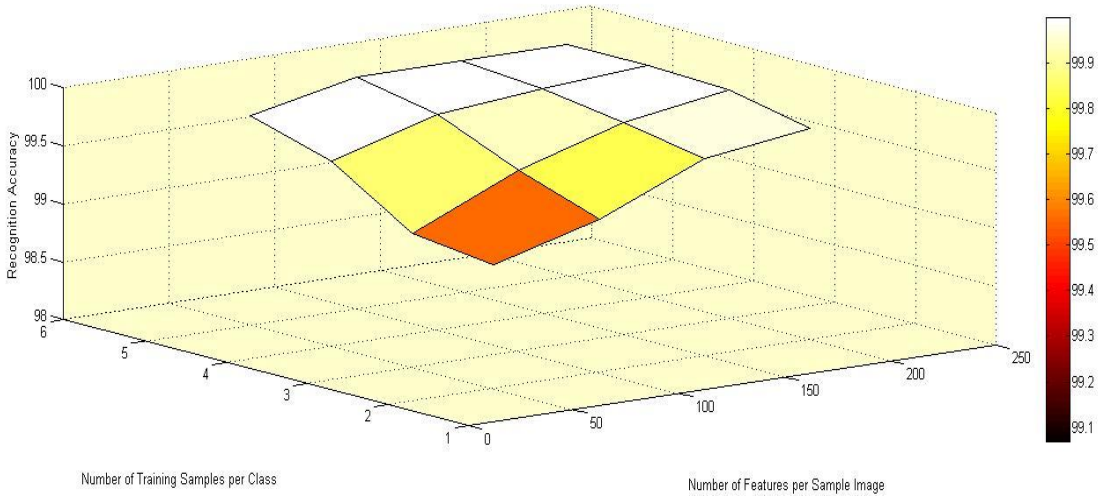


Figure 4.17. Graphical Representation of Recognition Accuracy Measured by Using NLSE

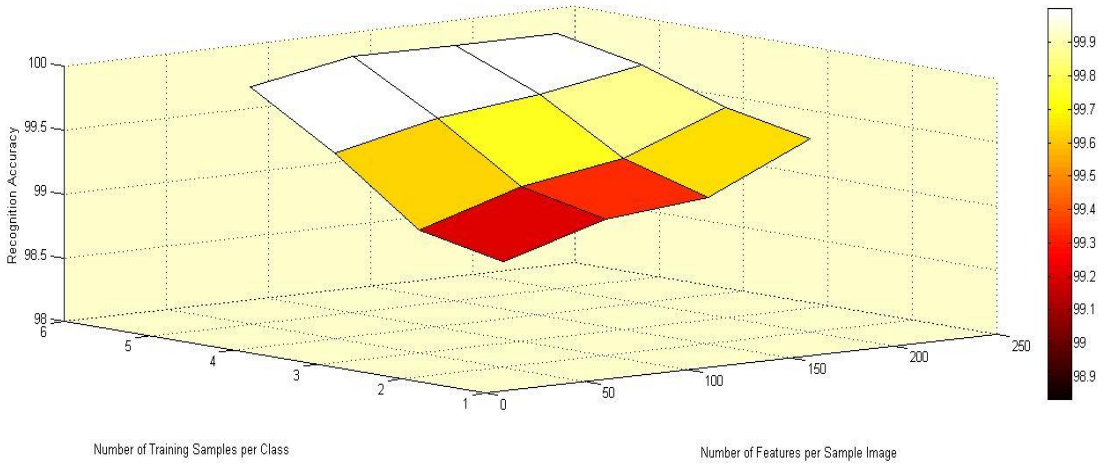


Figure 4.18. Graphical Representation of Recognition Accuracy Measured Using Correlation

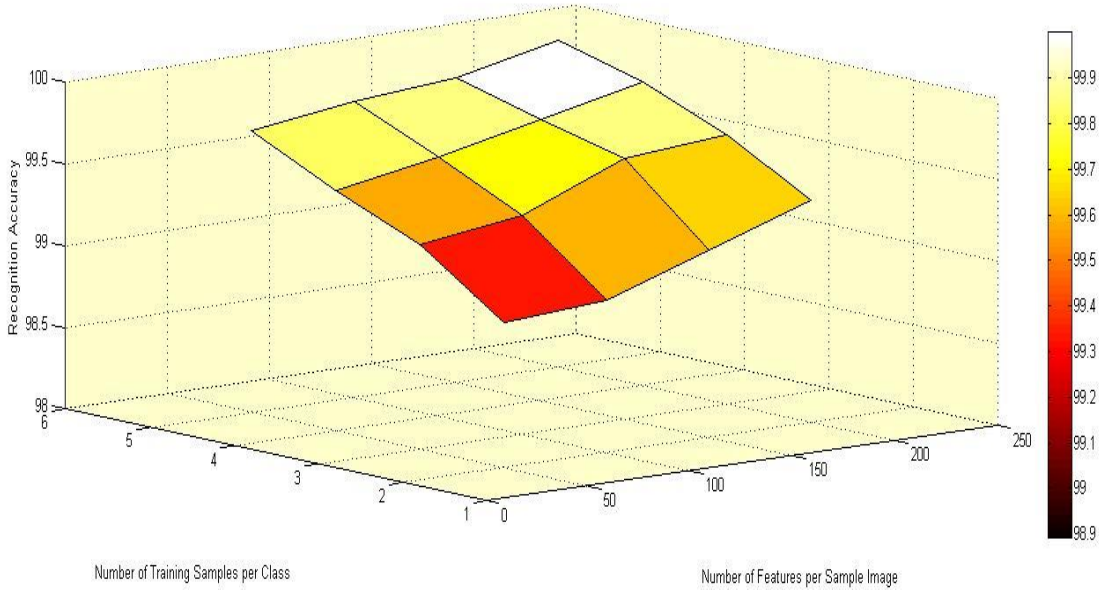


Figure 4.19. Graphical Representation of Recognition Accuracy Measured by Using PSNR

