Department of Mechanical Engineering
University of Malaga

Ph.D. Thesis

Robust Image Recognition Based on a New Supervised Kernel Subspace Learning Method

Doctoral Program:

Mechanical Engineering and Energy Efficiency

Presented by:

Ali Khalili Mobarakeh

Under Supervision of:

Prof. Juan Antonio Cabrera Carrillo
Dr.Juan Jesús Castillo Aguilar

Malaga, April 2019

D. JUAN ANTONIO CABRERA CARRILLO, Catedrático de Universidad del Área de Ingeniería Mecánica de la Universidad de Málaga y D. JUAN JESÚS CASTILLO AGUILAR, Profesor de Universidad del Área de Ingeniería Mecánica de la Universidad de Málaga.

CERTIFICA: que D. Ali Khalili Mobarakeh, Ingeniero Electrónico por la Universidad de Sains Malaysia, ha realizado bajo nuestra dirección la tesis doctoral que tiene por título "Robust Image Recognition Based on a New Supervised Kernel Subspace Learning Method". Y que ha alcanzado los objetivos de investigación propuestos, estando debidamente cualificada para su defensa.

Málaga a seis de septiembre de dos mil diecinueve.

Fdo: Juan Antonio Cabrera Carrillo

Fdo: Juan Jesús Castillo Aguilar

Acknowledgment

First of all, I am highly indebted to my supervisor Prof. Cabrera for his guidance, constant supervision and kind advice throughout my Ph.D. research studies. It was a real privilege and honor for me to work under his supervision and also of his extraordinary human qualities.

Second, I would like to express my great appreciation to my co-supervisor Doctor Juan Jesús Castillo Aguilar for his kind support, availability and constructive recommendations, especially during our last publication, which was determined for the accomplishment of this research work.

Third, I gratefully acknowledge the funding received from the University of Malaga towards my last article publication, which has been published in MDPI journals, Optical Sensors section.

Finally, I would like to express my gratitude to my lovely wife for her kind helps and being patient during my Ph.D. program and also to my darling parents for their kind support and encouragement which helped me a lot during this research.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

DNE	Discrimination Neighbor Embedding
KPCA	Kernel Principal Component Analysis
LPP	Locality Preserving Projection
LDA	Linear Discriminant Analysis
LDNE	Locality-Based Discriminant Neighborhood Embedding
NIR	Near-Infrared
PCA	Principal Component Analysis
ROI	Region of Interest
SKLDNE	Supervised Kernel Locality-Based Discriminant Neighborhood Embedding
SSS	Small Sample Size problem
UDP	Unsupervised Discriminant Projection

Robust Image Recognition Based on a New Supervised Kernel Subspace Learning Method

Abstract

Image recognition is a term for computer technologies that can recognize certain people, objects or other targeted subjects through the use of algorithms and machine learning concepts. Face recognition is one of the most popular techniques to achieve the goal of figuring out the identity of a person. This study has been conducted to develop a new non-linear subspace learning method named "supervised kernel locality-based discriminant neighborhood embedding," which performs data classification by learning an optimum embedded subspace from a principal high dimensional space. In this approach, not only is a nonlinear and complex variation of face images effectively represented using nonlinear kernel mapping, but local structure information of data from the same class and discriminant information from distinct classes are also simultaneously preserved to further improve final classification performance. Moreover, to evaluate the robustness of the proposed method, it was compared with several well-known pattern recognition methods through comprehensive experiments with six publicly accessible datasets. In this research, we particularly focus on face recognition however, two other types of databases rather than face databases are also applied to well investigate the implementation of our algorithm. Experimental results reveal that our method consistently outperforms its competitors across a wide range of dimensionality on all the datasets. SKLDNE method has reached 100 percent of recognition rate for Tn=17 on the Sheffield, 9 on the Yale, 8 on the ORL, 7 on the Finger vein and 11on the Finger Knuckle

respectively, while the results are much lower for other methods. This demonstrates the robustness and effectiveness of the proposed method.

1 INTRODUCTION

1.1 Overview

Reliable identification is a very important thing for many applications such as airport security and border control[1, 2]. Sometimes we hear about computer breakdown by hackers, bank security breaches and credit card hacking. A fundamental flaw in conventional access control systems was taken advantage of by criminals in most crimes. These systems cannot identify humans by “who we are” but by “what we have”. Such systems recognize humans through passwords and ID cards. Therefore, these systems are very unreliable. If you lose your ID card or credit card, they might get hacked. So we have to find solutions to this problem, which are temper-proof to assure security. The best choice, in this case, is to use something of the same person who is supposed to be identified or verified. Human faces are the most suitable means for this purpose [3, 4].

Face recognition is one of the most successful ways to reach the goal of figuring out who somebody is [5-7]. Several methods and algorithms for face recognition have been proposed recently. Among them, the most well-known subspace learning methods are[8], Principal Component Analysis (PCA)[8], Kernel Principal Component Analysis (KPCA), Discrimination Neighbor Embedding (DNE), Locality Preserving Projection (LPP) [9, 10], Unsupervised Discriminant Projection (UDP), Linear Discriminant Analysis (LDA)[11] [12, 13] and Locality-Based Discriminant Neighborhood Embedding (LDNE). In this particular thesis, these methods included our proposed recognition technique implemented in some different face databases to illustrate the most prominent method for face recognition in surveillance systems [14, 15].

1.2 Biometrics

Any technology which connects people's identity to their physical or behavioral characteristics to provide security and safety is called Biometrics[16]. Numerous behavioral or physical characteristics exist that can be recognized by biometric technology such as fingerprint and finger vein, iris, DNA, retina, voice. Two main factors in this field are time and accuracy which means that the highest priority in biometric systems is how to identify with maximum accuracy in a minimum of time. This shows why computers are applied to identify and verify people [11, 17].

Although access cards or passwords or a combination of both are very useful, they can easily be fraudulently used by criminals. Consequently, we cannot be totally sure of their safety and security. Therefore, to increase the level of security, a robust method must be found to solve this important problem related to these applications. As has already been mentioned, the best option is to use something of the person to be identified or verified. A human face is the most appropriate means for this purpose because of the following two main factors:

-Availability- photos of the person who has to be identified or verified are easily available.

-Convenience- it minimizes the issue of hygiene problems and photo can be taken without the person noticing it as it is contactless so it improves user acceptance.

1.3 Face recognition

To perform face recognition, many methods and techniques have been used. However, the extraction of important information and the dimensionality reduction of a photo without losing important information are the most common problems in all these techniques. Considering a general overview on how face recognition is done, it can be said that since the dimension of the original photo is too high and includes noise, it is better to determine merely the important factors which will consequently reduce the entire dimension. Extracting the important features from the photo is another important key to face recognition. Therefore, while these special features are extracted and the dimensionality is successfully reduced, then the comparison of the photos can easily be done and the scheme will inevitably determine which photo belongs to which category[18]. Figure 1.1 illustrates an example of how face recognition generally works[19].

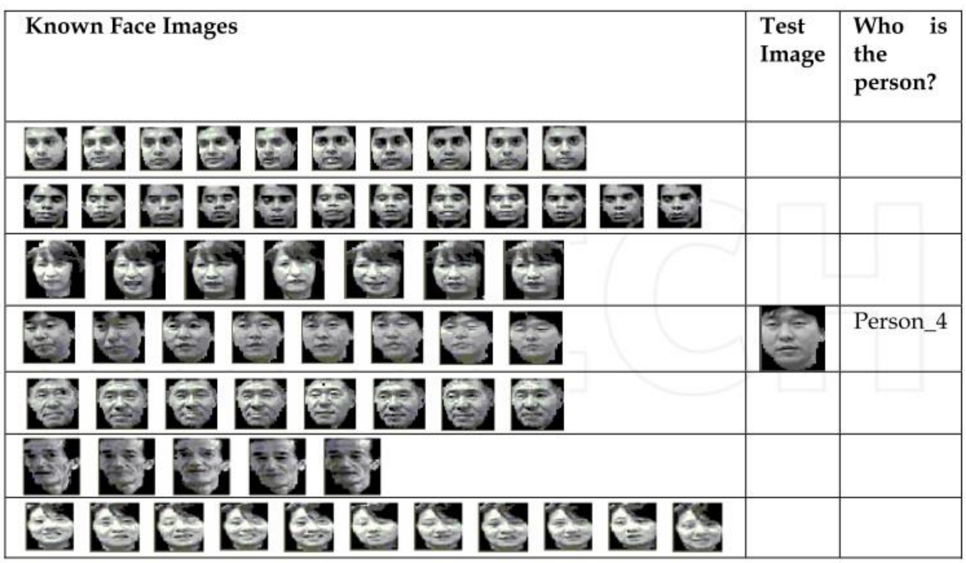


Figure 1.1. An example of face recognition scenario [19]

1.3.1 Identification and verification

It is very important to understand that identification and verification are two completely different fields. Identification means we don't know "who this person is" and we are trying to identify him/her. For example, when you see someone on the street and this person says" hi", then first you look at the person's face and your mind tries to recognize this person using the information which has previously been taken from that person's face. This process is the same in the biometric identification solution. It means that you have a lot of images stored in your mind (database). When you see an unknown person, you take a photo of this person and your biometric system tries to compare this picture with all the pictures in your database and return the information about this person to determine who this person is [20, 21]. Identification systems are the technology in which the image of a face is compared to all the images in the database to determine whose image the input data belong to which is called a "one-to-many" process. Verification is the process of verifying a person's identity. For instance, somebody claims that they are specific person and shows some information such as an ID card or passport. Then you try to compare their image with a specific person's image in your mind (database). Your mind will return a positive or negative response which indicates that a person is really who they claim to be [22]. In this thesis, merely identification is the field we purpose to focus on.

1.4 **Recognition methods**

As already mentioned, the most well-known subspace learning methods to be used in this project are, Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA)[23], Discrimination Neighbor Embedding (DNE), Locality Preserving

Projection (LPP), Unsupervised Discriminant Projection (UDP), Linear Discriminant Analysis (LDA) and Locality-Based Discriminant Neighborhood Embedding (LDNE).

PCA aims to preserve global geometric information for representation by maximizing the trace of the feature covariance matrix[24]. KPCA is an extension of PCA in which data is first mapped and then PCA is applied to the mapped data. One of the most widely used representations of face recognition is Eigenfaces, which is based on the principal component analysis. The Eigenface algorithm uses the principal component analysis (PCA) for dimensionality reduction and to find the vectors of those that best account for the distribution of face images within the entire face image spaces.

However, both PCA and KPCA can always suffer from the Small Sample Size (SSS) problem, especially in the case of excitant outliers, which dramatically decreases the final recognition rate. LDA aims to find global discriminant information for classification by maximizing the ratio between inter-class and intra-class scatters. However, LDA can also suff er from the Small Sample Size (SSS) problem. LPP is an unsupervised linear subspace learning method that finds graph embedding, which can well preserve local information for detecting the intrinsic manifold structure. Since the "over-learning of locality" problem still exists in LPP, the multi-manifolds for diff erent classes cannot be well achieved, which could degrade classification performance. UDP, which is a successful extension of LPP, is a linear subspace learning method to find graph embedding, which can well preserve local information for distinguishing the intrinsic manifold structure. Since the "overlearning of locality" problem still exists in UDP, which means the multi-manifolds for different classes cannot be well achieved and the classification performance can be degraded. Discrimination Neighbor Embedding (DNE) is also an effective dimensionality reduction technique although this method cannot preserve the local and geometrical structure information of data so the recognition rate will be highly degraded. Recently, a new supervised subspace learning method, called Locality-Based Discriminant Neighborhood Embedding (LDNE) has been proposed [25], which considers both the "locality" in LPP and the "discrimination" in DNE in an integrated modeling environment. The embedding yielded by LDNE cannot only preserve local structure information of data of the same class but can also obtain more discriminant information from diff erent classes which effectively improve classification performance.

1.5 Problem statement

In today's society identification plays an important role. Identity recognition tries to answer the aforementioned questions: whether or not the individual is really whom he/she claims to be; whether or not a specific person's records and information are available, whether or not a particular person has permission to enter the system. Face image is one of the most suitable means for this purpose. The highest priorities in these fields are actually how to recognize and identify with maximum accuracy. This explains why we tend to use a new method of classification to classify our data. In image recognition, dimensionality reduction is an effective technique to solve the "curse of dimensionality", and improve classification performance and computational efficiency in many applications. However, most of the existing dimensionally reduction techniques could suffer from the Small Sample Size (SSS) problem. Some of them also might fail to discover the essential nonlinear data structure hidden in the input space. "Overlearning of locality" and the "out-of-sample" are other existing problems in regards to pattern recognition.

In this thesis, to handle the aforementioned problems, a novel supervised subspace learning method named "supervised Kernel Locality-Based Discriminant Neighborhood Embedding" (SKLDNE) is proposed, in which not only is nonlinear and complex variation of face images effectively represented using nonlinear kernel mapping, but local structure information of data from the same class and discriminant information from distinct classes are also simultaneously preserved to further improve final classification performance.

1.6 **Research objectives**

Here are the main targets of this research:

- To develop an algorithm for face recognition utilizing our proposed supervised subspace learning method.
- To deal with complicated problems as many effective nonlinear data features may be lost during the classification process using linear techniques.
- To get benefits from the advantages of "locality" in LPP in which, due to the prior class-label information, geometric relations are preserved.
- To build a compact submanifold to preserve 'discrimination' information.
- To resolve the SSS problem, which is mostly faced by other techniques such as PCA, LDA, UDP, and LPP, as well as the "overlearning of locality" problem in the manifold learning.

- To investigate the performance of a new proposed algorithm compared with the state-of-the-art dimensionality reduction techniques such as PCA, KPCA, DNE, LPP, UDP, LDA and LDNE in six available databases.

In our novel SKLDNE, firstly we use nonlinear kernel mapping to map the data into an implicit feature space F, which is successfully used in the Support Vector Machine (SVM). Then we seek a linear transformation that can preserve within-class geometric structures in F. Thus, we can gain a nonlinear subspace that can approach the intrinsic geometric structure of the face manifold. Furthermore, both “discrimination” in DNE and “Locality” in LPP have been used in an integrated modeling environment for image recognition. Besides, to investigate the performance of our proposed method, we will compare it with the state-of-the-art dimensionality reduction techniques such as PCA, KPCA, LDA, UDP, LPP, DNE and LDNE in six different publicly available datasets.

1.7 **Thesis outlines**

This thesis is organized into five chapters as follows:

Chapter 2 introduces the literature review of the biometric system.

Chapter 3 describes the methodology of this research and also provides an overview of our proposed method.

Chapter 4 consists of some information about MATLAB software, an explanation of different databases and the latest experimental results as the obtained results are discussed to analyze the performance of the proposed method.

Chapter 5 is the last chapter and presents the conclusion and recommendations for future work.

2 LITERATURE REVIEW

2.1 Biometric systems

Recently, the adoption of biometric systems has ranked among the safest security measures to apply access control, also against attempts of identity theft [26]. This is due to the possibility to automatically discriminate people based on their physical or behavioral characteristics. Biometrics could be described as a research field that measures physical or behavioral human features to identify an individual[27]. These features, such as, facial image, Fingerprint [28], vein, iris [29, 30], DNA information, and voice are unique for each individual. Many types of research continue developing methods to characterize these features for each individual as it is very vital in many fields like access control, banking security and so on. Therefore, the popularity and reliability of biometric systems kept rising. They are presently used to chase high levels of security in different real-life applications, from video surveillance [31-33] to smartphone authentication and access control to restricted areas. A biometric system can be applied in both the verification and identification process [28], depending on which application is required [34]. There are generally two different classes for biometric characteristics (figure2.1)[35] [36]:

- Physiological characteristics, which are related to human body shapes such as face shape, finger shape and other parts of the human body. Face, fingerprint and palm recognition are some examples of the so-called biometric systems.
- Behavioral characteristics relating to human behavior such as voice and signature.

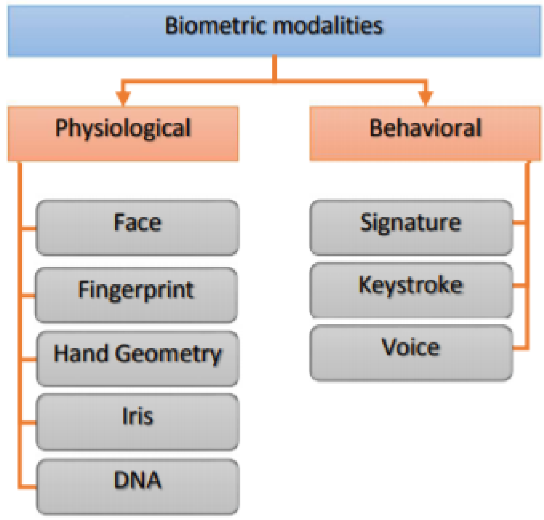


Figure 2.1. Principal biometric modalities

As crimes such as bank robbery and vehicle theft have been increasing substantially, it is very important to increase the security level in our society. Therefore, the number of governmental applications of biometric systems is growing fast to verify citizen identity. Consequently, industries have become more interested in producing biometric devices, aiming to enhance the level of security, such as surveillance systems control (system's control) access devices and so on. For instance, Apple surprised its customers all around the world when this company introduced its facial recognition system in its new iPhone production, called iPhone X, which is considerably more secure than other previous versions of Apple's Touch ID with fingerprint recognition system. Figure 2.1 shows the survey results of the question 'In your opinion what was the most exciting biometric modality in 2017?'[37].

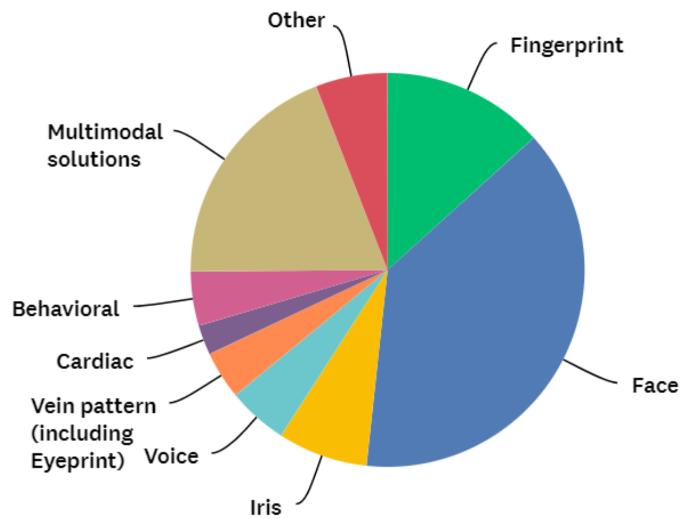


Figure 2.2. Over 200 people were surveyed, including executives from the world's leading biometrics companies on topics concerning the most exciting biometric modality in 2017

From the figure, it is clear that around 13 percent of respondents selected fingerprint modality as the most exciting, multimodality was chosen by 19 percent and facial recognition achieved around 38 percent as a topmost interesting modality among others.

Identification refers to the field in which the biometric system has already been trained with known data being taken from known users or people. Whenever the system receives an unknown input, it tries to match this input with one of the data in a database, to identify this unknown user, this new input should be compared with all training sets in the database one by one. It should be mentioned that the performance of the identification system is done without considering the subject having to claim an identity. The identification aim is to prevent a single person from using multiple identities [38].

In verification application, the biometric system captures the data from a person and compares this data with the data which has already been captured from this person to verify the individual's identity. The main aim in this field is preventing people or criminals from impersonating someone else's identity. In such a system, a person claims an identity and the system tries to determine whether this claim is true or not by proposing the following question 'Does this biometric data belong to this user or not?' Verification systems usually need a personal identification number (PIN), a smart card or a user name. By getting one of this information, the system tries to conduct a one-to-one comparison to verify the desired identity [7].

Figure 2.2 illustrates a general block diagram of identification and verification systems [7]. This system consists of four parts: a sensor used to capture the biometric data, a feature extractor applied to extract the main features from the input, a matcher to compare these features and decision module to indicate the response of accepting or rejecting.

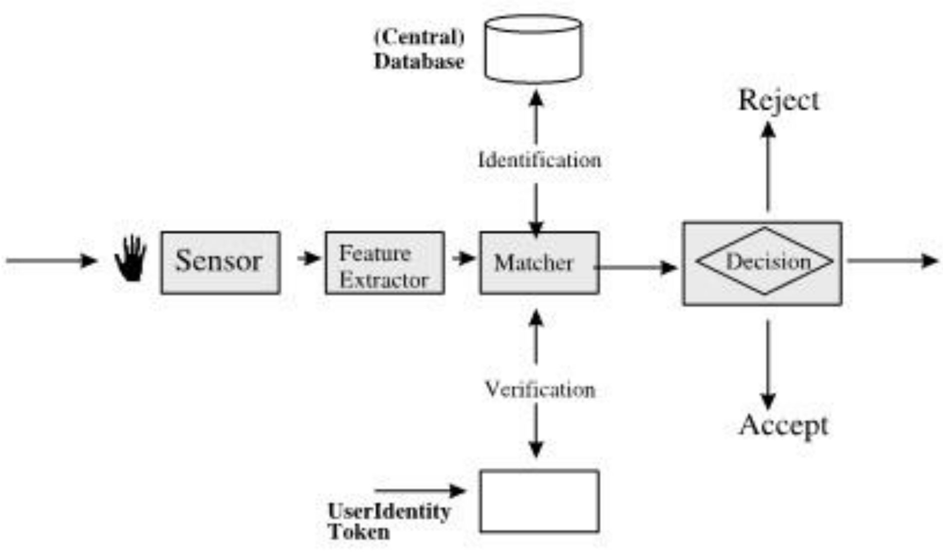


Figure 2.3. The general block diagram of identification and verification systems [7]

2.1.1 Face recognition

Face recognition, as a biometric authentication technique, is an important application field of artificial intelligence [39]. Its main advantage is that, unlike other biometric techniques such as finger print [40] , iris and speaker recognition [41], it does not require the applicant to spend time in the personal data acquisition process. For instance, facial recognition software, which is deployed in a public area where many different people pass by, can recognize faces of passers in a crowd and can help identifying a criminal. Its main disadvantage is the sensitivity to illumination variances, poses and occlusions which occur in unstructured environments. The issue of face recognition has been given a lot of attention by many researchers in pattern recognition, biometrics and computer vision [42]. Face recognition based on subspace analysis has been widely studied in recent years. There are some issues which should be considered when biometric features are applied in a practical biometric system [43]. The reliable biometric system should have the following properties:

- Universality: everyone should have this biometric characteristic.
- Acceptability: data from users can be taken easily without being noticed as it is contactless.
- Measurability: the characteristics should be measured easily.

Biometric systems should also include some other properties such as the following [44]:

- Performance: the biometric system should be able to achieve the desired accuracy and computational speed in a minimum of time by considering the operational factors that can affect the level of speed and accuracy.

- Circumvention: the biometric system should be very reliable and robust to prevent counterfeiting.

Based on the properties mentioned above, the comparison of different biometric technologies can be seen in Table 2.1 consisting of five factors and it is obvious that the face recognition method has a better position compared to the others, especially in the case of acceptability as user data can easily be taken without being noticed as it is totally contactless.

Table 2.1. Comparison of different biometric methods consisting of five factors [45]

Category	*Traits*	*Universality*	*Acceptability*	*Performance*	*Measurability*	*Circumvention*
Conventional	Face	H	H	M	H	H
	FP	M	M	H	M	M
	Iris	H	L	H	M	H
	Voice	M	H	L	M	L

H: High M: Medium L: Low

2.2 Dimensionally redaction

Since there are large volumes of high-dimensional data in numerous real-world applications, dimensionality reduction is a fundamental problem in many scientific fields. In the field of face recognition, many different dimensionality recognition approaches have been developed in recent times [15, 16, 19]. Dimensionality reduction is the main problem in numerous recognition techniques [1, 2, 46]. Dimensionality reduction techniques have been recommended by researchers to avoid "the curse of dimensionality," to amend the computational efficiency of image recognition [5, 14]. Generally, dimensionality reduction techniques can be classified into two main groups: i.e., linear and nonlinear. In linear methods, a significant low-dimensional subspace has to be discovered in the input data with high-dimensional space, where the embedded data in the input space have a linear structure [4, 6, 7, 47]. PCA is one of the famous linear methods [8, 11, 25, 48], which aims to retain global geometric information for data representation through enhancing the trace of the feature covariance matrix [8, 11, 49].

Linear discriminant analysis (LDA) is a linear technique that seeks to find out the discriminant information for data classification by enhancing the ratio between inter-class and intra-class scatters [11, 12]. Some of the limitations of both PCA and LDA are that they could suffer from the small sample size issue (SSS) [25] and that they may fail to recognize many important data structures that are nonlinear [13, 24]. Scholars have developed abundant practical nonlinear dimensionality reduction strategies [18] to address these problems. They can be classified into two types: manifold learning-based and kernel-based techniques [23, 50]. Manifold learning directly aims to discover the principal nonlinear data with low-dimensional structures that are concealed in the input space. Isometric Feature Mapping (ISOMAP) [42, 51] and Local Linear Embedding (LLE) [52, 53] are the most well-known manifold learning -based techniques to find inherent low-dimensional embedding of data [54]. Based on some experiments which have been done with these techniques, it has been proved that these methods can well discover meaningful embedded nonlinear data structures for face images. However, manifold learning-based techniques could suffer from two issues in terms of pattern recognition [25]. The first one is called “overlearning of locality,” [55] since manifold learning keeps locality data structures, but there is no straight connection with the classification. Out-of-sample is another issue that shows why most manifold-learning-based techniques are not appropriate for image recognition tasks [56-58]. These techniques can yield an embedding directly from a training data set, but they are often unable to find the sample's image in the embedding space when it is implemented in a new point. These problems cannot be overcome by the currently proposed manifold-learning methods. Although a few supervised forms have been proposed, they still suffer from these problems [59-61] because they are all based on "locality" characterization. Local quantity is sufficient for one manifold modeling, but it does not work well for classification tasks in multi-manifold modeling [49].

In contrast with manifold-learning-based techniques and to indirectly represent observed patterns in possibly much larger dimensional feature vectors, kernel-based techniques have been proposed by applying a kernelized nonlinear representation method. In this approach, the nonlinear data structure can be more separable in the observation space and become linear in the feature space. The representative strategies include the Kernel Fisher Discriminant (KFD) [50, 62] and the Kernel Principal Component Analysis (KPCA) [63-65]. Both have shown that they can be practical in many real-world functions, such as face recognition, to preserve the nonlinear data structure [66]. However, these kernel-based methods cannot directly consider the local data structure, which results in classification performance degradation. Recently, the Locality Preserving Projections (LPP) method [67] has been proposed as a linear subspace learning method to address the out of sample problem. LPP is an unsupervised linear subspace technique that has the remarkable advantage of being able to generate an explicit map. Similar to the one belonging to PCA and LDA, this map is linear and easy to compute and is also effective for many face recognition tasks. Although LPP has been designed based on “locality,” like most manifold learning methods, it still suffers from the “over learning of locality” problem, because there is no direct connection with the classification in its algorithm. Therefore, on some occasions, it cannot be guaranteed to map an appropriate projection for classification purposes [67].

Subsequently, to address this issue and delve into more influential projections for classification tasks, the Unsupervised Discriminant Projection (UDP) method [68] was developed as a simple version of LPP. UDP is considered a linear estimation of multi-manifold-based learning because it considers both the local and nonlocal scatter of data. In both LPP and UDP, data class label information is not considered, which may degrade their pattern classification performance. Furthermore, the Discriminant Neighborhood Embedding (DNE) method has been presented with the idea of using data class label information. Furthermore, this method [69] has been presented with the idea of using data class label information. DNE can find out a good embedding for classification considering intra-class absorption and inter-class separability. The main characteristic of DNE is called “discrimination”, meaning the ability to distinguish the same class from distinct classes. This specification of DNE can deal well with ‘out-of-sample’ and ‘small training sample size’ problems. Nevertheless, DNE cannot correctly preserve local information of data because it only concedes $+1$ and -1 to intra-class and inter-class neighbors [25]. Thus, a lot of the important geometrical structure information of data may be lost, and it might fail to discover the most significant sub-manifolds for pattern recognition. The locality-based discriminant neighborhood embedding method (LDNE) [25] has recently been proposed to tackle the problems existing in LPP and DNE. This method takes into account both “locality” and “discrimination” in a united modeling environment. However, many important non-linear data might be lost during the dimensionality reduction process, which dramatically influences classification accuracy.

According to the way dimensionality reduction algorithms “learn” about data to create predictions, they can be categorized into two different classes: supervised and unsupervised learning methods. Among these two, supervised machine learning is used more prevalently in which the data scholar acts as a guide to instruct the algorithm regarding what results should be found by [70]. The most well-known supervised algorithms include supervised LPP [69] , Local Discriminant Embedding (LDE) [71], Neighborhood Discriminant Projection (NDP) [72], Discriminant Locality Preserving Projections (DLPP) [73], Locally Discriminating Projection (LDP) [74] and Geometry Preserving Projections (GPP) [75] It is obviously clear that the aforementioned supervised techniques are generally applied to class label information to amend the dimensionality reduction. On the other hand, the unsupervised LPP-based algorithms generally aim to improve locality preserving and discriminating capabilities to further enhance the final performance of classification. Graph-Optimized Locality Preserving Projections (GOLPP) [25, 76], Orthogonal Locality Preserving Projection (OLPP) [77] and UDP [68] are some examples of unsupervised LPP-based methods. Table 2.2 shows the summary of the main articles on recent works stated in the literature review.

Table 2.2. The summary of the main articles on recent works stated in the literature review.

