SCALE-SPACE AND WAVELET DECOMPOSITION BASED SCHEME FOR FACE RECOGNITION USING NEAREST LINEAR COMBINATION

by

Farhana Hoque

B.Sc., University of Dhaka, 2008
M.Sc., University of Dhaka, 2009

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
COMPUTER SCIENCE

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

February 2014

UMI Number: 1525709
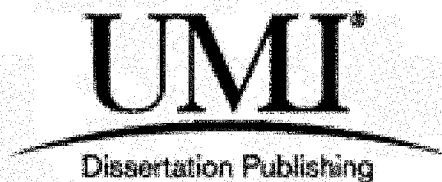
UMI

Dissertation Publishing

UMI 1525709

ProQuest

Abstract

Face recognition has attracted much attention from Artificial Intelligence researchers due to its wide acceptability in many applications. Many techniques have been suggested to develop a practical face recognition system that has the ability to handle different challenges. Illumination variation is one of the major issues that significantly affects the performances of face recognition systems. Among many illumination robust approaches, scale-space decomposition based methods play an important role in reducing the lighting effects in facial images. This research presents a face recognition approach for utilizing both the scale-space decomposition and wavelet decomposition methods.

In most cases, the existing scale-space decomposition methods perform recognition, based on only the illumination-invariant small-scale features. The proposed approach uses both large-scale and small-scale features through scale-space decomposition and wavelet decomposition. Together with the Nearest Linear Combination (NLC) approach, the proposed system is validated on different databases. The experimental results have shown that the system outperforms many recognition methods in the same category.

Contents

List of Tables

List of Figures

Acknowledgement

Firstly, I would like to express my sincere thanks to my supervisor, Dr. Liang Chen, for his continued support and guidance throughout my studies at the University of Northern British Columbia. Without his assistance I would not be able to complete my thesis. Words can not express my deepest sense of gratitude to him. I would also like to thank my supervisory committee members Dr. Desanka Polajnar and Dr. Jianbing Li for their guidance and encouragement.

I would like to take this opportunity to thank the University of Northern British Columbia for providing constant support throughout the last two years. I would like to thank Marva Byfield, Administrative Assistant of the Computer Science Department. I would also like to give thank to Dr. Jean Wang, HPC Senior Lab Instructor, for providing technical help on my research work. Also, I would like to thank all my friends, especially Dhawal, Negar, Adiba, Raman and Giridhar for their support and help.

Finally, I give special thanks to my parents for their unconditional love and encouragement. It is only because of their endless help, I have reached this point. I can not exclude thanking my husband Kaisur Rashed, for his encouragement and support throughout my studies.

Chapter 1

Introduction

1.1 Motivation

Face recognition is a major and well-known research area in computer vision. It is also closely related to pattern recognition, image processing, neural networks, computer graphics, and psychology [112]. It is a biometric method that is used to identify or verify persons from a digital image or video, based on different facial features.

Face recognition has several advantages over other biometric methods, which make it the most favourable choice for identifying people today. Face images can be easily captured from a distance by a camera without much involvement of the user, and this is useful for security and surveillance purposes [50]. The presence of noise and variation of illumination in obtained face images can be removed by using proper preprocessing and recognition algorithms.

Due to these advantages, face recognition has many important applications in biometrics, access control, security [47], video surveillance, law enforcement [112] and

1

entertainment [49]. The increasing demand of commercial and security applications has inspired the researchers to focus on developing various machine face recognition systems.

However, to design a robust and practical face recognition system, researchers have to overcome different limitations, such as variation in lighting or illumination, head pose, facial expression, accessory, aging, and so on [45]. The changes that occur in facial appearance under abnormal lighting conditions can severely degrade the performance in face recognition. To solve the problem of face recognition under varying illumination, various approaches have been proposed in the past. Among these approaches, scale-space decomposition based methods have attracted a lot of attention because of their simplicity and efficiency. Scale-space decomposition methods roughly decompose a facial image into large-scale features (like illumination, shadow cast etc.) and small-scale features (like eyes, nose and mouth). These methods use only the illumination-insensitive small-scale features for recognition and ignore the large-scale features [45]. However, according to various researches [44] [106] [105], large-scale features are also useful for good recognition performance. The goal of this research is to use both large- and small-scale features of facial images for the purpose of recognition.

1.2   Contribution

The proposed approach is developed based on utilizing facial features of different scales. The major contribution of this research is that it uses a combination of scale-space decomposition and wavelet decomposition, to extract the facial features from images. Then for the classification, the representation of an individual is generated as a linear combination of decomposed feature vectors by using the Nearest Linear Combination (NLC) approach. Results obtained from experiments on several benchmark

2

databases have shown that the proposed approach is better than various standard face recognition methods. Also, the approach reduces the error rates of several scale-space decomposition based methods.

1.3 Overview of the Thesis

The conducted research presented in this thesis is organized into several chapters. Chapter 2 contains the basic ideas and background of the face recognition problem. It also reviews the existing approaches that are related to the proposed method. Chapter 3 provides an overview of different illumination invariant approaches. It also introduces the scale-space decomposition and wavelet decomposition methods. Chapter 4 briefly describes the design of the proposed methodology. Chapter 5 demonstrates the experimental procedure, the results, and a comparative study with other methods to evaluate the effectiveness of the proposed approach. Finally, Chapter 6 concludes the thesis, and summarizes the proposed approach and provides suggestions for exploration in the future.

Chapter 2

Literature Review

Many researchers have tried to use knowledge of the human recognition process in order to solve the automatic face recognition problem.

Recent studies [14] [70] have found that the ability to recognize face is an exceptional feature of our brains and a special part of the brain is used to do this. However, researchers are still working to find accurate information about how the brain performs face recognition. It is useful to observe that when the special area of the brain responds to faces, many neural networks become involved in processing information from the faces. Here, this mechanism uses almost hundreds of billions of neurons. Therefore, it may be impractical to implement the same process through the machine face recognition system.

Furthermore, the human brain can easily recognize a face even if the appearance of the face is changed by illumination, expression, pose, or accessories. However, an automatic face recognition system sometimes may not be successful in this case.

Considering these facts, it is very difficult to design a face recognition system

which simulates the human face recognition process absolutely. To solve the problem, various approaches have been developed with different mathematical tools to compute the similarity between faces instead of using human-relevant features.

2.1 Face Recognition System

Face recognition systems simply take an image or video as an input, and identify the person or people from the image or video. To recognize a face, a system can be designed using these three steps [112] as shown in Figure 2.1.
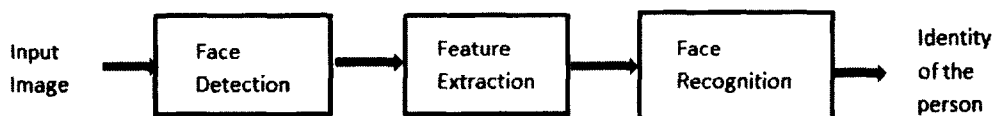
Figure 2.1: Face Recognition System.

The first stage involves locating and extracting all faces in the image using different algorithms. Performance of the face recognition partially depends on how perfectly the faces are detected and this step plays a vital role in the recognition systems. Most of the recognition approaches were evaluated by using different image databases, where faces are already detected, aligned, and cropped from the original picture. Machine recognition systems work with two set of face images: firstly, a gallery image set, where the faces are known. Secondly, a test or probe image that needs to be identified. The gallery image set contains one or more than one image for each person with variation in expression, pose, lighting, or accessory.

The second step of the recognition system is designed to obtain significant features from the gallery face image set using various feature extraction techniques. These techniques represent the gallery images and the test image through lower dimensional

feature vectors [67].

Finally, the system uses a classification algorithm that compares the gallery image representations and the test image vector (both obtained from the second step) to compute the similarity between them and the recognition result is reported. Numerous face recognition approaches have been proposed based on the combination of the last two steps. The issues of different feature extraction, as well as classification algorithms will be highlighted in this chapter.

2.2 Feature Extraction Approaches

Humans can easily recognize a person even when wearing sunglass, hat, or other accessories. Age of the person also does not affect the recognition ability of humans. All these appear as challenging factors for the machine face recognition system and these effects can be handled through the use of efficient feature extraction techniques. The goal of feature extraction is to obtain a small set of relevant features from a face, which will decrease the false positive ratio in classifiers [51]. Feature extraction is also useful for dimension reduction, which in turn reduces the computation time and memory requirement. Therefore, feature extraction is treated as a core part of the face recognition problem.

There are many algorithms for feature extraction and these approaches are divided into two major categories: geometrical local feature based methods and holistic template matching or appearance based approaches [50] [17] [41] [46]. Holistic approaches use the overall representation from the whole facial image while local feature based approaches use some prominent features of the face to perform recognition. An overview of these two approaches is given below.

6

2.2.1 Local Feature based Approach

In this approach, discrete facial features like eyes, mouth, nose, hair, and chin etc. are extracted and then geometric relationships including areas, distance, and angles among those features are measured [50] [63]. Face images are compared through these geometric feature representations using various standard recognition techniques [50] [76]. The Elastic Bunch Graph Matching (EBGM) method [102], Deformable Templates [110], and Hough Transform methods [74] are well-known local feature based approaches. Advantages of this scheme are that it is insensitive to pose, illumination, or size variation in the input images [50]. Although it provides a condensed feature representation of a face image that makes the matching faster [16], the convenience of these features depends on the efficiency of the feature detection stage [50] [30] [63]. In practise, the low accuracy and high computational complexity of the available feature detection techniques degrade the performance of the local feature based approaches [57]. Hence, nowadays holistic approaches are extensively used in solving most of the face recognition problems.

2.2.2 Holistic Template or Appearance based Approach

This approach represents a face image as a two-dimensional array of intensity variation and compares the image to several templates of other faces to classify the image [50]. In earlier solutions, face matching was performed using direct correlation based approaches. Due to very high dimensional space, these approaches were not able to overcome the limitations of face size, orientation, lighting, and noise [50]. The common solution is to apply dimension reduction tools that extract the basis information required for classification. These projection methods create a lower dimensional face space, where faces of different classes reside in separate regions in that space.

The general idea behind the dimension reduction method is stated below.

In projection based approaches, a training data G can be represented as a matrix of N rows of mn dimensional image vectors. After the dimension reduction step, a lower dimensional subspace denoted as P of size $mn \times k$ is obtained. To obtain the most significant information from the images, G is projected into the subspace as following:

$$X_p = GP \tag{2.1}$$

where X_p is a $N \times k$ matrix that is the basis of the low dimensional projection of G.

From a test image Y of size $m \times n$, we obtain, $Y_p = YP$. Finally, a comparison between Y_p and X_p is performed using some measures to find the highest similarity.

Many statistical projection based procedures are used as feature extraction methods, such as PCA [98], LDA [12], ICA [10], weighted PCA [101], two dimensional PCA [108] [64], kernel PCA [71], and so on. Some feature extraction approaches, like Discrete Cosine Transform [29] and Discrete Wavelet Transform [60] are used to downsample images into lower resolution components. Mostly these feature extraction approaches are used as a portion of face recognition systems with different classification techniques. The basic principles of Principal Component Analysis (PCA) are explained in the following section, which is used in the proposed approach to extract features, as well as to reduce dimension.