The graphical representation of performance comparison for various measuring methods with variation in number of training samples per class and the comparison among various measuring methods with variation in the number of features per sample are also given below in figures 4.20. to 4.27.:

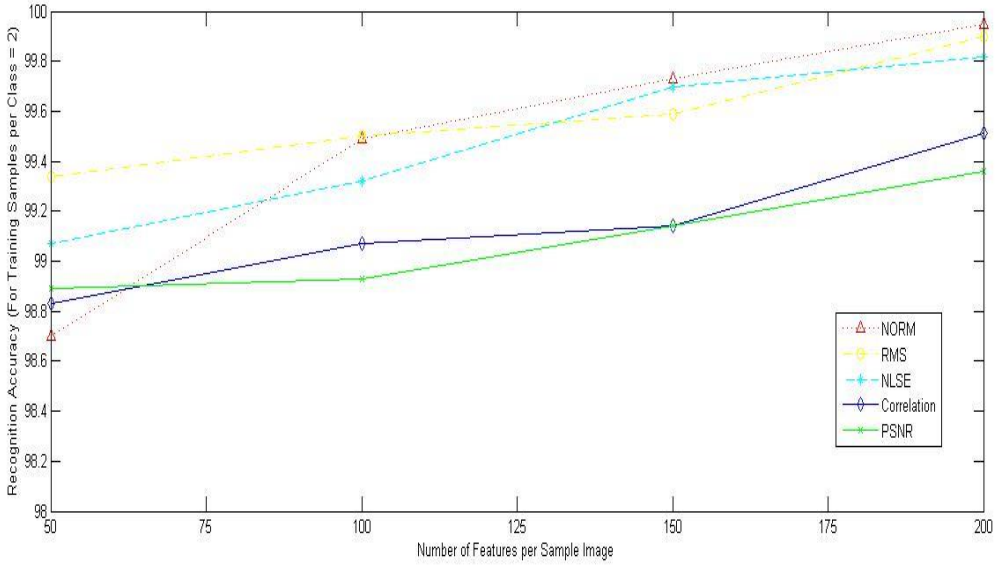


Figure 4.20. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 2)

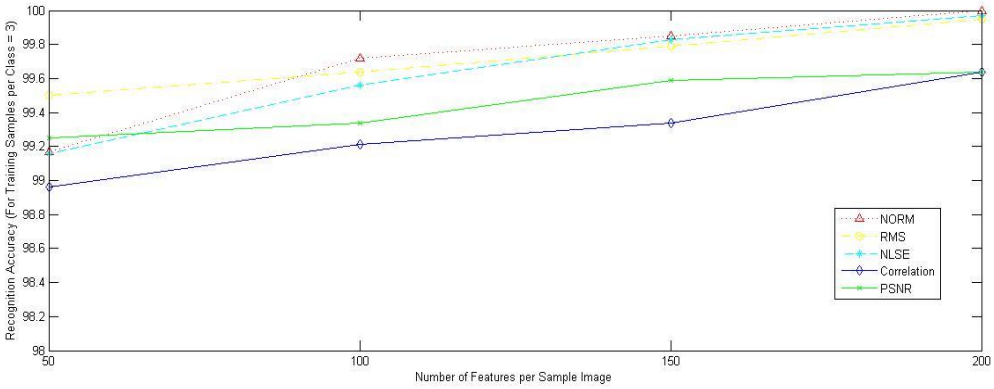


Figure 4.21. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 3)

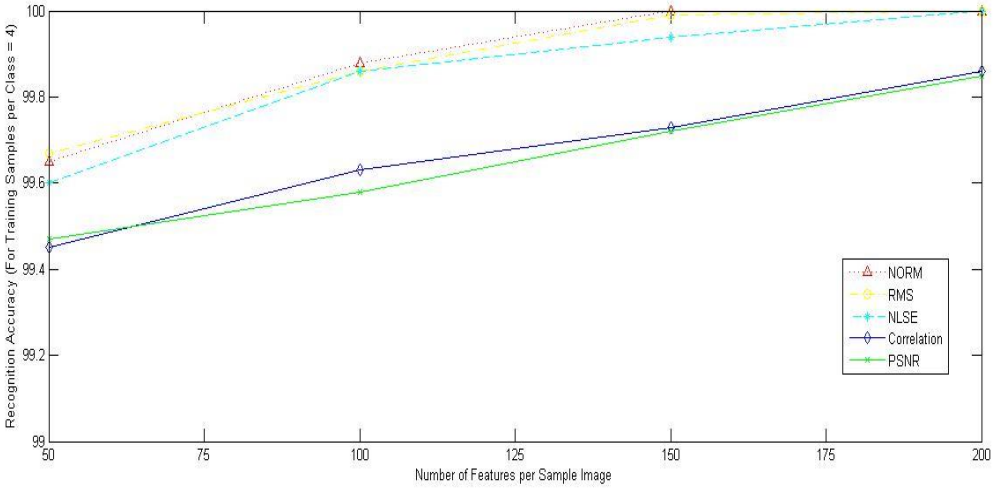


Figure 4.22. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 4)