Authors	Title (method)	Weakness	Robustness
Damavandinejadmonfared, S., et al [11]	Finger vein recognition using PCA-based methods	small sample size problem	low computational complexity
Kim, K.I., et al [65]	Face recognition using kernel principal component analysis	losing the local data structure	preserving the nonlinear data structure
Yu, H., et al [12]	A direct LDA algorithm for high-dimensional data	Cannot recognize important nonlinear data structures	Preserving the discriminant information
Blackburn, J., et al [51]	Human motion recognition using isomap and dynamic time warping	“overlearning of locality,” problem	discovering meaningful embedded nonlinear data structures
Lu, J.,et al [67]	Regularized locality preserving projections and its extensions for face recognition	“over learning of locality” problem	designed based on “locality,”
Wang, T., et al [75]	Geometry preserving projections algorithm for predicting	“overlearning of locality,” problem	Using class label information
Schölkopf, B., et al [62]	Nonlinear component analysis as a kernel eigenvalue problem	losing the local data structure	preserving the nonlinear data structure
Deng, W., et al., [68]	Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Application to Face and Palm Biometrics	class label information is not considered	considering both the local and nonlocal scatter of data
Chen, H.-T., et al [71]	Local discriminant embedding and its variants	over learning of locality” problem	Using class label information
You, Q., et al [72]	Neighborhood discriminant projection for face recognition	over learning of locality” problem	Using class label information
Roweis, S.T., et al [52]	Nonlinear dimensionality reduction by locally linear embedding	“overlearning of locality,” problem	discovering meaningful embedded nonlinear data structures
Zhang, L., et al [76]	Graph-optimized locality preserving projections. Pattern	“overlearning of locality,” problem	discriminating abilities
Zhang, W., et al., [69]	Discriminant neighborhood embedding for classification	cannot correctly preserve the local information	dealing well with ‘out-of-sample’ and ‘small training sample size’ problems
Shao, J., et al [77]	Generalized orthogonal locality preserving projections for nonlinear fault detection and diagnosis.	“overlearning of locality,” problem	discriminating abilities
Gou, J., et al [25]	Locality-based discriminant neighborhood embedding	Losing important nonlinear data structures	taking into account both “locality” and “discrimination”
Lu, G., et al [73]	Face recognition using discriminant locality preserving projections based on maximum margin criterion.	over learning of locality” problem	Using class label information
Kishore, K., et al [74]	Hybrid face recognition with locally discriminating projection	over learning of locality” problem	Using class label information

In this project, a new supervised subspace learning algorithm named ‘Supervised Kernel Locality-Based Discriminant Neighborhood Embedding’ (SKLDNE) is proposed, in which not only the nonlinear data structure can be preserved by applying a kernelized nonlinear mapping method, but also both “locality” and “discrimination” of data in an integrated modeling environment are considered simultaneously. It should be noted that this technique is supervised through a direct connection with classification to well guide the procedure of dimensionality reduction. Due to its kernel-weighting, it is very influential in reducing the negative influence of outliers in the projection directions, which effectively handles the drawbacks of the linear model and makes it more robust to outliers. To obtain a reliable and powerful comparison, the efficiency of the proposed SKLDNE technique is compared with PCA, KPCA, LDA, UDP, LPP, DNE and LDNE through a wide range of experiments on different publicly available face datasets, i.e., Yale face, ORL face, Head Pose, and Sheffield. Moreover, Finger Vein and Finger Knuckle databases are also applied to well investigate the implementation of our algorithm in other types of databases rather than face databases.

2.2.1 Principal Component Analysis (PCA)

PCA is one of the fundamental and effective methods in the case of dimensional reduction [78]. This kind of transformation method is used to simplify data analysis. Dimensionality reduction and feature extraction of images are the main proposals of PCA [79].

2.2.1.1 Background of mathematics

Some elementary mathematical background skills which require for the understanding of the PCA process are given in this section [80]. The following topics have been covered independently from each other.

2.2.1.2 Standard Deviation

Before calculating the standard deviation, the mean of the samples must be obtained by the given formula [80]:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{2.1}$$

Then the standard deviation (*SD*) can be calculated as follows [80]:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}} \tag{2.2}$$

Where, n is the total number of data set and x is the set value. The standard deviation of a data set is the spread measure of the different data.

2.2.1.3 Variance

Variance is also used to measure the spread of data. However, the standard deviation is the most common one, but the variance is sometimes used. In fact, it is almost identical to the *SD*. The formula can be seen below:

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n_x - 1)(n_y - 1)} \tag{2.3}$$

2.2.1.4 Covariance

Standard deviation and variance can only operate in one dimension. However, many data sets have more than one dimension and the aim is usually to see the relationship between these dimensions. Covariance is measured between two dimensions. The variance formula and covariance are very similar to each other. It means that if you try to find the covariance of x by itself, it will give you the variance. Here is the formula for covariance:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{(n-1)} \tag{2.4}$$

2.2.1.5 Eigenvectors

Eigenvectors are a special method to multiply two matrices together. It should be mentioned that the eigenvectors can only be found for square matrices. For example, observing two multiplications between a matrix and a vector in equation 2.5 and equation 2.6, for equation 2.5 the result vector is not an integer multiple of the original vector, but in equation 2.6 the result is a multiplication of integer value 4 by the original vector. Thus, number 4 is called the eigenvalue, $\binom{3}{2}$ is the eigenvector and equation 2.5 also shows the non-eigenvector [80].

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix} \tag{2.5}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} \tag{2.6}$$

2.2.1.6 Advantages of PCA

As it has been mentioned before, PCA is a way of identifying patterns in data and also expresses data to highlight their differences and similarities. In high dimension data, the pattern in the data is hard to find and the PCA can be applied to analyze these data.

Another important advantage of PCA is reducing the number of dimensions without losing information [81]. This technique is also used to compress the image.

2.2.1.7 Mathematics of PCA

PCA can analyze the 1-D images since face images are 2-D. The first step is the dimensional reduction to present the 1-D images. Assume M vectors of size N (i.e. rows of the image multiplied by columns of image), where p is the pixel value:

$$\mathcal{X}_i = [p_1 \dots p_N]^T, i = 1, \dots, M \tag{2.7}$$

Based on PCA methods, mentioned in previous sections, the next step is computing the mean center of images:

$$m = \frac{1}{M} \sum_{i=1}^{M} \mathcal{X}_i \tag{2.8}$$

To calculate the mean centered image the following formula is used:

$$w_i = \mathcal{X}_i - m \tag{2.9}$$

The covariance matrix should be obtained to determine a set of eigenvectors and eigenvalues:

$$C = WW^T \tag{2.10}$$

Where W is a matrix composed of column vectors w_i placed side by side.

If we assume λ as an eigenvector, v as an eigenvalue and considering proven equation $\lambda v = Cv$, which shows the multiplication of integer value C (covariance) and original vector, we can obtain the following equation.

$$WW^T(Wv) = \lambda(Wv) \tag{2.11}$$

It should be mentioned that this equation is obtained by multiplying both sides of the given equation by w and the substitution of C

It indicates that the first M- 1 eigenvectors λ and eigenvalues v can be obtained by calculatingWW^T.

After finding the M eigenvectors and eigenvalues, images can be projected onto the $L \ll M$ dimensions by computing Ω which is the projected value and could be calculated by the following formula:

$$\Omega = [v_1 v_2 \ldots v_L]^T \tag{2.12}$$

To determine which finger vein images provide the best description of an input image; the Euclidean distance should be calculated as follows:

$$\in_k = \|\Omega - \Omega_k\| \tag{2.13}$$

Where the minimum value of $\in_k$ decide the unknown data into the k class.

2.2.2 Kernel Principal Component Analysis (KPCA)

The KPCA method to extract features was proposed after PCA. KPCA is a nonlinear extension of PCA, which computes the principal components in a high-dimensional feature space *F*, which is nonlinearly related to the feature space. PCA is a linear method that ensures that the transformed data is uncorrelated and insensitive to the dependencies of multiple features in the patterns. To overcome this problem, KPCA is proposed.

2.2.2.1 The idea of KPCA

The basic idea of KPCA is to first map the input data into feature space F via nonlinear mapping Q. Once we have done the nonlinear mapping, the input data, the linear PCA, is performed on the mapped data.

2.2.2.2 Mathematics of KPCA

As it has already been mentioned, the basic idea of KPCA is to nonlinearly map input data X into feature space F. When the input data is mapped by nonlinear mapping Φ, a linear PCA is performed in F. Assuming that F is centered, $\sum_{i=1}^{M} \Phi(X_i) = 0$where M is the number of input data. The covariance matrix of F can be defined as

$$C = \frac{1}{M} \sum_{i=1}^{M} \Phi(X_i). \Phi(X_i)^T \tag{2.14}$$

To do so, equation $\lambda\, v = Cv$, which is the eigenvalue equation, should be solved for eigenvalues $\lambda \geq 0$ and eigenvectors $v \in$ F.

As Cv = $(1/M) \sum_{i=1}^{M} (\Phi(X_i). v)\Phi(X_i)$, solutions for v with $\lambda \neq 0$ lie within the span of $\Phi(X_1), \ldots, \Phi(X_M)$, coefficients $\alpha_i (i = 1, \ldots, M)$ are obtained in such a way

$$\mathrm{V} = \sum_{\mathrm{i=1}}^{\mathrm{M}} \alpha_{\mathrm{i}}\, \Phi(\mathrm{X_i}) \tag{2.15}$$

The equations can be considered as follows

$$\lambda(\Phi(X_i). V) = (\Phi(X_i). Cv) \quad for\ all\ i = 1, \ldots, M \tag{2.16}$$

Having $M \times M$ matrix K by $K_{ij} = k\left(X_i, X_j\right) = (\Phi(X_i). \Phi\left(X_j\right))$, causes an eigenvalue problem.

The solution to this is as follows:

$$M\ \lambda\alpha = K\alpha \tag{2.17}$$

By selecting the kernels properly, different mappings can be achieved. One of these mappings can be achieved by taking the d-order correlations, which is known as ARG, between the entries, X_i, of input vector X. The required computation is prohibitive where $d > 2$.

$$\left(\Phi_d(X).\Phi_d(y)\right) = \sum_{i_1,\dots,i_{d=1}}^{N} \mathcal{X}i_1 \ \dots.\mathcal{X}i_d \ .\mathcal{Y}i_1 \ \dots.\mathcal{Y}i_d \ = \left(\sum_{i=1}^{N} \mathcal{X}_\mathrm{i}\,.\mathcal{Y}_\mathrm{i}\right)^d = \ (x.y)^d.$$

(2.18)

2.2.3 Locality Preserving Projection

Locality preserving projection (LPP) [23] is the linear dimensionality reduction algorithm that finds graph embedding of data sets, which directly models the manifold structure by constructing the nearest-neighbor graph that discloses neighborhood relations of data points to preserve the local structure of the input data in the projection. Although LPP has been applied effectively as feature extraction in many circumstances, it might be unsuitable for pattern recognition as it is a linear method. It often fails to retain with-in class local structure images which are subjected to involved nonlinear changes because of large expression, pose or illumination variations. It also suffers from 'over-learning of locality' which dramatically degrades classification performance.

LPP works based on a linear approximation of the Laplacian Eigen Map, which searches transformation P, in which a high-dimensional input data X= $[x_1, x_2, \dots, x_n]$ could project into low-dimensional subspace Z while the local structure of the input data is preserved. To calculate linear transformation T, an objective function should be minimized as follows:

$$min_P \sum_{i,j=1}^{n} \left\| z_i - z_j \right\|^2 H(i,j) \tag{2.19}$$

Where weight matrix H (called the heat kernel) is obtained by the nearest-neighbor graph and $z_i = P^T x_i$.

$$H(i,j) = e^{-\frac{\left\| x_i - x_j \right\|^2}{t}} \tag{2.20}$$

Where the parameter t is an appropriate constant number. Otherwise,$S(i,j) = 0$. On the other hand, when x_i and x_j are the nearest neighbors, the weight matrix H could clearly be set as $H(i,j) = 1$. Otherwise,$H(i,j) = 0$. The optimal transformation matrix can be calculated by using the minimization problem to solve the generalized eigenvalue problem

$$XLX^T P = \lambda XDX^T P, \tag{2.21}$$

Where $L = D - H$ is the Laplacian Matrix and $Dii = \sum_j H(i.j)$ is a diagonal matrix.

2.2.4 Discriminant Neighborhood Embedding

Discriminant Neighborhood Embedding (DNE) is proposed, based on an intuition of a dynamics theory. DNE is a supervised learning method which modulates an optimum low dimensionality embedding of multi-class data points in a high dimensional space for classification. Furthermore, DNE effectively avoids the complication of the singularity matrix as the inverse matrix does not need to be calculated anymore. Based on the main characteristics of DNE, it can present a good solution for the small-sample-size (SSS) and out-of-sample problems. Although this comprehensive technique is effective in pattern classification, it still cannot uphold the local and geometrical structure information of data. The main steps of the DNE algorithm [69] come next:

1- Adjacent matrix $\overline{H}$ of graph *G* which refers to the underlying supervised manifold structure is as follows:

$$\overline{H}_{ij} = \begin{cases} -1, & x_i \in knn(j) \ or \ x_j \in knn(i) and \ (c_i \neq c_j) \\ +1, & x_i \in knn(j) \ or \ x_j \in knn(i) and (c_i = c_j) \\ 0\,, & otherwise \end{cases} \qquad (2.22)$$

Where, ci shows the class label of xi and knn(i) illustrates the set of k nearest-neighbors of *xi*. Note that each edge is weighed +1 or −1 respectively, to determine the local intra-class attraction and inter-class repulsion between neighboring points.

2- The optimal transformation of matrix *P* can be defined as follows:

$$min \sum_{ij} \left\| z_i - z_j \right\|^2 \overline{H}_{ij} \qquad (2.23)$$

The minimization problem can be reduced to:

$$argmin \ tr(P^T X \overline{L} X^T P) \qquad (2.24)$$

Subject to $P^T P = I$, where $\overline{L} = \overline{D} - \overline{H}$. $\overline{D_{ii}} = \sum_j \overline{H_{ij}}$ is a diagonal matrix.

Like LPP, parameter *P* (projection matrix) could be optimized by calculating the minimum eigenvalue solution to the generalized Eigenvalue problem as follows:

$$X\bar{L}X^T P = \lambda P \, , \tag{2.25}$$

Where P is constituted by r eigenvectors corresponding to its first smallest negative d eigenvalues of *d*, i.e., $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_d < 0 \leq \lambda_{d+1}$ (2.26)

2.2.5 Unsupervised Discriminant Projection (UDP)

UDP is a linear projective map in which the neighborhood structure of data sets can be preserved where the lower-dimensional manifold embedded the desired space should be obtained from high dimensional data sets. UDP considers local and nonlocal scatters simultaneously while seeking to detect a projection maximizing the ratio of the non-local scatters to local scatters. UDP performs more intuitively than LPP for classification tasks as non-local information is utilized. Nevertheless, in both LPP and UDP, the class label information of data is not considered, which might dramatically reduce pattern classification performance. Adjacent matrix *H* is defined as follows [82]:

$$H_{ij} = \begin{cases} 1, & \left\| X_i - X_j \right\|^2 < \delta \\ 0, & otherwise \end{cases} \tag{2.27}$$

Where the mean square of the Euclidean distance between any pair of the projected sample points that are within any local δ-neighborhood ($\delta > 0$). Specifically, two samples *xi* and *xj* are viewed within a local δ-neighborhood provided that $\left\| X_i - X_j \right\|^2 < \delta$.

The K-nearest neighbors' method can obtain the following adjacent matrix *H*:

$$H_{ij} = \begin{cases} 1, & if\ X_j\ \ is\ among\ K\ nearest\ nieghbors\ of\ X_i \\ & and\ X_i\ \ is\ among\ K\ nearest\ nieghbors\ of\ X_j \\ 0, & otherwise \end{cases} \tag{2.28}$$

2.2.6 Linear Discriminant Analysis (LDA)

LDA simply finds the global discriminant information for classification by maximizing the ratio between inter-class and intra-class scatters. Linear Discriminant Analysis (LDA) seeks those vectors with the best discriminant among classes. Mathematically described LDA defines two measures, for all the samples of all classes. The first called the within-class scatter matrix which is, as follows:

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} \left(X_i^j - y_j\right) \left(X_i^j - y_j\right)^T \tag{2.29}$$

Where X_i^j is the *i*th sample of class j, y_j is the mean of class j, c is the number of classes, and N_j is the number of samples in class. The second measurement is called the between-class scatter matrix:

$$S_b = \sum_{j=1}^{c} \left(y_j - y\right) \left(y_j - y\right)^T \tag{2.30}$$

Where y represents the mean of all classes.

However, LDA suffers from the Small Sample Size (SSS) problem in data classification, which means that when a small training data set is used, there is no guarantee that LDA performs well.

2.2.7 Locality- Based Discriminant Neighborhood Embedding

In the multi-class classification assignment, N data points should be classified. The problem that arises here is finding a circumlocutory manifold embedded subspace. Based on the DNE, there are two classifications for the important characteristic of manifold structure, namely inter-class compactness and intra-class scatters. These two classes can be defined as follows:

- Intra-class absorption: it is the interaction between pairs of neighbours from the same class.
- Inter-class abhorrence: it is the interaction between pairs of neighbours from different classes.

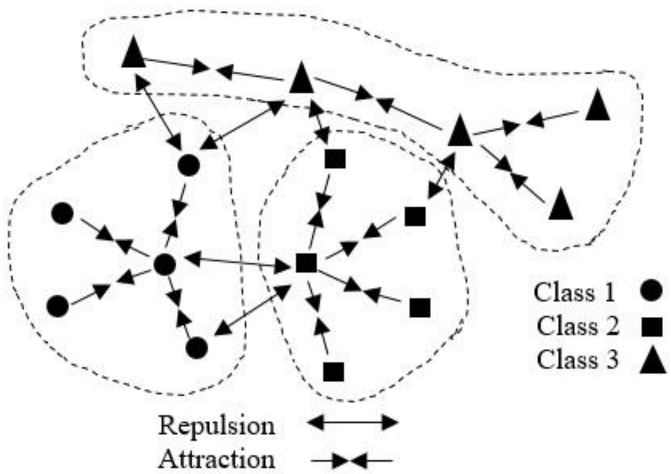


Figure 2.4. The interactions by attraction and repulsion for the points between different classes

Significantly, it is possible to classify all data points based on absorption interaction or distracting behavior using these two classes (figure 2.3). As a result, neighbors from the same class are absorbed while neighbors from the different class become separable in the subspace. To formulate the method first we consider that x_i is a data point, $N^s(x_i)$ demonstrates the intra-class neighbors of x_i, $N^d(x_i)$ denotes the inter-class neighbor of x_i and $N(x_i)$ represents all the x_i neighbors. Thus, to carry out this task as the purpose of understanding these two classes, the edges between x_i , inter-class neighbors and intra-class neighbors are indicated using different weights. Denoted weights are calculated with a kernel function whose functioning is based on the dissimilarity between x_i and its neighbors. The neighborhood, including inter-class and intra-class neighbors, can be called the discriminant neighborhood. Discriminant adjacent graph *G* can be obtained by the ϵ-neighborhood or k-neighborhood. Discriminant adjacent weight matrix (DAWM) *S* of *G* using the k-neighborhood is defined as [20]:

$$S_{ij} = \begin{cases} -\exp\left(-\frac{\|x_i - x_j\|^2}{t}\right), & x_i \in N_k^s(x_i) \; or \; x_i \in N_k^s(x_j) \\ +\exp\left(-\frac{\|x_i - x_j\|^2}{t}\right), & x_j \in N_k^d(x_i) or \; x_i \in N_k^d(x_j) \\ 0\,, & otherwise \end{cases} \tag{2.31}$$

Where, *t* is the regulator, $N_k^s(x_i)$ is the intra-class neighbor of x_i and $N_k^d(x_i)$ is the inter-class neighbor of x_i in the k-neighborhood.

It is obvious that different samples lead to different classification results. In view of the fact that the individual feature space location of the sample indicates its conditions, a parameter was defined to regulate adjacent weight between pairs of neighbors. This regulator can be formulated as:

$$t = \frac{1}{k} \Sigma_{j=1}^{k} \left\|x_i - x_j\right\|^2 \tag{2.32}$$

To gain intra-class compactness and inter-class scatters in the transformation space, it is recommended to apply a linear mapping method to project intra-class absorption and inter-class abhorrence of the input data points. As a result, the new low dimensional space can be defined as:

a. Intra-class compactness:

$$\Phi(P) = \Sigma_{ij}\left\|y_i - y_j\right\|^2 W_{ij} = \Sigma_{ij}\left\|P^T x_i - P^T x_j\right\|^2 W_{ij} \tag{2.33}$$

$x_j \in N_k(x_i)$ or $x_i \in N_k\left(x_j\right) and$ $(c_i = c_j)$

b. Inter-class scatters:

$$\Psi(P) = \Sigma_{ij}\left\|y_i - y_j\right\|^2 W_{ij} = \Sigma_{ij}\left\|P^T x_i - P^T x_j\right\|^2 W_{ij} \tag{2.34}$$

$x_j \in N_k(x_i)$ or $x_i \in N_k$ and $(c_i \neq c_j)$

Finally, the difference between the weighted distance from each data point to the inter-class neighbors in $N^d(x_i)$ and those from x_i to the intra-class neighbors in $N^s(x_i)$ in the mapped space must be calculated and, by maximizing this measurement, we can obtain the optimum result. This measurement can be referred to as a margin, which is calculated as follows:

$$\Theta(P) = \Psi(P) - \Phi(P) \tag{2.35}$$

Thus, if the original data points are close together this margin can keep the projected data points as close as possible. However, we can prevent x_i and x_j from being mapped far apart if they are close by defending the retribution generation:

$$\Theta(P) = \Sigma_{ij}\left\|P^T x_i - P^T x_j\right\|^2 F_{ij} \tag{2.36}$$

3 METHODOLOGY

3.1 Introduction

In this chapter, first, the overall view of the proposed method is explained with mathematics. Then, all databases used in this research are introduced

3.2 Main idea of the proposed method

In this project, we have proposed a novel supervised subspace learning method named "Supervised Kernel Locality-Based Discriminant Neighborhood Embedding" (SKLDNE) which is presented based on the following main ideas:

-To apply a kernel trick as an instance-based learner in a nonlinear kernel feature space: As many effective nonlinear data features would be lost during the classification process using the linear technique, applying a nonlinear method can improve the recognition performance.

-To obtain the advantages of "locality" in LPP: in which geometric relations are preserved due to the prior class-label information.

-To use "discrimination" from DNE: in which the compact submanifold for data from the same class is formed in the embedded low dimensional subspace.

3.3 Supervised Kernel Locality-Based Discriminant Neighborhood Embedding

In our novel SKLDNE, first nonlinear kernel mapping is applied to map the data into implicit feature space F. Therefore, a nonlinear subspace that can approach the intrinsic geometric structure of the face manifold can be obtained. Then, we seek a linear transformation in which both "locality" and "discrimination" of data are successfully preserved in the manifold learning phase. In fact, the proposed SKLDNE is modeled to capture nonlinear data in the feature space while the important "locality" , as well as the "discrimination" of data, are simultaneously preserved. Suppose X= $[x_1, x_2,. . . , x_n$] is a set of d-dimensional input samples and this input data is projected onto a higher dimensional feature space F via nonlinear mapping Ø : R^n→ F. Then, manifold learning is carried out on the projected samples $\emptyset(X) = [\emptyset(x_1), \emptyset(x_2), \ldots, \emptyset(x_n)]$. Now assume that we are to find the projection transformation $V_\emptyset$ in F. The optimization problem can be expressed as:

$$Maximize\ \sum_{ij}^{n} \left\| z_i - z_j \right\|^2 F_{ij} \tag{3.1}$$

Subject to $V_\emptyset{}^T V_\emptyset = I$, where I denotes the identity matrix, $z_i = v_\emptyset^T \emptyset(x_i)$ and

$z_j = v_\emptyset^T \emptyset(x_j)$ are the projection of $\emptyset(x_i)$ and $\emptyset(x_j)$ with respect to $V_\emptyset$ and F_{ij} represents the relationship between of x_i and x_j. The optimization problem can be kernelized as

$$\sum_{ij}^{n} \left\| z_i - z_j \right\|^2 F_{ij} = \sum_{ij}^{n} \left\| v_\emptyset^T \emptyset(x_i) - v_\emptyset^T \emptyset(x_j) \right\|^2 F_{ij} \tag{3.2}$$

This equation can be rewritten from the square of the norm in Eq. (3.2) into the trace form

$$\sum_{ij}^{n} \left\| v_\emptyset^T \emptyset(x_i) - v_\emptyset^T \emptyset(x_j) \right\|^2 F_{ij} = \mathrm{tr}\ \{\sum_{ij}^{n} (v_\emptyset^T \emptyset(x_i) - v_\emptyset^T \emptyset(x_j))(v_\emptyset^T \emptyset(x_i) - v_\emptyset^T \emptyset(x_j))^T F_{ij}\} \tag{3.3}$$

$=\text{tr}\ \{v_{\emptyset}^{T} \sum_{ij}^{n}(2\emptyset(x_i)F_{ij}\emptyset(x_j)^{T} - \left(2\emptyset(x_j\right)F_{ij}\emptyset(x_i)^{T}\ F_{ij})V_{\emptyset}\}$

The linear transformation should lie in the span of $\emptyset(x_1), \emptyset(x_2), \ldots, \emptyset(x_n)$,

$\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_n]$ consists of expansion coefficient vectors and

$$V_{\emptyset} = \sum_{i=1}^{n} \alpha_i\, \emptyset(x_i) = \emptyset(X)\ \alpha\ . \tag{3.4}$$

Substituting (3.4) into (3.3), we obtain

$U_{\emptyset}=2\text{tr}\{v_{\emptyset}^{T}\emptyset(X)(\mathrm{D}-\mathrm{F})\ \emptyset(X)^{T}V_{\emptyset}\}$

$=2\text{tr}\{v_{\emptyset}^{T}\emptyset(X)\mathrm{L}\ \emptyset(X)^{T}V_{\emptyset}\}$

$$=2\sum_{l=1}^{m}\{v_{\emptyset l}^{T}\emptyset(X)\mathrm{L}\ \emptyset(X)^{T}V_{\emptyset l}\} \tag{3.5}$$

Where $D_{ii}=\sum_j F(i,j)$ is the diagonal matrix and L=D-F, L and D are the symmetric matrix and represent the number of eigenvalues

So with some effort optimization problem can be rewritten as

$Maximize\ \text{tr}\ \{v_{\emptyset}^{T}\emptyset(X)\mathrm{L}\ \emptyset(X)^{T}V_{\emptyset}\}$

$$=\alpha^{T}KLK\alpha \tag{3.6}$$

Subject to $\alpha^{T}K\alpha = I$

Where k is the kernel matrix with $\mathrm{k(xi, xj)} = [\ \emptyset(x_i).\emptyset(x_j)]$ and a kernel in the matrix form is

$$K = \emptyset(X)^{T}\emptyset(X) \tag{3.7}$$

The corresponding generalized eigenvalue problem can be obtained by computing the maximum eigenvalues in $\emptyset(X)\mathrm{L}\,\emptyset(X)^T V_\emptyset = \lambda V_\emptyset$ where the generalized eigenvector corresponding to the largest eigenvalue is the main interest. So we need to compute the dot product via kernel and find its nearest neighbor in the embedding space.

3.4 The main algorithm of the SKLDNE

The detailed steps of the algorithm are summarized as follows:

SKLDNE Algorithm

Input: high-dimensional input X= [$x_1, x_2, . . . , x_n$]

Output: Low dimensional subspace

1 **For** i=1**To** N

2 Construct a nonlinear kernel mapping to map the data into an implicit feature space

3 Obtain Discriminant adjacent graph G by the ϵ-neighborhood or k-neighborhood.

4 $\boldsymbol{If}\ x_i \in N_k^s(x_i)\ or\ x_i \in N_k^s(x_j)\ \boldsymbol{Then}\ - exp\left(-\frac{\|x_i-x_j\|^2}{t}\right)$

5 $\boldsymbol{If}\ \ x_j \in N_k^d(x_i) or\ x_i \in N_k^d(x_j)\ \boldsymbol{Then} + exp\ \left(-\frac{\|x_i-x_j\|^2}{t}\right)$

6 $\boldsymbol{If\ Not}$ 0

7 Optimized the eigenvalues via the generalized eigenvalue.

8 Apply the transformation matrix to reduce the dimension from the original space to a new subspace.

9 Classify the transformed data point.

10 **end**

3.5 **Benefits of the SKLDNE method**

It is worthwhile to highlight several characteristics of the proposed approach here:

(1) SKLDNE has been designed successfully with some effort to retain local geometric relations of the within-class samples, which are very important for image recognition. Generally, the categorization strength of methods with a linear learning algorithm is restricted. They fail to deal with complicated problems. Many effective nonlinear data features may be lost during the classification process using linear techniques such as LDNE, LDA, DNE, and LPP. Therefore, applying a nonlinear method can effectively improve classification performance.

(2) This technique is a supervised learning method, as the data scholar acts as a guide to instruct the main algorithm whose conclusion should be found. SKLDNE considers class label information of neighbors in which there is a direct connection with the classification to enhance final recognition performance.

(3) It benefits from the advantages of "locality" in LPP in which, due to the prior class-label information, geometric relations are preserved.

(4) Not only can it build a compact submanifold by minimizing the distance between the same points in the same class, but it also expands the gaps among submanifolds of distinct classes simultaneously, which is called 'discrimination'.

(5) SKLDNE can resolve the SSS problem, which that is mostly faced by other aforementioned techniques such as PCA, LDA, UDP, and LPP, as well as the "overlearning of locality" problem in the manifold learning.

(6) Due to its kernel weighting, it is very efficient in reducing the negative influence of outliers on the projection directions, which effectively handles the drawbacks of linear models and makes it more robust to outliers.

Figure 3.1 briefly shows the main aforementioned benefits of the SKLDNE method which are already explained.