2.2.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a common statistical technique for deriving features from multi-dimensional data so that the features contain most of the variance of the data. In addition to feature extraction, PCA is also used to reduce dimension, eliminate redundancy, and have a compact representation of huge data

without much loss of information. Turk and Petland [98] first introduced the technique to solve the problem of face recognition. They proposed the Eigenfaces approach based on the fact that compressed but important patterns from a facial image could improve performance of the face recognition approaches.

The Eigenfaces approach consists of two stages: training and recognition.

Assume the training set contains M images of size $m \times n$ and each image is converted to one-dimensional column vector of size $mn \times 1$ or $N \times 1$ (let $N = mn$). These column vectors are inserted into a matrix in row wise. Direct use of this high dimensional space to the face recognition task makes it very complex. Therefore, to perform recognition, we need to derive a lower dimensional space that retains as much variance of the images as possible [5]. In this method, maximum nonconformity of each face from the mean face image is computed to form the covariance matrix of all faces. The variance of face images is then represented through eigenvectors obtained from the covariance matrix [9].

Let X be the resultant training data matrix of size $M \times N$. Firstly, we have to calculate the mean vector u along each dimension $k = 1, 2, ..., M$.

$$u[k] = 1/N \sum_{i=1}^{N} X[k, i] \tag{2.2}$$

Then the mean vector is subtracted from each column of X to create the mean-centred

matrix X_m, where

$$X_m = \begin{bmatrix} X_{11} - u_1 & X_{12} - u_1 & \cdots & X_{1N} - u_1 \\ X_{21} - u_2 & X_{22} - u_2 & \cdots & X_{2N} - u_2 \\ \vdots & \vdots & \vdots & \vdots \\ X_{M1} - u_M & X_{M2} - u_M & \cdots & X_{MN} - u_M \end{bmatrix} \qquad (2.3)$$

Next step is to find the $N \times N$ covariance matrix C of X_m.

$$C = X'_m X_m \qquad (2.4)$$

Then the matrices of eigenvectors and related eigenvalues are computed from the covariance matrix. Both of them will have dimension $N \times N$. This computation causes high memory and time consumption. An alternative solution to this problem is to check if $M < N$, then eigenvectors and eigenvalues are calculated using the following covariance matrix C of size $M \times M$, which can lessen the computational complexity.

$$C = X_m X'_m \qquad (2.5)$$

In this case, eigenvectors and eigenvalues will have dimension $M \times M$.

Computed eigenvectors indicate a set of features or eigenfaces that express the direction of maximum variance among faces and the corresponding eigenvalues denote the value of variance [98]. Eigenvectors are usually arranged in decreasing order according to eigenvalues.

To reduce the dimension, let P eigenvectors be selected from the eigenfaces, to

10

form a set V that has the P highest eigenvalues. These P eigenfaces are used to create a P-dimensional subspace or face space that describes the most variance within the set of face images [98].

Afterwards, the training data matrix X_m is projected into the face space to obtain the matrix X_p of size $M \times P$

$$X_p = X_m V \tag{2.6}$$

where X_p is the representation of training images in the face space. The advantage of this transformation is that each training image is now converted into a vector of size $1 \times P$ instead of $1 \times N$, where $P < N$.

In the recognition stage, a probe face image is taken that contains the same dimension of the training images, and converted into a matrix Y of size $1 \times N$ (where $N = m \times n$), which comprises the mean subtracted values in its columns. To get the feature vector from the probe image, Y is projected into the same subspace to get a row vector Y_p consisting of P columns.

$$Y_p = YV \tag{2.7}$$

Classification is done by using the Nearest Neighbour (NN) classifier. The probe feature vector is compared to every training projection and the identity of the training image that has the best match to the input face is chosen to be the result of recognition. Recognition is done by computing the Euclidean distance D from the probe image representation to all training image representations.

$$D = \|Y_p - X_p\| \tag{2.8}$$

There will be a total of M distances for M training images and the unknown probe image is classified to the class of the most similar training image, which has the minimum distance with the input image.

With the introduction of the Eigenfaces approach, other subspace projection techniques can be explored. In 1997, Belhumeur [12] proposed the Fisherfaces approach based on Fisher's Linear Discriminant Analysis (LDA) [31]. It creates a lower dimensional space to project the input data, where patterns belonging to different classes are well separated from each other while patterns of the same class stay relatively close. This task is accomplished by extracting a set of vectors, which best discriminates among the classes in order to minimize the within-class scatter and maximize the between-class scatter [26]. PCA is used as a preprocessing step to reduce the computational complexity. The recognition of face images is done in the same way as in the Eigenfaces approach. The limitations of LDA are that it requires more than one gallery image per person to work properly and it is also sensitive to different training data sets [50].

Independent Component Analysis (ICA) is another subspace projection method proposed by Barlett [10], which is used to find basis vectors that are spatially localized and statistically independent [26]. This approach was tested using two different representations on the FERET [79] database. Although pre-application of PCA increases the performance of ICA by reducing dimension, it still requires more time to compute ICA basis vectors compared to PCA.

Instead of having other subspace projection approaches, PCA is considered one of the most popular and efficient approach for feature extraction. This approach is embedded into the proposed research to extract features. However, in PCA representation, two images of the same person and two images of different persons are

treated almost in the same way, which characterizes the poor discriminatory power of PCA [30] [63] [86]. This problem can be solved by using Linear Discriminant Analysis (LDA) [12], but this requires a large number of images per class. The expensive process of finding eigenvectors is also a disadvantage of PCA. If the number of training images M is larger than N (the number of pixels in training images), then it involves very high computational complexity that depends on N. To reduce the complexity, eigenvectors are calculated through an alternative way using (2.5) when $M < N$. However, the complexity will be increased in polynomial order with the increase of M. The problem can be solved by combining the Discrete Wavelet Transform (DWT) with PCA to extract features, which reduces the computational load, and also improves the discriminatory power of PCA [30].

2.3 Classification Approaches

The classification approaches are applied on the feature space obtained from the feature extraction step. The goal of the classification problem is to properly separate the patterns belonging to different classes by decision boundaries. The decision boundaries are determined through the probability distribution of features of each class [87]. Classification is performed based on the similarity of the input image with the gallery images, which is measured through minimizing the error between them. Different classifiers are usually used in face recognition algorithms, such as the Nearest Neighbour (NN) [3], Nearest Linear Combination (NLC) [58], Bayesian classifier [81], Neural Networks [66], Support Vector Machines [43], and Decision tree [28]. To solve the classification problem in this proposed study, the Nearest Linear Combination (NLC) classifier is preferred to use.

2.3.1 Nearest Linear Combination

Nearest Linear Combination (NLC) is a pattern classification approach which is successfully applied in face recognition [58]. The main concept of this technique is that patterns belonging to a same subject or class are assumed to be the elements of one linear subspace [12] [11] [48]. Based on this concept, a linear combination of the gallery image vectors from an individual is developed to represent a probe or query image vector of that person [58]. The linear combination is termed as NLC since it is closest to the probe face. The NLC establishes relation between the gallery images and the probe image. For the classification, distances between the given probe image and all the linear combinations that represent classes are computed. The class of the probe image will be decided in favour of the class with minimum distance [48].

Let N be the number of classes and m be the number of prototypes for each class. Assume that $\{x_1, x_2, ..., x_m\}$ is a set of prototypical vectors of gallery images from a face class C. In the training stage, each prototypical vector in n-dimensional space can be represented as, $x_k = (x_{k1}, ..., x_{kn})$ where $x_k \in \mathbb{R}^{m \times n}$, $k = 1, 2, \ldots, m$.

A linear combination $Y_p \in \mathbb{R}^{n \times 1}$ of the prototypical vectors of the same class is formed as

$$Y_p = x_1\beta_1 + x_2\beta_2 + \cdots + x_m\beta_m = X\beta \tag{2.9}$$

where $X \in \mathbb{R}^{n \times m}$ is the class specific model or predictor of the class C that can be developed by piling each n-dimensional vectors of that class and $\beta \in \mathbb{R}^{m \times 1}$ is the vector of parameters or weights of the linear model. The expression can be written

in matrix notation.

$$
\begin{bmatrix} Y_{p1} \\ Y_{p2} \\ \vdots \\ Y_{pn} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} \tag{2.10}
$$

For a given unknown probe image vector $Y \in \mathbb{R}^{n \times 1}$, weights can be estimated using the least-square estimation [48] [88] [84].

$$
\beta = (X'X)^{-1}X'Y \tag{2.11}
$$

With the calculated weights β and predictors X, the linear combination or response vector Y_p for the class C can be determined using (2.9) and (2.11).

$$
Y_p = X(X'X)^{-1}X'Y \tag{2.12}
$$

At the recognition step, the Euclidean distance between the response vector Y_p and the probe image vector Y is measured using the following formula [48]:

$$
Dist(Y, Y_p) = \|Y - Y_p\|_2 \tag{2.13}
$$

There will be N linear combinations to represent N classes and therefore, a total of N distances need to be computed to classify the probe image Y.

$$
Dist_i(Y, Y_{pi}) = \|Y - Y_{pi}\|_2, i = 1, 2, \ldots, N. \tag{2.14}
$$

15

The probe image is classified into one of N classes, with which it has minimum distance.

Compared to the Nearest Neighbour (NN), the NLC approach is more suitable for the case if there exists more than one training image per class [58]. In NN classification, it needs to compare a probe image to every training image, where the expression and illumination variations in training images affect the result. But unfortunately in most cases, sufficient numbers of prototypes are not available for a face class that can account for all the changes in images. To overcome this limitation of the NN, the NLC approach makes a linear combination of the prototypical vectors of each class and compares the probe image to each linear combination. The changes in estimated parameters that form the linear combinations are considered to cover all the possible variations in face images [58]. In this way, it provides a large number of representative prototypical points for each class, which characterize more changes in images than the available prototypes [59].

Li [58] proposed the NLC approach that used the Eigenfaces technique [98] in the feature extraction step. The approach achieved considerably better recognition result than the standard Eigenfaces technique and the NN method when experimented on two custom databases.

Later in 1999, Li and Lu [59] proposed another classification method, called the Nearest Feature Line (NFL) approach, based on the idea of the NLC approach. In the NFL method, a linear model with a pair of images of a person is used to form a feature line (FL) that passes through the two prototype feature points of the class and expands the two points. Then the probe image is compared with each feature line for recognition purpose. Experiments on a combined dataset reported that the error rate of the NFL method was 43.7% to 65.4% lower than the standard Eigenfaces

method. Also it showed better performance compared to the Eigenfaces method on the ORL [8] dataset.

In 2010, Naseem, Tongeri and Bennamoun [48] suggested an approach to make linear combinations of downsampled gallery images. The efficiency of the proposed method was evaluated by performing experiments on five well-known databases: the AT&T [85], Georgia Tech [1], FERET [79], Extended YaleB [35] [55], and AR [61]. The experimental results showed that this approach improved the performance compared to several standard approaches, like the Eigenfaces, Fisherfaces [12], Kernel eigenfaces [108], 2DPCA [108], and ICA [108] [10].

Moon and Lee [92] proposed an illumination invariant face recognition approach based on linear combination of face prototypes. The prototypes are synthesized from the photometric stereo images of each subject by using PCA. After that, a linear combination of the exemplars is formed to represent a new image of a subject and an input image is classified by determining the minimum distance between the input image and the linear combinations of prototypes. The approach was evaluated using the MPI [15] face database and showed better performance when compared with the Eigenfaces and Bilinear analysis approach [32].

The efficiency of the NLC approach is thus evaluated using several standard databases. To classify the face images, this novel classification method is used in the proposed research.