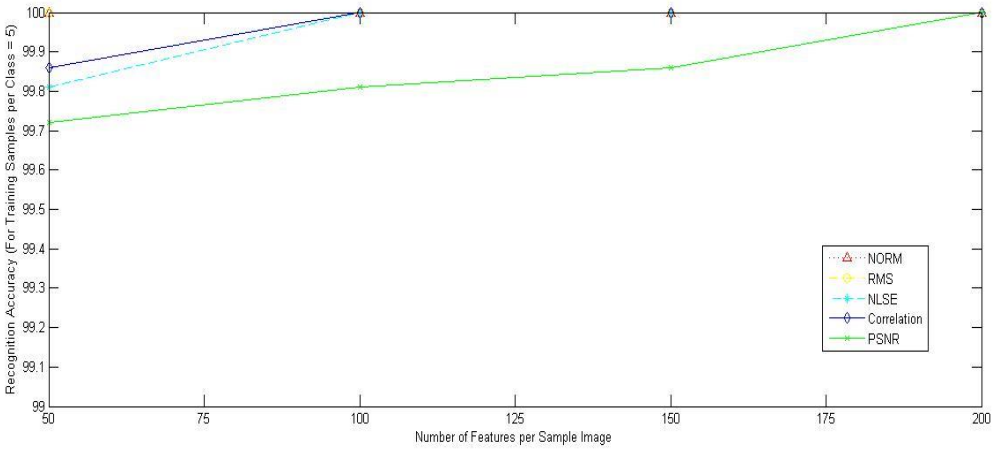


Figure 4.23. Performance Comparison among Measuring Methods with Varying Numbers of Features per Sample Image When Number of Training Samples per Class Is Fixed (= 5)

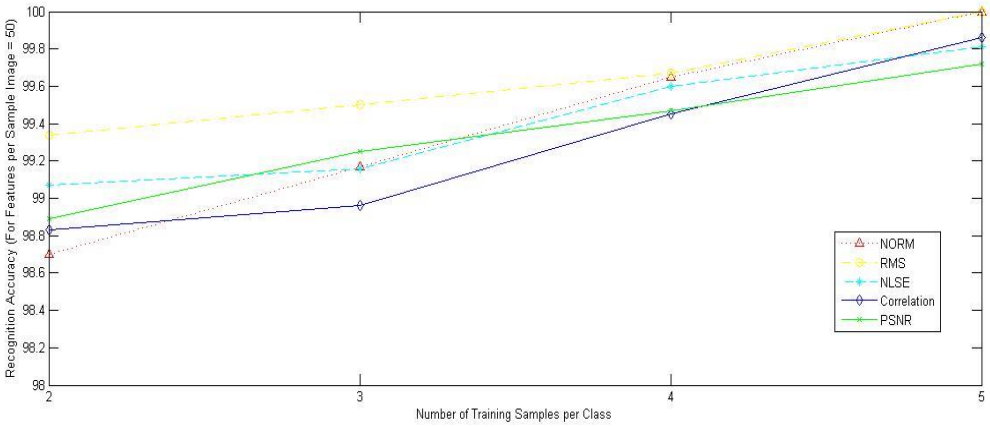


Figure 4.24. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 50)

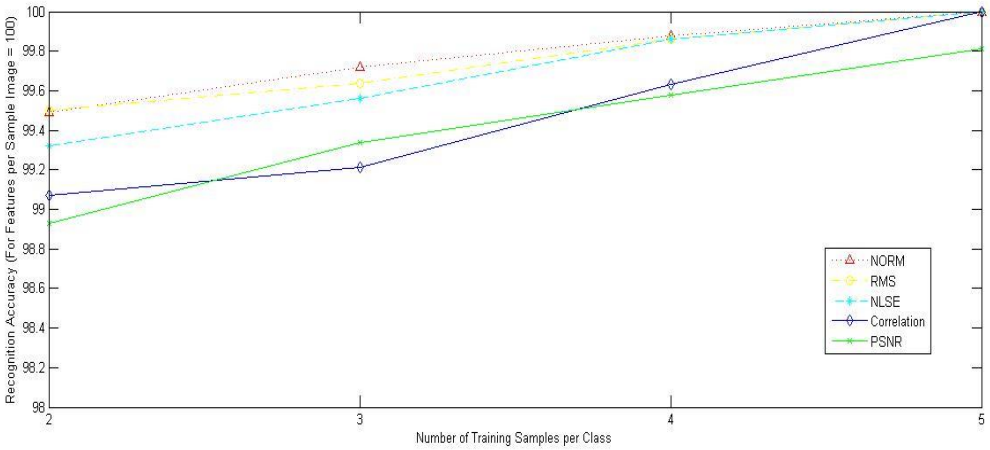


Figure 4.25. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 100)