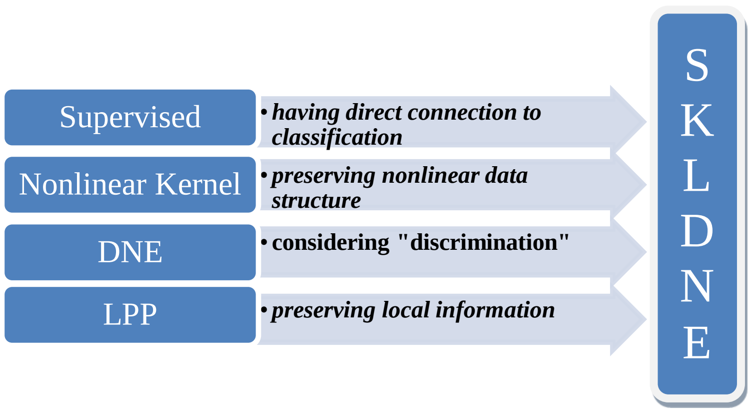


Figure 3.1. The basic benefits of the proposed SKLDNE method

3.6 Methodology and algorithm

As already mentioned, DNE cannot correctly preserve local information of data because it only concedes +1 to intra-class and −1 to inter-class neighbors, so it might fail to find out the most significant submanifold for pattern classification. In addition, LPP is designed based on 'locality' since it has no direct connection with classification, and it still suffers from the 'over-learning of locality' problem. Therefore, LDNE has been proposed to overcome the problems existing in LPP and DNE, considering both 'locality' and 'discrimination' in a unified modeling setting. However, it does not guarantee an appropriate projection for classification purposes, because many important non-linear data might be lost during its dimensionality reduction process. In some cases, LDNE cannot distinguish inter-class and intra-class neighbors properly to conduct projection for all points either, which can degrade the classification performance. To address these problems, we propose a new supervised subspace learning method named “Supervised Kernel Locality-Based Discriminant Neighborhood Embedding” (SKLDNE). Combined with a nonlinear data structure, locality, and discrimination information, SKLDNE can yield an optimal subspace that best finds the indispensable submanifold-based structure.

In our proposed SKLDNE, we first use nonlinear kernel mapping to represent the input data in implied feature space F. Afterwards, a linear transformation is searched to retain within-class geometric structures in the feature space. Hence, we can achieve a nonlinear subspace that can estimate the essential geometric structure of the face manifold. The proposed SKLDNE is modeled to take the nonlinear data in the feature space while important features of data including “locality” and “discrimination” are simultaneously preserved. To well elucidate the performance of our SKLDNE, we have compared it with several dimensionality reduction techniques including PCA, KPCA, LDA, UDP, LPP, DNE, and LDNE on six different publicly available datasets. As we can see from Figure 3.2, all aforementioned techniques have been divided into two classes. Linear and nonlinear. Figure 3.3 shows the overall view of the main recognition algorithm including Database, Image preprocessing, identification process, and decision-making part. Image preprocessing part includes ROI extraction, image resizing, and image enhancement. The identification process, which is the main purpose of this research, includes feature extraction, classification and comparison part. This part will be explained in further details in the next sections.

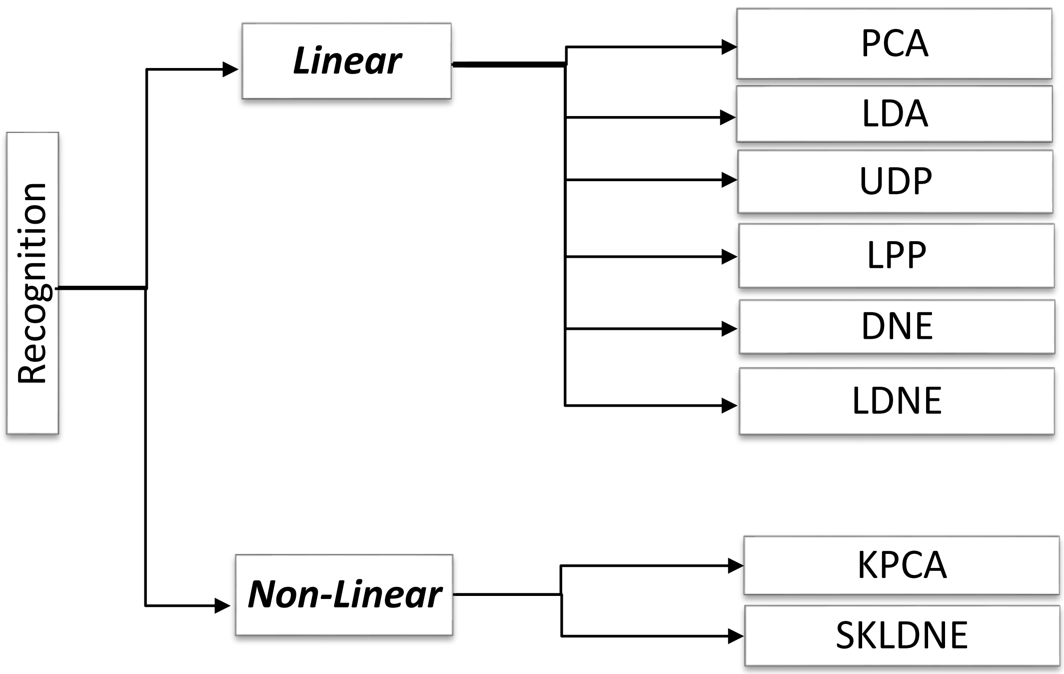


Figure 3.2. Different recognition techniques applied in this project

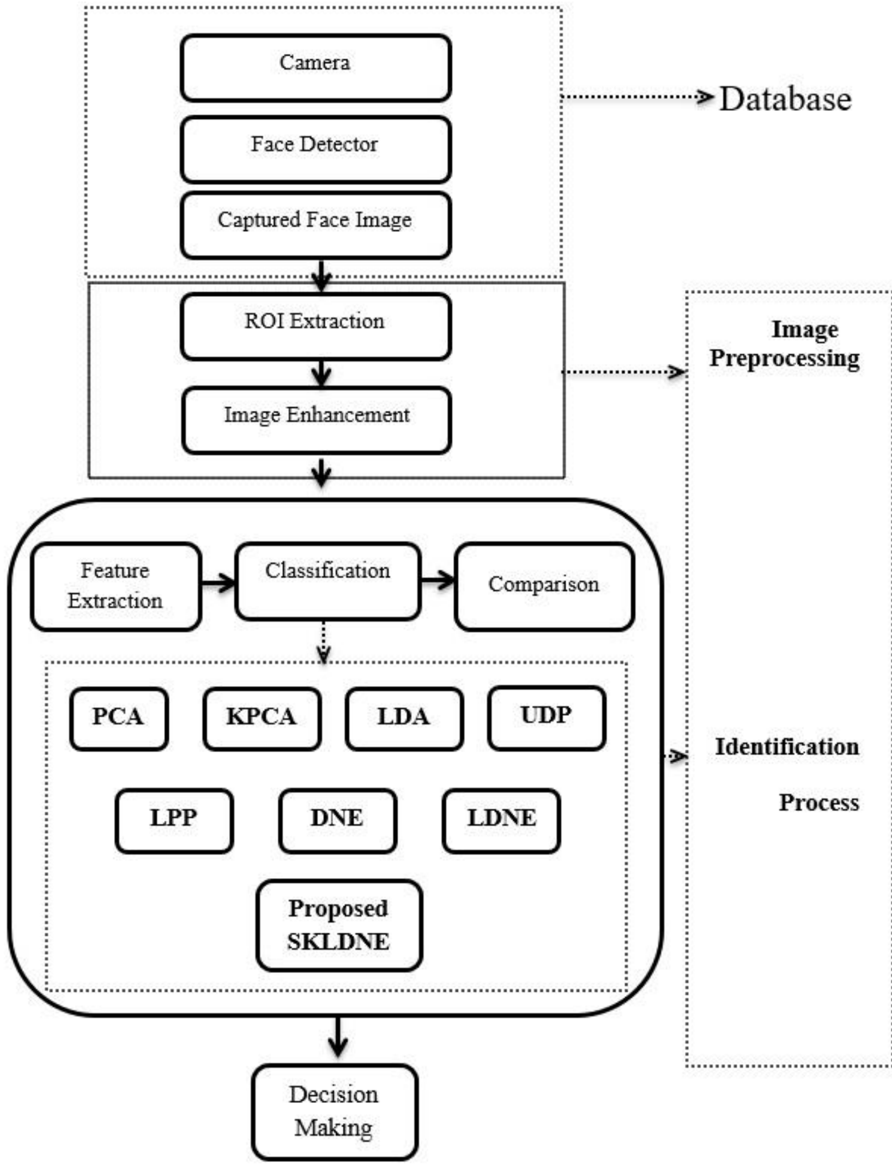


Figure 3.3. The overall view of main recognition algorithm

3.7 Databases

To obtain a reliable and powerful comparison, the performance of the proposed SKLDNE method is compared with PCA, KPCA, LDA, UDP, LPP, DNE and LDNE in extensive experiments on different publicly available face datasets, i.e., the Yale Face, ORL Face, Head Pose and Sheffield (Normal and pre-cropped). Moreover, Finger Vein and Finger Knuckle databases are also applied to well examine the performance of our method in other types of databases rather than face databases. For each dataset, depending on the number of data for each class, some samples of each class are randomly selected as training samples, while the remaining ones of that class are chosen for testing. Furthermore, the nearest neighbor (NN) classifier with the Euclidean distance is used in the recognition phase. In all the experiments, for fair comparisons, parameter K selected in all methods is chosen as a fixed number of K=Tn-1 where Tn denotes the number of training samples of each class. Some information about the MATLAB software (which is used in our implementations) and database is provided in the next section. Besides, how to gain accuracy is described as well. Analyses and discussions are based on the experimental results for each database separately.

3.8 MATLAB software

MATLAB is a very useful program to develop the algorithm, visualization, data analysis, and numerical computing. All codes used in this research are programmed by MATLAB software. The MATLAB version is R2016, Natick, MA, USA. It is included in the whole processing, matching algorithm and performance evaluation.

3.9 How to find accuracy?

As mentioned, the comparison between methods is done at different numbers of training and testing finger vein images. If a test image is correctly identified, the number of correct identifications is increased by one and if a test image is not correctly recognized, it will not contribute to the number of correct identifications so it will increase the number of errors. The final accuracy can be calculated by the following simple formula:

$$Accuracy\% = \frac{the\ number\ of\ correct\ identifications}{total\ number\ of\ identifications} \times 100$$

3.10 Summary

This chapter first presented the proposed method of this research and related algorithm and benefits. In addition, the SKLDNE algorithm is discussed more providing its implementation flow. Finally, to have fair comparisons, 7 different methods have been chosen for implementation on six publicly available datasets. In the next, chapter all experimental results corresponding to each database will be discussed in detail.

4 RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, the performance of SKLDNE has been evaluated on six different publicly available datasets, i.e., the Yale Face, ORL Face, Head Pose, Sheffield(normal and pre-cropped), Finger Vein and Finger Knuckle and compared with the performances of PCA, KPCA, LDA, UDP, LPP, DNE and LDNE. The performance of our proposed method is evaluated by comparing it with the other aforementioned dimensionality reduction methods.

4.2 Experiments and discussion

To have a fair investigation of the performance of the proposed SKLDNE method, it has been compared with PCA, KPCA, LPP, UDP, LDA, DNE and LDNE in extensive experiments on six different data sets, i.e., the Sheffield (original and pre-cropped database), ORL, Yale, Finger Vein, Finger Knuckle databases. For each experiment, the first Tn samples have been chosen from each class as training samples and the rest of each class have been used for testing. To simplify and improve the recognition result, the nearest neighbor (NN) classifier using the Euclidean distance has been used in the recognition phase. The k-neighborhood parameter k for calculating the weight matrix is denoted by Wk in the following discussions. In all the experiments, for fair comparisons, Wk has been selected as Wk=Tn-1 (where Tn is the number of training samples per class) because, based on our experiments, in this value of Wk, the aforementioned methods have achieved the optimal recognition rate.

4.2.1 Experiment using the Sheffield database

The Sheffield Face Database includes a total of 20 individuals with 564 images of them (mixed race/gender/appearance) in which each individual is illustrated in a range of poses from profile to frontal views. The images are all in PGM format, with a size of approximately 220*220 pixels with a 256-bit grayscale. Figure 4.2 shows a sample for different poses of one subject contained in the Sheffield Face Multi-View and figure 4.3 illustrates a sample of pre-cropped face images on the Sheffield Face. In our experiments, each image was resized to 112*92 pixels. Figure 4.1 demonstrates the different types of implementations by the SKLDNE method on the Sheffield database.

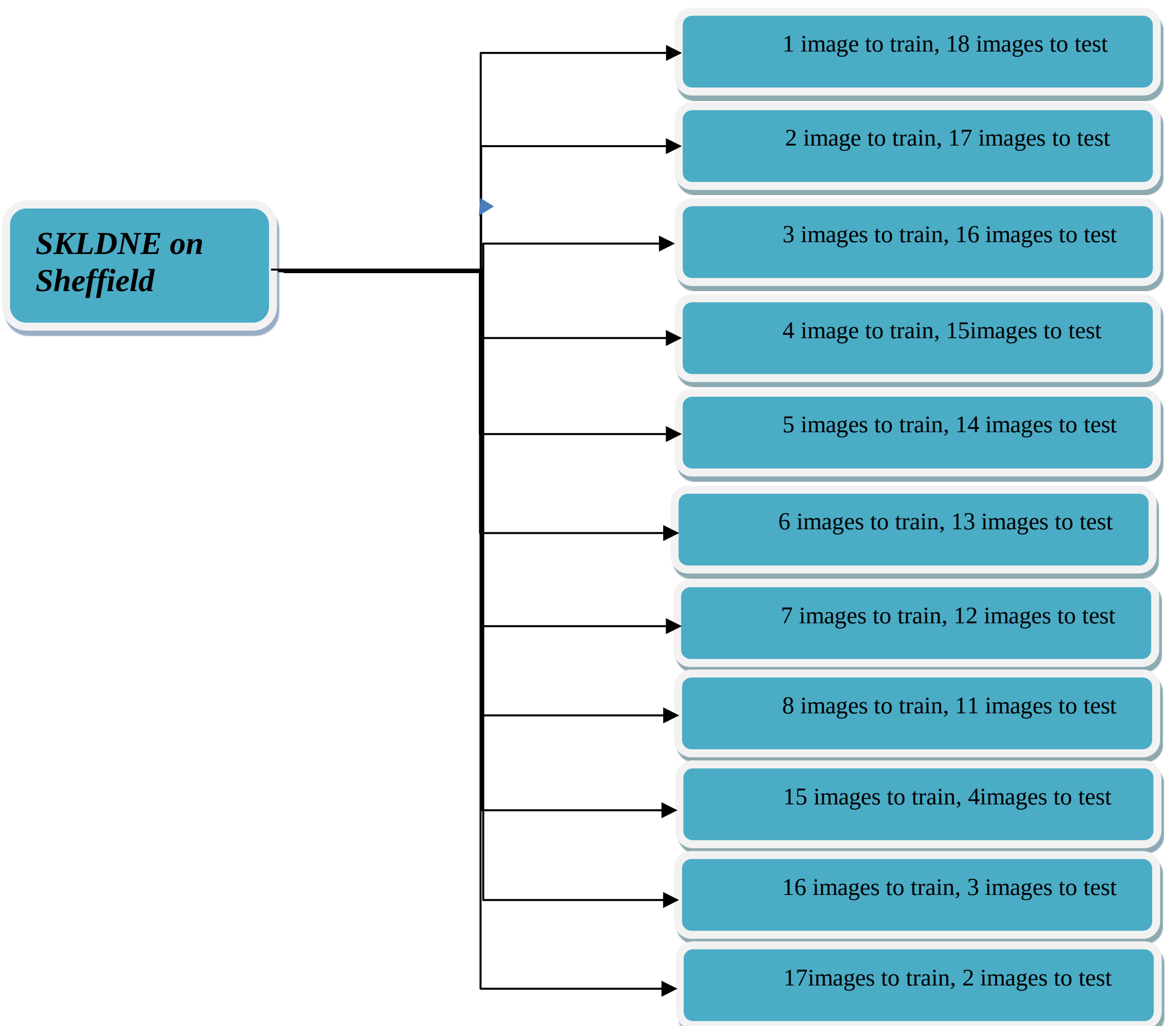


Figure 4.1. Different types of implementations by the SKLDNE method in the Sheffield database

Figure 4.2. A Sample of one subject with different poses from the Sheffield Face Multi View

Figure 4.3. A Sample of pre-cropped face image in the Sheffield Face

The maximal rate of recognition of each method and the related dimension implemented in the Sheffield database are illustrated in Table 4.1 and for that of a pre-cropped face image in Table 4.2. Note that in both tables the best performance among other methods is assigned in boldface. In addition, in all experiments due to the large number of implementations, it is decided to select some number of training and testing images that are more challenging for classification task to investigate the performance of aforementioned directionally reduction techniques in these critical areas, so considering the small training sample size problem (SSS), first we have selected a few small training samples and then some large numbers to examine the performance of our SKLDNE methods in some common existing problems of dimensionality reduction techniques such as SSS problem, over learning and out of sample problem.

Table 4.1. Maximum recognition accuracies (in percentage terms) of supervised kernel locality-based discriminant neighborhood embedding (SKLDNE) and other methods for the different numbers of training and testing images in the Sheffield Face and corresponding dimensions (shown in parentheses).

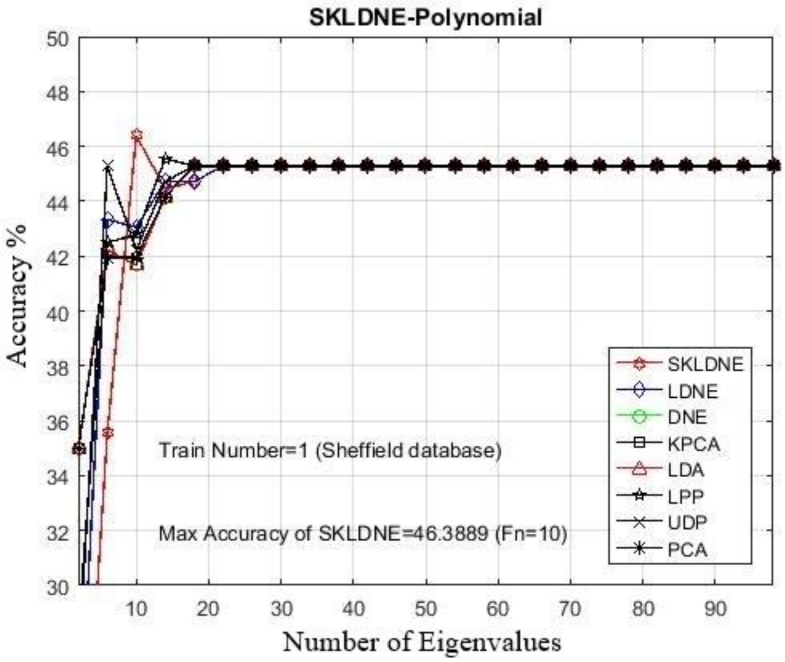
DATABASE	SHEFFIELD FACE										
TN	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**15**	**16**	**17**
PCA	45.1 (18)	47.94 (18)	49.68 (18)	49 (18)	49.64 (30)	51.15 (26)	52.5 (30)	54.54 (26)	85.5 (18)	93 (28)	92 (10)
KPCA	45.2 (18)	48.2 (22)	50.31 (30)	50.33 (38)	50.7 (42)	51.92 (58)	54.15 (50)	55.9 (69)	87.5 (38)	93 (30)	92 (14)
UDP	45.33 (10)	48.1 (22)	48.43 (30)	51.66 (30)	50.71 (45)	51.9 (66)	55.41 (54)	57.27 (50)	87 (90)	95 (86)	92.5 (14)
LPP	45.55 (14)	48.2 (26)	52.81 (12)	50.66 (34)	56.07 (38)	52.3 (54)	54.16 (46)	55.9 (62)	90 (74)	93.33 (42)	95 (34)
LDA	45.2 (18)	50.29 (26)	48.43 (14)	58.66 (34)	56.07 (22)	59.23 (6)	60 (42)	61.36 (50)	92.5 (22)	93 (30)	97.5 (18)
DNE	45.2 (18)	48.23 (34)	50.31 (14)	51.33 (74)	51.78 (58)	51.9 (42)	54.58 (74)	56.36 (78)	87.5 (38)	93.33 (30)	92.5 (20)
LDNE	45.27 (22)	50.58 (9)	56.87 (14)	58.66 (14)	65 (14)	73.07 (34)	76.66 (18)	76.81 (14)	90 (22)	96.1 (14)	97.5 (18)
SKLDNE	**46.38** (10)	**52.94** (10)	**59.06** (14)	**62.66** (10)	**69.64** (14)	**78.46** (30)	**83.75** (10)	**80.45** (10)	**93.75** (34)	**98.33** (10)	**100** (10)

Table 4.2. Maximum recognition accuracies (in percentage terms) of supervised kernel locality-based discriminant neighborhood embedding (SKLDNE) and other methods for the different numbers of training and testing images in the Sheffield Face (pre-cropped) and corresponding dimensions (shown in parentheses).

DATABASE

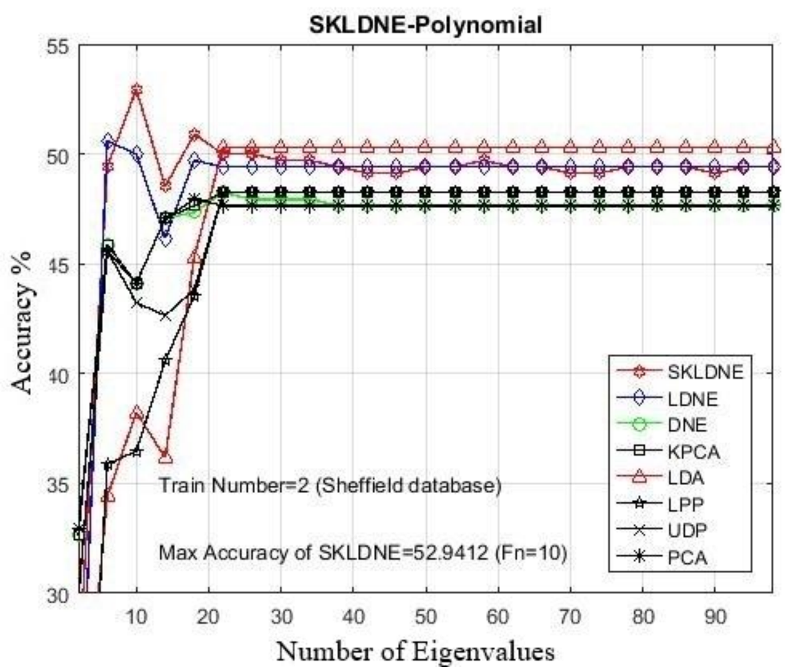
TN	***1***	***3***	***15***	***16***	***17***
PCA	*46.5* (18)	*48.75* (22)	*85* (14)	*93.3* (30)	*92.5* (14)
KPCA	*46.8* (14)	*49.37* (22)	*87* (30)	*93.3* (30)	*95* (14)
UDP	*46.9* (17)	*47.81* (38)	*87.5* (82)	*93.3* (28)	*92.5* (30)
LPP	*48.05* (10)	*50* (30)	*87.5* (38)	*93* (38)	*95* (78)
LDA	*46.94* (14)	*49.06* (22)	*82. 5* (74)	*93.3* (22)	***100*** (30)
DNE	*48.33* (22)	*49.37* (22)	*87.5* (30)	*93.33* (30)	*92.5* (14)
LDNE	*48.33* (22)	*57.81* (18)	*88.75* (14)	*95* (36)	***100*** (13)
SKLDNE	***49.16*** (18)	***60.93*** (22)	***91.25*** (62)	***98.33*** (30)	***100*** (10)

In the following, the output figures of MATLAB with comparative recognition results are plotted with changing the dimensionality of the transformation matrix for each given training number Tn on each data.



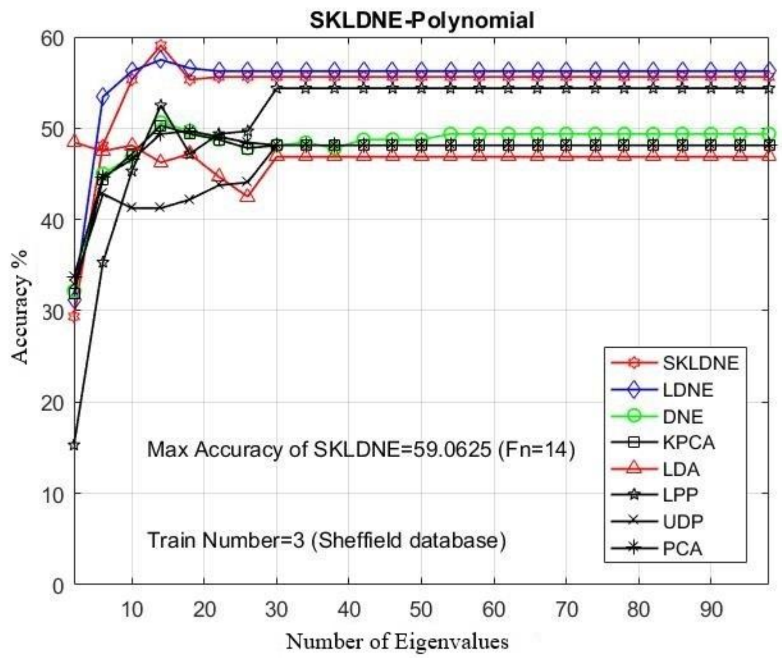
SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
Train Number=1 (Sheffield database)
Max Accuracy of SKLDNE=46.3889 (Fn=10)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

a



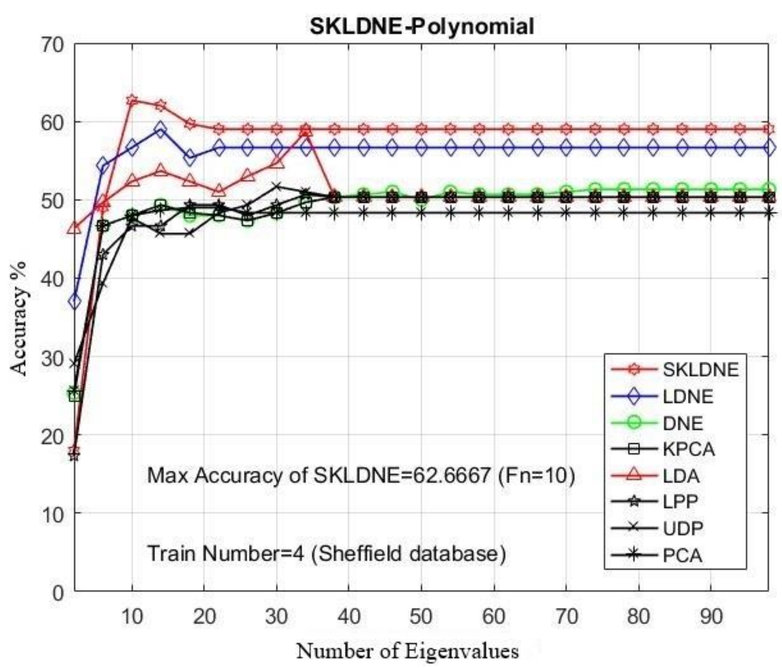
SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
Train Number=2 (Sheffield database)
Max Accuracy of SKLDNE=52.9412 (Fn=10)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

b



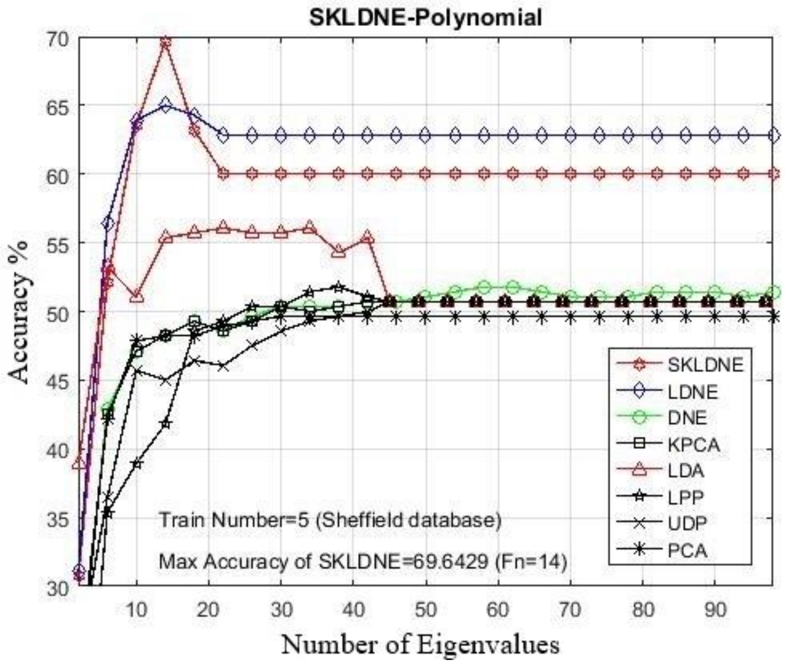
SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=59.0625 (Fn=14)
Train Number=3 (Sheffield database)

c

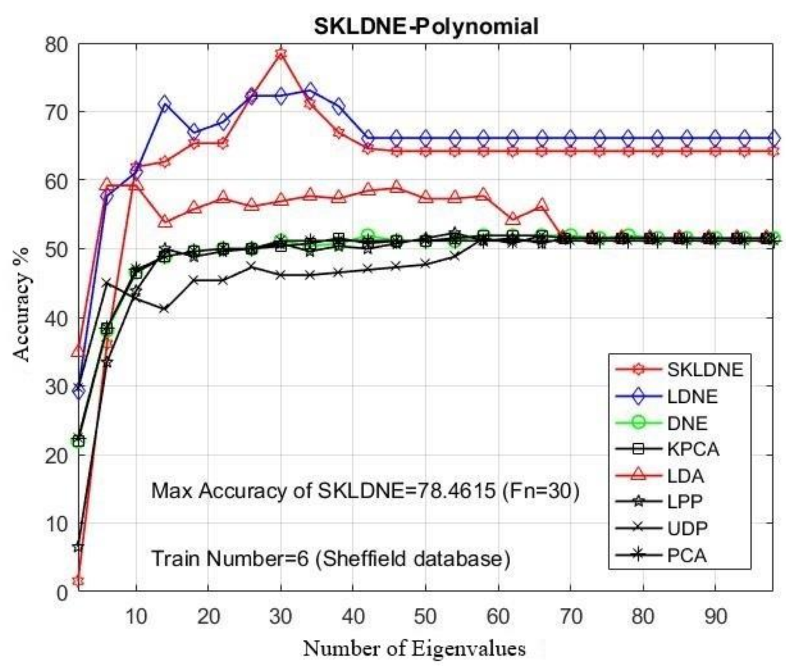


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=62.6667 (Fn=10)
Train Number=4 (Sheffield database)