2.4 Summary

In this chapter, the basic steps of face recognition system are introduced. It also provides a detailed description of a well-known feature extraction method and a classification technique that are utilized in the proposed approach. The main difficulty

caused by illumination variation in face images and the different approaches that can be used to handle the problem will be stated in the next chapter. A brief overview of the Discrete Wavelet Transform (DWT) is also provided in the next chapter.

Chapter 3

Illumination Invariant Approaches and Wavelet Decomposition

Illumination variations in images usually appear if the images are taken under uncontrolled lighting environments. The illumination effect can severely degrade the performance of various face recognition approaches [106]. Sometimes due to the extreme lighting situation, the whole face or part of the face can be in a shadow, which makes it very difficult to extract any feature from the face or the part of the face. It has been proven both experimentally [4] and theoretically [111] that the variation in face images due to changing illumination is more significant than the differences among different individuals under same illumination conditions. The effect of lighting variation on face images in the CMU PIE [96] database is shown in Figure 3.1, where face images of same person having illumination variations appeared very different compared to images in normal lighting condition.

In recent years many face recognition algorithms have been proposed to tackle the illumination problem. Some of the approaches are summarized in the following
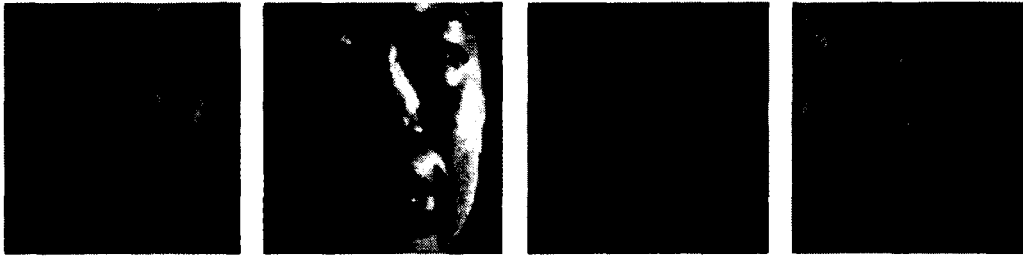
Figure 3.1: Face images under different illumination conditions in CMU PIE Database.

sections.

3.1 Statistical Approach

In several studies [89] [38], statistical methods have been used in illumination invariant face recognition approaches. Linear subspace algorithms including Linear Discriminant Analysis (LDA) [12] [95] and Bayesian methods showed better performance than the Eigenfaces [98] approach in terms of illumination variation, but still they are not successful in practical applications. To improve the performance, nonlinear methods, such as Kernel PCA [71], and Kernel LDA [82], are used. In many cases, these algorithms struggle to deal with huge differences in lighting among face images [78].

3.2 Light Modelling Approach

To develop illumination invariant face recognition approaches, researchers are now focusing on establishing different methods to balance lighting in images, such as Illumination Cone methods [13] [35], spherical harmonic based representations [11] [83], and quotient image based approaches [91] [90] [93].

3.2.1 Illumination Cone Methods

The Illumination Cone method proposed by Belhumeur and Kriegman [13] showed that a set of images under varying illumination but with a fixed pose can be used to create a convex cone to represent each individual. The illumination cone is estimated through low dimensional linear-subspaces that are built from the training images by using traditional approaches. The test image is recognized according to the identity of the illumination cone that is nearest to the image. The method is based on the assumption that the objects' surfaces have Lambertian reflectance functions, as well as convex shapes to avoid cast shadows. The objects are illuminated using arbitrary number of point light sources and the illumination cone is built by using very few images. Georghiadas and Belhumeur [35] proposed another variation of the Illumination Cone method, where an illumination cone is generated for each pose of a face class and the illumination cones for all poses of a face class are combined to represent each face class. For the recognition, the Euclidean distance between a test image and each face class representation is computed and then the closest representation is determined based on the minimum distance. The drawback of the illumination cone based approach is that to create a linear subspace from the training data, it requires multiple images of a face class with uneven lighting condition [39].

3.2.2 Quotient Image (QI) based Approaches

Quotient Image (QI) based approaches extract an illumination invariant image for each object class from a small set of images. The QI approach was first proposed by Shashua and Riklin-Raviv [91]. Initially, this method generated a linear combination of the set of face examples from each object class. Then a lighting invariant quotient image was obtained by computing the ratio of albedo between an image of arbitrary

face class and the computed linear combination. Performance of the method is better than the Eigenfaces approach, but it requires a large training data set [33]. S. Shan et al [90] presented a Quotient Image Relighting (QIR) approach using some models of the QI method. The proposed approach used some face images having abnormal lighting condition to obtain a face image that is re-lighted using normal lighting. This approach improved the performance compared with other illumination normalization approaches; however, one of the limitations was that it requires prior knowledge about the lighting system of the images.

3.2.3 Spherical Harmonic based Approaches

R. Basri and David Jacobs [11] suggested the use of spherical harmonic based representations to describe lighting. The set of images of an object that are seen in the harmonic lighting are defined as harmonic images and these harmonic basis images form a low-dimensional linear subspace model of that object. A given unknown image with arbitrary lighting conditions can be recognized by comparing the image with each linear subspace model that represents an object. This method showed some limitations in dealing with cast shadow, specularity, and pose variation in images. Since the representation of the harmonic image set is dependent on the way of lighting, accuracy can be improved by selecting appropriate lighting.

From the above analysis, it can be said that most of these approaches have some constraints. For example, they may need assumption of the lighting mode, or they require a large number of training images that are considered as major obstacles to be applied in a practical system. To deal with illumination variation in practical applications, recently researchers have paid much attention to other approaches that are based on image scale-space decomposition by using total variation image models. One of the most well-known total variation image models is $TV+L^1$ model [21].

3.3 The TV+L^1 Model for Additive Signal Decomposition

To extract useful feature components from an image by scale driven decomposition is an important task in image analysis. This decomposition can be accomplished by applying Total Variational image models. A 2-dimensional grey-scale image can be denoted by a function f, which is defined on a bounded image domain Ω of $\mathbb{R}^2$ and Ω is generally a rectangle in $\mathbb{R}^2$ [75] [103] [20] [19]. In the Total Variation (TV) regularization framework, the observed image f is considered as an additive signal that is composed of image cartoon u and texture v, where u and v are functions in Ω [23]. Cartoon u mainly contains extrinsic features or large-scale features, like background hues, shadow and important boundaries, such as sharp edges. Texture v covers small-scale patterns or intrinsic features and some noise.

The image decomposition problem is viewed as splitting an image f into u and v components. Image f can be decomposed by using the Total Variation regularization with an L^1 fidelity term, also known as TV+L^1 model. Since u is more regular than v, u can be obtained from image f by solving the following variational regularization problem [21] [24]:

$$min \int_\Omega |\nabla u| + \lambda \|f - u\|_{L^1} \tag{3.1}$$

The TV+ L^1 model involves two different terms and a scalar constant λ (lambda) that adjusts the truncation scale between u and v [23]. The first term $\int_\Omega |\nabla u|$ denotes the total variation of u over the image domain Ω and the second term states the L^1 norm $\|f - u\|$, which is the degree of closeness between f and u [21] [23]. Since $f = u + v$, with u solved from (3.1) the small-scale component v can be calculated as, $v = f - u$.

In Figure 3.2, an example of image decomposition using the TV+L^1 model is

23

shown along with corresponding cartoon (u) and texture (v). It is easily observable that with different λ values, u and v parts are showing varying large-scale and small-scale features respectively.
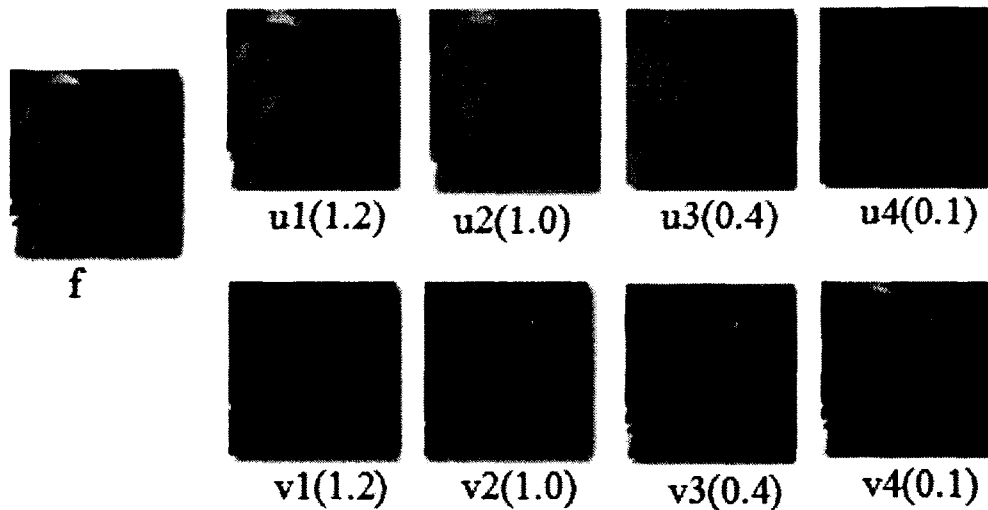
Figure 3.2: Image decomposed into cartoon (u) and texture (v) using varying value of λ.

To solve (3.1), different approaches have been proposed, such as PDE-based gradient descent technique [21], Second-order Cone Program (SOCP) [6] [36], and Parametric Maximum Flow (PMF) algorithm [37]. The PDE approach can be implemented easily and requires less memory, although it sometimes creates numerical difficulties [24]. To escape from this problem, the SOCP approach can be used that finds a solution of (3.1) using modern interior-point methods [6] [36]. It provides better accuracy but a significant amount of memory is needed for this approach [24]. Using the PMF algorithm, the user can obtain approximately the same result as the SOCP approach [106]. However, it can perform computation much faster. This influences many researchers to accept this algorithm for solving (3.1).

3.3.1   Parameter λ and Image Feature Scale

By changing the value of the parameter λ, we can determine which image features need to be kept based on their scale. The scale of a feature is defined as the ratio of the area of the feature to its perimeter [94]. For example, the scale of a circle with radius r can be written as following:

$$s = \frac{\pi r^2}{2\pi r} = \frac{r}{2} \tag{3.2}$$

D. Strong and T. Chan [94] proved that the intensity change of any image feature, caused by regulation, is proportional to truncation scale α and inversely proportional to the scale of image feature. Here truncation scale (α) determines the closeness between u and f by controlling regularization and it is inversely proportional to λ. For a smaller truncation scale or larger value of λ, regularization is minimized and u becomes close to f by retaining the sharp edges or large-scale features in u. The opposite effects will occur when a larger truncation scale or smaller value of λ is selected.

According to the definition of scale, the intensity change δ is defined as following [94]:

$$\delta = \frac{\alpha}{scale} \tag{3.3}$$

Now, the parameter λ can be related to the features in different scales that need to be preserved. For a small truncation scale (α), the change of large-scale features will be less and they will be kept in u. Thus, u will be similar to f with a larger value of λ because only a little amount of small-scale features will be removed in v. If we need to remove most of the small-scale features to v, then we need to choose a large truncation scale and a smaller λ. This effect is illustrated in Figure 3.2, where with

the decrease of the value of λ from larger to smaller, almost all the small-scale features have been extracted in v, and only some large-scale features, such as shadow, lighting effect, and background hues, are kept in u.