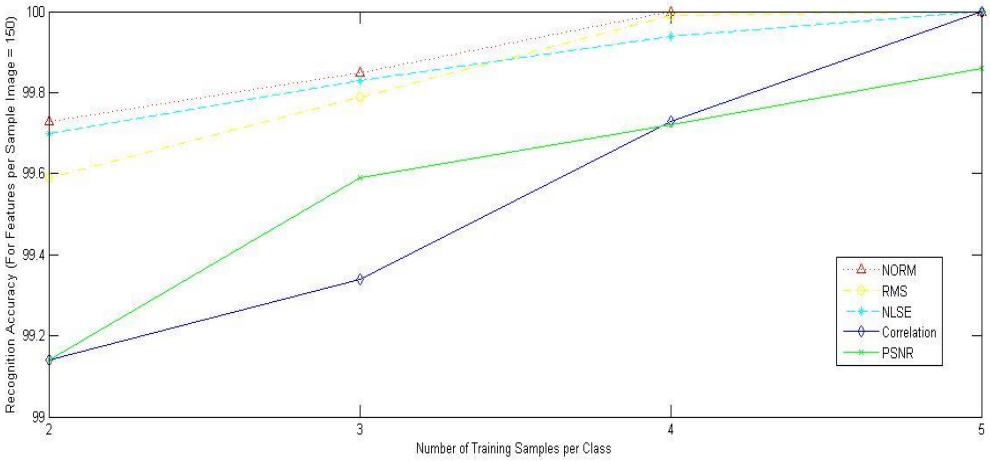


Figure 4.26. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 150)

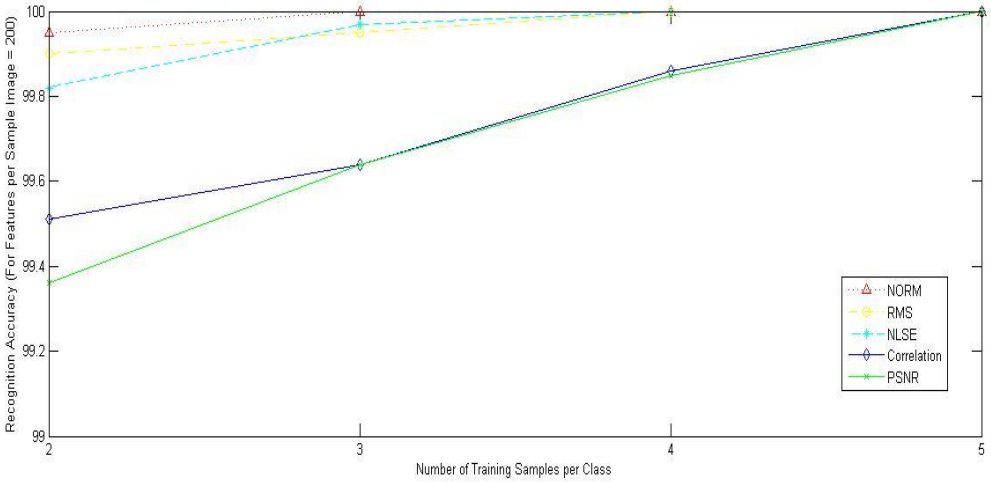


Figure 4.27. Performance Comparison among Measuring Methods with Varying Number of Training Samples per Class When Number of Features per Sample Image Is Fixed (= 200)

From above experiments it can be easily seen that FLDE give a good recognition accuracy for NIR image of palm dorsa vein pattern. And this good recognition accuracy is because of addition of fuzzy membership to each data by using k-nearest neighbourhood.

It can also observe that for all measuring algorithm i.e. for Norm, RMS method, NLSE method, Correlation method, and PSNR method as the training samples per class is increased, the recognition accuracy is increased. Recognition accuracy is also increased with the increasing number of features per sample image. We can also observe that increasing the no of training sample increases recognition accuracy more.

From above experiments it is clear that from all the above measuring methods by using Norm, Root Mean Square (RMS), Normalized Least Square Error (NLSE), and Correlation for measuring we get the best performance i.e. best recognition accuracy.

4.6 Summery:

The present chapter has shown how a state-of-the-art variant of LDE algorithm called FLDE algorithm can be effectively employed for thermal imaging of PDVP based biometric systems. Various performance measuring methods like Norm, RMS, Correlation, NLSE, and PSNR are used for measuring the recognition accuracies and their performance comparisons are also carried out.

In this methodology the membership degree of every data in every classes using fuzzy k nearest neighbourhood (FKNN) is computed. And this membership degree is incorporated into the

definition of the laplacian scatter matrix to obtain the fuzzy laplacian scatter matrices. And we can see that FLDE gives very good recognition accuracy in vein pattern based biometric authentication. Among all measuring methods (NORM, RMS, NLSE, Correlation, PSNR), gives better recognition accuracy.

CHAPTER 5

Conclusions and Future Scopes of Work

This chapter represent an eventual elucidation and provide a route to future research works.

CHAPTER 5

Conclusions and Future Scopes of Work

In this thesis some of the most exigent and contemporary issues related to general biometric pattern recognition have been addressed and analysed This thesis has effectively demonstrated how several state-of-the-art variants of local discriminant embedding (LDE) algorithms can be effectively utilized for dimensionality reduction and subsequent biometric authentication and recognition systems, developed using thermal imaging of palm dorsa vein patterns (PDVP). To be specific, the thesis has concentrated in developing such useful solutions utilizing two contemporary improved variants of LDE called Stable Orthogonal Local Discriminant Embedding (SOLDE) algorithm and Fuzzy Local Discriminant Embedding (FLDE) algorithm.