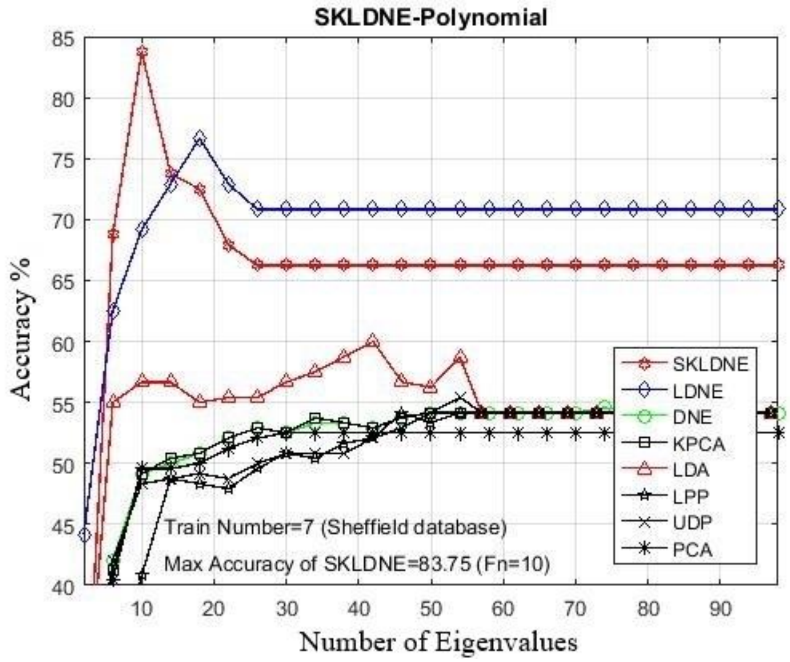
d


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Train Number=5 (Sheffield database)
Max Accuracy of SKLDNE=69.6429 (Fn=14)

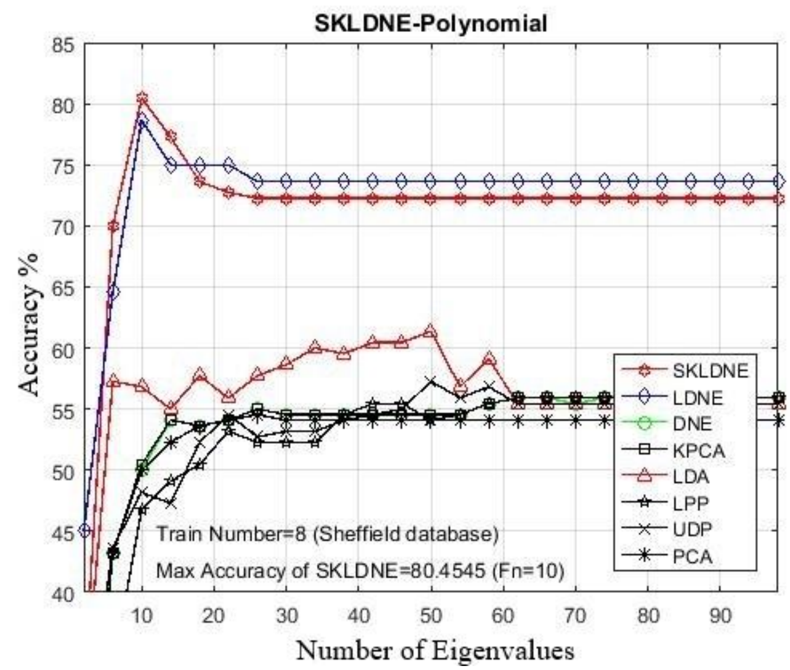
e


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=78.4615 (Fn=30)
Train Number=6 (Sheffield database)

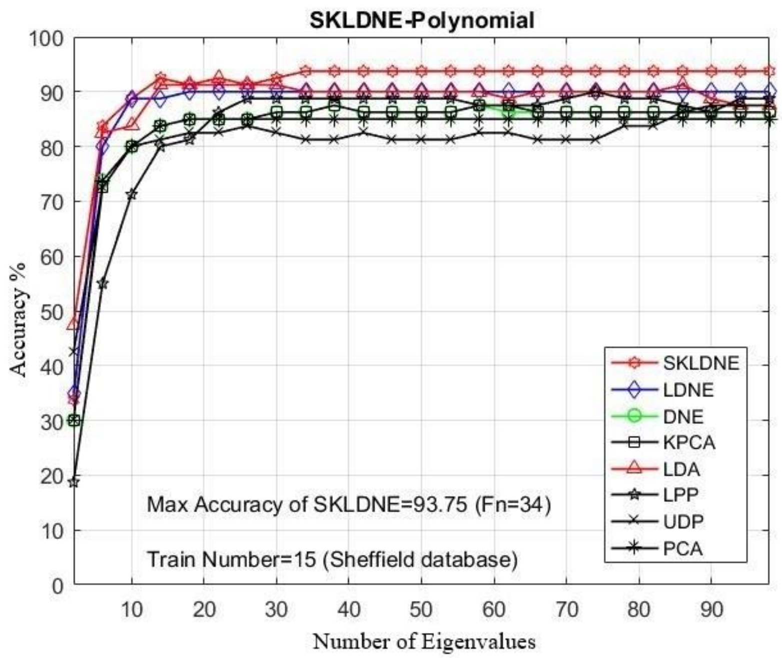
f


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Train Number=7 (Sheffield database)
Max Accuracy of SKLDNE=83.75 (Fn=10)

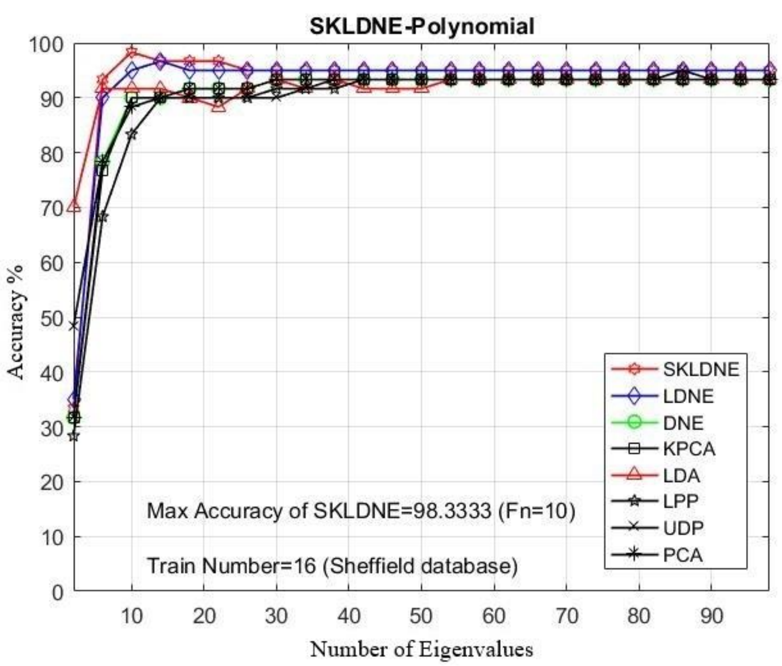
g


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Train Number=8 (Sheffield database)
Max Accuracy of SKLDNE=80.4545 (Fn=10)

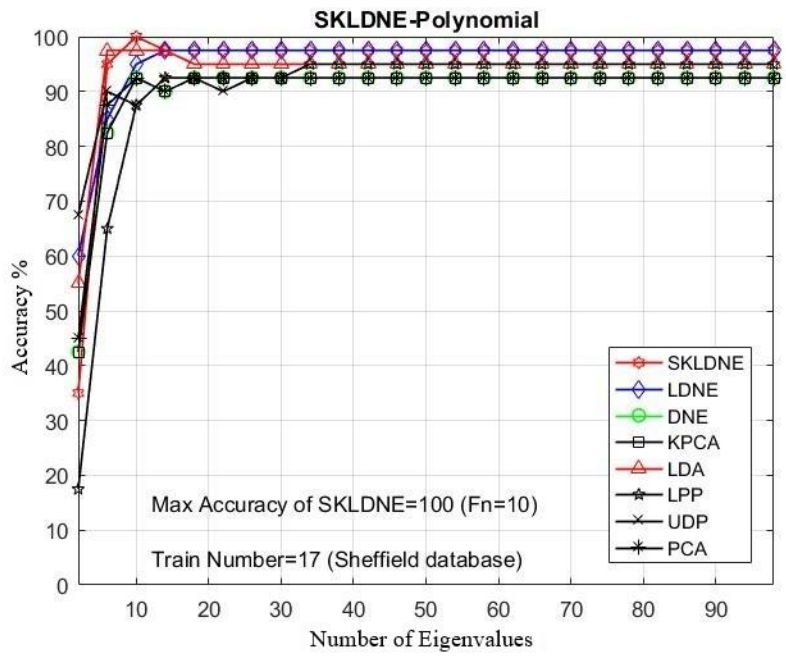
h


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=93.75 (Fn=34)
Train Number=15 (Sheffield database)

i

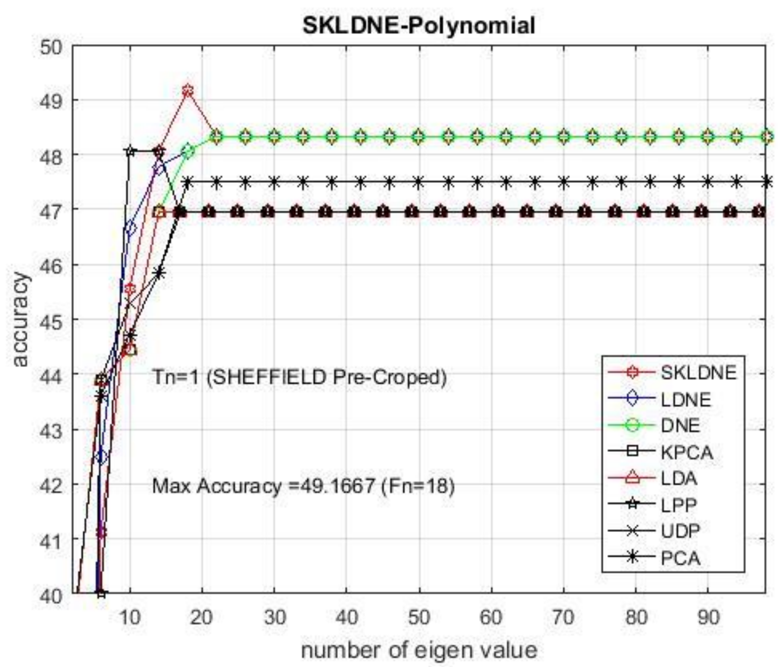

SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=98.3333 (Fn=10)
Train Number=16 (Sheffield database)

j

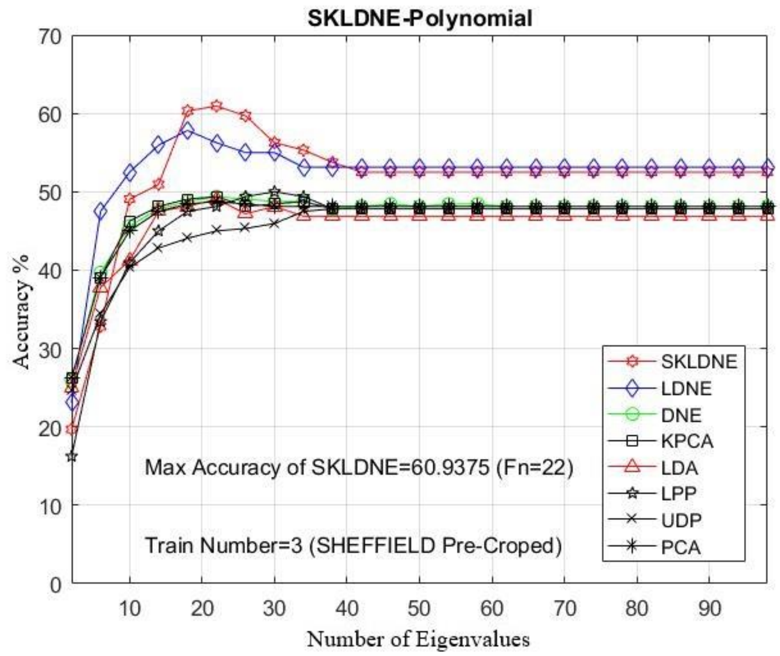


k

Figure 4.4. (a-k). The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn on each data (Sheffield Face database).

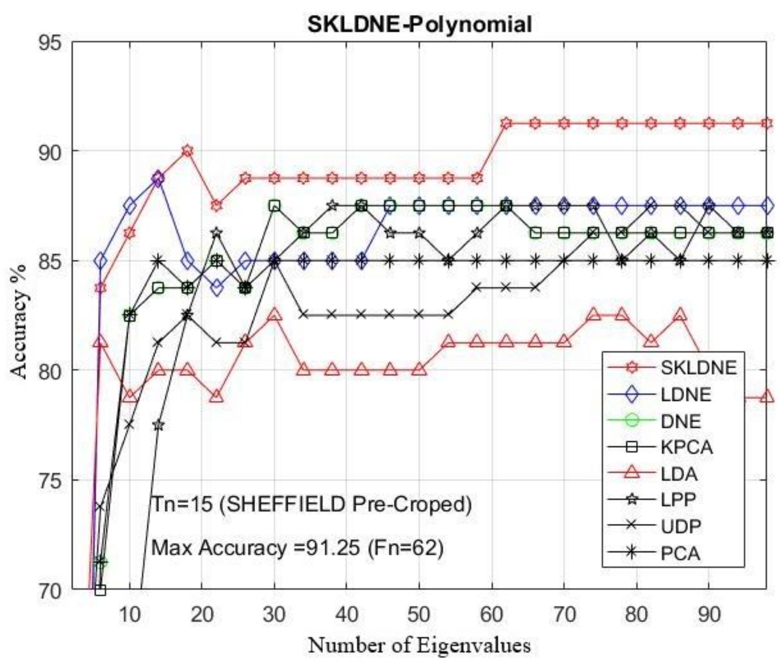


a



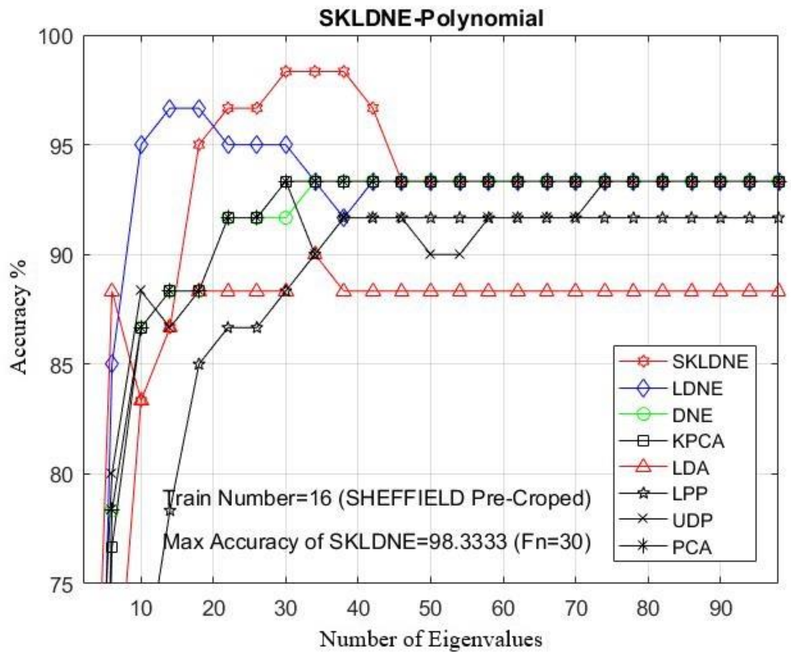
SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=60.9375 (Fn=22)
Train Number=3 (SHEFFIELD Pre-Croped)

b

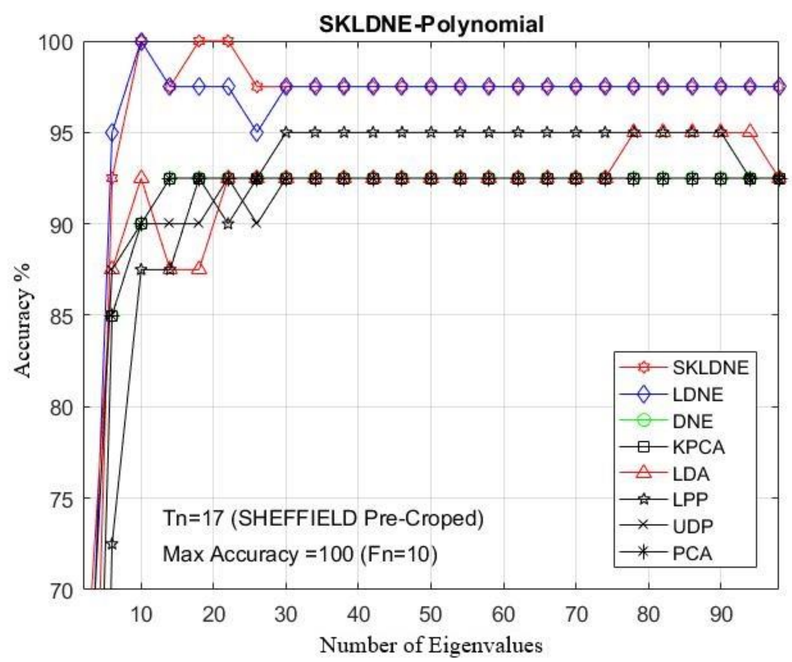


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=15 (SHEFFIELD Pre-Croped)
Max Accuracy =91.25 (Fn=62)

c



d



e

Figure 4.5. (a-e). The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn on each data (Sheffield pre-cropped database).

The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn on Sheffield face and Sheffield pre-cropped database are shown in Figure 4.4a-k and Figure 4.5a-e respectively. According to Table 4.1 and Table 4.2, three main conclusions can be drawn. First, SKLDNE significantly outperformed other methods (PCA, KPCA, UDP, LPP, LDA, DNE and LDNE) over an extensive range of dimensionality for all the different numbers of training and testing images, whether the training sample size was large or small. As can be observed, when the training sample number was small, SKLDNE clearly behaved more efficiently than all other recognition techniques which proves the robustness of this method in a case of small training sample size problem (SSS). Secondly, it is obvious that the recognition rates of all implementations are better when more training samples are used. Third, when the dimensionality increases to about 20, the recognition accuracy of each method first surges rapidly and then roughly becomes stable. The differences between the recognition rate of SKLDNE and other methods are obvious when the training sample number is very small. However, for the larger training sample, the mentioned differences of classification rate increase which shows the superiority of our technique. For instance, in training number of 1, 2, 3, SKLDNE has achieved much better results than others. Besides, for Tn=17 on Sheffield, SKLDNE has reached 100 percent of recognition rate while the results are much lower for other methods. The Accuracy of SKLDNE for Tn=16, 15, 8, 7, 6, 5, 4 is 2.2%, 3.7%, 3.6%, 7%, 5.4%, 4.6% and 4% more than LDNE respectively.

SKLDNE can effectively yield an optimal embedding subspace that finds a substantial submanifolds-based structure with lower dimensionality. The within-class local structure, which is very important for face recognition, can be preserved simultaneously in a nonlinear kernel feature space. SKLDNE can solve the “out-of-sample" problem and the "overlearning of locality" problem in manifold learning, which other aforementioned methods often fail. To explain the superiority of our method compared to its main competitors (LDNE, LPP, and DNE), we should discuss their differences and similarities. LPP is an unsupervised subspace learning that preserves locality without considering class label information of neighbors. Unlike LPP, SKLDNE not only takes into account locality with kernel weighting but also utilizes class label information. For multi-class classification problems, LPP could improperly take inter-class repulsion as intra-class attraction, which may result in that the inter-class neighbors might have the same representations as intra-class neighbors in the transformed space, and could further degrade classification performance. In contrast, SKLDNE divides the neighborhood of a data point into inter-class and intra-class neighborhoods to distinguish points from different classes in the new subspace by analyzing inter-class repulsion and intra-class attraction. It can simultaneously preserve intra-class and interclass geometrical information and have more discriminating power than LPP. Besides, due to the small sample size (SSS) problem, the generalized Eigen equation of LPP cannot be directly solved, but this problem does not exist in our method. SKLDNE, LDNE, and DNE are supervised subspace learning methods. LDNE and DNE are designed to use class information to distinguish points from different classes in the transformed space. However, their projection method might not be effective for preserving locality and nonlinear features. Moreover, because DNE ignores similarities between a point and its neighbors, its simple weight assignment scheme could be inadequate for the analysis of intra-class compactness and inter- class scatterness in the embedded space, which could result in the degradation of classification performance.

Therefore, based on these results, it can be concluded that the recommended SKLDNE technique is a promising technique to be used for dimensionality reduction with very satisfactory performance in classification to deal with high-dimensional data.

4.2.2 Experiment using the Yale database

The Yale face database [83] contains 165 grayscale images in GIF format from 15 individuals. Under different facial expressions and lighting conditions include 11 images for each subject with different facial expression or configuration as following: center-light, wearing glasses, happy, left light, wearing no glasses, normal, right-light, sad, sleepy, surprised and wink. In the experimental results, each image was cropped and resized to 32*32 pixels. Figure 4.6a-b shows sample images of one person in the Yale database and corresponding cropped images and Figure 4.7 demonstrates the different types of implementations with all the different number of training and testing image on the Yale database.

a

b

Figure 4.6 (a). A subset of original YALE database, (b) a subset of cropped images

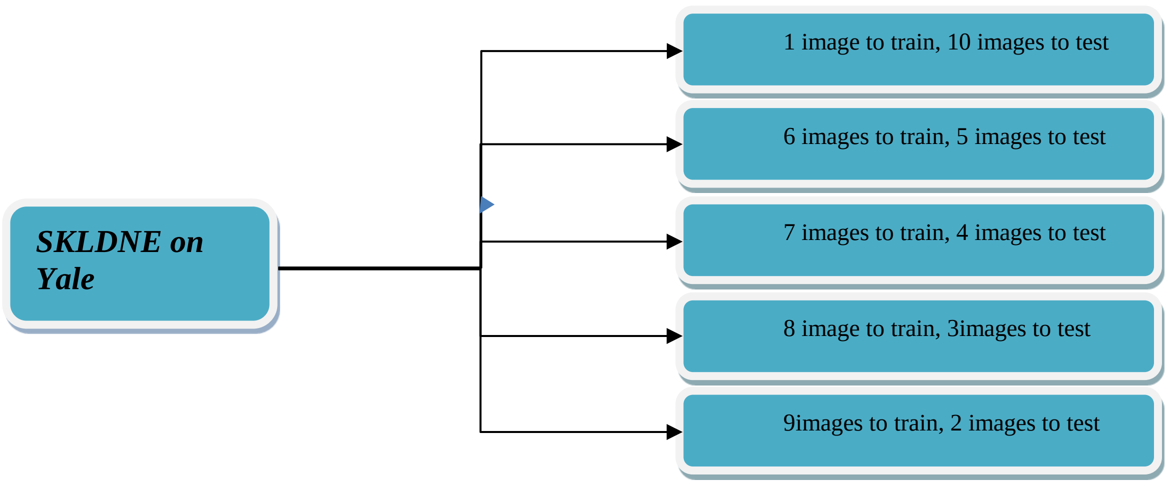
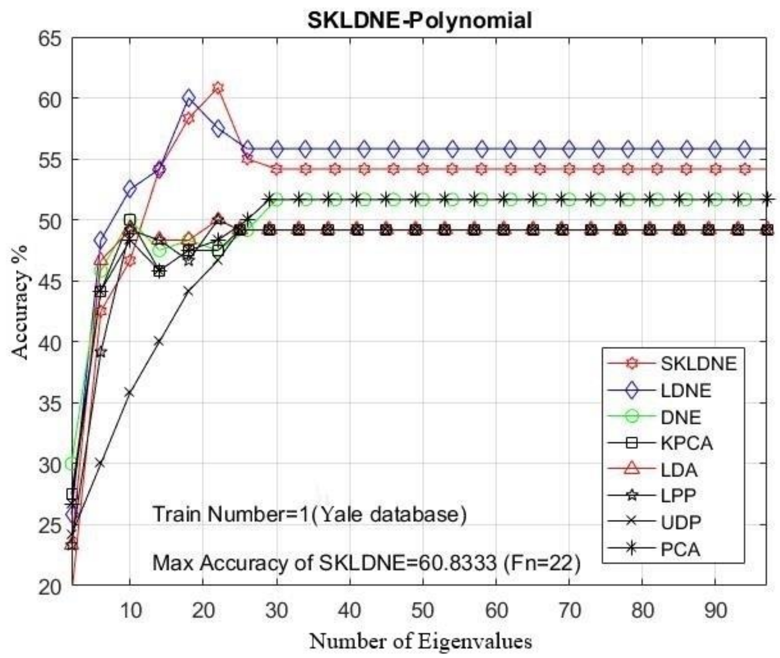


Figure 4.7. Different types of implementations of the SKLDNE method in the Yale Database

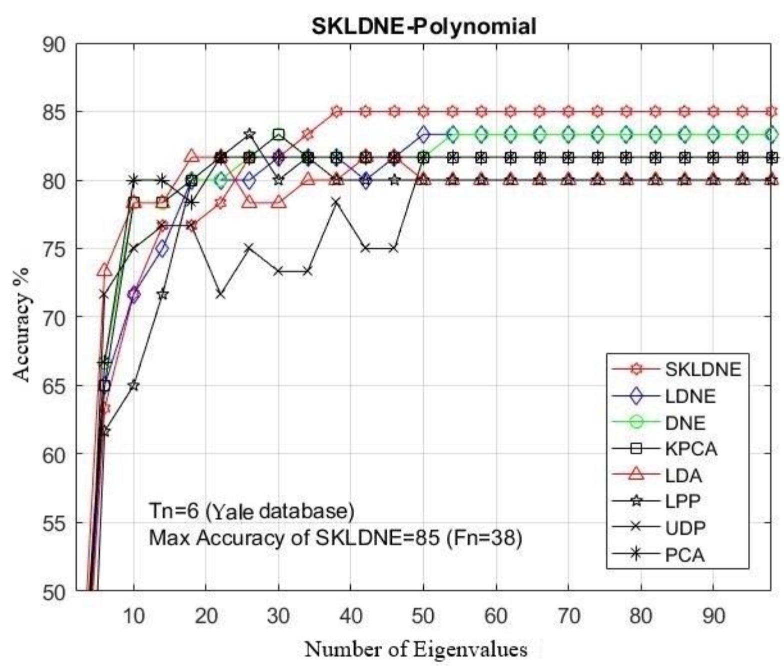
Note that PCA was used in all methods for feature extraction, and all methods include a PCA phase. The optimum rate of recognition of each technique and the equivalent dimension implemented in the Yale database is illustrated in Table 4.3. Furthermore, in all experiments, due to a large number of implementations, it was decided to select some training and testing images that were more challenging for the classification task to clarify the performance of the aforementioned dimensionality reduction techniques in these critical areas. Considering the small training sample size problem (SSS), we first selected a training number of 1 sample and then some larger numbers from 6 to 9 to evaluate the performance of our SKLDNE method in some common existing problems such as the SSS problem and the out-of-sample problem.

Table 4.3. Maximum recognition accuracies (in percentage terms) of SKLDNE and other methods for the different number of training and testing images in the Yale Face database and corresponding dimensions (shown in parentheses).