Thus, the parameter λ plays an important role in extracting image features in different scales through multi-scale image decomposition. Based on the size of an image, a proper λ can be selected from some recommended values or by exploration [23]. To use the TV+L^1 model for different face image sizes, the following recommended λ values can be used: 100×100 (pixels): $\lambda = 0.7$ to 0.8, 200×200: $\lambda = 0.35$ to 0.4, 400×400: $\lambda = 0.175$ to 0.2 [24]. According to [44], for handling most of the lighting variations in face images of size 64×80, λ can be selected in the range of [0, 1.2].

3.3.2   Properties of the TV+L^1 Model

The TV+L^1 model is popular for its important merit of preserving edges [94]. This property can be analysed in a simple way. Firstly, minimizing the term $\int_\Omega |\nabla u|$ attempts to reduce the variation of u over Ω [23]. Secondly, minimizing the L^1 term $\|f - u\|$, which has a tendency to keep u almost the same as f, and reserves the sharp object edges of image f in u [23]. In this way, the entire sharp edges of shadow cast on faces can be preserved in u by reducing $\int_\Omega |\nabla u| + \lambda \|f - u\|_{L^1}$ with an appropriate large λ. These large-scale features do not affect the recognition process, so after removing u, the impact of lighting can be reduced in f. Finally, the features in small-scale band (i.e. shape, contours, the size and relative position of eyes, nose, mouth, eyebrows) are obtained in v. These features are more important to distinguish different persons, because classification is made based on comparison of the positions of small-scale facial features in two face images [23] [24].

Another exceptional property of the TV+L^1 model is the ability to perform scale-

dependent and intensity-independent decomposition [24]. For a simple 2-dimensional image f, this property can be illustrated by using the approach described in [103] [23] [24].

Let $f = cB_{r(y)}(X)$ be a function that has intensity $c = 0$ everywhere except at all the points inside the disk $B_{r(y)}$ with center y and radius r, where the intensity $c > 0$. For this f, Chan and Esedoglu [21] showed that after solving (3.1), we get the following:

$$
u = \begin{cases}
0 & if \ 0 \leq \lambda < \frac{r}{2} \\
c'f \ for \ any \ c' \in [0.1] & if \ \lambda = \frac{r}{2} \\
f & if \ \lambda > \frac{r}{2}
\end{cases}
\tag{3.4}
$$

From the above analysis of the example, we can infer two important facts. First, the value of parameter λ, which controls whether u_λ will be equal to 0 or f, depends on the radius of the disk only, and intensity c does not have any impact on λ. Secondly, the signal of the disk or f is either completely preserved in u_λ or v_λ for all values of λ (except $\lambda = \frac{r}{2}$), which depicts the least signal corruption property of the TV+L^1 model [103] [23]. An example of scale-dependent image decomposition property of the TV+L^1 is presented in Figure 3.3. It is observed that signals with different scales are separated into either u or v, but not into both u and v. Based on these properties, the TV+L^1 model with an appropriate λ can reduce the lighting effects from face images and extract the illumination invariant small-scale features that can improve the performance of the face recognition algorithms. The effect of applying the TV+L^1 model on some input images with illumination variation from the Yale [107] database is shown in Figure 3.4.
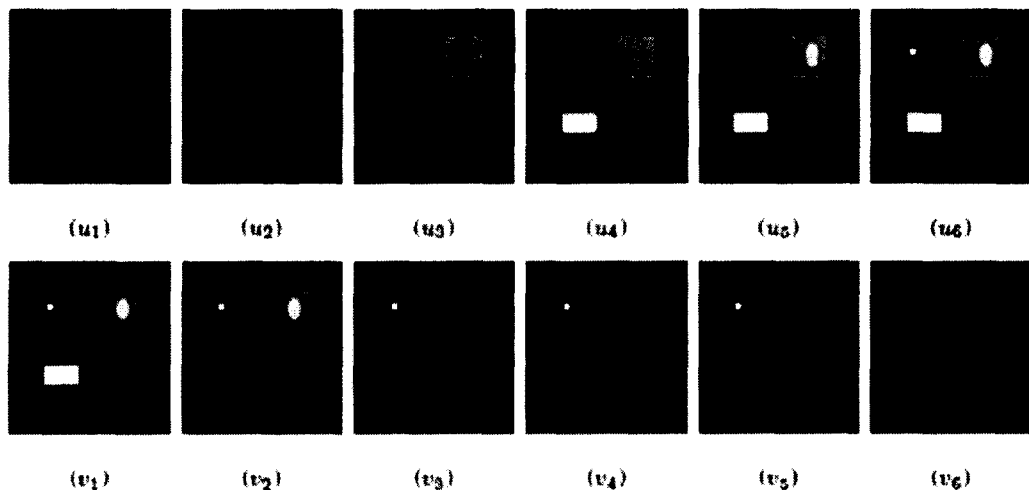
Figure 3.3: Extraction of signals with different scales by the TV+L^1 model [109].
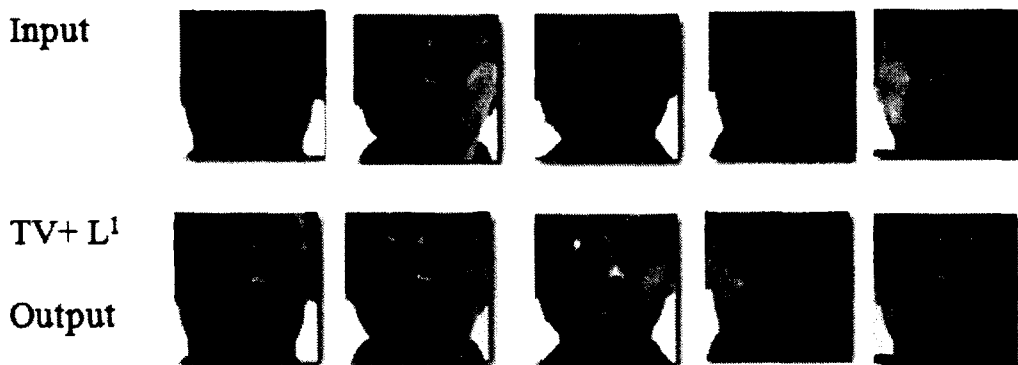
Figure 3.4: Input images and their v components as the result of the TV+L^1 model.

3.3.3 Application of the TV+L^1 Model

The TV+L^1 model has been successfully applied to many other image processing tasks based on its properties. Alliney [7] and Nicolova [72] [73] analysed the properties and applied the 1-dimensional or 2-dimensional version of the TV+L^1 model to different signal processing tasks. Nicolova [73] suggested the use of the TV+L^1 model to noise removal and outlier identification. Chan and Esedoglu [21] first utilized its cartoon-texture decomposition property. By taking the advantage of this property, W. Yin and T. Chen applied this model to correct the irregular background for cDNA

28

microarray images [104] and digital microscope images [23].

Based on the basic properties of the $TV+L^1$, various improvements have been made over it. W. Yin and T. Chen [23] [24] enhanced the $TV+L^1$ model to be used as a multiplicative model that can factorize images under multiplicative illumination effect, and can provide the lighting robust facial features. They proposed the Total Variation based Quotient Image (TVQI) model [23] and Logarithmic Total Variation (LTV) model [24] to normalize the illumination condition in face images. In the TVQI model, initially an image f is decomposed using the $TV+L^1$ model; and then, f is divided by u to get the illumination normalized component, v $(v = \frac{f}{u})$, where small-scale features are evenly emphasized. This approach showed significant performance improvement over the other normalization techniques on the Yale face database B [35]. The LTV model [24] applies the logarithmic transformation on the image f at first, to convert it from the multiplicative image model to additive, and then uses the $TV+L^1$ model on logarithm of the image. To compare the performance of the model with other algorithms, it was tested on the Yale face database B [35] and CMU PIE [96] database, and it showed better recognition rate for both of the databases.

Xie et al. [105] [106] suggested an approach using both small- and large-scale features. After decomposing a face image into large- and small-scale feature images, illumination normalization is performed on the large-scale feature image and smoothing operation is performed on the small-scale feature image to get better visual results. Finally, the normalized large-scale feature and smoothed small-scale feature images are combined to form a normalized face image, which is classified using the Nearest Neighbour (NN) classifier. The approach was evaluated on the Extended YaleB [35] and CMU PIE [96] databases, where it outperformed other existing methods.

H. Han et al. [44] proposed Separability Oriented Preprocessing (SOP) approach

that utilizes the facial features of different scales to pre-process a face image for illumination invariant face recognition. In this approach, features are extracted using the $TV+L^1$ model. The extracted features are linearly combined to infer an illumination robust image, which is used with different face recognition approaches. While forming the linear combination, the weights are determined from the training images by maximizing the Fisher separability of different subjects' faces. In the SOP formulation step, the solution of the optimization problem is obtained by using simulated annealing [54]. From the conducted experiment, SOP with the Fisherfaces [12] approach achieved 75% recognition rate on the Multi-PIE [40] database and 84.1% on the Extended YaleB [35] database, that were higher than other preprocessing methods. To evaluate SOP with different face recognition approaches other than the Fisherfaces, it was combined with the Eigenfaces [98] and LBP [97] and showed significant improvement on their performances when experimented on the FRGC Ver2.0 Exp.4 [77] dataset.

From the above mentioned applications, it is clear that the $TV+L^1$ model can be a good choice for face recognition under lighting variation due to its ability to preserve the illumination invariant facial structure from a single face image. The model has other advantages; it does not require any lighting assumption, and it needs to set only one parameter, λ. Due to these benefits, the $TV+L^1$ model is embedded into the proposed approach for scale-space decomposition of images.

3.4 Wavelet Transform

The wavelet transform [25] is a popular image decomposition approach that provides both spatial (location in time) and frequency information of an image simultaneously [34]. It is more suitable than Fourier transform methods to analyse signals that contain small discontinuities. The Fourier transform methods analyse all

frequencies in the signal through the same window or resolution while the wavelet transform uses multi-resolution technique. The multi-resolution technique examines different frequencies with different sized windows to indicate the exact position (in time) of discontinuity. The wavelet transform uses small window for high frequencies to provide good time resolution, but poor frequency resolution. On the other hand, low frequencies are viewed through large window that gives poor time, but good frequency resolution [52].

The wavelet transform provides information of the image in the form of a small number of wavelet coefficients that can be stored and analysed for further use [53]. There are two classes of wavelet transform: continuous and discrete. Continuous Wavelet Transform (CWT) is defined by the following equation [65]:

$$C(scale, position) = \int_{-\infty}^{\infty} x(t)\psi(scale, position, t)\, dt \qquad (3.5)$$

where C represents the set of wavelet coefficients that is obtained from the signal $x(t)$. Many wavelet coefficients are generated by applying wavelet function $\psi(t)$ for each time interval over the signal through changing position and scaling (stretching or compressing). Low frequencies in the signal are stretched and overall information about the signal is represented. On the other hand, compression is done on the high frequency parts of the signal that provides detail information in the signal. Since wavelet coefficients are computed at every possible scale and shifting position, CWT requires a lot of time and resources to store redundant information.