In chapter 1, a brief review of different types of biometric security system and their corresponding features, advantages and disadvantages are discussed. Thermal image based biometric and their features are explored in this chapter. In addition to these, features of an ideal biometric and mode of operation of a biometric security system is also given in this chapter.

In chapter 2, various dimension reduction methods like Linear Discriminant Analysis (LDA), Local Discriminant Embedding (LDE) are discussed. As high dimensional data increases the required space, computational burden and time, so reduction of feature dimension is very much necessary in image processing scenario. Variants of LDE like SOLDE, FLDE, and Kernel LDE are also given in brief in this chapter.

In chapter 3, Stable Orthogonal Local Discriminant Embedding is used for feature extraction and pattern classification based on real-time palm dorsa subcutaneous vein pattern. General LDE uses only the similarity between the data pairs belongs to same class. But SOLDE uses both the similarity and diversity between the data pairs belongs to same class in local neighbourhood. This extra additional feature make it a good approach for pattern recognition. Vein biometric based pattern recognition using SOLDE is probably first to use vein pattern in SOLDE. JU-NIR-V2: NIR database is used for experimental purpose. Experimental results of SOLDE shows that SOLDE gives a good recognition accuracy using vein pattern and also it is observed that for all measuring algorithm i.e. for Norm, RMS method, NLSE method,

Correlation method, and PSNR method as we increase the training sample per class the recognition accuracy is increased. Recognition accuracy is also increased with the increasing number of feature vectors per sample. Among above mentioned measuring methods NLSE gives better recognition accuracy than others measuring methods.

The present thesis work carried out can pave way for potential future works in several directions. For example, in future, the possible research scopes can pursue directions like:

i. More extensive performance evaluations of these algorithms for an enhanced database with more number of individuals.
ii. More extensive study on the weightage of the diversity matrix can be carried out and how the parameter α can be optimized for better performance, can also be studied.
iii. The experimentation of the SOLDE algorithm has been done here only for NIR images. We can experiment this on other types of infrared images e.g. FIR images.
iv. Gao *et al.* used SOLDE [69] for face based biometric authentication. In this thesis it has been utilized for thermal imaging of PDVP. Other biometric features like iris, fingerprint, ear, knuckle, and palm print etc. can also be explored for further future research.

In chapter 4, another variants of LDE i.e. Fuzzy Local Discriminant Embedding is discussed and used for real time palm dorsa subcutaneous vein pattern based biometric authentication system. In this methodology the membership degree of every data in every classes using fuzzy k nearest neighbourhood (FKNN) is computed. And this membership degree is incorporated into the definition of the Laplacian scatter matrix to obtain the fuzzy Laplacian scatter matrices. Experimental result shows that FLDE gives very good recognition accuracy in vein pattern based biometric authentication.

The future scopes for research in this context can be as follows:

i. We can also study the performance of this algorithm for an enhanced database with more number of individuals.
ii. Yang *et al.* used FLDE [77] for face based biometric authentication. In this thesis it has been utilized for thermal imaging of PDVP. Other biometric features like iris, fingerprint, ear, knuckle, and palm print etc. can also be explored for further future research.

iii. Different variants of FKNN can also be explored to develop new variants of FLDE algorithm which can be utilized to develop effective biometric solutions in future.

In addition to these research directions, development of similar biometric solutions using other variants of LDE e.g. Kernel LDE, two dimensional LDE etc. can also be explored for biometric recognition purposes.

REFERENCES

[1] B. Victor, K. Bowyer, and S. Sarkar, “An Evaluation of Face and Ear Biometrics”, Proceedings of the 16th IEEE International Conference on Pattern Recognition (ICPR). Vol. 1, pp. 429 - 432, 2002.

[2] M. De Marsico, M. Nappi, and D. Riccio, “FARO: Face Recognition against Occlusions and Expression Variations”, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol. 40, issue 1, pp. 121 - 132, January 2010.

[3] P. Melin, O. Mendoza, and O. Castillo, “Face Recognition with an Improved Interval Type-2 Fuzzy Logic Sugeno Integral and Modular Neural Networks”, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 41, issue 5, pp. 1001 - 1012, September 2011.

[4] W. Di, D. Zhang, D. Zhang, and Q. Pan, “Studies on Hyper-Spectral Face Recognition in Visible Spectrum With Feature Band Selection”, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, issue 6, pp. 1354 - 1361, November 2010.

[5] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, “Robust Face Recognition for Uncontrolled Pose and Illumination Changes”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 43, issue 1, pp. 149 - 163, January 2013.

[6] V.N. Boddeti and B.V.K.V. Kumar, “Extended-Depth-of-Field Iris Recognition Using Unrestored Wavefront-Coded Imagery”, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, issue 3, pp. 495 - 508, May 2010.

[7] N.D. Kalka, Z. Jinyu, N.A. Schmid, and B. Cukic, “Estimating and Fusing Quality Factors for Iris Biometric Images”, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, issue 3, pp. 509 - 524, May 2010.

[8] [Online]. http://en.wikipedia.org/wiki/Iris_recognition

[9] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, “Handbook of Fingerprint Recognition”, Springer Professional Computing, 2009.