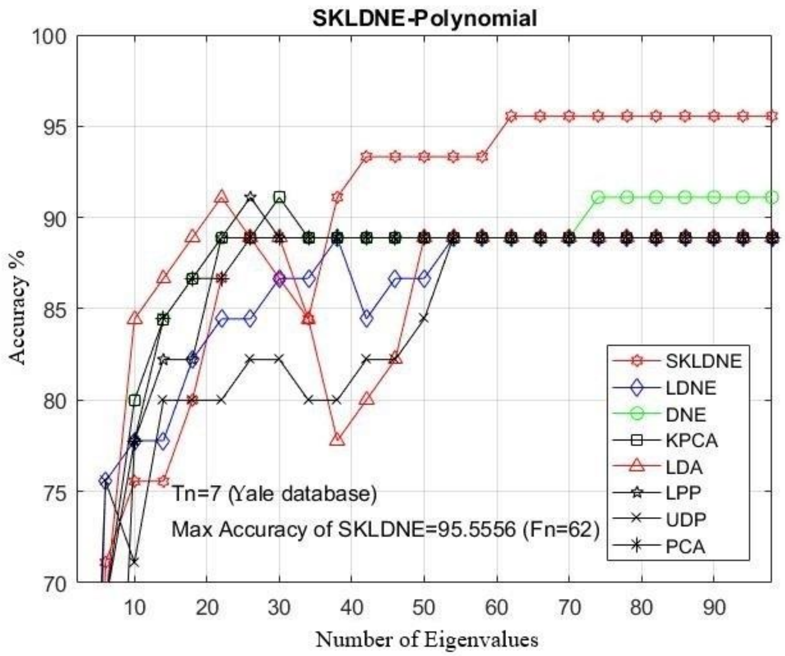
Database	**Yale Face**				
Tn	1	6	7	8	9
PCA	51.66 (29)	81.66 (22)	88.88 (26)	86.6 (26)	93 (10)
KPCA	50 (10)	83.3 (30)	91 (30)	86.66 (90)	93.3 (10)
UDP	49.16 (25)	81.66 (50)	88.8 (54)	90 (28)	92.9 (18)
LPP	51 (22)	83 (26)	91.1 (30)	93 (34)	93.3 (18)
LDA	50 (22)	81.66 (18)	91.1 (22)	93.3 (98)	93.3 (50)
DNE	51.66 (30)	83.3 (30)	91 (30)	90 (66)	93 (10)
LDNE	60 (19)	83.3 (50)	88.88 (57)	93.33 (48)	**100** (42)
SKLDNE	**60.83** (22)	**85** (38)	**95.55** (52)	**96.66** (46)	**100** (26)


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Train Number=1(Yale database)
Max Accuracy of SKLDNE=60.8333 (Fn=22)

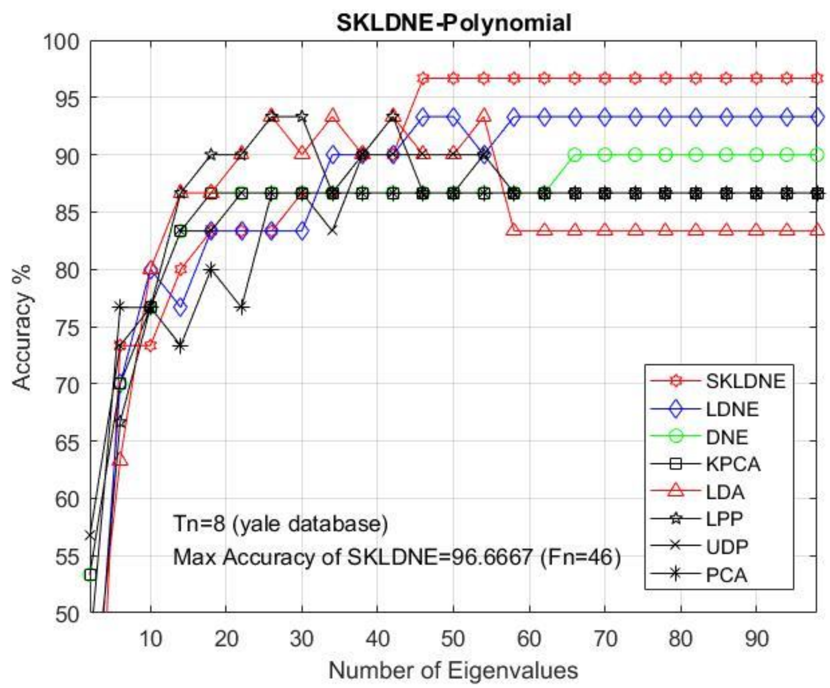
a


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=6 (Yale database)
Max Accuracy of SKLDNE=85 (Fn=38)

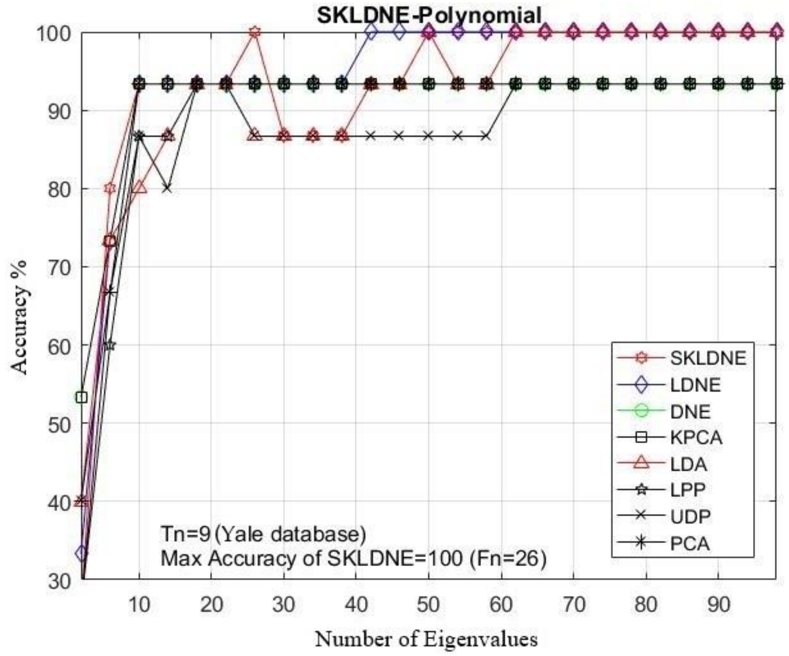
b


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
Tn=7 (Yale database)
Max Accuracy of SKLDNE=95.5556 (Fn=62)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

c


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
Tn=8 (yale database)
Max Accuracy of SKLDNE=96.6667 (Fn=46)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

d



e

Figure 4.8. (a-e). The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn on each data

Table 4.3 shows that the SKLDNE method achieved the highest accuracy in 100% of the implementations in the Yale Database. In Figure 4.8a–e, the comparative classification accuracies are plotted for each given Tn (training number) in each dataset through changing the dimensions of the transformation matrix. As is shown, the recognition rate of each technique increased promptly until the dimensionality was almost 40, and then it stabilized. It can be observed in Table 4.3 that SKLDNE was implemented more efficiently than others among a wide variety of dimensionality in the Yale Face Database. Meanwhile, the best implementation of SKLDNE was achieved at smaller dimension values in most of the training numbers for each data set compared to LDNE. Moreover, differences in the classification between SKLDNE and other methods are very clear, especially when the training number was small, for instance, Tn = 1. For training number 7, SKLDNE yielded an improvement of around 4.5% compared with DNE, LPP, LDA, and KPCA, and 6.6% in comparison with LDNE, UDP, and PCA respectively. For training number 9, both SKLDNE and LDNE gained 100% accuracy, while accuracies in other methods with the same training number were much lower. To explain the superiority of the proposed method, our SKLDNE first mapped the data in the kernel space to capture the substantial extracted data and then both geometrical and discriminant information of the data were taken, benefiting from a significant form of the affinity weight matrix to embed the graph. Although LPP, DNE, and LDNE outperforming PCA, KPCA, and UDP demonstrates that the discriminant and local data structure-based methods are more suitable for face recognition, our SKLDNE had more nonlinear data representation, locality preservation, and discriminating power than other methods, and consequently achieved the best recognition accuracy. Therefore, based on the mentioned characteristics of SKLDNE, it can be concluded that our SKLDNE can overcome the “SSS,” “out-of-sample,” and “overlearning” problems.

4.2.3 Experiment using the ORL database

The ORL face database [84] contains a set of face images capturing between 1992 and 1994 in the AT&T lab in collaboration with the Robotic Group of the Cambridge University Engineering Department [85]. There are ten different grayscale images from 40 distinct subjects. All 10 face images of each subject were captured at different times, with changes in the lighting, facial details (with glasses or no glasses) or facial expressions (smiling/not smiling, open/closed eyes,), against a dark homogeneous background, and with straight and frontal views. The size of all images is equal to 92 × 112 pixels in PGM format. The images were captured in 40 directories [86]. It should be noted that preprocessing was used and all original images were already cropped and resized. In this project, the size of 32 × 32 pixels was chosen for all ORL images. Figure 4.9 illustrates three different subjects (each with 10 images) from the ORL database.

Figure 4.9. The three different subjects (each with 4 images) from ORL database.

Figure.4.10 below shows the different types of implementations of the SKLDNE method in the ORL Face Database include one small training sample size and the rest of larger size.

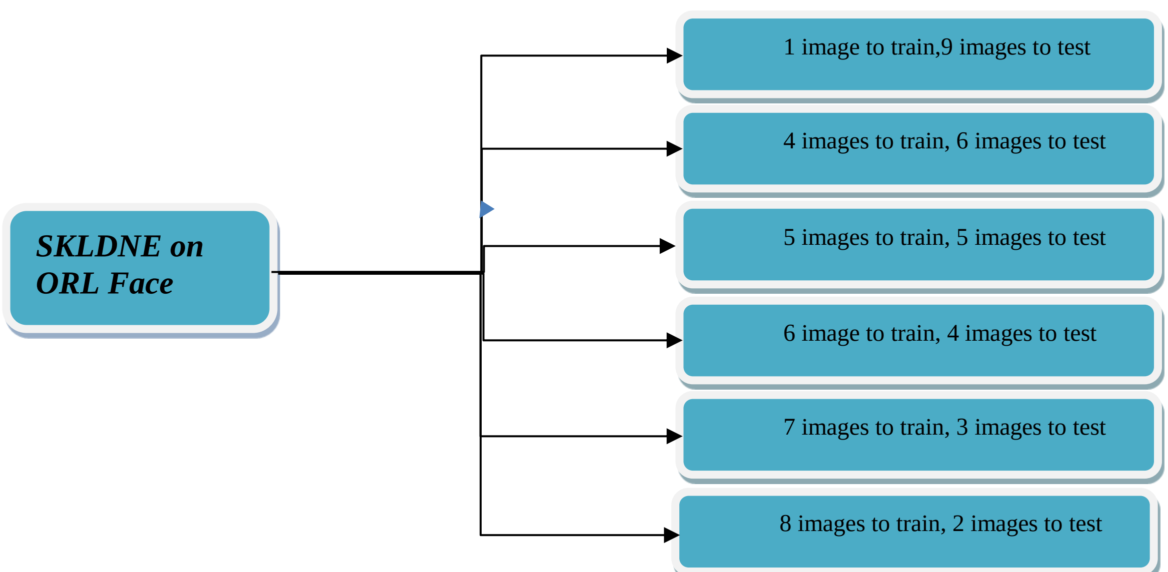
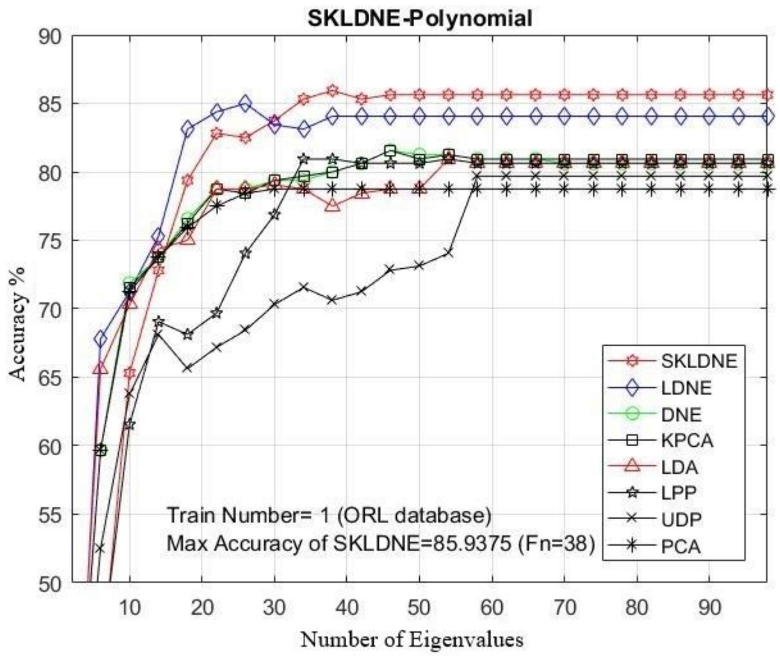


Figure.4.10. The different types of implementations of the SKLDNE method in the ORL Face Database

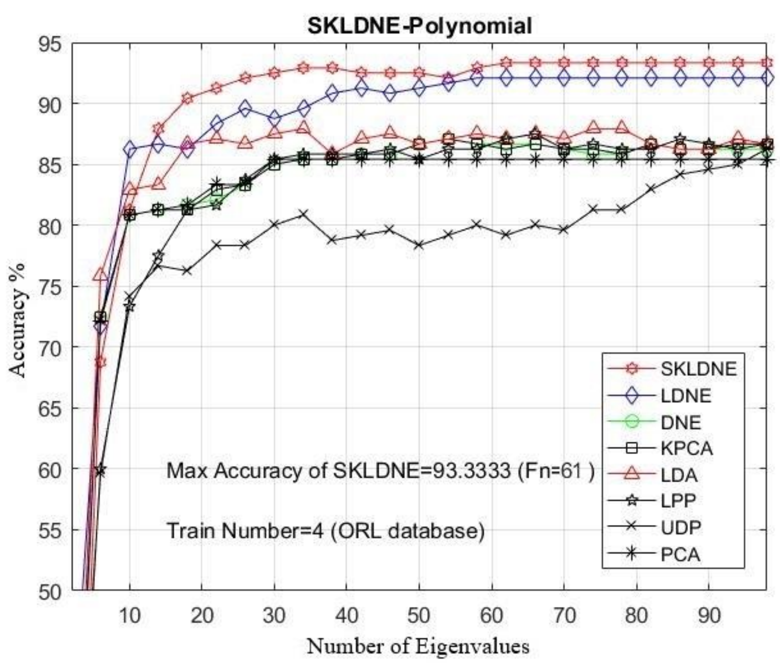
In our experiments, the number of training samples Tn = 1, 4, 3, 4, 5, 6, 7, 8 were chosen from the dataset related to each subject to make the training sample set. The other numbers of images are applied as a testing sample set. As already mentioned, PCA was used in the classification phase in all methods. The maximal average accuracy (in percentage terms) and its corresponding dimension, followed by the alteration in the training sample sizes, are illustrated in Table 4.4. It should be noted that the best performance among other methods is indicated in boldface. From Table 4.4 and Figures 4.11a-f, it can be observed that SKLDNE generally outperformed LDNE, whether the number of training sample size was small or not, in almost smaller numbers of dimensions. Moreover, as a supervised method, SKLDNE also significantly outperformed other techniques (KPCA, LPP, DNE, UDP, and LDA) regardless of the change in the training sample size. Compared to other techniques, SKLDNE performed better in small training sample size case. Furthermore, when the training number was equal to 8, SKLDNE had a zero error rate compared to that of PCA (4.1%), KPCA (3.75%), UDP (4%), LPP (2.5%), LDA (4%), DNE (3.75), and LDNE (3.5%). The Accuracy of SKLDNE for Tn=7, 6, 5, 4 is 2.5%, 1.2%, 2%, and 1.3%more than LDNE respectively. SKLDNE can simultaneously discover inter-class and intra-class geometrical information and have more nonlinear data representation, locality preservation, and discriminating power than other techniques. Therefore, SKLDNE does have merit over other techniques in terms of resolving classification problems in face recognition. This characteristic of SKLDNE in small sample size cases is indeed important to improve the recognition rate in practice since face recognition is commonly a small sample size problem. Normally, a small number of images of each person are accessible in many real-world tasks.

Table 4.4. Maximum recognition accuracies (in percentage terms) of SKLDNE and other methods for the different number of training and testing images in the ORL Face database and corresponding dimensions (shown in parentheses).

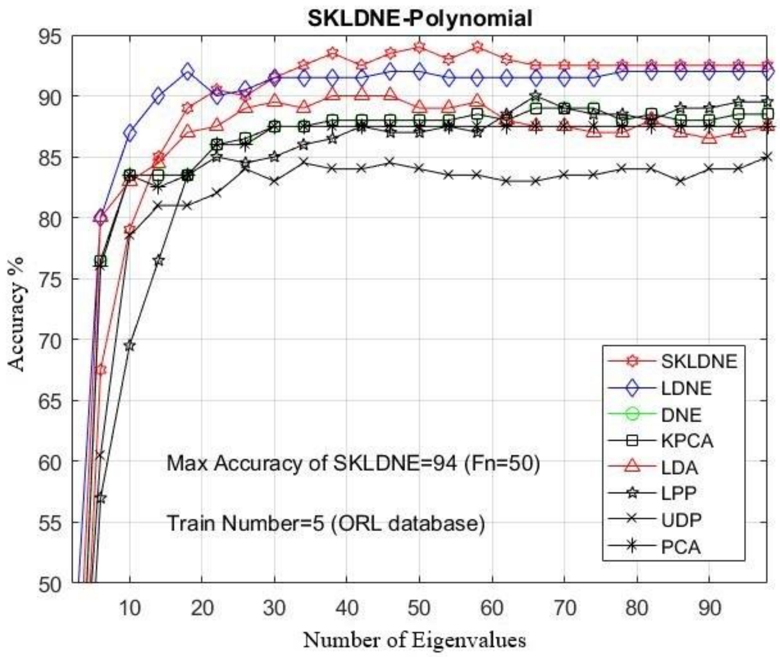
DATABASE	ORL FACE					
TN	**1**	**4**	**5**	**6**	**7**	**8**
PCA	78.75 (30)	85.41 (30)	87.5 (26)	95.62 (20)	95.83 (10)	95.9 (10)
KPCA	81.56 (46)	87 (54)	89 (66)	96.2 (34)	96.66 (34)	96.25 (20)
UDP	80 (54)	86.66 (90)	89.5 (98)	94.75 (38)	96.6 (18)	96 (14)
LPP	80.62 (58)	87.5 (86)	90 (94)	95 (34)	95.8 (30)	97.5 (62)
LDA	80.93 (54)	87.91 (34)	90 (38)	96.25 (22)	95.83 (34)	96 (18)
DNE	81.56 (46)	87.08 (54)	89 (66)	96.2 (34)	96.66 (34)	96.25 (10)
LDNE	85 (26)	92 (62)	92 (78)	95.6 (30)	95 (24)	96.5 (54)
SKLDNE	**85.93** (38)	**93.33** (61)	**94** (50)	**96.87** (66)	**97.5** (22)	**100** (18)


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Train Number= 1 (ORL database)
Max Accuracy of SKLDNE=85.9375 (Fn=38)

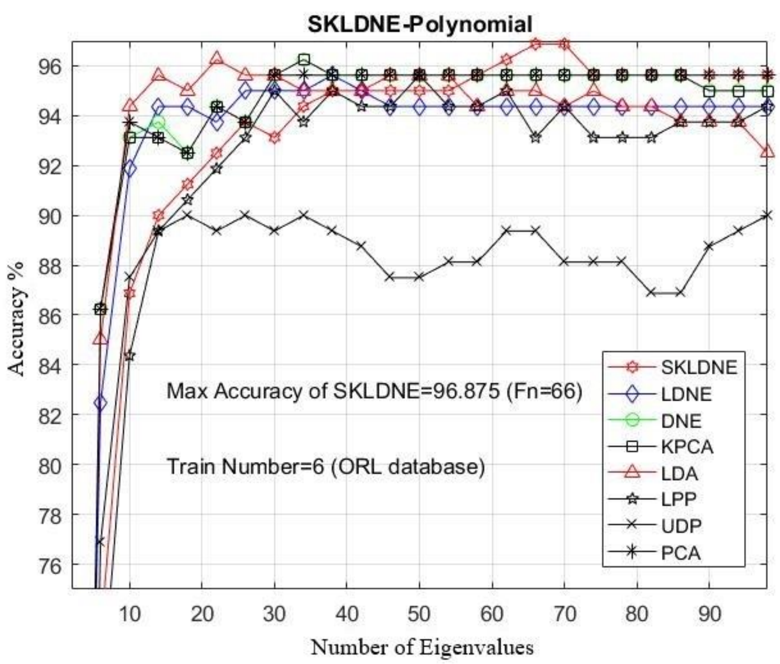
a


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Max Accuracy of SKLDNE=93.3333 (Fn=61)
Train Number=4 (ORL database)

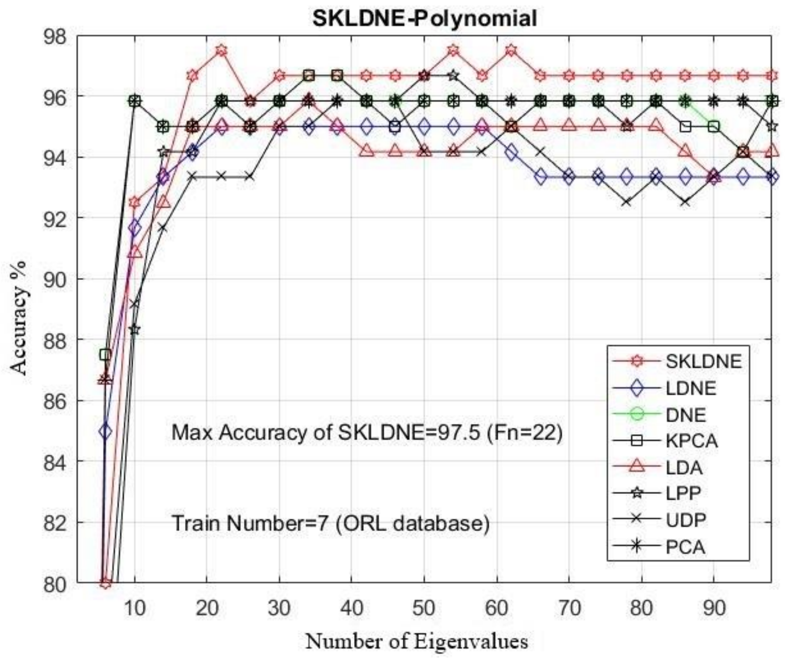
b


SKLDNE-Polynomial
Max Accuracy of SKLDNE=94 (Fn=50)
Train Number=5 (ORL database)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Accuracy %
Number of Eigenvalues

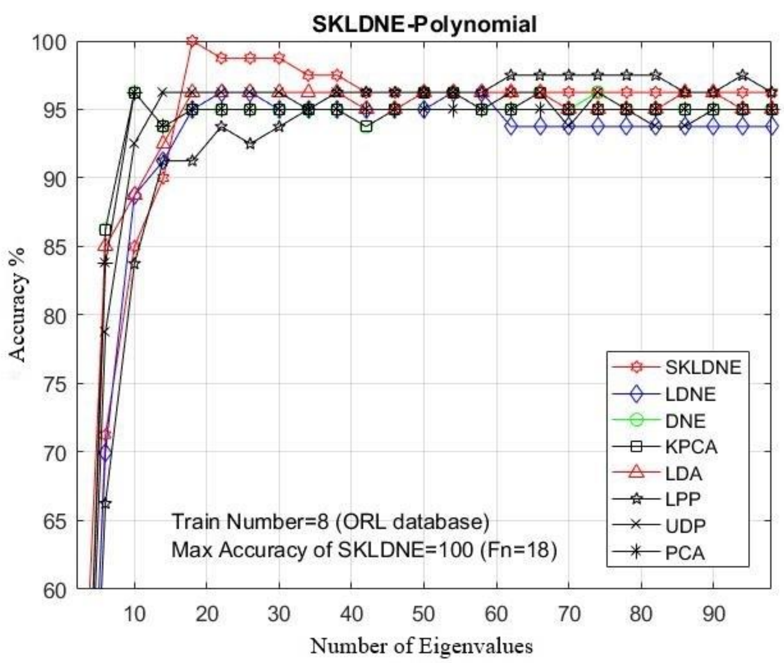
c


SKLDNE-Polynomial
Max Accuracy of SKLDNE=96.875 (Fn=66)
Train Number=6 (ORL database)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Accuracy %
Number of Eigenvalues

d



e



f

Figure 4.11. (a-f). The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn on each data(ORL database)

4.2.4 Experiment using the Head Pose database

Head Pose database[87, 88] contains 2790 face images of 15 individuals with variation of pan and tilt angles from -90 to +90 degrees. For every person 2 series of 93 images (93 different poses) were taken. Figure 4.12 illustrates a subset of images of one subject from head pose image database.

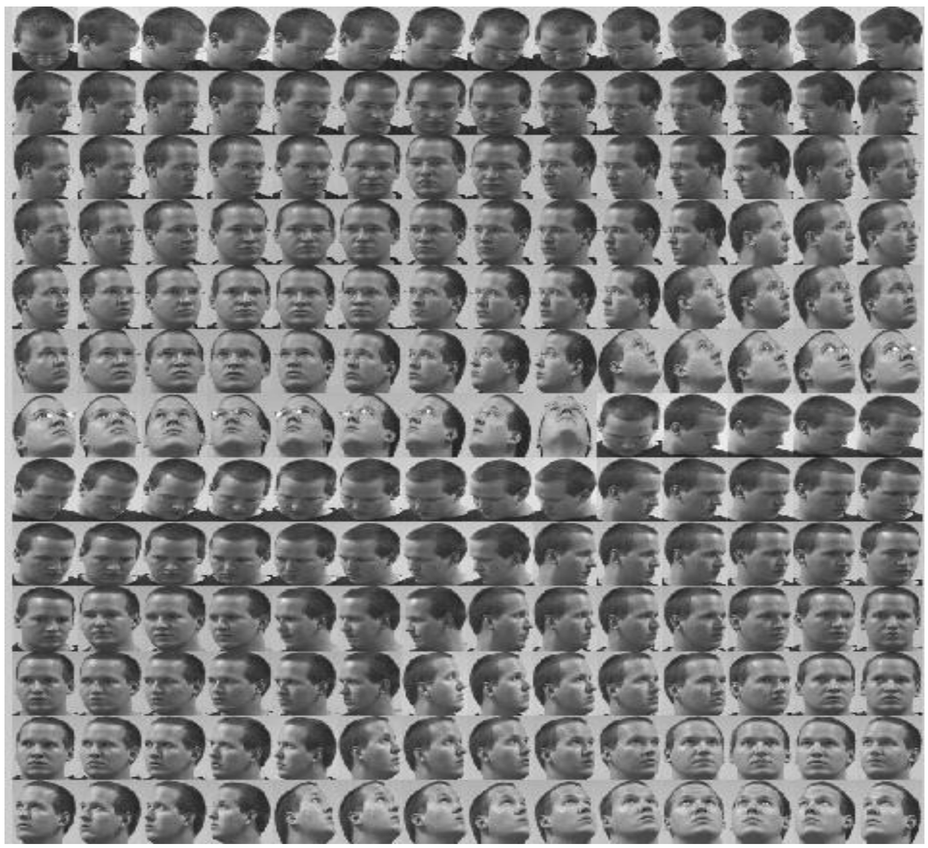


Figure 4.12. A subset of images of one subject from the Head Pose database

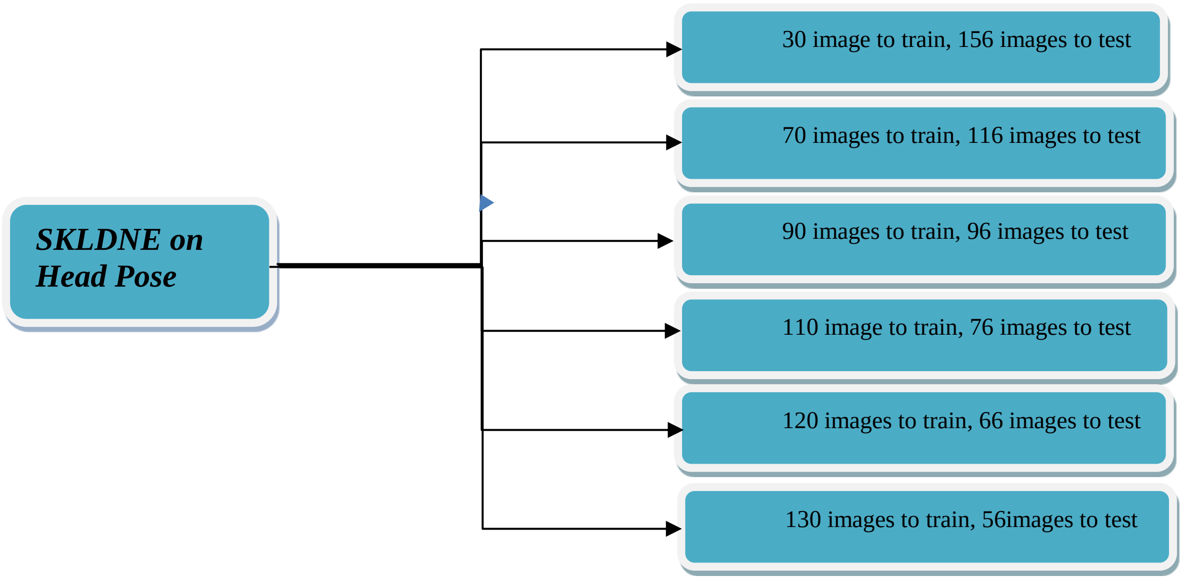
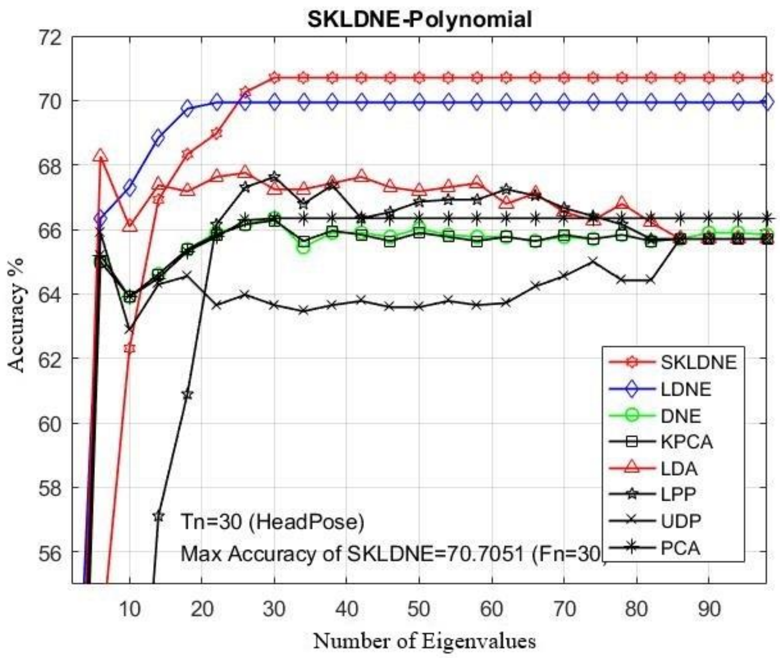


Figure 4.13. The different types of implementations of the SKLDNE method in the Head Pose Database

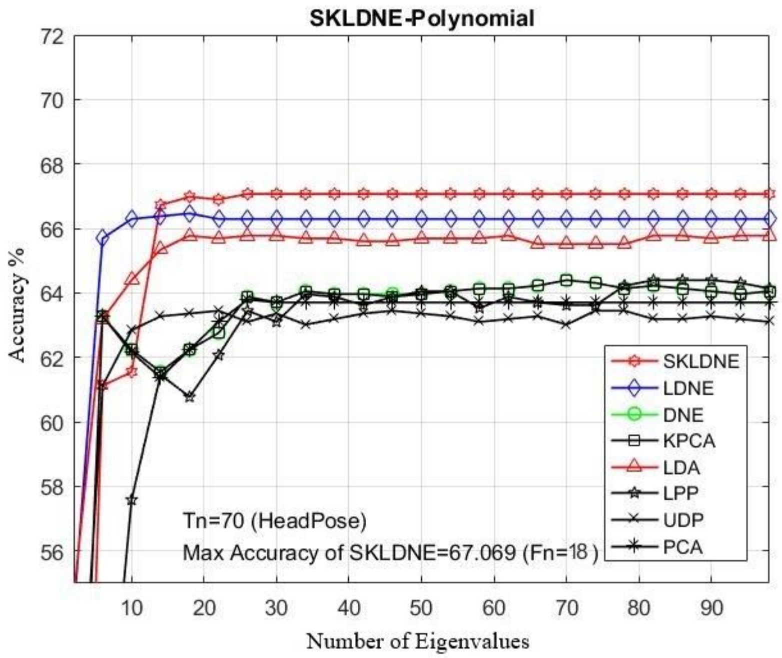
As can be observed in Table 4.5 and Figure 4.14a-f, SKLDNE performed better than the other seven methods, regardless of the variation in the training sample size. The maximal recognition rate of SKLDNE when Tn = 130 was up to 99.28%, while for other methods it was much lower. This reveals that, when the given training sample size for each class gets larger, SKLDNE can obtain much better results than other methods. Two more points can also be outlined. First, our supervised method with kernel weighting can notably enhance the class classification performance, but applying the kernel trick has no significant influence on PCA performance. Second, SKLDNE achieves optimal recognition rates at an almost smaller number of dimensions as the recognition rate of SKLDNE retains the best results as the dimension varies from 14 to 30. Compared to the other techniques, SKLDNE preserves the more discriminating and local features of face images. It also preserves more local geometric relations of the within-class samples by nonlinear kernel mapping. It should be noted that linear methods such as LPP, LDA, LDNE, DNE, and UDP often fail to deliver good classification performance when face images are subject to complex nonlinear changes such as expression, lighting, pose and so on. Figure 4.14a-f indicates that the recognition implementations of all methods first sharply increase while the projected dimensions are added, and then, after obtaining the optimum, they tend to become stable.