In contrast, the Discrete Wavelet Transform (DWT) is performed only on a subset of scales and positions that results in faster and easier computation of wavelet coefficients. The DWT decomposes a signal into coarse approximation (low frequency) and detail (high frequency) information by using a low pass and a high pass filter respec-

31

tively. After the filtering step, downsampling operation is performed that halves the resolution by discarding every second sample. The process of one-dimensional DWT is shown in Figure 3.5.
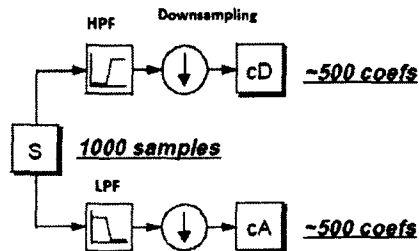
Figure 3.5: Discrete Wavelet Transform (DWT) [65].

Considering the advantages of DWT, two-dimensional DWT is generally preferred for performing wavelet decomposition of two-dimensional images. It is performed by two steps: firstly, one-dimensional DWT is applied on each row of an image to produce a row transformed image and then the same one-dimensional DWT is performed on the columns of the new image in the second step. As a result, the image is decomposed into four sub-band images corresponding to various frequency ranges. The process is illustrated in Figure 3.6.
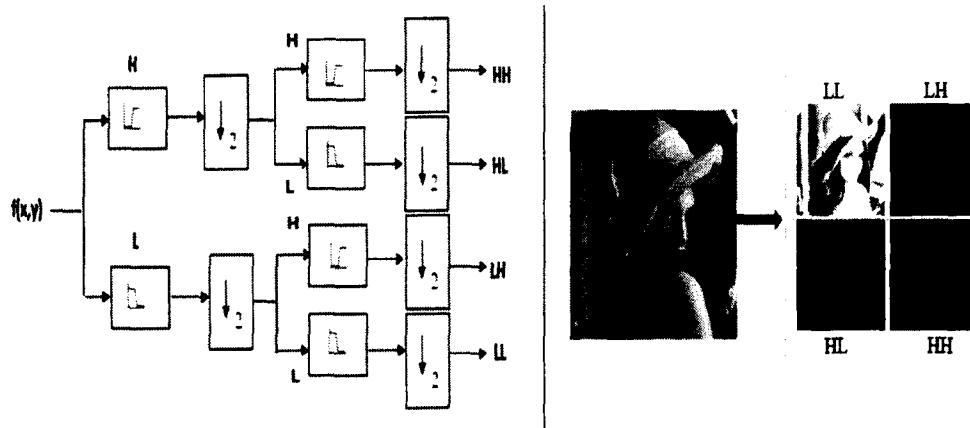
Figure 3.6: An image is decomposed into four images after one level DWT.

As shown in Figure 3.6, the sub-band images are denoted as low-low (LL), low-high

32

(LH), high-low (HL) and high-high (HH). The LL sub-band reflects global information (approximation) and represents the visual mood of the original image. The high frequency components, like the LH band denotes horizontal details, the HL band provides vertical details, and the HH band is the representative of diagonal details and noise [52] [27] [42]. Each of these sub-band images represents a smaller version of the original image. If the size of the original image is $N \times N$, then four $N/2 \times N/2$ sub-band images are obtained after the first level decomposition. The LL band image represents the identity of the original image and contains the most prominent information. Further levels of decomposition will be carried on the LL sub-band as in Figure 3.7. The maximum number of levels is chosen depending on the size of the image or type of application.

To implement the DWT, a mother wavelet function is needed as a source and all wavelet functions in the transformation can be produced from this original function through shifting and scaling. Efficiency of the wavelet transform depends on the use of appropriate wavelet function. The wavelet function is selected based on the need of application. A number of wavelet functions are currently available, such as Haar, Daubechies, Coiflet1, Symlet2, Meyer, Morlet, and Mexican Hat etc. Among these wavelet functions, Daubechies wavelet [25] is usually used in wavelet transform because of its ability to retain most of the information of the data.

In wavelet decomposition based recognition process, the selection of a sub-band image plays an important role in the recognition performance. Nastar et al. [68] [69] found that, high frequency part of the image is affected by the facial expression and small accessories while variation in pose affects the low frequency part. By removing the high frequency components, the effects of various facial expressions and occlusions in the images of the same person can be reduced [100]. Therefore, through the use of the low frequency sub-band image (LL band image) for recognition purpose, it
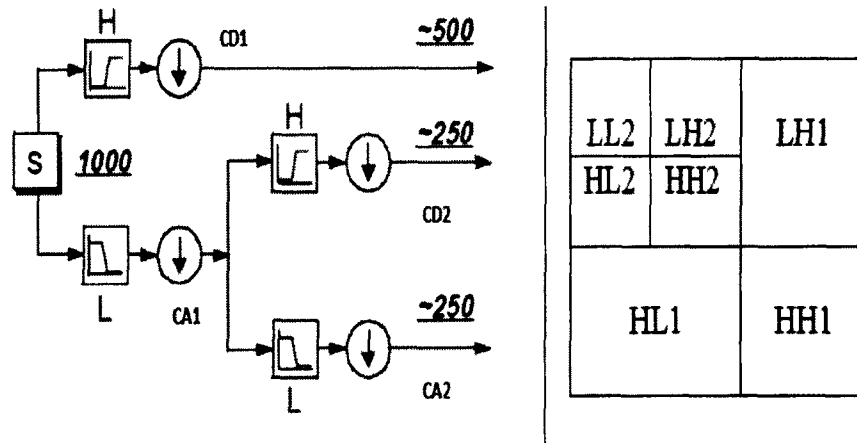
Figure 3.7: Two level wavelet decomposition [65].

can improve the matching of different images of the same person. Furthermore, the selected LL band images will have lower resolution than the original image dimension, which reduces the computational load in computing eigenvectors by PCA.

From the above analysis, it can be said that the following advantages make the DWT a widely used tool in face recognition [30] [63] [86]:

1. In DWT, by discarding the high frequency information from the low frequency data at each level, the size of the sub-band images are reduced by 2^n times of the original image after n level decomposition. Therefore, by extracting features from a compressed sub-band image, the computational work can be reduced.

2. Time and frequency information can be located more exactly by the advantage of mutiresolution decomposition, which is useful to attenuate distortion from an image.

Numerous studies on face recognition including [30] [63] [86] [52] [53] [42] [100] [62] have used the DWT to improve performance. In [100], different types of wavelet transform were applied on the ORL [8] dataset to evaluate the recognition performance

34

with different distance measures. The benefits of feature extraction through PCA from the wavelet sub-bands were investigated in [30] [86] [62]. In [63] [42], a combination of wavelet decomposition and the Eigenfaces method was used to compute features from images and the recognition task is performed by using the Support Vector Machines (SVM) classifier.

For the multilevel dimension reduction property, wavelet transform is combined with the TV+L^1 model in the proposed approach. This is done to reduce the computational load of extracting features from images.

3.5   Summary

Different algorithms for solving illumination problems are presented in this chapter. The basic concepts of the TV+L^1 model and Discrete Wavelet Transform (DWT) that play an important role in the proposed approach are also stated concisely. A brief overview of the proposed approach is illustrated in the next chapter.

Chapter 4

Proposed Approach

The proposed face recognition approach is developed based on scale-space decomposition by the TV+L^1 model and wavelet decomposition. Instead of using only small-scale features for recognition, the approach presented here takes into consideration the importance of both large-scale and small-scale features. It is designed to utilize the features at different scales that are extracted from face images using the TV+L^1 model. After that, the wavelet decomposition followed by PCA is applied on the extracted components of various scales to find the compact and important features that improve the performance of the classification step.

Some of the factors that can be considered as important to design this approach are:

The TV+L^1 model, which is used for scale-space decomposition, can be solved using the following equation:

$$min \int_{\Omega} |\nabla u| + \lambda \|f - u\|_{L^1} \qquad (4.1)$$

36

Here the observed image f is decomposed into image cartoon u and texture v. The illumination component u contains large-scale features, such as background hues and shadow, or lighting effect. The small-scale component v contains intrinsic features, like eyes, nose, and mouth. The parameter λ (lambda) adjusts the truncation scale between u and v.

Analysis of the traditional uses of the Total Variational model (see section 3.3.1) reveals that it is sometimes hard to determine the truncation scale, which is inversely proportional to λ. According to the analysis of parameter λ in [21], face images with normal lighting will need a larger truncation scale to reduce less illumination components from image f, and a smaller λ should be set for that purpose. This will remove very small amounts of large-scale features to u and the small-scale component v will be close to f. In case of abnormal lighting, more large-scale features need to be removed from f to u. Therefore, a small truncation scale is used with a larger λ. Thus, the result will not be desirable if a single λ is chosen for face images with different lighting attributes.

Another important aspect of the TV+L^1 based face recognition method (see section 3.3.3) is that it assumed that only features in small-scale component are sufficient to identify a person. Therefore, the features in large-scale component are ignored, but large-scale features can also be helpful to classify different faces [44] [106] [105]. Large-scale features are important to explain the sex information and they also represent the overall description of the individual, which is useful for feature-based recognition [112]. Moreover, it is very difficult to produce a frontal-illuminated image having a good visual quality without large-scale features [106]. On the other hand, small-scale features provide finer details that facilitate the identification process. Therefore, both small-scale and large-scale features have importance in face recognition process.

To overcome the above-mentioned constraints, the proposed approach is designed with the combination of multiscale decomposition and wavelet decomposition. In multiscale decomposition method using the TV+L^1 model, the values of λ are linearly selected within a range that will cover all the lighting variations. Then the features at different scales of face images are exploited with the help of wavelet decomposition and PCA. Finally, the use of the Nearest Linear Combination (NLC) approach for classification, expands the representational capacity of images, which is expected to reduce the margin of error.

4.1 Face Recognition System based on the Proposed Approach

The proposed approach is designed in two stages: training and recognition. In the training stage, a set of training or gallery images is used to derive training image representations that are required in the recognition step. After getting probe image representations, similarity between the probe and training image representations is computed to identify the probe image. The block diagram of the proposed approach is shown in Figure 4.1.

4.1.1 Training Stage

The training stage can be divided into 4 steps. A brief overview of each step is stated below:

(I) **Multiscale Decomposition:** The first step in the training stage is to extract facial features of different scales through multiscale decomposition. To address the issue of lighting variations in images, a gallery image of a person is decomposed using the TV+L^1 model into pairs of large-scale and small-scale components: $\{u_k, v_k\}; k =$
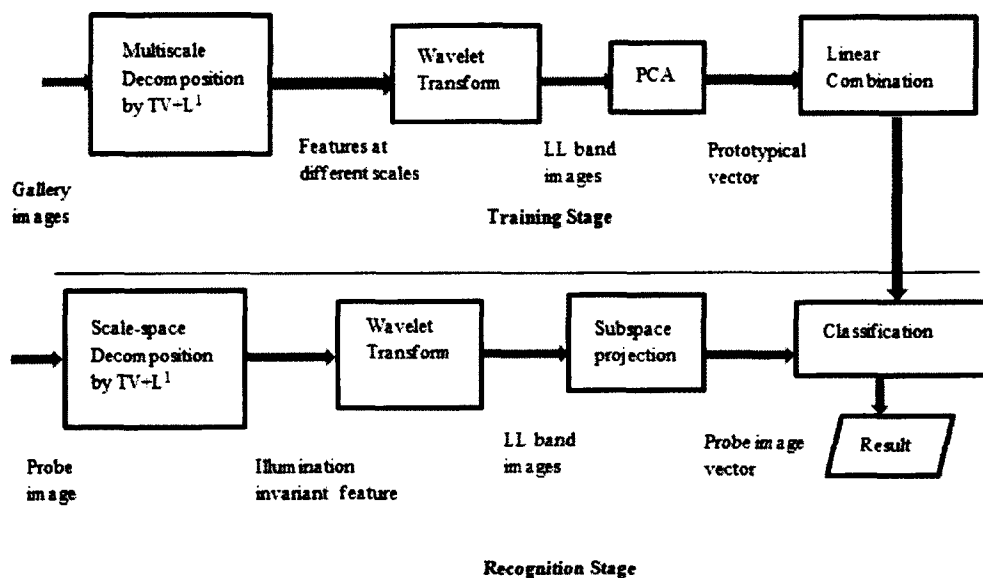
Figure 4.1: Block diagram of the proposed face recognition system.