[10] M. Kawagoe, and A. Tojo, “Fingerprint Pattern Classification”, Pattern Recognition, vol. 17, issue 3, pp. 295 - 303, 1984.

[11] [Online]. http://en.wikipedia.org/wiki/Fingerprint_recognition

[12] Z. Waheed, A. Waheed, and M. U. Akram, "A Robust Non-Vascular Retina Recognition System using Structural Features of Retinal Image", Proceeding of 13th International Bhurban Conference on Applied Science and Technology (IBCAST), pp. 101 - 105, 2016.

[13] [Online]. http://en.wikipedia.org/wiki/Retinal_scan

[14] A. Kumar, M. Hanmandlu, Vamsi K. Madasu, and Brian C. Lovell, "Biometric Authentication Based on Infrared Thermal Hand Vein Patterns", Digital Image Computing: Techniques and Applications, pp. 331 - 338, December 2009.

[15] S. Joardar, A. Chaterjee, A. Rakshit, "A Real-Time Palm Dorsa Subcutaneous Vein Pattern Recognition System Using Collaborative Representation-Based Classification", IEEE Transactions on Instrumentation and Measurement, vol. 64, issue 4, April 2015.

[16] S. Chakraborty, I. Bhattacharya, and A. Chatterjee, "A Palm-Print Based Biometric Authentication System Using Dual Tree Complex Wavelet Transform", Measurement, vol. 46, issue 10, pp. 4179 - 4188, December 2013.

[17] B. Zhang, W. Li, P. Qing, and D. Zhang, "Palm-Print Classification by Global Features," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 43, issue 2, pp. 370 - 378, August 2013.

[18] J.D. Bustard, and M.S. Nixon, "Toward Unconstrained Ear Recognition from Two Dimensional Images", IEEE Transactions Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, issue 3, pp. 486 - 494, May 2010.

[19] S. Banerjee, and A. Chaterjee, "Ear Recognition Using Force Field Transformation and Collaborative Representation-Based Classification with Single Training Sample per class", Intelligent Computing and Applications, vol. 343, pp. 511 - 517, 2015.

[20] A. Gangopadhyay, O. Chatterjee, and A. Chatterjee, "Hand Shape Based Biometric Authentication System Using Radon Transform and Collaborative Representation Based

Classification", Proceeding of 2013 IEEE 2nd International Conference on Image Information Processing (ICIIP – 2013), pp. 635 - 639, 2013.

[21] K. Usha, and M. Ezhilarasan, "Finger Knuckle Biometrics – A review", Computers and Electrical Engineering, vol. 45, pp. 249 - 259, July 2015.

[22] Lin Zhang, Lei Zhang, David Zhang, and H. Zhu, "Online Finger-Knuckle-Print Verification for Personal Authentication", Pattern Recognition, vol. 43, issue 7, pp. 2560 - 2571, July 2010.

[23] M. S Arya, N. Jain, J. Sisodia, and N. Sehgal, "DNA Encoding Based Feature Extraction for Biometric Watermarking", International Conference on Image Information Processing (ICIIP), 2011.

[24] G. Gautam, and D. kumar, "Biometric System from Heart Sound Using Wavelet Based Feature Set", International Conference on Communication and Signal Processing, pp. 551 - 555, April 2013.

[25] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, "A New Biometric Authentication System Using Heart Sounds Based on Wavelet Packet Features", IEEE International Conference on Electronics, Circuits, and Systems (ICECS), December 2015.

[26][Online].https://www.google.co.in/search?q=heart+sound+biometric+recognition&espv=2&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjyvM7Cp9fMAhWHPo8KHazzAzUQ_AUIBygB&biw=1366&bih=643#imgrc=w7rkv3npQMYfKM%3A

[27] A. Godil, P. Grother, and S. Ressler, "Human Identification from Body Shape", Proceedings of the Fourth International Conference on 3-D Digital Image and Modelling, pp. 386 - 392, October 2003.

[28] S. L. Wang, and A. W. C. Liew, “ICA-Based Lip Feature Representation for Speaker Authentication”, Proceedings of the Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, pp. 763 - 767, December 2008.

[29] A. Mehra, M. Kumawat, and R. Ranjan, “Expert System for Speaker Identification Using Lip Features With PCA”, Proceedings of the 2nd International Workshop on Intelligent Systems and Applications (ISA), pp. 1 - 4, 2010.

[30][Online].https://www.google.co.in/search?q=lip+biometrics&espv=2&source=lnms&tbm=isch&sa=X&ved=0ahUKEwikq9_K7qzMAhVLG44KHUUKAAAQ_AUICCgC&biw=1366&bih=599#imgrc=NkZ4uZzmdxmkAM%3A

[31] W. Ahmad, H. Karnick, and R. M Hegde, “Voice Based Biometric Authentication Using Collapsing Classes Discriminative Space Transform”, IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 1 - 6, 2014.

[32] E. Maiorana, P. Campisi, and A. Neri, “Template Protection for Dynamic Time Warping Based Biometric Signature Authentication”, Proceeding of the 16th International Conference on Digital Signal Processing, pp. 1 - 6, 2009.

[33] D. Muramatsu, K. Yasuda, and T. Matsumoto, “Biometric Person Authentication Method Using Camera-Based Online Signature Acquisition”, Proceedings of the 10th International Conference on Document Analysis and Recognition, pp. 46 - 50, 2009.