Table 4.5. Maximum recognition accuracies (in percentage terms) of SKLDNE and other methods for the different number of training and testing images in the Head Pose database and corresponding dimensions (shown in parentheses).

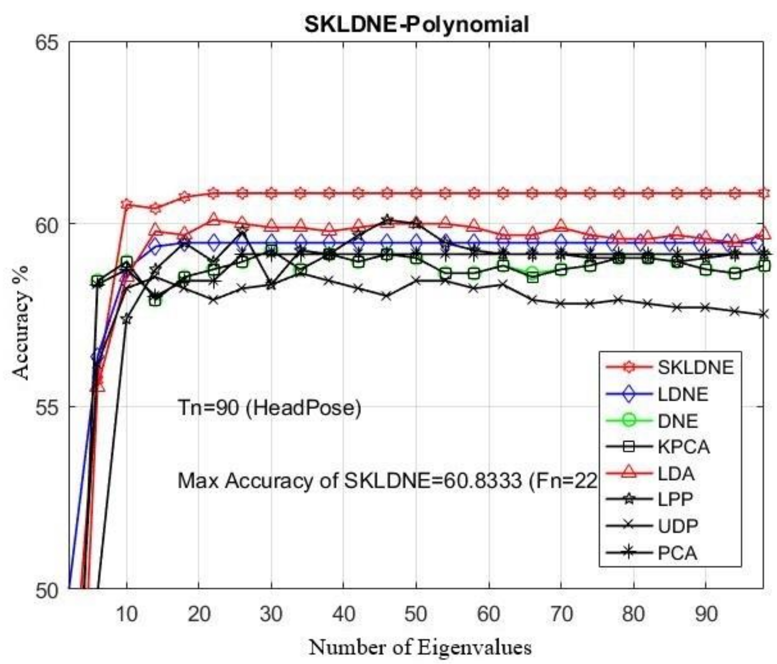
DATABASE	**HEAD POSE**					
TN	**30**	**70**	**90**	**110**	**120**	**130**
PCA	66.21 (30)	63.69 (26)	50.16 (26)	79.73 (26)	83.6 (74)	82.67 (26)
KPCA	66.38 (30)	64.31 (70)	59.37 (30)	84.86 (90)	86.22 (22)	85.71 (62)
UDP	65.83 (6)	64.39 (22)	58.75 (34)	85.52 (98)	87.21 (94)	85.71 (78)
LPP	68.01 (30)	64.4 (82)	59.7 (26)	85.39 (90)	88.36 (62)	87.85 (50)
LDA	68.21 (6)	65.14 (18)	60 (22)	86.57 (98)	87.04 (98)	88.57 (46)
DNE	66.28 (30)	64.5 (70)	58.5 (30)	84.63 (90)	86.22 (74)	85.71 (62)
LDNE	69.7 (20)	66.2 (19)	59 (18)	96.9 (24)	98 (18)	98.02 (18)
SKLDNE	**70.7** (30)	**67.06** (18)	**60.83** (22)	**98.94** (22)	**99.01** (18)	**99.28** (14)


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=30 (HeadPose)
Max Accuracy of SKLDNE=70.7051 (Fn=30)

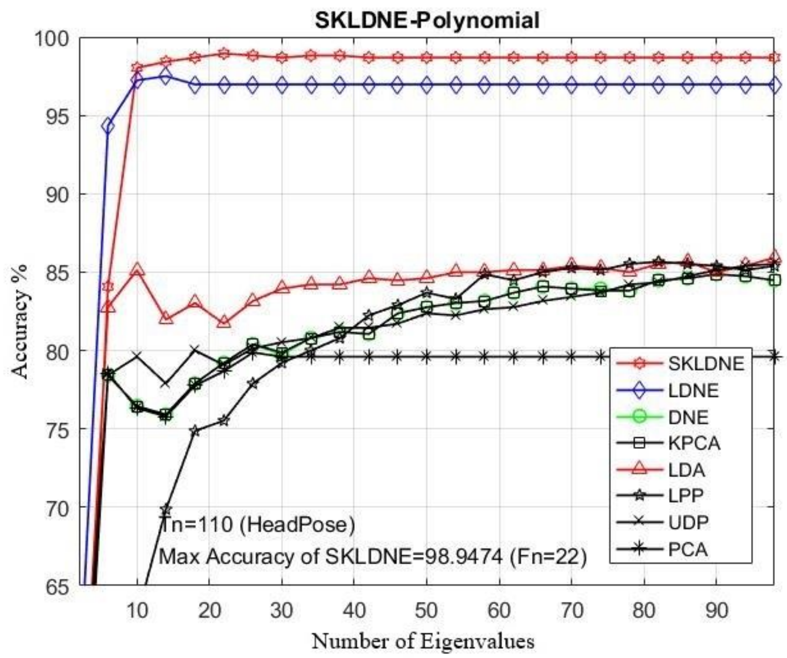
a


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=70 (HeadPose)
Max Accuracy of SKLDNE=67.069 (Fn=18)

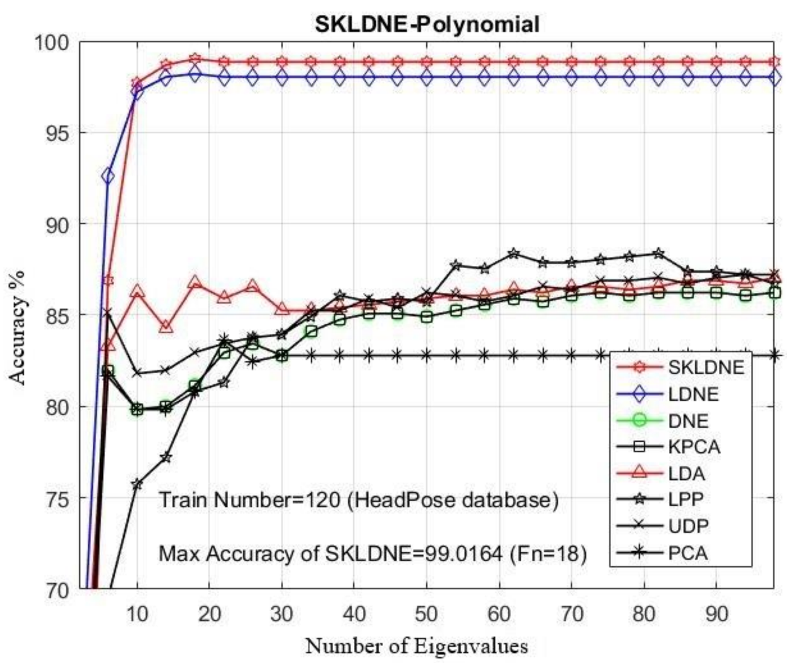
b


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=90 (HeadPose)
Max Accuracy of SKLDNE=60.8333 (Fn=22

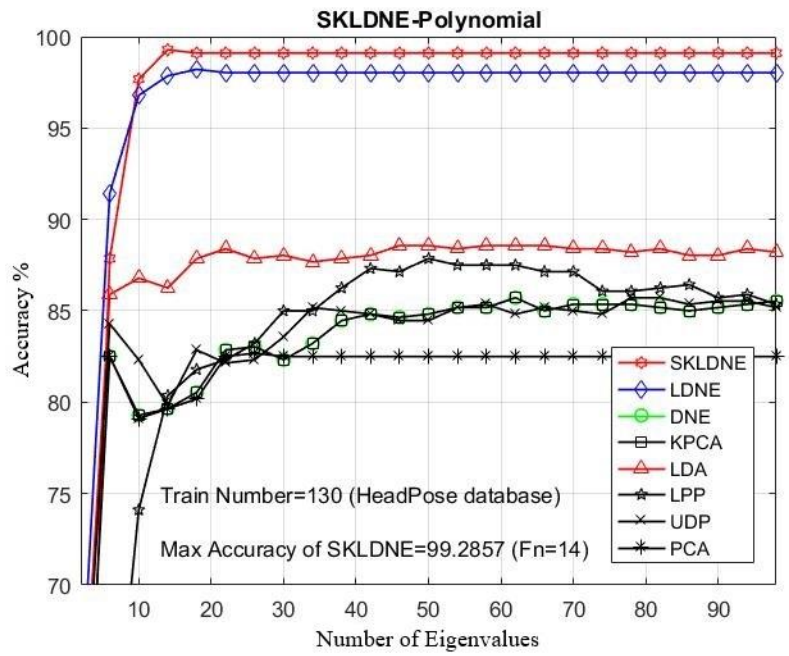
c


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=110 (HeadPose)
Max Accuracy of SKLDNE=98.9474 (Fn=22)

d


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Train Number=120 (HeadPose database)
Max Accuracy of SKLDNE=99.0164 (Fn=18)

e



f

Figure 4.14. (a-f). The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn on each data(Head Pose)

4.2.5 Experiment using the Finger Vein and Finger Knuckle Print databases

To examine the performance of our method on other databases rather than face database, it has been decided to do all implementations on Finger Vein and Finger Knuckle Database which are two famous public databases. The following is a brief explanation of both databases. The Finger Vein database used in this project was collected from 51 individuals (male and female) who were aged between 21 and 56 [89]. 10 images were captured from each subject. Four fingers were used for capturing, including right and left middle finger and right and left index finger. There are 204 different fingers in the database, and the data consist of 2040 images in total, in which each finger image originally had a dimension of 480 × 160 pixels. In our implementations, each image was resized to 32 × 32. The captured images from one person can be seen in Figure 4.15.

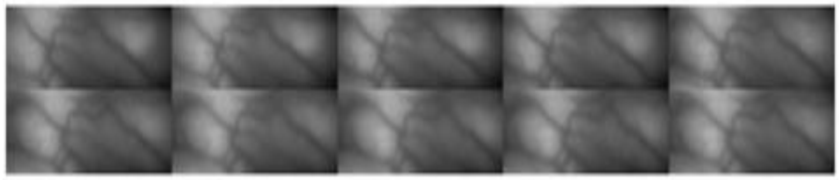


Figure 4.15. The captured images from one person in the Finger Vein Database

The Finger Knuckle Print (FKP)[90] database has been provided by Hong Kong Polytechnic University and is freely available online. Based on the database description, Finger Knuckle Print (FKP) images were collected from 165 individual volunteers (males and females) [91]. The samples were collected in two distinct sessions and, in each one, 6 images were captured from 4 fingers (including left and right index finger and the left and right middle finger). Therefore, 7920 finger images in total were taken from 660 different fingers. In our experiments, the original image of the database was cropped and was then resized to 32 × 32 pixels. Figure 4.16 shows a cropped sample of the FKP database.

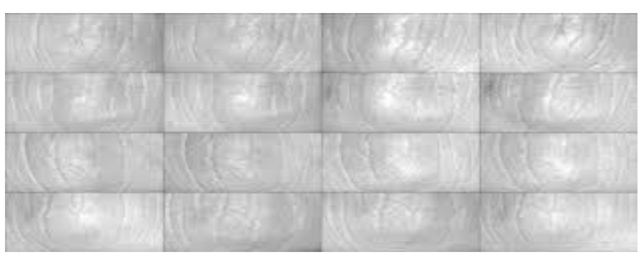


Figure 4.16. A cropped sample of the FKP database.

Figure 4.17 and Figure 4.18 illustrate the different types of implementations by the SKLDNE method in the Finger Vein and Finger Knuckle databases respectively.

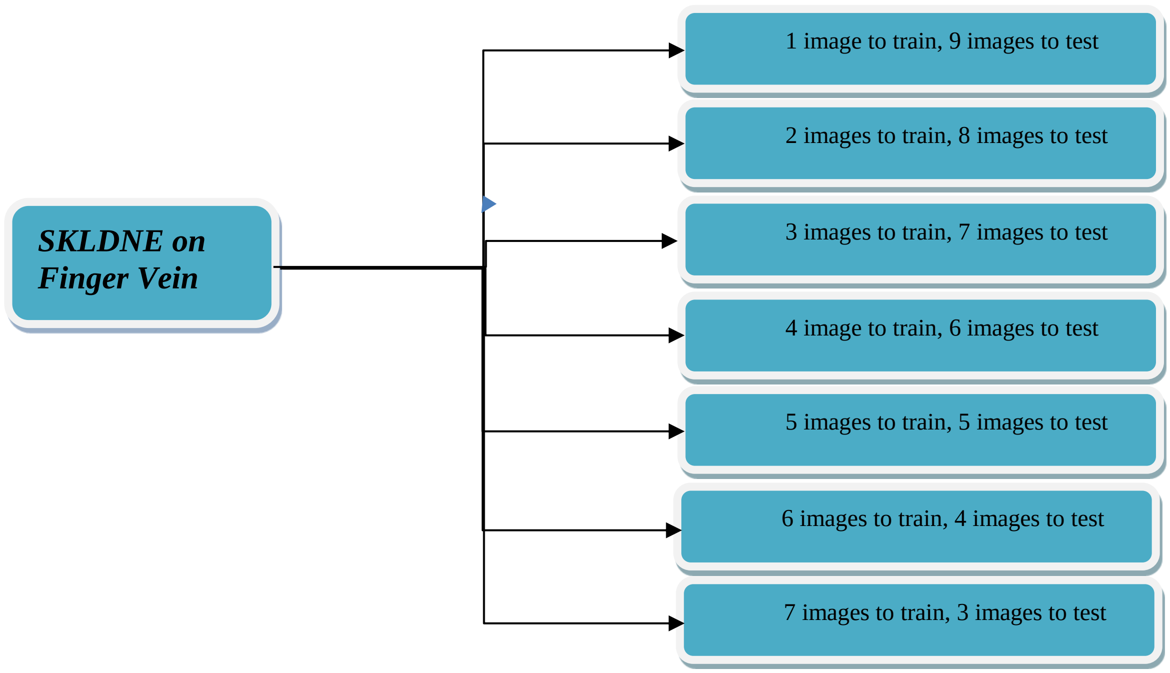


Figure 4.17 .The different types of implementations of the SKLDNE method in the Finger Vein

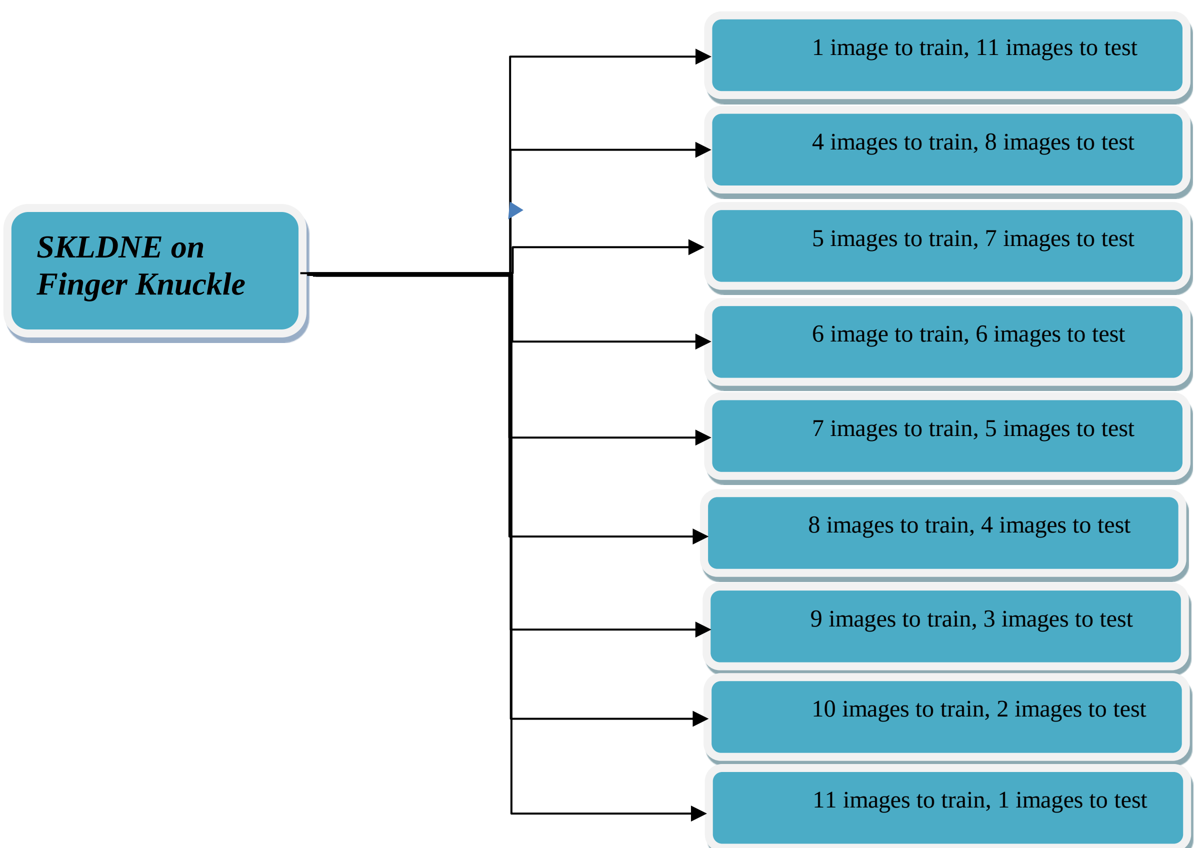


Figure 4.18. The different types of implementations of the SKLDNE method in the Finger Knuckle

In this section, the performance of each method is explored by changing the dimensionality of the transformation matrix and the related best recognition rate with the corresponding dimension on each database. Besides, the first Tn images of each subject in the dataset are used for training and the remaining images for a test. Furthermore, the PCA classifier using Euclidean distance is applied in the recognition phase. The best classification rates of all methods implemented in both databases are demonstrated in Table 4.6 and Table 4.7 (in the Finger Vein and Finger Knuckle databases respectively). Based on the experiment results shown in Table 4.6 and Figure 4.19a-g, our SKLDNE gained the best recognition rate among all the different training numbers in the Finger Vein database, which proves that it has a convincing performance compared to other advanced methods. Regarding the small training sample size case, the SKLDNE method still showed that it performed significantly better than other techniques, as the maximal recognition accuracy rate of SKLDNE in training number 2 was almost 5% more than PCA, KPCA, UDP, LPP, and DNE, and 1.2% more than LDNE. The Accuracy of SKLDNE for Tn=5 is 2.8%, 2.4%, 1%, 2.6%, 2.2%, 2.8% and 2.7% more than LDNE, DNE, LDA, LPP, UDP, KPCA, and PCA respectively. Our SKLDNE method was always able to represent its optimal embedding space with a lower value of dimensions in comparison with the other seven techniques. For example, the SKLDNE results for Tn = 7 achieved 100% recognition accuracy in the smallest value of projected dimensions (26). This conveys that our approach is more effective, due to its significant characteristics, being able to represent not only nonlinear and complex variations in images but also to model both localities of LPP and discrimination of DNE simultaneously. This fact demonstrates the good performance of our proposed method.

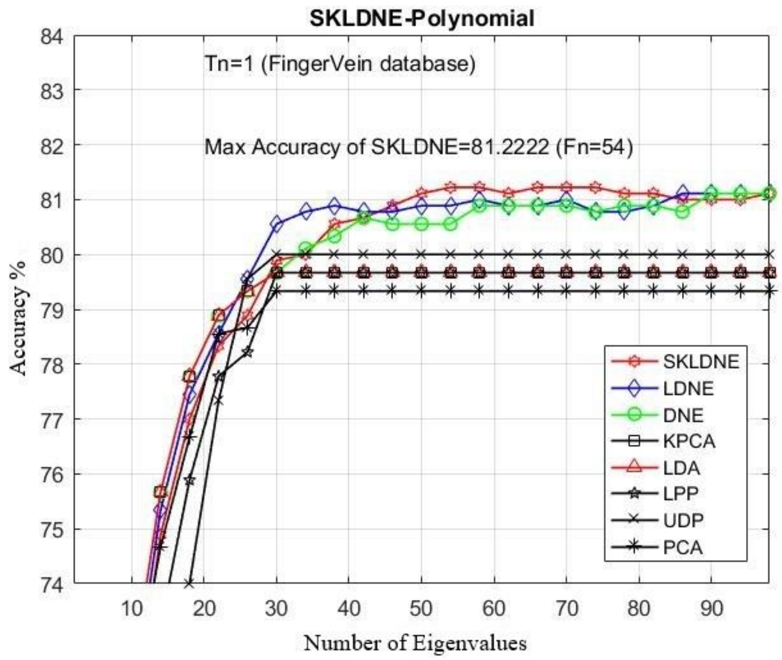
Table 4.7 shows that the SKLDNE recognition performance was significantly more efficient than other techniques, regardless of the variation in the training sample size, in the Finger Knuckle database. The recognition rate of SKLDNE when Tn = 11 was equal to 100%, while for the LDNE it was equal to 97.2%. Another point that is worth mentioning on the recognition performance of SKLDNE compared to other methods is related to the small training sample size case, as SKLDNE had the best performance in this respect. Again, all the best recognition performances of SKLDNE were mostly achieved at smaller dimension values on every Tn per data set. It can also be observed that, when the given training sample size of each class became larger, SKLDNE achieved much better results than other techniques. For example, for Tn = 6, the accuracy of SKLDNE was around 8% and 21% more than LDNE and PCA respectively. The Accuracy of SKLDNE for Tn=1, 4, 5, 7, 8, 9, and 10 is 2.4%, 2.6%, 3.4%, 6%, 5.2%, and 5.7% more than LDNE respectively. To well explain the superiority of SKLDNE, it should be noted that it can preserve more effective nonlinear features and more geometrical and discriminant information so, it can tackle the small sample size, the out-of-sample, and the “overlearning of locality" problems. Hence, it can be concluded that the SKLDNE approach is a promising supervised technique with satisfactory classification performance when applied in the Finger Knuckle database. The recognition rates in comparison with the variety of dimensions for each Tn in the Finger Knuckle Print database are shown in Figure 4.20a-i.

Table 4.6. Maximum recognition accuracies (in percentage terms) of SKLDNE and other methods for the different numbers of training and testing images in the Finger Vein database and corresponding dimensions (shown in parentheses).

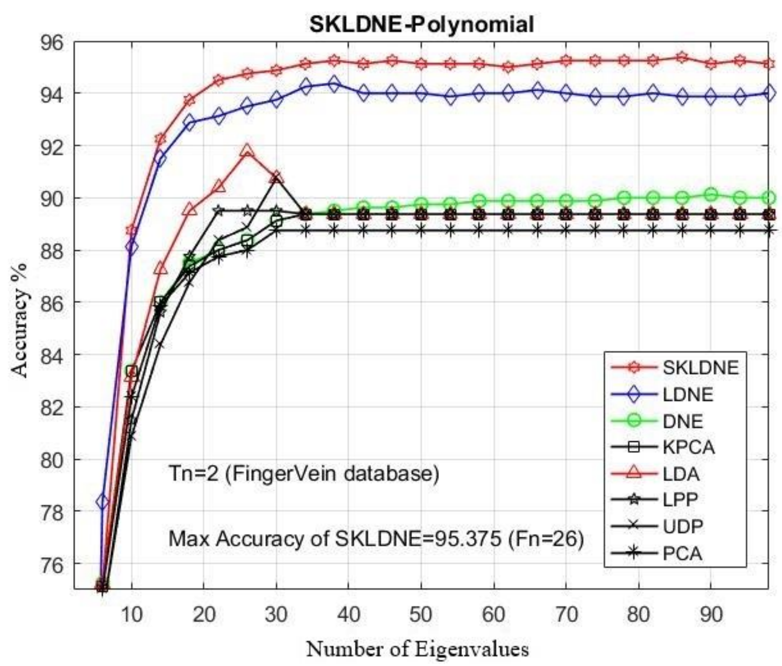
DATABASE	**FINGER VEIN**						
TN	**1**	**2**	**3**	**4**	**5**	**6**	**7**
PCA	79.66 (30)	89.37 (34)	94.85 (34)	96.3 (26)	96.41 (26)	99 (26)	99.54 (27)
KPCA	79.33 (30)	88.75 (30)	94.57 (30)	96.33 (30)	96.4 (26)	99.2 (26)	99.5 (27)
UDP	80 (30)	90.75 (30)	95.85 (30)	96.8 (34)	97 (30)	99 (34)	99.5 (30)
LPP	79.66 (30)	89.5 (22)	94.85 (34)	96.83 (26)	96.6 (22)	99 (34)	99.56 (34)
LDA	79.66 (30)	91.75 (26)	96.71 (30)	97.16 (30)	98.2 (30)	99.5 (22)	99.55 (29)
DNE	81 (90)	90.12 (90)	95.42 (62)	97 (86)	96.8 (66)	99.15 (54)	**100** (86)
LDNE	80.75 (86)	94.17 (38)	96.85 (62)	97.5 (76)	96.4 (74)	99.25 (66)	99.2 (42)
SKLDNE	**81.22** (54)	**95.38** (26)	**97.71** (26)	**98.5** (34)	**99.2** (34)	**99.75** (26)	**100** (26)

Table 4.7. Maximum recognition accuracies (in percentage terms) of SKLDNE and other methods for the different numbers of training and testing images in the Finger Knuckle database and corresponding dimensions (shown in parentheses).