$1, 2, ..., M$ by changing the value of λ. It is observed that λ in the range of $[0,1.2]$ will be sufficient to handle most of the lighting variations for face images of size 64×80 [45]. The range is dependent on the size of the image. To include all possible lighting variations, it is necessary to find a way to determine the values of parameter λ so that the values remain in that range.

Therefore, to decompose images cleanly into texture and cartoon components by the $TV + L^1$ model, a schedule for the parameter λ is determined as the following:

$$\lambda_j = 2j\lambda_1, \quad j = 2, 3, ..., M \tag{4.2}$$

Here the value of regularization parameter λ will increase linearly. For initial value; $\lambda_1 = 0.1$ and the number of decomposition; $M = 8$, the resulting values of λ will remain in the range of $[0.1, 1.6]$ which is not far from $[0, 1.2]$.

After getting the set of λ values using (4.2), the values are arranged in decreasing

order. As a result of multiscale decomposition using the values, eight different pairs of large-scale and small-scale features $\{u_k, v_k\}; k = 1, 2, ..., M$ can be obtained that are sufficient to cover almost all of the features of different bands (from fine to coarse). Initially, there will be only small-scale facial features in v_1. With the decrease of λ from the largest to smallest value in that range, it will continue to add more face shape information from v_1 to v_M.

Since the image is decomposed into u and v based on an additive model, the relationship among all v_j can be written as below [44]:

$$v_1 \subset v_2 \subset ...... \subset v_M \tag{4.3}$$

At first, the v_1 component will contain very little small-scale information. After that, the features that are already in v_1 will be included in v_2 along with some additional features. Following in the process, more and more features will be added from v_2 to v_M.

Facial features are extracted from these small-scale components. The features of different bands in scale-space can be obtained after decomposition as shown below [44]:

$$S_j = \begin{cases} v_1 & if \ j = 1 \\ v_j - v_{j-1} & otherwise \end{cases} \tag{4.4}$$

where S_j, $j = 1, 2, ..., M$ are the extracted facial features that include both small-scale and large-scale features. Here, the S_1 band contains the smallest-scale features, like eyes and mouth. The largest-scale features, such as shape and shading, are contained in the S_M band. In Figure 4.2, some examples of the extracted features are presented along with the small-scale components.
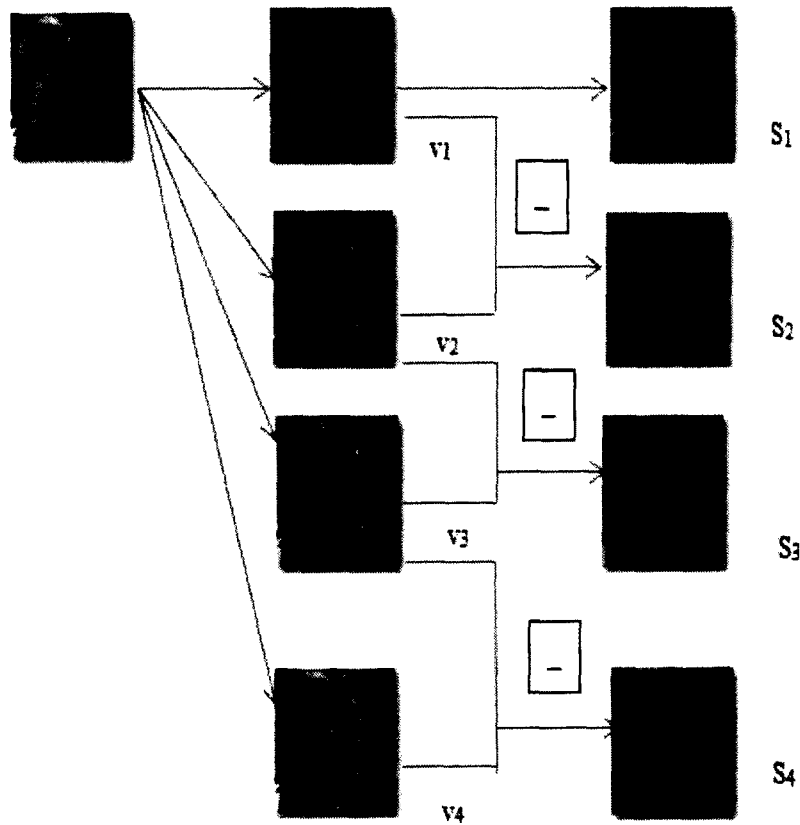
Figure 4.2: Extraction of features in different bands after scale-space decomposition of an image.

(II) Wavelet Decomposition: In the second step, Discrete Wavelet Transform (DWT) is applied on the extracted components to reduce the dimension of the resultant images. After each level of DWT, there will be four sub-band (three high frequency sub-band and one low frequency sub-band) images that represent a wide range of variance and its sum is equal to that of the whole image [30]. An outline of the wavelet decomposition process is provided in Chapter 3 (see section 3.4).

According to the analysis in [68] [69] [100], the effects of various facial expressions and occlusions in the images of the same person can be reduced by using the low frequency sub-band (LL sub-band) image. Therefore, it can improve the performance

41

of recognition. Also, the lower resolution of the LL sub-band images can induce less computational load in computing eigenvector by PCA.

(III) Principal Component Analysis (PCA): PCA (see section 2.2.2.1) is used on the LL sub-band images to achieve the most of the image information, using the minimum number of representative vectors. As described in Chapter 2, the computational complexity in obtaining eigenvectors depends on the dimension (denoted by R) of the training images, which can have a standard value of 64 × 64. This induces a high computational load in computing eigenvectors, which is usually made lower by using the number of images N when $N < R$. However, the complexity is increased non-linearly with the increase of N [30] [86]. In the proposed approach, to obtain the features in different scale bands through multiscale decomposition, an image is decomposed into M images. There will be total Q ($Q = N \times M$) images instead of N, which can induce significantly high computational weight in the feature extraction step of PCA.

By using the LL sub-band images obtained after DWT, the complexity in computing eigenvectors can be minimized significantly. For example, by using second level decomposition on 64 × 64 sized training images, LL sub-band images with resolution 21× 21 ($R = 21 \times 21 = 441$) are obtained. Suppose, there are 120 ($N = 120$) images in the training data set and each image is decomposed into 8($M = 8$) feature bands. Since now, the total number of training images, Q (120 × 8 = 960), is greater than the resolution ($R = 441$) of the images (LL sub-band); therefore, the computational complexity will depend on the reduced resolution of 441. In this way, it makes a substantial reduction in the calculation load when the number of training images is larger than resolution 441. This is considered normal for most of the practical applications.

The result of PCA is a set of eigenvectors that will form an eigen subspace and

the gallery image representations or prototypical vectors are computed by projecting the gallery or training images into the subspace.

(IV) Linear Combination of Extracted Features: Using the synthesized features of the different bands, suppose m prototypical vectors are formed for each person where, $m = k \times M$ with k is the number of images per person in the training data set and M is the number of extracted features as described in the above analysis.

For classification using the Nearest Linear Combination (NLC) method, the prototypical vectors of images from an individual are linearly combined to form a response vector Y_p, which represents the individual

$$Y_p = X_1\beta_1 + X_2\beta_2 + \cdots + X_m\beta_m \tag{4.5}$$

where X_1, X_2, ..., X_m are the prototypes of the same class and β_1, β_2, ..., β_m are the weights that are determined using the NLC approach, which is stated in Chapter 2 (see section 2.3.1).

4.1.2 Recognition Stage

The recognition task is also performed using 4 steps. Initially, the test image is decomposed using the TV+L^1 model to obtain the illumination invariant or small-scale facial structure (v component) from the image. According to the analysis in [45], use of the TV+L^1 model with a large λ value is appropriate for images with abnormal lighting condition, but it might cause negative impact if the input image is taken under normal lighting condition.[1] To avoid this situation, decomposition is performed using a smaller λ value in the range [0, 1.2], which is sufficient to remove the lighting effects

---

[1] Most of the small-scale features will be removed from the image for a larger λ value.

from the probe image.

In the next step, wavelet decomposition is conducted on the extracted component to transform the test image to the LL sub-band image that will be of the same resolution of the training images.

The resultant LL sub-band image is represented as probe image vector Y by projecting the LL images into the subspace which is obtained in the training stage.

The probe image representation and the computed linear combinations or response vectors that represent all the classes are used in the final step to measure the similarity between them. If there are N_c classes, then the probe image needs to be compared with a total of N_c linear combinations to determine whether the image is matched to any of the classes. Similarity is measured by using the following formula that computes the distance or difference between the probe image vector (Y) and all the response vectors (Y_p):

$$D_i(Y, Y_{pi}) = \|Y - Y_{pi}\|_2; \quad i = 1, 2, \ldots, N_c. \tag{4.6}$$

The probe image will be recognized to a class if it has a minimum distance with that class.

$$Decision = arg \underbrace{min}_{i} D_i(Y, Y_{pi}); \quad i = 1, 2, \ldots, N_c. \tag{4.7}$$

4.2 Pseudo-code of the Approach

The overall algorithm can be presented in the following pseudo-code:

A. Training Stage:

Selection of λ values: Form a schedule (denoted by L) of λ values using (4.2) so

44

that it can change linearly. Then arrange the λ values of L from the largest to smallest value within a specific range.

Require: A gallery or training image set G that contains $n \times n$ sized face images.

1. For each image f in G

    (i) Obtain the decomposed pairs of large-scale (u) and small-scale (v) components by using the TV+L^1 model with the λ values in L.

    (ii) Extract the features of different scales from the v components using (4.4). Let S be the set of features of various bands.

    (iii) Perform wavelet decomposition on each extracted feature component in S and obtain the LL sub-band images.

2. Apply PCA on the LL sub-band images of all gallery images and generate an eigen subspace.

3. Calculate prototypical vectors by projecting the LL sub-band images into the subspace.

4. For each individual in G
    generate a linear combination (Y_p) of the prototypical vectors that is used as a representation of that individual.

B. Recognition Stage:

Require: T is an $n \times n$ sized probe image.

1. Decompose T by the TV+L^1 model using a single λ to reduce the illumination effects from the small-scale component (v_t).

45

2. Obtain the LL sub-band image after applying wavelet decomposition on v_t.

3. Compute the probe image vector Y by projecting the LL sub-band image into the face space.

4. Compute the distance between the probe image representation and all the linear combinations according to equation (4.6).

5. Find the class C, which has a minimum distance with Y, and classify Y as the class C.

4.3 Summary

An overview of all of the steps of the proposed method is given in this chapter. Of most importance is the use of a multiscale decomposition process to extract components representing the features of multiple scales. The performance evaluation of the proposed approach using different databases is provided in the next chapter.