[34] T. Scheidat, C. Vielhauer, and J. Dittmann, “Biometric Hash Generation and User Authentication Based on Handwriting Using Secure Sketches”, Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis, pp. 89 - 94, 2009.

[35] Md. L. Ali, C. C. Tappert, M. Qiu, J. V. Monaco, “Authentication and Identification Methods Used in Keystroke Biometric System”, 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 17th International Symposium on Cyberspace Safety (CSS), and 2015 IEEE 12th International Conference on Embedded Software and Systems(ICESS), pp. 1424 - 1429, 2017.

[36] D. Gafurov, E. Snekkenes, and P. Bours, “Spoof Attacks on Gait Authentication System”, IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 491 - 502, September 2007.

[37] J. Zhang, J. Pu, C. Chen, and R. Fleischer, “Low-Resolution Gait Recognition”, IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 40, no. 4, pp. 986 - 996, August 2010.

[38] N. K. Negied, “Human Biometrics: Moving towards Thermal Imaging”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277 - 3878, vol. 2, issue 6, pp. 73 - 77, January 2014.

[39] S.-Yeung Cho, L. Wang, and W. J. Ong, “Thermal Imprint Feature Analysis for Face Recognition”, pp. 1875-1880, IEEE International Symposium on Industrial Electronics (ISlE 2009) Seoul Olympic Parktel, July 2009.

[40] A. M. Guzman, M. Goryawala, J. Wang, A. Barreto, J. Andrian, N. Rishe, and M. Adjouadi, “Thermal Imaging as a Biometrics Approach to Facial Signature Authentication”, IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 1, pp. 214 - 222, January 2013.

[41] I. Pavlidis, P. Buddharaju, C. Manohar, P. Tsiamyrtzis, “Biometrics: Face Recognition in Thermal Infrared”.

[42][online].http://spie.org/Images/Graphics/Newsroom/Imported-2010/003230/003230_10_fig1.jpg

[43] A. Kumar, M. Hanmandlu, V. K. Madasu, B. C. Lovell, “Biometric Authentication Based on Infrared Thermal Hand Vein Patterns”, Digital Image Computing: Techniques and Applications, pp. 331 - 338, December 2009.

[44] A. Yuksel, L. Akarun, and B. Sankur, “Hand Vein Biometry Based on Geometry and Appearance Methods”, IET Computer Vision, vol. 5, issue 6, pp. 398 - 406, November 2011.

[45][Online].https://www.google.co.in/search?q=gait%20biometric&source=lnms&tbm=isch&sa=X&ved=0ahUKEwi-r-rj87DMAhUH5KYKHe3aAscQ_AUICCgC&biw=1024&bih=667#imgdii=BoxDoSlx5_aKqM%3A%3BBoxDoSlx5_aKqM%3A%3B34plIpWdiN6m6M%3A&imgrc=BoxDoSlx5_aKqM%3A.

[46][Online].https://www.google.co.in/search?q=vein%20pattern%20based%20on%20thermal%20image&biw=1024&bih=667&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjGhICe7bDMAhXHGaYKHSxPBqcQ_AUIBigB#imgrc=xXhI8ij1ogUGeM%3A.

[47] L. I. Smith, “A Tutorial on Principal Components Analysis”, February 26, 2002.

[48] [Online]. https://en.wikipedia.org/wiki/Principal_component_analysis

[49] [Online]. https://en.wikipedia.org/wiki/Linear_discriminant_analysis

[50] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, Science, vol. 290, no. 5050, pp. 2319 - 2323, 2000.

[51] [Online]. https://en.wikipedia.org/wiki/Isomap

[52] M. Belkin, and P. Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”, Neural Computation, vol. 15, no. 6, pp. 1373 - 1396, 2003.

[53][Online].https://en.wikipedia.org/wiki/Nonlinear_dimensionality_reduction#Laplacian_eigenmaps.

[54] S. T. Roweis, and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” Science, vol. 290, no. 5050, pp. 2323 - 2326, 2000.

[55] [Online]. https://en.wikipedia.org/wiki/Nonlinear_dimensionality_reduction

[56] H. T. Chen, H. W. Chang, and T. L. Liu, "Local Discriminant Embedding and Its Variants", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 846 - 853, 2005.

[57] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph Embedding and Extensions: A General Framework for Dimensionality Reduction", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, issue 1, pp. 40 - 51, January 2007.

[58] [Online]. https://en.wikipedia.org/wiki/Multidimensional_scaling

[59] M. Sugiyama, "Dimensionality Reduction of Multimodal Labelled Data by Local Fisher Discriminant Analysis", The Journal of Machine Learning Research, vol. 8, no. 27, pp. 1027 - 1061, 2007.

[60][Online].https://en.wikipedia.org/wiki/Nonlinear_dimensionality_reduction#Manifold_learning_algorithms

[61] C. Hou, F. Nie, C. Zhang, and Y. Wu, "Learning an Orthogonal and Smooth Subspace for Image Classification", IEEE Signal Processing Letters, vol. 16, no. 4, pp. 303 - 306, April 2009.

[62] J. Ye and T. Xiong, "Null Space versus Orthogonal Linear Discriminant Analysis", Proceedings of the 23rd International Conference on Machine Learning, pp. 1073 - 1080, 2006.