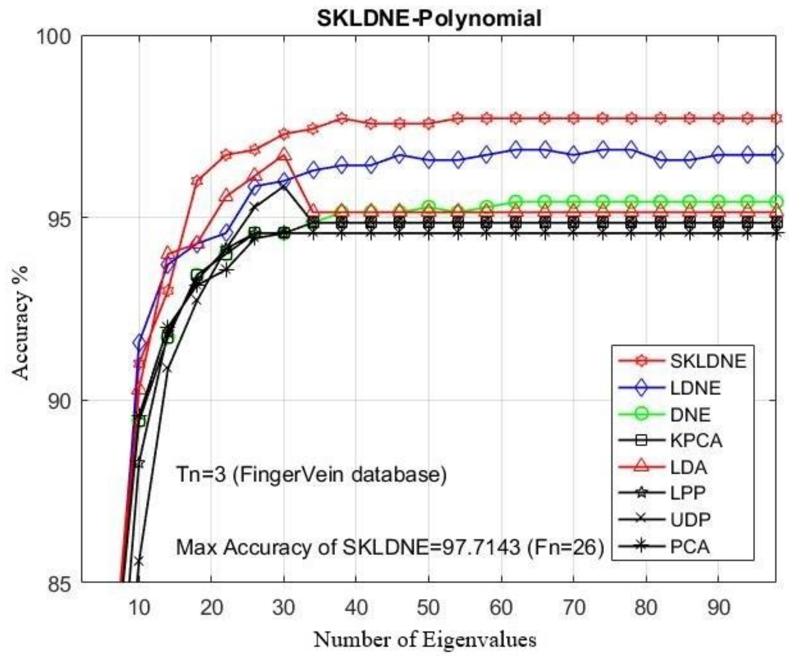
DATABASE	**FINGER KNUCKLE**								
TN	**1**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
PCA	50.18 (30)	67.5 (78)	68.28 (94)	59.66 (30)	75.2 (26)	92 (82)	94 (90)	93 (38)	97.5 (40)
KPCA	52.9 (50)	61.87 (60)	63.42 (39)	59.66 (37)	80.2 (30)	87.75 (25)	89.66 (26)	93 (27)	97.15 (20)
UDP	56.18 (62)	65.25 (98)	67.14 (90)	63.88 (90)	82.2 (98)	92 (98)	93.3 (90)	96.5 (70)	97 (35)
LPP	55 (62)	71 (94)	72 (94)	67.5 (98)	82.2 (94)	92.5 (98)	94.2 (98)	96 (74)	98 (86)
LDA	53 (62)	72.12 (94)	72.14 (98)	68.83 (98)	84.8 (90)	92.7 (74)	94.33 (90)	97 (82)	98 (26)
DNE	53.81 (98)	67.5 (88)	68.3 (94)	65.33 (94)	80.4 (94)	92 (82)	94 (90)	93 (38)	97 (20)
LDNE	53.9 (86)	76.75 (87)	78 (86)	72.70 (94)	84.6 (76)	92.75 (34)	94.66 (58)	93.3 (46)	97.2 (26)
SKLDNE	**56.36** (22)	**79.37** (86)	**81.42** (98)	**80.16** (90)	**90.6** (66)	**98** (66)	**98.66** (26)	**99** (22)	**100** (18)


SKLDNE-Polynomial
Tn=1 (FingerVein database)
Max Accuracy of SKLDNE=81.2222 (Fn=54)
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

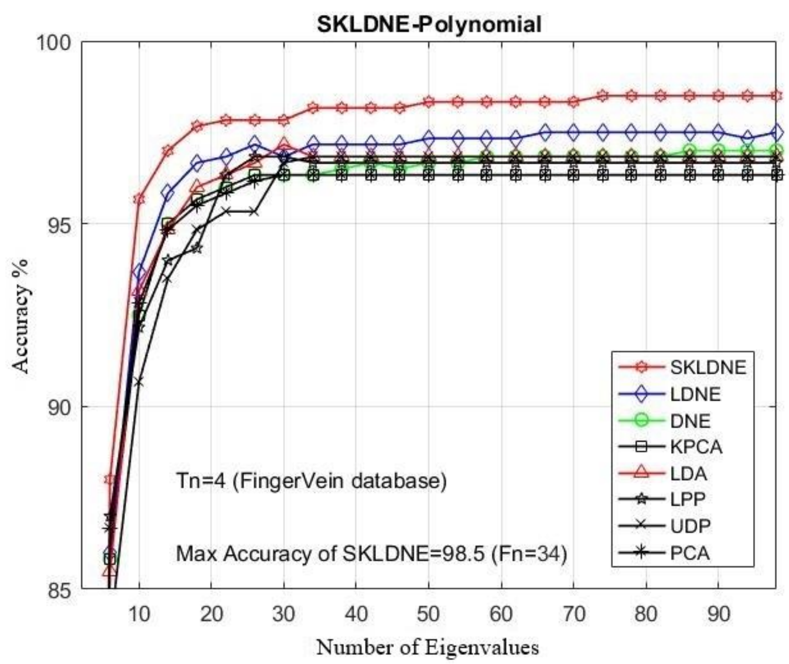
a


SKLDNE-Polynomial
Tn=2 (FingerVein database)
Max Accuracy of SKLDNE=95.375 (Fn=26)
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

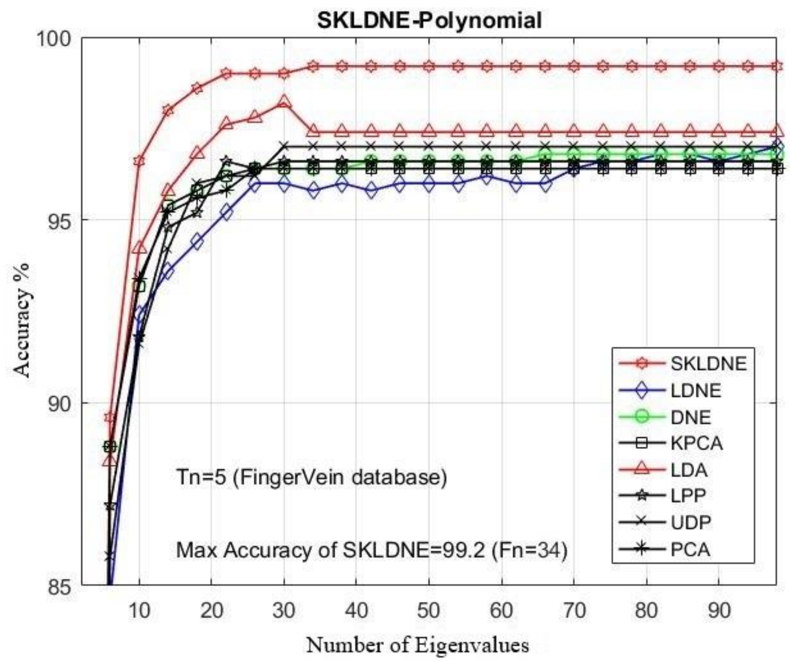
b


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=3 (FingerVein database)
Max Accuracy of SKLDNE=97.7143 (Fn=26)
100
95
90
85
10
20
30
40
50
60
70
80
90

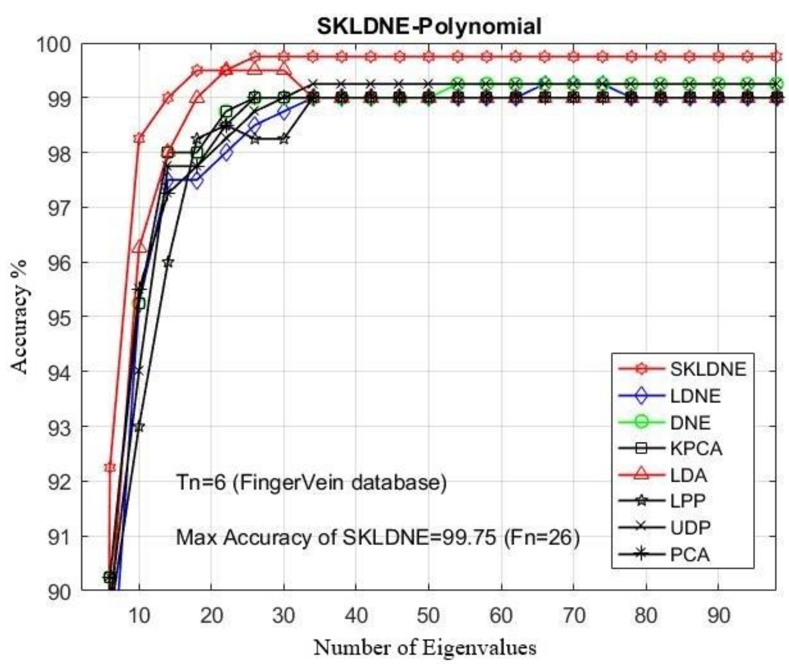
c


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=4 (FingerVein database)
Max Accuracy of SKLDNE=98.5 (Fn=34)
100
95
90
85
10
20
30
40
50
60
70
80
90

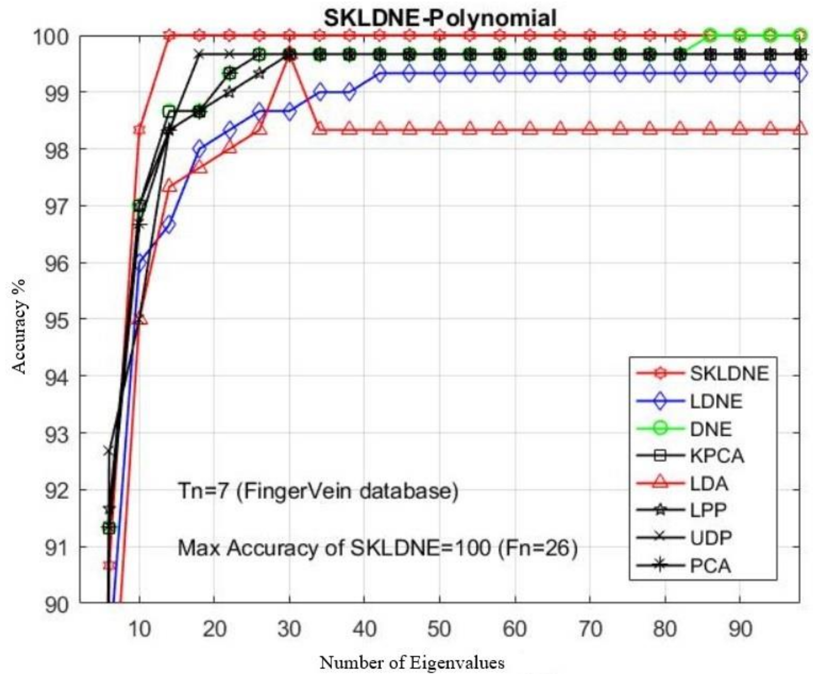
d


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
Tn=5 (FingerVein database)
Max Accuracy of SKLDNE=99.2 (Fn=34)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

e

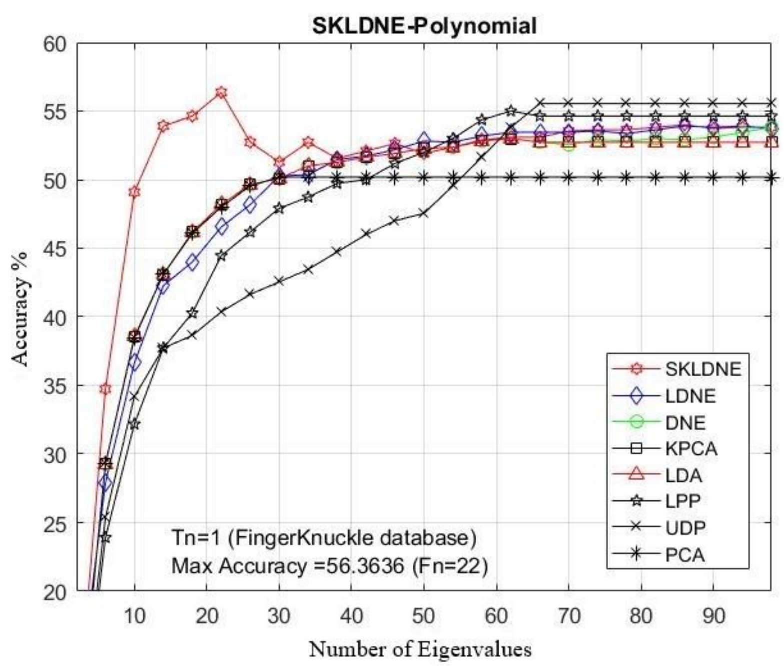

SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
Tn=6 (FingerVein database)
Max Accuracy of SKLDNE=99.75 (Fn=26)
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA

f

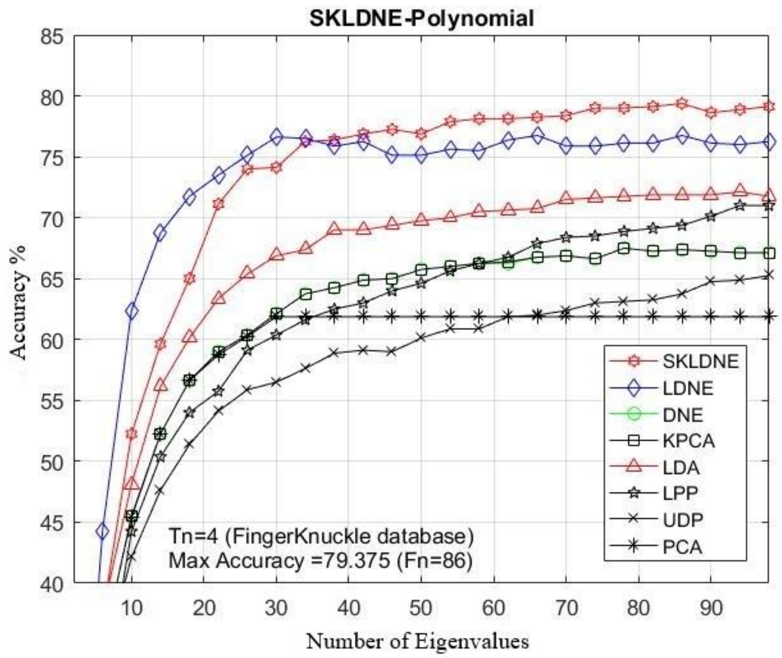


g

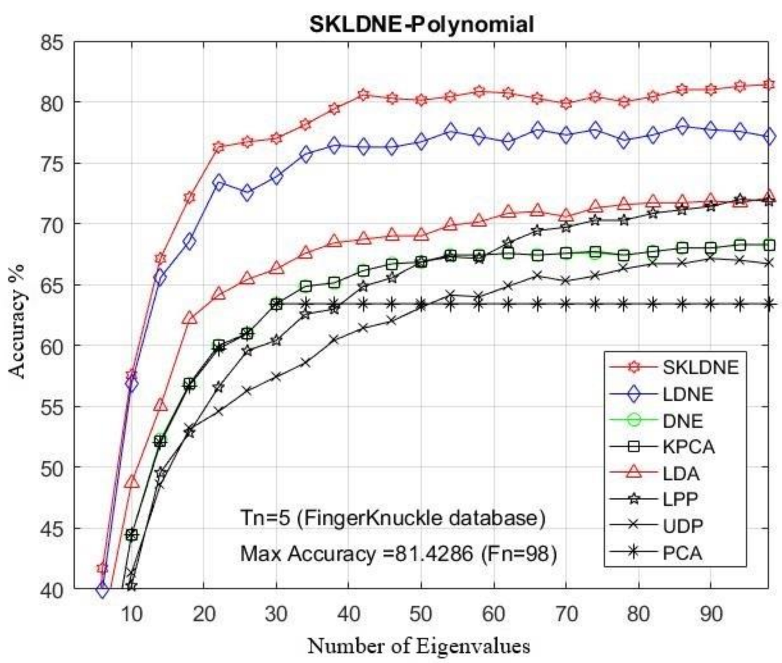
Figure 4.19. (a-g). The comparative recognition results, by changing the dimensionality of the transformation matrix for each given training number Tn in the Finger Vein Database



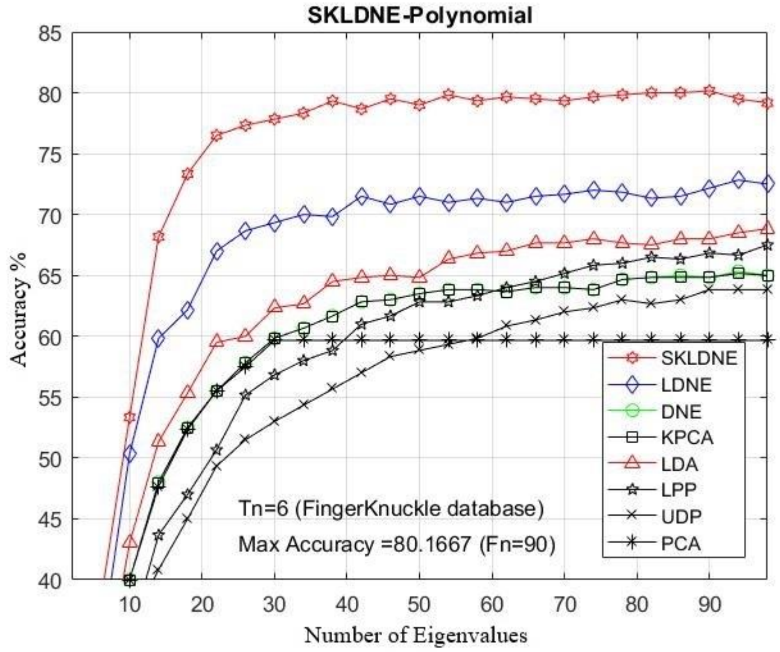
a


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=4 (FingerKnuckle database)
Max Accuracy =79.375 (Fn=86)

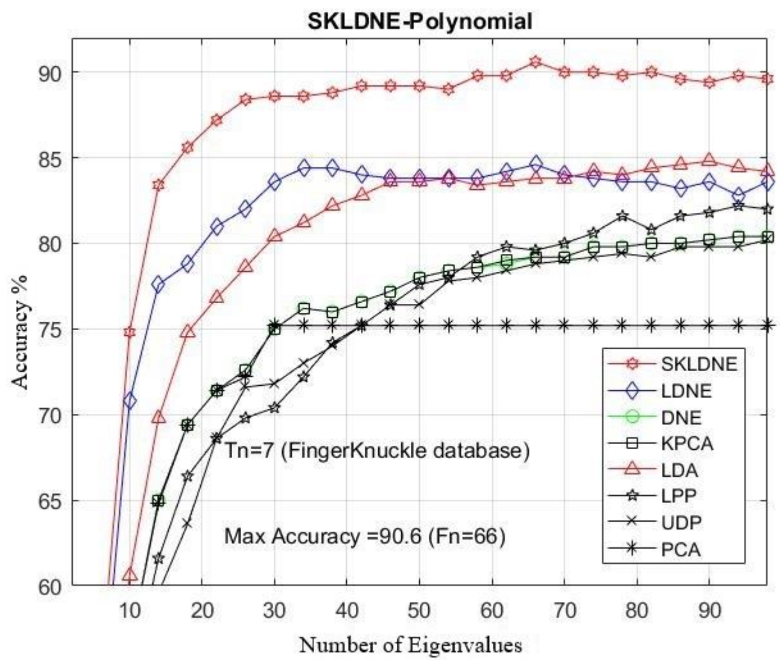
b


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=5 (FingerKnuckle database)
Max Accuracy =81.4286 (Fn=98)

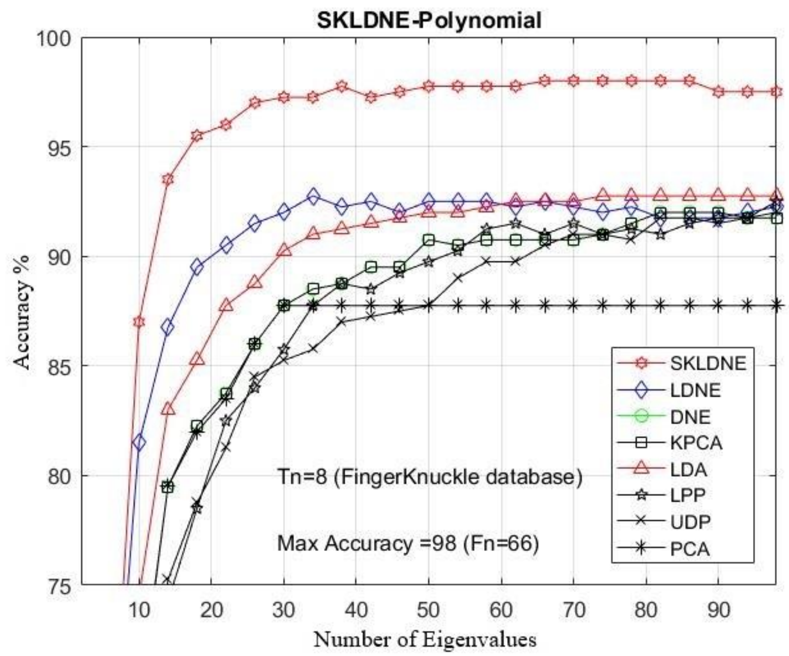
c


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=6 (FingerKnuckle database)
Max Accuracy =80.1667 (Fn=90)

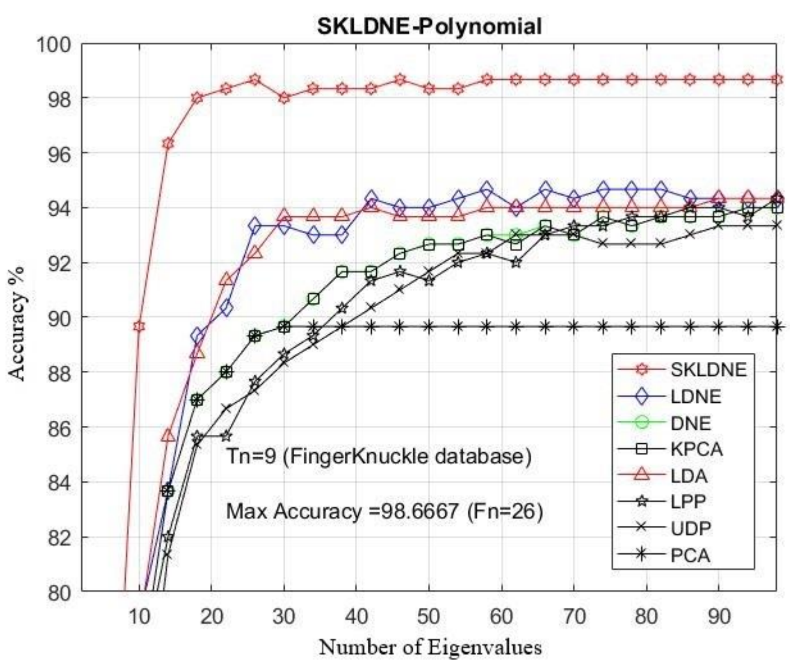
d


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=7 (FingerKnuckle database)
Max Accuracy =90.6 (Fn=66)

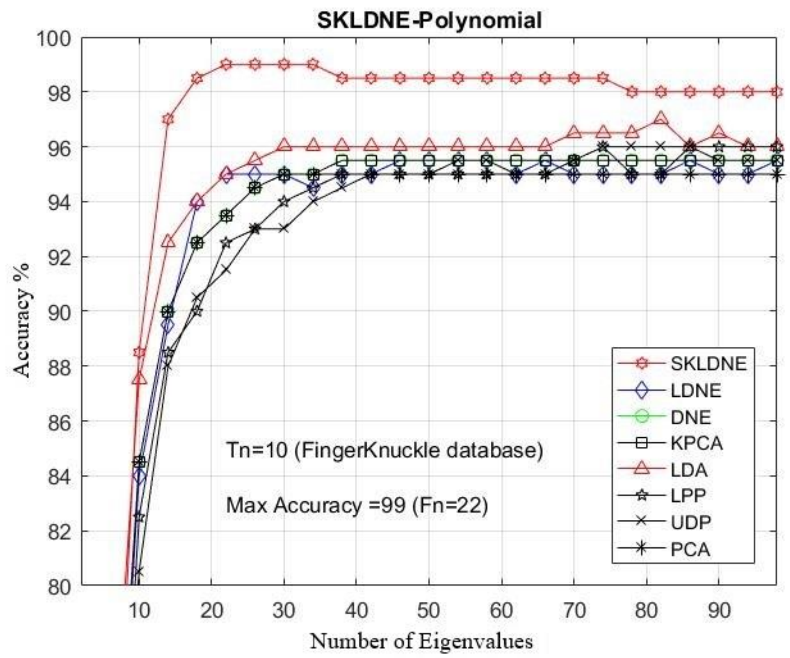
e


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=8 (FingerKnuckle database)
Max Accuracy =98 (Fn=66)

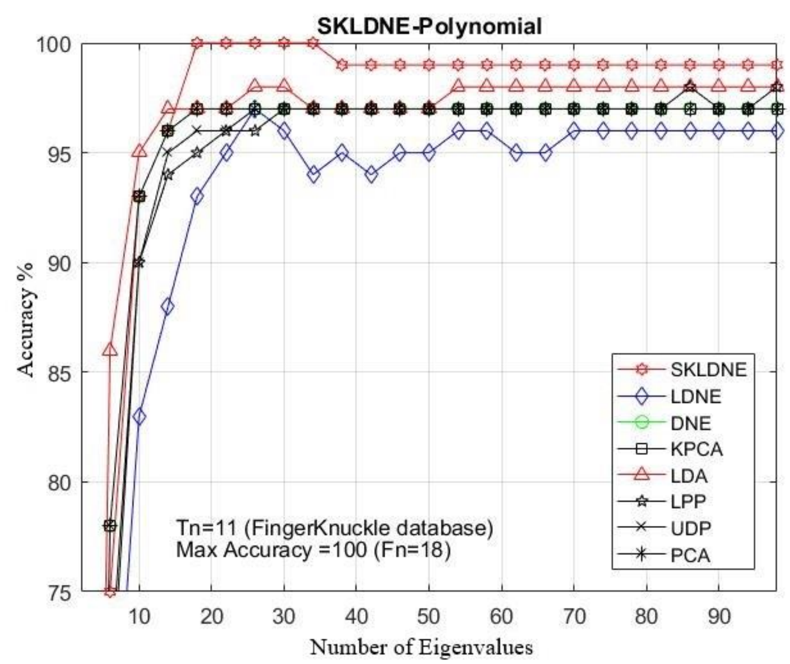
f


SKLDNE-Polynomial
Accuracy %
Number of Eigenvalues
SKLDNE
LDNE
DNE
KPCA
LDA
LPP
UDP
PCA
Tn=9 (FingerKnuckle database)
Max Accuracy =98.6667 (Fn=26)

g



h



i

Figure 4.20. (a-i). The comparative recognition results,by changing the dimensionality of the transformation matrix for each given training number Tn in the Finger Knuckle Database

4.2.6 Standard evaluation metrics used in this research

In this section, the classifier performance of our method is evaluated based on Precision, F score, and Recall (or true positive rate), considering the same implementation setting for all methods. Confidence intervals of error (CI) are also calculated and listed. In Tables 4.8, 4.9, 4.10, 4.11 and 4.12, we have summarized the results. The proposed method achieved the best precision, 95.5% for Yale, 60.8% for Head pose, 99.2% for finger vein and 79.1% for finger knuckle, among the compared methods. It was around 3.6%, 1.4%, 0.4%, and 6.6% better than LDNE on Head pose, Yale, Finger vein and Finger knuckle database respectively. This proved that the proposed method consistently outperformed others, especially in higher precision ranges. Besides, the highest f-score and recall are achieved by SKLDNE. For example on the Yale database, the f-score of SKLDNE is 96.1% which is 3%, 3.9%, 6%, 13.4%, 14%, 18.6% and 11.4% more than LDNE (93.1%), DNE (92.2%), PCA (90.2%), LDA (82.7%), LPP (82.1%), UDP (77.5%) and KPCA (84.7%) respectively. It can also be observed in tables that the values of the confidence interval of error for SKLDNE are very smaller than other techniques.

To well explain the superiority of SKLDNE, it should be noted that it can preserve more effective nonlinear features and more geometrical and discriminant information. SKLDNE can yield an optimal subspace that best finds the indispensable submanifolds-based structure. It has been designed successfully to preserve local geometric relations of the within-class samples, which are very important for image recognition. Many effective nonlinear data features may be lost during the classification process using linear techniques such as LDNE, LDA, DNE, and LPP. Therefore, applying a nonlinear method can effectively improve classification performance. This technique is a supervised learning method, as the data scholar acts as a guide to instruct the main algorithm whose conclusion should be found. SKLDNE considers class label information of neighbors in which there is a direct connection with the classification to enhance final recognition performance. Besides, it also benefits from the advantages of "locality" in LPP in which, due to the prior class-label information, geometric relations are preserved. SKLDNE can resolve the SSS problem, which that mostly faced by other aforementioned techniques as well as the "overlearning of locality" problem in the manifold learning. Due to its kernel weighting, it is very efficient in reducing the negative influence of outliers on the projection directions, which effectively handles the drawbacks of linear models and makes it more robust to outliers.