Chapter 5

Experimental Results

Several experiments were conducted on different datasets to demonstrate the effectiveness of the proposed approach. In the experiments, the gallery face images were decomposed into different feature bands by using the $TV+L^1$ model. To solve the $TV+L^1$ model, the Parametric Maximum Flow (PMF) algorithm [37] was adopted, and the code of PMF algorithm was embedded from CAAM's website of Rice University [99]. To determine the λ values for multiscale decomposition using the $TV+L^1$ model, the λ-schedule stated in Chapter 4 (see section 4.1.1) was used. To obtain illumination insensitive small-scale facial features from the test images, the $TV+L^1$ model was used with λ (approximately 0.1). For wavelet decomposition, Daubechies wavelet D4 [25] was used, as the experimental results in [30] [62] revealed that Daubechies wavelet D4 was better for recognition performance and computation time compared to other wavelets [27]. The implementation code of the baseline PCA algorithm was downloaded from Deng Cai's website [18].

To evaluate the proposed approach, four different datasets were selected to perform experiments: the Yale database [107], the Olivetti Research Laboratory (ORL)

database [8], the Carnegie Melon University Pose, Illumination, and Expression (CMU PIE) database [80] and the Facial Recognition Technology (FERET) database [78]. The performance of the proposed approach was compared with four different methods: the standard Eigenfaces or PCA [98] (see section 2.2.2.1), Nearest Linear Combination (NLC) [58] (see section 2.3.1) and combination of the TV+L^1 model [21] (see section 3.1) with PCA (TV+L^1&PCA), as well as NLC (TV+L^1&NLC). For the NLC approach, the images were first converted into prototypical vectors using PCA and then classified using the NLC classifier. To combine the TV+L^1 model with PCA and the NLC, the images were first decomposed with a smaller λ (approximately 0.1) to extract small-scale features. For feature extraction, PCA was performed on the small-scale feature images and then images were classified using the Nearest Neighbour (NN) classifier (for TV+L^1&PCA) and NLC classifier (for TV+L^1&NLC). To make an easier presentation, the proposed approach is called the Nearest Linear Combination of large and small-scale feature images (NLC+LS). The following sections will cover descriptions of the datasets used, the results, the analysis, as well as evaluation of the results.

5.1   Experiments on Yale, ORL and PIE Datasets

To develop a reliable face recognition system, different databases of facial images and a testing procedure to evaluate the system need to be used [78]. As a part of testing, Yale, ORL, and PIE databases were used to verify the accuracy of the proposed approach initially. A short description of these datasets is provided below.

5.1.1   Datasets: Yale, ORL and PIE

The Yale dataset from Yale University contains 165 grayscale images of 15 persons. There are a total of 11 images per person that have variations in facial expression and

configuration: centre-light, with glasses, happy, left-light, with no glasses, normal, right-light, sad, sleepy, surprised, and wink [107]. The set of face images that belong to one of the persons in the Yale dataset is shown in Figure 5.1.

The ORL dataset comes from Cambridge AT&T Laboratory, previously Olivetti Research Laboratory. This dataset contains 40 different subjects and each subject has 10 distinct images. Thus, there are total 400 grayscale images in this dataset that were taken at different times. The images have variations in lighting, facial expressions (open / closed eyes, smiling/not smiling), and facial details (glasses/ no glasses) [63]. Figure 5.2 shows face images of three persons (in three rows) in the ORL dataset.
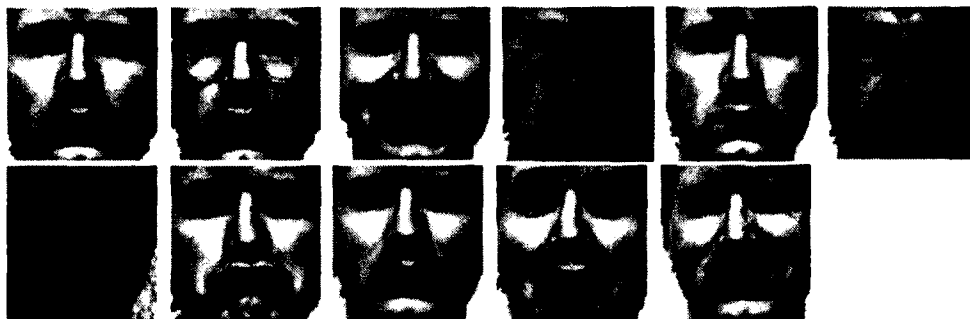
Figure 5.1: Images from Yale Database.

The CMU PIE database from Carnegie Melon University consists of 41,368 images of 68 persons. There are 13 different poses, 43 different illumination conditions, and 4 different facial expressions for each person [80]. However, only the five near frontal poses (C05, C07, C09, C27, C29) were used in the experiments, which makes a total of 11,554 face images. There are 170 images for each person, except subject 38, which has 164 images. Some of the face images of three persons in the PIE database are shown in Figure 5.3.

On Deng Cai's website [18], face images for all three datasets have been manually
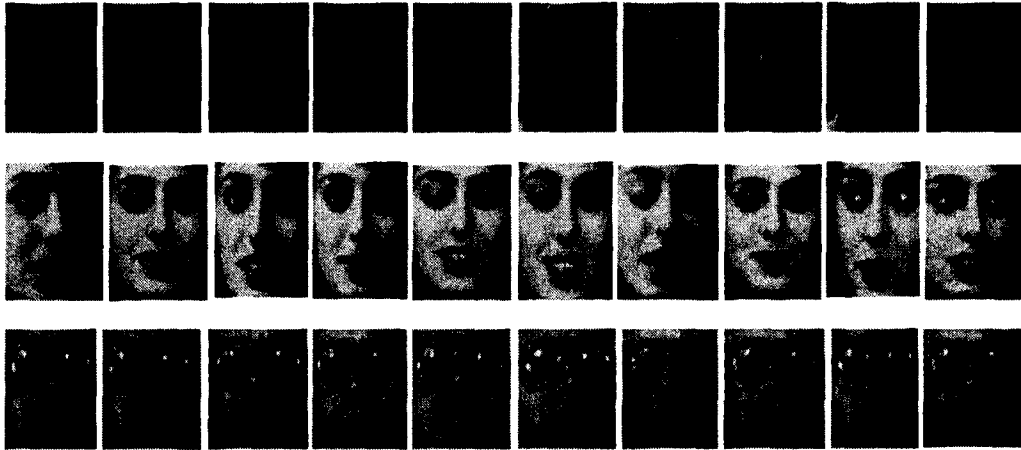
Figure 5.2: Face images of 3 persons in ORL Database.
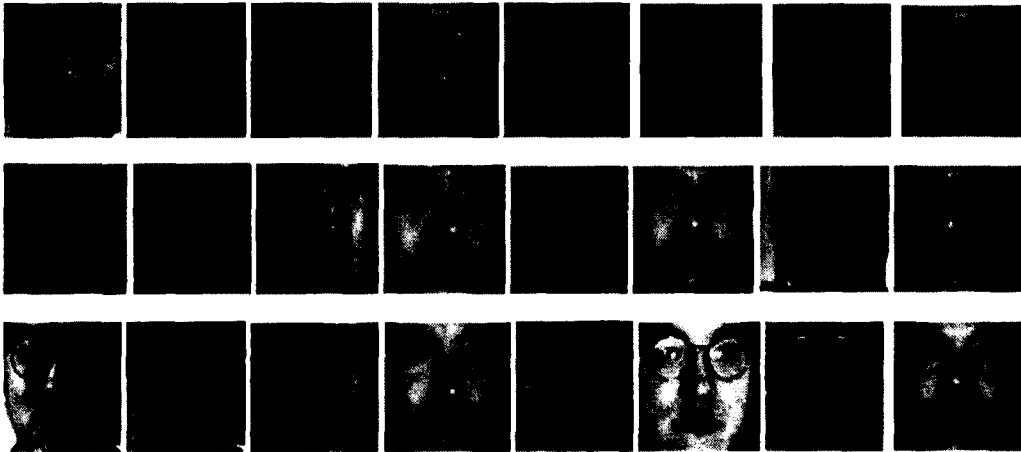
Figure 5.3: Eight different face images of 3 persons in PIE Database.

aligned by pupils and were cropped to 64×64 pixels, with 256 gray levels for each pixel [22]. The datasets were downloaded from the website and the experiments were carried out on exactly the same data.

5.1.2 Results on Yale, ORL and PIE Datasets

The experiments were conducted on each k train (k = 2, 3, 4, 5, 6, 7, 8 for Yale, k = 2, 5, 8 for ORL and k = 5, 30, 80 for PIE), where k indicates the number of images per

person in the gallery or training images. The rest were used as test images. For each k train, a total of 50 random splits are available in the original versions of these three datasets. For both Yale and ORL datasets, to test the performance of the proposed approach, the first 25 random splits were used for each k train. Average, as well as standard deviations (STD) of error rates were calculated for these two datasets. Due to the large size of PIE dataset, only the first random split was used and error rates were estimated.

For the Yale and ORL databases, features were extracted from 8 different bands through multiscale decomposition. For the PIE database, the images were decomposed into 3 feature bands. For PCA approach, the number of reduced dimension was chosen as min (M, S)-1, where M is the total number of gallery images (the number of extracted features from gallery images) and S is the number of pixels in each image [22].

The results for different training sets on the Yale and ORL databases are given in Table 5.1 and Table 5.2 respectively. Table 5.3 illustrates the results on the PIE database for 5, 30 and 80 training sets.

Table 5.1: Average error rates and STD on Yale Database with different Train/Test sizes.

Train Size	PCA	NLC	TV+L^1&PCA	TV+L^1&NLC	NLC+LS
2 Train	45.72±4.82	41.63±4.84	42.55±6.17	39.79±5.69	**35.88±5.12**
3 Train	43.13±4.44	36.33±3.72	39.60±4.72	35.57±4.02	**30.83±3.95**
4 Train	39.92±3.43	34.36±4.08	35.96±3.34	33.52±3.48	**28.42±4.09**
5 Train	35.60±3.32	30.27±4.80	30.53±3.77	28.71±4.91	**24.44±4.94**
6 Train	35.63±4.61	28.59±4.84	31.36±4.05	29.23±4.38	**25.87±4.30**
7 Train	34.47±4.36	30.33±4.88	28.47±5.05	30.07±5.02	**26.47±4.99**
8 Train	32.80±6.20	29.60±7.92	28.27±5.81	29.16±5.81	**26.58±6.33**

Table 5.2: Average error rates and STD on ORL Database with different Train/Test sizes.

Train Size	PCA	NLC	TV+L^1&PCA	TV+L^1&NLC	NLC+LS
2 Train	33.01±2.95	29.36±2.51	25.62±2.03	22.82±2.11	**22.47±2.62**
5 Train	14.34±2.32	8.40±1.83	10.78±2.31	7.43±1.84	**6.20±1.25**
8 Train	8.80±2.11	3.70±1.71	7.27±2.12	4.12±1.46	**3.10±1.39**

Table 5.3: Error rates on PIE Database with different Train/Test sizes.

	PCA	NLC	TV+L^1&PCA	TV+L^1&NLC	NLC+LS
Train Size	Error Rate	Error Rate	Error Rate	Error Rate	Error Rate
5 Train	68.50	49.92	68.52	53.51	**45.85**
30 Train	25.81	6.32	25.81	6.52	**5.76**
80 Train	7.34	**3.07**	7.34	3.39	3.33

It is easy to see that the proposed approach has shown encouraging performance on the Yale, ORL and PIE datasets. It persistently achieves better recognition accuracy than all the other methods on different train sizes of the Yale dataset (see Table 5.1). For the Yale dataset, the best result of the NLC+LS is reported on 5 train where it outperforms the Eigenfaces, NLC, TV+L^1&PCA, and TV+L^1&NLC methods by 11.16%, 5.83%, 6.09% and 4.27% respectively.