[63] D. Cai, X. He, J. Han, and H. Zhang, "Orthogonal Laplacianfaces for Face Recognition", IEEE Transactions on Image Processing, vol. 15, issue 11, pp. 3608 - 3614, November 2006.

[64] T. Zhang, K. Huang, X. Li, J. Yang, and D. Tao, "Discriminative Orthogonal Neighbourhood-Preserving Projections for Classification", IEEE Transactions System, Man, and Cybernetics, Part B (Cybernetics) , vol. 40, no. 1, pp. 253 - 263, February 2010.

[65] Q. Gao, H. Xu, Y. Li, and D. Xie, “Two-Dimensional Supervised Local Similarity and Diversity Projection”, Pattern Recognition, vol. 43, no. 10, pp. 3359 - 3363, 2010.

[66] K. Q. Weinberger, and L. K. Saul, “An Introduction to Nonlinear Dimensionality Reduction by Maximum Variance Unfolding”, Proceedings of the 21st Association of Advance Artificial Intelligence, pp. 1683 - 1686, 2006.

[67] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, “Face Recognition Using Laplacianfaces”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, issue 3, pp. 328 - 340, March 2005.

[68] J. Duchene, and S. Leclercq, “An Optimal Transformation for Discriminant and Principal Component Analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 6, pp. 978 - 983, June 1988.

[69] Q. Gao, J. Ma, H. Zhang, X. Gao, and Y. Liu, “Stable Orthogonal Local Discriminant Embedding for Linear Dimensionality Reduction”, IEEE Transactions on Image Processing, vol. - 22, no. - 7, pp. 2521-2531, July 2013.

[70] S. Wang, J. Lu, X. Gu, H. Du, J. Yang, “Semi-Supervised Linear Discriminant Analysis for Dimension Reduction and Classification”, Pattern Recognition, vol. 57, pp. 179 - 189, 2016.

[71] S. Balakrishnama, A. Ganapathiraju, “Linear Discriminant Analysis - A Brief Tutorial”.

[72] [Online]. https://en.wikipedia.org/wiki/Linear_discriminant_analysis

[73] B. Lia, C. H. Zhengc, D. S. Huanga, “Locally Linear Discriminant Embedding: An Efficient Method for Face Recognition”, Pattern Recognition, pp. 3813 - 3821, 2008.

[74] H. T. Chen, H. W. Chang, and T. L. Liu, “Local Discriminant Embedding and Its Variants”, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 846 - 853, 2005.

[75] G. Hua, M. Brown, and S. Winder, “Discriminant Embedding for Local Image Descriptors”, Proceedings of the IEEE 11th International Conference on Computer Vision, pp. 1 - 8, October 2007.

[76] K. C. Kwaka, and W. Pedrycz, “Face Recognition Using a Fuzzy Fisherface Classifier”, Pattern Recognition 38, pp. 1717 - 732, 2005.

[77] M. Wan, G. Yang, Z. Lai, Z. Jin, “Feature Extraction Based on Fuzzy Local Discriminant Embedding with Applications to Face Recognition”, The Institution of Engineering and Technology, pp. 301 - 308, 2011.

[78] Zadeh L.A., “Fuzzy sets”, Information and Control 8, vol. 8, issue 3, pp. 338 - 353, 1965.

[79] J.M. Keller, M.R. Gray, and J.A. Givens, “A Fuzzy k-Nearest Neighbour Algorithm”, IEEE Transactions on Systems, Man, and Cybernetics, vol. 15, issue 4, pp. 580 - 585, 1985.

[80] S. Joardar, “Biometric Security Systems Using Systems Using Far and Near Infrared Imaging of Palm Dorsa Vein Pattern” MEE thesis, 2014.

[81] S. Banerjee, “Ear Based Biometric Recognition Employing Sparse and Collaborative Representation”, MEE thesis, 2015.

[82] G. K. O. Michael, T. Connie, and A. B. J. Teoh, “A Contactless Biometric System Using Palm Print and Palm Vein Features”, Advanced Biometric Technologies, Dr. G. Chetty (Ed.), InTech Publishers, DOI: 10.5772/19337, pp. 155 - 178, 2011.

[83] S. Joardar, A. Chaterjee, A. Rakshit, “Real-time NIR imaging of Palm Dorsa subcutaneous vein pattern based biometrics: An SRC based approach”, IEEE Instrumentation & Measurement Magazine, vol. 19, issue 2, pp. 13 – 19, April 2016.

[84] [Online]. "User Manual Model: FI8918W Indoor Pan/Tilt Wireless IP Camera", http:// foscam.us/downloads/FI8918W%20User%20Manual.pdf

[85] D.G. Lowe, "Distinctive image features from scale-invariant key-points", International Journal of Computer Vision, vol. 60, issue 2, pp. 91 - 110, November 2004.

[86] [Online]. "Introducing the Raspberry Pi 2 Model B". https://learn.adafruit.com/downloads/pdf/introducing-the-raspberry-pi-2-model-b.pdf

[87] [Online]. "N-150 4-Port Wireless Router JNR 1010 User Manual" http://www.downloads.netgear.com/files/GDC/JNR1010/JNR1010_UM_10Aug12.pdf

[88] [Online]. http://www.sonel.pl/sites/default/files/en/datasheet/datasheet_KT-384_en_v3.pdf)