Table 4.8. Comparison of standard evaluation metrics (Precision (%), Recall (%), F-Score and Confidence Interval of error (CI) (%)) on Head pose Database

Method	Precision	Recall	F-score	CI
SKLDNE	0.608333	0.708815	0.654742	[0.360789-0.422545]
LDNE	0.594792	0.687863	0.637951	[0.374153-0.436264]
DNE	0.588542	0.541744	0.564174	[0.380329-0.442588]
PCA	0.578531	0.533155	0.554842	[0.390571-0.452483]
LDA	0.596875	0.566870	0.581486	[0.372095-0.434155]
LPP	0.589583	0.548630	0.568370	[0.379299-0.441534]
UDP	0.575000	0.530431	0.551817	[0.393728-0.456272]
KPCA	0.591667	0.556076	0.573319	[0.377240-0.439427]

Table 4.9. Comparison of standard evaluation metrics (Precision (%), Recall (%), F-Score and Confidence Interval of error (CI) (%)) on Yale Database

Method	Precision	Recall	F-Score	CI
SKLDNE	0.955556	0.966667	0.961079	[0.000000-0.104657]
LDNE	0.919178	0.943532	0.931058	[0.005649-0.162027]
DNE	0.911112	0.933333	0.922088	[0.005739-0.172038]
PCA	0.888889	0.916667	0.902564	[0.019288-0.202934]
LDA	0.830243	0.833333	0.827740	[0.065063-0.289485]
LPP	0.822222	0.820000	0.821110	[0.066070-0.273485]
UDP	0.755556	0.796349	0.775416	[0.118878-0.370011]
KPCA	0.844444	0.850000	0.847213	[0.049660-0.261451]

Table 4.10. Comparison of standard evaluation metrics (Precision (%), Recall (%), F-Score and Confidence Interval of error (CI) (%)) on Finger Vein Database

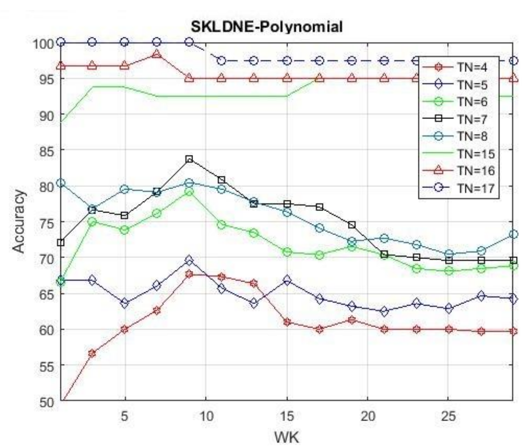
Method	Precision	Recall	F-score	CI
SKLDNE	0.992000	0.993333	0.992666	[0.000191-0.015809]
LDNE	0.988000	0.990619	0.989308	[0.002456-0.021544]
DNE	0.968000	0.974095	0.971038	[0.016573-0.047427]
PCA	0.954000	0.961734	0.954676	[0.019977-0.053428]
LDA	0.974000	0.978810	0.976399	[0.012051-0.039949]
LPP	0.966000	0.973571	0.969771	[0.018115-0.049885]
UDP	0.972000	0.978393	0.975186	[0.013540-0.042460]
KPCA	0.964000	0.970762	0.967369	[0.019671-0.052329]

Table 4.11. Comparison of standard evaluation metrics (Precision (%), Recall (%), F-Score and Confidence Interval of error (CI) (%)) on Finger Knuckle Database

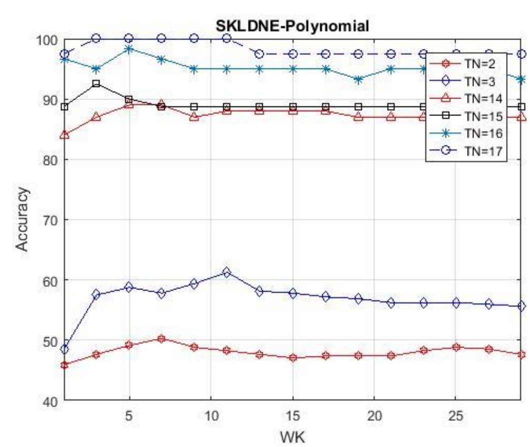
Method	Precision	Recall	F-score	CI
SKLDNE	0.791667	0.840189	0.815207	[0.175837-0.240829]
LDNE	0.725000	0.787221	0.754831	[0.239271-0.310729]
DNE	0.650000	0.673473	0.661528	[0.311834-0.388166]
PCA	0.650000	0.673851	0.661711	[0.311834-0.388166]
LDA	0.688333	0.715652	0.701727	[0.274605-0.348728]
LPP	0.675000	0.717619	0.695657	[0.287522-0.362478]
UDP	0.638333	0.700059	0.667773	[0.323220-0.400113]
KPCA	0.596667	0.602185	0.599413	[0.364080-0.442587]

4.2.7 Classification performance

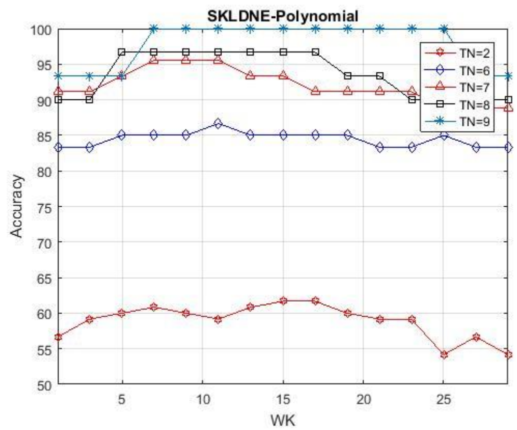
In this section, utilizing a new group of experiments, we evaluated the SKLDNE performance through changing k-neighborhood variation Wk (from 1 to 30 with a scale of 2). The number of training samples selected for each database was Tn = 4, 5, 6, 7, 8, 15, 16, 17 in Sheffield, Tn=2, 3, 14, 15, 16, 17 in Sheffield Pre-cropped, Tn = 2, 6, 7, 8, 9 in Yale, Tn = 5, 6, 7, 8, 9 in ORL, Tn = 120, 130, 140, 150, 160 in Head Pose, Tn = 5, 6, 7, 8, 9 in Finger Vein, and Tn = 7, 8, 9, 10, 11 in the Finger Knuckle databases. It should be noted that the rest of each class in each dataset was used for testing. The training number samples were randomly selected. The maximum recognition rates of SKLDNE in comparison with Wk for the different numbers of training samples are indicated in Figure 4.21a-g.



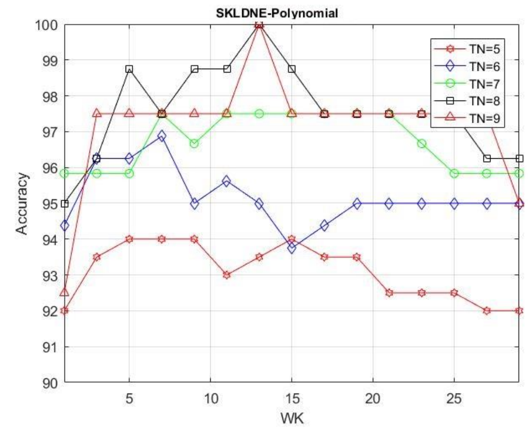
a.(Sheffield)



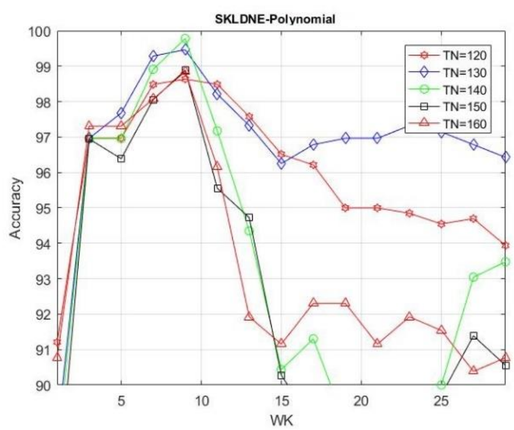
b.(Sheffield-Pre-cropped)



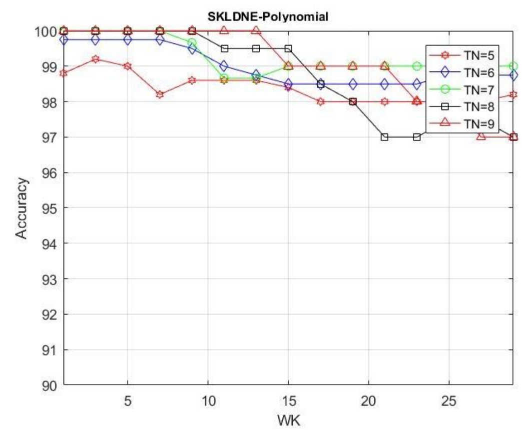
c.(Yale)



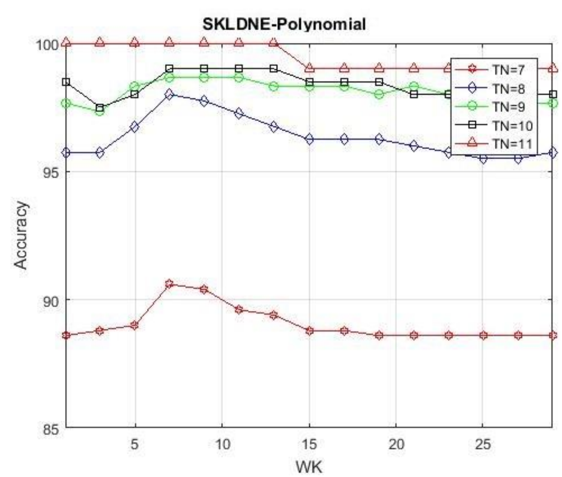
d.(ORL)



e.(Head Pose)



f.(Finger Vein)



g.(Finger Knuckle)

Figure 4.21. (a-g). Maximum recognition rate of SKLDNE versus Wk for the different numbers of training samples on Sheffield, Yale, ORL, Head Pose, Finger Vein and Finger Knuckle Databases.

It can be seen that the classification performance became better as the number of training samples increased. The classification performance of SKLDNE improved first with an increase of Wk until almost Wk = 9 in the Sheffield and Head Pose database and then it decreased dramatically. In the Yale, ORL, Finger Vein, and Finger Knuckle data sets, it can easily be observed that the recognition performance of SKLDNE enhanced rapidly when Wk varied from 1 to 7 in Yale, 1 to 10 in ORL, 1 to 9 in Finger Vein, from 1 to 6 in Finger Knuckle, and then it decreased when Wk became larger, since large values of the k-neighborhood variable Wk have an effect on creating the adjacent weight matrix. It has already been proved that the k-neighborhood selected for data points might contain more outliers belonging to other classes at a large number of Wk when a dataset includes many classes with a small number of samples for each class. Thus, the constructed adjacent weight matrix does not have sufficient discrimination for image recognition. These numbers of Wk with the best recognition performances will be chosen further on to be used in SKLDNE.

4.2.8 Computational cost

The experiment of analyzing computational cost was carried out on an Intel (R) Core i5-4200U CPU, 2.3 GHz, 10 GB RAM machine using MATLAB (R2016, Natick, MA, USA). The computational costs of the different classification methods using the Yale database are listed in Table 4.8.

Table 4.12. The computational costs (Time(s)) of the different classification methods using the different databases.

Method	SKLDNE	LDNE	DNE	KPCA	LDA	LPP	UDP	PCA
Yale	0.35	0.4	0.37	0.012	0.04	0.36	0.06	0.02
UIMST	0.39	0.43	0.87	0.021	0.034	0.056	0.054	0.04
Sheffield	0.65	0.77	0.82	0.25	0.16	0.25	0.48	0.035
ORL	0.44	0.5	0.51	0.22	0.029	0.10	0.10	0.017
Head pose	77.33	83.23	64.71	0.61	0.57	6.85	6.86	0.59
Finger vein	1.61	2.45	1.73	0.1	0.14	0.62	0.63	0.16
Finger Knuckle	5.75	20.86	20.98	0.21	0.42	2.52	2.40	0.22

From the results shown in Table 4.8, it can be observed that the proposed SKLDNE method was faster than its main competitors, such as LDNE, DNE, and LPP. The processing times of PCA, KPCA, UDP, and LDA were lower. However, the recognition rate results illustrate that these methods were much less accurate than the SKLDNE method.

4.2.9 Comparison with other previously reported results

Now, to have a more reliable comparison, we briefly compare our recognition results of the proposed method with previously published works, including a deep learning method named Deep Belief Networks (DBNs) [92, 93] with a traditional multilayer perceptron model (MLP) in the used facial expression databases, i.e., the JAFFE Database[94]. As a deep learning method, DBNs have an unsupervised feature learning ability. The JAFFE database includes 10 individuals (Japanese women) with 7 different expressions and has around 3 or 4 images for each expression. There are 213 images in total in this database. Each image has a resolution pixel of 256 × 256. In detail, we divided all image samples into 10 parts, 90% of which were applied to training, and the remaining were applied to testing. Table 8 illustrates the recognition performance comparisons in the JAFFE database when dealing with three different image resolutions of 16 × 16, 32 × 32, and 64 × 64. We can see that the proposed SKLDNE method achieved the best recognition performance (100% in all cases), in comparison with the other previously reported results, which are much lower. This is attributed to the main characteristics of SKLDNE, which effectively represents more nonlinear data structures and has more locality and discrimination information preserving power. The results (Table 4.9) again show the robustness of SKLDNE for facial expression recognition.

Table 4.13. Maximum recognition accuracies (in percentage terms) of SKLDNE and other methods in the JAFFE database.

Method	16 × 16	32 × 32	64 × 64
MLP	64.76	84.76	86.19
DBNs + MLP	88.57	89.05	90.95
SKLDNE	**100**	**100**	**100**

4.3 Summary

In this chapter first MATLAB software and database properties, which are used in our thesis, are explained. Then the optimum size of images that can be selected for use in our experimental implementations is discussed. Besides, the concept of accuracy is discussed and the related formula to determine the accuracy is provided. From the wide range of experimental results of tests conducted in 6 different databases, the SKLDNE classifier outperforms other states of the art dimensionality reduction techniques like PCA, KPCA, UDP, LDA, DNE, LPP and LDNE classifier in all the different numbers of training sets and testing sets. The experimental results demonstrate that our proposed method has satisfactory classification behavior regardless of varying the training sample size and dimensions.

5 CONCLUSION

5.1 Summary

In this study, the performance of several well-known pattern recognition strategies was analyzed to clarify which techniques are best suited to be applied in face recognition. We also evaluated the weakness and robustness of each technique. As already mentioned, DNE cannot correctly preserve local information of data because it only assigns +1 to intra-class and −1 to inter-class neighbors, so it might fail to discover the most significant submanifolds for pattern classification. LPP is designed based on "locality" since it has no direct connection with classification, and it still suffers from the "over learning of locality" problem. LDNE has been proposed to overcome the problems existing in LPP and DNE; however, it does not guarantee an appropriate projection for classification purposes because many important non-linear data might be lost during its dimensionality reduction process. Besides, in some cases, LDNE cannot distinguish inter-class and intra-class neighbors well either to conduct projection for all points. This can degrade classification performance.

To address these problems, we have proposed a new supervised subspace learning algorithm named "Supervised Kernel Locality-Based Discriminant Neighborhood Embedding". Combined with nonlinear data structures, locality, and discrimination information, SKLDNE can yield an optimal subspace that best finds the indispensable submanifolds-based structure. SKLDNE has been designed successfully to preserve local geometric relations of the within-class samples, which are very important for image recognition. Many effective nonlinear data features may be lost during the classification process using linear techniques such as LDNE, LDA, DNE, and LPP. Therefore, applying a nonlinear method can effectively improve classification performance. This technique is a supervised learning method, as the data scholar acts as a guide to instruct the main algorithm whose conclusion should be found. SKLDNE considers class label information of neighbors in which there is a direct connection with the classification to enhance final recognition performance. Besides, It also benefits from the advantages of "locality" in LPP in which, due to the prior class-label information, geometric relations are preserved. SKLDNE can resolve the SSS problem, which is mostly faced by other aforementioned techniques as well as the "overlearning of locality" problem in the manifold learning. Due to its kernel weighting, it is very efficient in reducing the negative influence of outliers on the projection directions, which effectively handles the drawbacks of linear models and makes it more robust to outliers.

Six publicly available datasets, i.e., Yale face, ORL face, Sheffield, Head Pose, Finger Vein and Finger Knuckle, were used to illustrate the significance of the proposed technique. Based on the experimental results, SKLDNE outperforms and demonstrates the potential to be implemented in real-world systems compared to other advanced dimensionality reduction methods by obtaining the highest recognition rates in all experiments. Representing complex nonlinear variations makes SKLDNE more powerful and more intuitive than LDNE and other aforementioned techniques in terms of classification. It had the best performance compared to others at smaller numbers of projected dimensions in each number of training samples per data set. Moreover, when the given training sample size for each class grew larger, SKLDNE also achieved much better results than other techniques. The overlearning of the locality problem and the out-of-sample problem in manifold learning can be avoided by applying our developed classifier. Compared to the other state-of-the-art techniques, such as PCA, KPCA, LDA, LPP, UDP, DNE, and LDNE, our SKLDNE method is more robust and effective for classification, which has been illustrated by the highest recognition rates. Experimental results reveal that our method consistently outperforms its competitors as SKLDNE has reached the 100 percent of recognition rate for Tn=17 in Sheffield, 9 in Yale, 8 in ORL, 7 in Finger vein and 11in Finger Knuckle respectively, while the results are much lower for other methods. Therefore, it can be concluded that the proposed SKLDNE technique is a promising technique to be used for dimensionality reduction with a very satisfactory classification performance when dealing with high-dimensional data.

5.2 Future work

In this thesis, the performance of the proposed method in six different databases is examined and our SKLDNE outperformed other methods in all experiments with a gray level format. As a plan, we are going to develop this classifier to be directly applied to two-dimensional data to effectively reduce computational cost. We also want to evaluate the effectiveness of our method for robot vision by improving its algorithm to be able to utilize different multiple face patterns. In particular, we intend to develop a planning function for data collection and use the actual implementation of our method in a robot. Furthermore, in some cases, there is a problem to distinguish inter-class and intra-class neighbors for all the points during the projections, which can degrade the classification performance. To overcome the problem, a new weight function can be designed for constructing an adjacent weight matrix. However, this is out of the scope of this article, and we will discuss it in future work.

References

1. Yang, G., X. Xi, and Y. Yin, *Finger vein recognition based on a personalized best bit map.* Sensors, 2012. 12(2): p. 1738-1757.
2. Rosdi, B.A., C.W. Shing, and S.A. Suandi, *Finger vein recognition using local line binary pattern.* Sensors, 2011. 11(12): p. 11357-11371.
3. Sebestyen, G., *Pattern recognition by an adaptive process of sample set construction.* Information Theory, IRE Transactions on, 1962. 8(5): p. 82-91.
4. Webb, A.R., *Statistical pattern recognition*. 2003: John Wiley & Sons.
5. Turk, M.A. and A.P. Pentland. *Face recognition using eigenfaces*. in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. 1991. IEEE.
6. Liu, Q., et al. *Face recognition using kernel-based fisher discriminant analysis*. in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. 2002. IEEE.
7. Schouten, B. and B. Jacobs, *Biometrics and their use in e-passports.* Image and Vision Computing, 2009. 27(3): p. 305-312.
8. Mobarakeh, A.K., et al. *Finger Vein Recognition Using Local Mean Based K-Nearest Centroid Neighbor Classifier*. in *Advanced Materials Research*. 2013. Trans Tech Publ.
9. Heidari, M., et al., *Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm.* Physics in Medicine & Biology, 2018. 63(3): p. 035020.
10. Heidari, M., et al. *Applying a machine learning model using a locally preserving projection based feature regeneration algorithm to predict breast cancer risk*. in *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*. 2018. International Society for Optics and Photonics.
11. Damavandinejadmonfared, S., et al., *Finger vein recognition using PCA-based methods.* World Academy of Science, Engineering and Technology, 2012. 66.
12. Yu, H. and J. Yang, *A direct LDA algorithm for high-dimensional data—with application to face recognition.* Pattern recognition, 2001. 34(10): p. 2067-2070.
13. Yang, J. and J.-y. Yang, *Why can LDA be performed in PCA transformed space?* Pattern recognition, 2003. 36(2): p. 563-566.
14. Zheng, Z., et al., *Gabor feature-based face recognition using supervised locality preserving projection.* Signal Processing, 2007. 87(10): p. 2473-2483.
15. Belhumeur, P.N., J.P. Hespanha, and D.J. Kriegman, *Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*. 1997, Yale University New Haven United States.
16. Yang, M.-H., N. Ahuja, and D. Kriegman. *Face recognition using kernel eigenfaces*. in *Image processing, 2000. proceedings. 2000 international conference on*. 2000. IEEE.
17. Zargari, A., et al., *Prediction of chemotherapy response in ovarian cancer patients using a new clustered quantitative image marker.* Physics in Medicine & Biology, 2018. 63(15): p. 155020.
18. Pratihar, D.K., *Non-Linear Dimensionality Reduction Techniques*, in *Encyclopedia of Data Warehousing and Mining, Second Edition*. 2009, IGI Global. p. 1416-1424.
19. Yanagawa, Y., M. Sakuragi, and Y. Minato, *Face identification device*. 2010, Google Patents.
20. Damavandinejadmonfared, S., W.H. Al-Arashi, and S.A. Suandi. *Pose invariant face recognition for video surveillance system using kernel principle component analysis*. in

Fourth International Conference on Digital Image Processing (ICDIP 2012). 2012. International Society for Optics and Photonics.

21. Buriro, A., et al. *Hold and sign: A novel behavioral biometrics for smartphone user authentication*. in *2016 IEEE Security and Privacy Workshops (SPW)*. 2016. IEEE.
22. Hitachi. *Identification, Verification, Authentication, Authorization*. 2009; Available from: http://www.360biometrics.com/blog/difference-between-identification-authentication/.
23. Shawe-Taylor, J. and N. Cristianini, *Kernel methods for pattern analysis*. 2004: Cambridge university press.
24. Mobarakeh, A.K., et al., *Face Recognition using a Newly Developed Linear Subspace Learning Method.* Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2017. 9(1-5): p. 141-144.
25. Gou, J. and Z. Yi, *Locality-based discriminant neighborhood embedding.* The Computer Journal, 2013. 56(9): p. 1063-1082.
26. Barra, S., et al., *A hand-based biometric system in visible light for mobile environments.* Information Sciences, 2019. 479: p. 472-485.
27. Zhang, D., Z. Guo, and Y. Gong, *Multispectral biometrics systems*, in *Multispectral biometrics*. 2016, Springer. p. 23-35.
28. Peralta, D., et al., *A survey on fingerprint minutiae-based local matching for verification and identification: Taxonomy and experimental evaluation.* Information Sciences, 2015. 315: p. 67-87.
29. De Marsico, M., et al., *Insights into the results of MICHE I - Mobile Iris CHallenge Evaluation.* Pattern Recognition, 2018. 74: p. 286-304.
30. Wang, Q., et al. *Mobile Iris Recognition via Fusing Different Kinds of Features*. in *Biometric Recognition*. 2017. Cham: Springer International Publishing.
31. Xu, Z., et al., *The big data analytics and applications of the surveillance system using video structured description technology.* Cluster Computing, 2016. 19(3): p. 1283-1292.
32. Luo, Y., et al., *Anonymous subject identification and privacy information management in video surveillance.* International Journal of Information Security, 2018. 17(3): p. 261-278.
33. Neves, J., et al., *Biometric recognition in surveillance scenarios: a survey*. Artificial Intelligence Review, 2016. 46(4): p. 515-541.
34. Jain, A.K., *An introduction to biometric recognition.* 2004. 14(1): p. 4-20.
35. Elhoseny, M., et al. *Cascade multimodal biometric system using fingerprint and Iris patterns*. in *International Conference on Advanced Intelligent Systems and Informatics*. 2017. Springer.
36. Bhattacharyya, D., *Biometric Authentication: A Review.* Science and Technology, 2009. 2: p. 13-28.
37. *Available on : [https://findbiometrics.com/year-review-modalities-part-3-501180/]*. .
38. Wayman, J.L., *"Fundamentals of biometric authentication technologies.* 2001: p. 93-113.
39. Ranjan, R., et al. *An All-In-One Convolutional Neural Network for Face Analysis*. in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. 2017.
40. Malhotra, A., H. Dhaliwal, and F. Gerlach, *Fingerprint sensors and systems incorporating fingerprint sensors*. 2019, Google Patents.
41. Snyder, D., et al. *X-Vectors: Robust DNN Embeddings for Speaker Recognition*. in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018.
42. Ning, X., et al., *BULDP: Biomimetic Uncorrelated Locality Discriminant Projection for Feature Extraction in Face Recognition.* IEEE Trans. Image Processing, 2018. 27(5): p. 2575-2586.

43. Jain, A.K., A.A. Ross, and K. Nandakumar, *Introduction to biometrics*. 2011: Springer Science & Business Media.
44. Araabi, B.N. and Z. Gharibshah. *Face recognition with manifold-based kernel discriminant analysis*. in *The 2012 international joint conference on neural networks (IJCNN)*. 2012. IEEE.
45. Shuxu, L.X.a.G., *The Fourth Biometric - Vein Recognition*. 2008. p. 537-546.
46. Mobarakeh, A.K., et al. *Applying Weighted K-nearest centroid neighbor as classifier to improve the finger vein recognition performance*. in *Control System, Computing and Engineering (ICCSCE), 2012 IEEE International Conference on*. 2012. IEEE.
47. Jain, A.K., R.P. Duin, and J. Mao, *Statistical pattern recognition: A review.* IEEE Transactions on pattern analysis and machine intelligence, 2000. 22(1): p. 4-37.
48. Damavandinejadmonfared, S., et al., *Evaluate and determine the most appropriate method to identify finger vein.* Procedia Engineering, 2012. 41: p. 516-521.
49. Niu, B., *Statistical pattern recognition : locality preserving embedings and ensemble of rules*. 2008.
50. Schölkopf, B., A. Smola, and K.-R. Müller. *Kernel principal component analysis*. in *International Conference on Artificial Neural Networks*. 1997. Springer.
51. Blackburn, J. and E. Ribeiro, *Human motion recognition using isomap and dynamic time warping*, in *Human motion–understanding, modeling, capture and animation*. 2007, Springer. p. 285-298.
52. Roweis, S.T. and L.K. Saul, *Nonlinear dimensionality reduction by locally linear embedding.* science, 2000. 290(5500): p. 2323-2326.
53. Belkin, M. and P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation.* Neural computation, 2003. 15(6): p. 1373-1396.
54. Tang, L., et al. *Dimensionality reduction based on minimax risk criterion for face recognition*. in *Neural Networks (IJCNN), The 2010 International Joint Conference on*. 2010. IEEE.
55. Huang, P., et al., *Feature extraction using local structure preserving discriminant analysis.* Neurocomputing, 2014. 140: p. 104-113.
56. Bengio, Y., et al. *Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering*. in *Advances in neural information processing systems*. 2004.
57. Cheng, J. and H. Lu, *Local Structure Preserving Based Subspace Analysis Methods and Applications*, in *Subspace Methods for Pattern Recognition in Intelligent Environment*. 2014, Springer. p. 105-121.
58. Du, Y., et al. *A performance comparison of low-and high-level features learned by deep convolutional neural networks in epithelium and stroma classification*. in *Medical Imaging 2018: Digital Pathology*. 2018. International Society for Optics and Photonics.
59. Kouropteva, O., O. Okun, and M. Pietikäinen. *Supervised locally linear embedding algorithm for pattern recognition*. in *Iberian Conference on Pattern Recognition and Image Analysis*. 2003. Springer.
60. De Ridder, D., M. Loog, and M.J. Reinders. *Local fisher embedding*. in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. 2004. IEEE.
61. Vlassis, N., Y. Motomura, and B. Kröse, *Supervised dimension reduction of intrinsically low-dimensional data.* Neural Computation, 2002. 14(1): p. 191-215.
62. Schölkopf, B., A. Smola, and K.-R. Müller, *Nonlinear component analysis as a kernel eigenvalue problem.* Neural computation, 1998. 10(5): p. 1299-1319.
63. Baudat, G. and F. Anouar, *Generalized discriminant analysis using a kernel approach.* Neural computation, 2000. 12(10): p. 2385-2404.

64. Yang, J., et al., *KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition.* IEEE Transactions on pattern analysis and machine intelligence, 2005. 27(2): p. 230-244.
65. Kim, K.I., K. Jung, and H.J. Kim, *Face recognition using kernel principal component analysis.* IEEE signal processing letters, 2002. 9(2): p. 40-42.
66. Li, S.Z., *Encyclopedia of Biometrics: I-Z.* Vol. 2. 2009: Springer Science & Business Media.
67. Lu, J. and Y.-P. Tan, *Regularized locality preserving projections and its extensions for face recognition.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2010. 40(3): p. 958-963.
68. Deng, W., et al., *Comments on" Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Application to Face and Palm Biometrics".* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008. 30(8): p. 1503-1504.
69. Zhang, W., et al., *Discriminant neighborhood embedding for classification.* Pattern Recognition, 2006. 39(11): p. 2240-2243.
70. Kotsiantis, S.B., I. Zaharakis, and P. Pintelas, *Supervised machine learning: A review of classification techniques.* Emerging artificial intelligence applications in computer engineering, 2007. 160: p. 3-24.
71. Chen, H.-T., H.-W. Chang, and T.-L. Liu. *Local discriminant embedding and its variants*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.
72. You, Q., et al. *Neighborhood discriminant projection for face recognition*. in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. 2006. IEEE.
73. Lu, G.-F., Z. Lin, and Z. Jin, *Face recognition using discriminant locality preserving projections based on maximum margin criterion.* Pattern Recognition, 2010. 43(10): p. 3572-3579.
74. Kishore, K. and K.H. Kumar. *Hybrid face recognition with locally discriminating projection*. in *Signal Acquisition and Processing, 2010. ICSAP'10. International Conference on*. 2010. IEEE.
75. Wang, T., T. Xia, and X.-m. Hu, *Geometry preserving projections algorithm for predicting membrane protein types.* Journal of theoretical biology, 2010. 262(2): p. 208-213.
76. Zhang, L., L. Qiao, and S. Chen, *Graph-optimized locality preserving projections.* Pattern Recognition, 2010. 43(6): p. 1993-2002.
77. Shao, J.-D., G. Rong, and J.M. Lee, *Generalized orthogonal locality preserving projections for nonlinear fault detection and diagnosis.* Chemometrics and Intelligent Laboratory Systems, 2009. 96(1): p. 75-83.
78. Delac, K., M. Grgic, and S. Grgic, *Independent comparative study of PCA, ICA, and LDA on the FERET data set.* International Journal of Imaging Systems and Technology, 2005. 15(5): p. 252-260.
79. Saul, L.K., et al., *Spectral methods for dimensionality reduction.* Semisupervised learning, 2006: p. 293-308.
80. Smith, L.I., *A tutorial on Principal Components Analysis*. 2002. p. 12-22.
81. Jain, A., R. Bolle, and S. Pankanti, *Introduction to Biometrics.* Editors. 2002, Springer US. p. 1-41.
82. Yang, J., et al., *Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics.* IEEE transactions on pattern analysis and machine intelligence, 2007. 29(4): p. 650-664.
83. *The Yale face database, Available at: http://vision.ucsd.edu/datasets/yale_face_dataset_original/yalefaces.zip.*

84. *The ORL face database*, . Available at: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.
85. Zhang, T., et al., *Topology preserving non-negative matrix factorization for face recognition.* IEEE Transactions on Image Processing, 2008. 17(4): p. 574-584.
86. Wan, H., et al. *A new linear discriminant analysis method to address the over-reducing problem*. in *International Conference on Pattern Recognition and Machine Intelligence*. 2015. Springer.
87. *The UMIST Face Database, Available at: http://www.sheffield.ac.uk/eee/research/iel/research/face.*
88. *http://www-prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html*.
89. Mahri, N., S.A.S. Suandi, and B.A. Rosdi. *Finger vein recognition algorithm using phase only correlation*. in *Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), 2010 International Workshop on*. 2010. IEEE.
90. Zhang, T., et al., *Multimodal biometrics using geometry preserving projections.* Pattern Recognition, 2008. 41(3): p. 805-813.
91. Zhang, L. and H. Li, *Encoding local image patterns using Riesz transforms: With applications to palmprint and finger-knuckle-print recognition.* Image and Vision Computing, 2012. 30(12): p. 1043-1051.
92. Hinton, G.E., S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets.* Neural computation, 2006. 18(7): p. 1527-1554.
93. Zhao, X., X. Shi, and S. Zhang, *Facial expression recognition via deep learning*. IETE Technical Review, 2015. 32(5): p. 347-355.
94. Lyons, M., et al. *Coding facial expressions with gabor wavelets*. in *Proceedings Third IEEE international conference on automatic face and gesture recognition*. 1998. IEEE.