The proposed NLC+LS approach experimented on the ORL dataset (see Table 5.2) outperforms the traditional PCA and NLC approaches and it also improves the accuracy of the TV+L^1&PCA approach. It is fairly comparable to the TV+L^1&NLC method for all train sizes of the ORL dataset.

As shown in Table 5.3, the NLC+LS approach achieves impressively better performance than the Eigenfaces and TV+L^1&PCA methods on the PIE dataset. It attains low error rates of 5.76% and 3.33% for PIE 30 train and 80 train respectively, which

are favourably comparable to accuracies achieved by the NLC and TV+L^1&NLC methods.

To obtain a clear view of the experimental results, the effects of combining the large-scale and small-scale features by the NLC+LS method on the Yale and ORL datasets are depicted in Figure 5.4 and Figure 5.5 respectively.
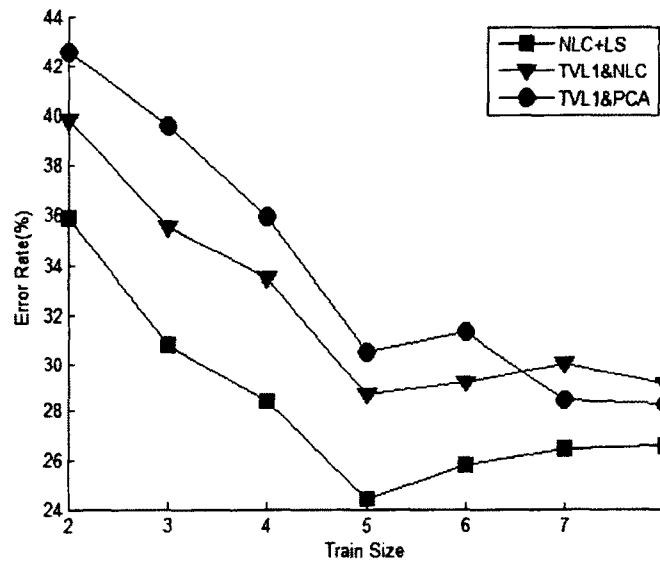
Figure 5.4: Comparison of error rates of the NLC+LS with two other methods for Yale Database.

These figures show the extent to which the performances of the standard TV+L^1 based methods, TV+L^1&PCA and TV+L^1&NLC, can be improved by the NLC–LS method. These two methods (TV+L^1&PCA and TV–L^1&NLC) only used the small-scale features for recognition and they discarded the large-scale features. The better performances of the NLC–LS approach compared to these two approaches validate the fact that only using the small-scale features of face images is not enough for recognition. Large-scale features also contain useful information and should not be discarded.
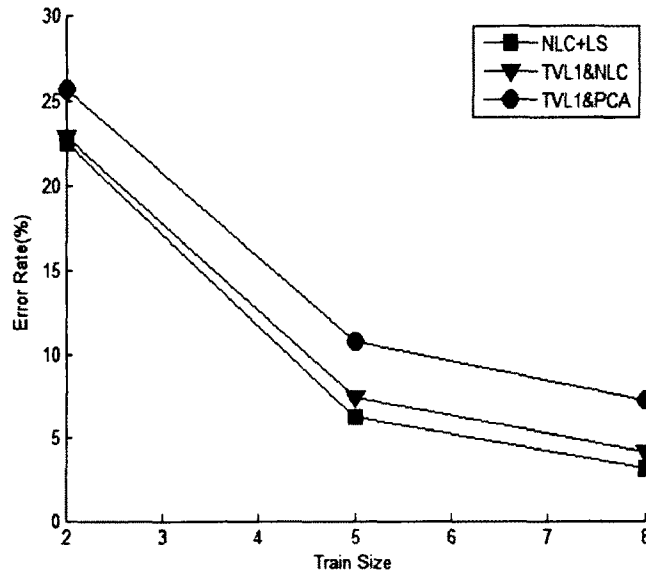
53

Figure 5.5: Comparison of error rates of the NLC+LS with two other methods for ORL Database.

From these figures, it is observed that the error rates of the methods decrease with the increase of the size of the training sets. However, one exception is noticed in Figure 5.4, where the accuracy of the $TV-L^1$&NLC method begins to drop for 7 train and 8 train. This fact can be supported by the analysis in [58], which states that the performance of the NLC approach sometimes may not improve with the increase of the number of images per person in the training dataset.

In Figure 5.4, it is easy to see that the addition of features of different bands by the NLC+LS method has significantly improved the accuracies of the $TV-L^1$&PCA and $TV+L^1$&NLC.

From Figure 5.5, it is perceived that the NLC+LS method is more effective in improving performance than the $TV-L^1$&PCA method and it also outperforms the $TV-L^1$&NLC by a small margin for the ORL dataset.

Based on the above analysis, it can be observed that the proposed approach has

better performance with small train size, which is true with conclusion of the study in [58].

5.1.3 Experiment on FERET Dataset

The facial images in the FERET database were collected by NIST from 1993 to 1997 [56]. As described in [78], the standard FERET database consists of total 14051 images from 1199 persons. It includes five subsets (one gallery set and four probe sets) that were used in this experiment for recognition test. The gallery set Fa contains 1196 images from 1196 subjects; therefore, it has one image per subject. The set of probe images is divided into four partitions (subsets) based on expression variation, lighting change, and the time difference from the gallery images. The Fb probe set consists of 1195 images of the subjects in Fa, but with different facial expressions. In the probe set Fc, there are 194 images of the same subjects of Fa, taken under different lighting environments. The Dup1 (Duplicate 1) probe set contains 722 images that are captured any-time between one minute and 1031 days after the matching gallery image was taken. The Dup2 (Duplicate 2) probe set, a subset of the Dup1, consists of 234 images, where the images are taken at least 18 months after the corresponding image in Fa. Figure 5.6 shows different images of a subject from the FERET database. Some images in the Fc probe set are shown in Figure 5.7.

The images were resized into 70×60 from the actual size of 150×130. To compute the eigenvectors from the extracted features initially, the minimum value between the total number of extracted images and the resolution of the images was chosen. Then, according to the FERET recommendation [56] [26], the subspace was formed by using the first 40% of the eigenvectors that would contain about 98% of the information stored in the eigenvectors. Since the gallery set Fa contains one image per subject, the same result was obtained for the Eigenfaces and NLC approaches. To avoid redun-
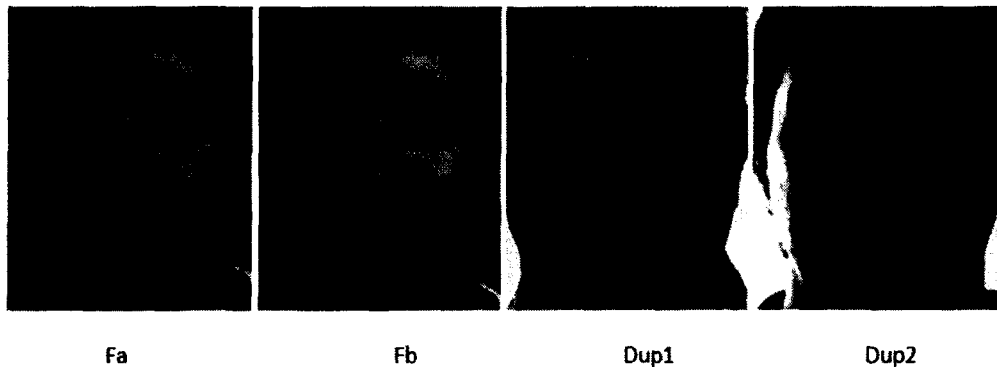
| Fa | Fb | Dup1 | Dup2 |

Figure 5.6: Gallery image and different types of probe images of a subject from FERET Database.
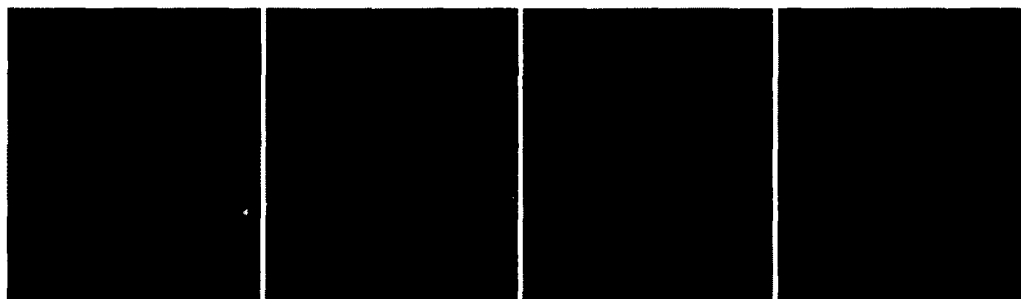
Figure 5.7: Example of probe images from Fc probe set.

dancy, the comparison result for the NLC, TV+L^1&NLC, and NLC+LS approaches on the FERET dataset is shown in Table 5.4.

From the results in Table 5.4, it is observed that the NLC–LS copes well with the problem of illumination variations by achieving comparatively better result for the Fc probe set. This is the toughest probe set, as it contains images of the subjects under considerably different lighting conditions [56]. In case of the Fc probe set, the NLC+LS outperforms the NLC approach and TV+L^1&NLC by approximately 7.7% and 8.8% respectively, which states the better tolerance of the NLC+LS for varying illumination compared to the NLC and TV+L^1&NLC. However, it shows slightly degraded performance compared to the NLC for the expression variation of the Fb probe set. For the Dup1 and Dup2 probe sets, the accuracies achieved by the

56

Table 5.4: Error rates on different Probe sets of FERET Database.

	NLC	TV+L^1&NLC	NLC+LS
Probe Set	Error Rate	Error Rate	Error Rate
Fb+Fc+Dup1+Dup2	41.53	43.07	**41.28**
Fb	**15.32**	18.66	16.40
Fc	66.49	67.53	**58.76**
Dup1	63.30	63.02	**62.60**
Dup2	87.61	**85.90**	88.03
Fb+Fc+Dup1	36.43	38.32	**36.09**

NLC+LS are fairly comparable to the NLC and TV+L^1&NLC.

As stated by [48] [58], it needs to use more than one gallery image per person to form a reliable linear subspace by the Nearest Linear Combination (NLC) approach. Also the gallery images should represent variation in viewing, expressions, and illumination conditions to consider all possible changes in the probe images. Since the gallery set Fa contains a single image for each person; the predictor of a class was built by using the extracted feature images from a single image to match with the probe image. However, the formed predictor is not sufficient to cover all of the required variations, which is essential for improving the performance by the NLC.

The results in Table 5.4 clearly reflect that the recognition performance of the NLC+LS is at least comparable to the NLC and TV+L^1&NLC methods for almost all of the cases of the FERET dataset.

5.2   Computational Complexity

The time complexity of the proposed approach can be determined in the following way:

To find the time complexity of the testing stage, at first, the complexity of solving the TV+L^1 model by the Parametric Maximum Flow (PMF) algorithm is considered. According to [37], the TV+L^1 model can be solved in $O\left(mnlog(\frac{n^2}{m})\right)$ time, where m and n are the height and weight of the image. Wavelet decomposition is performed with complexity $O\left(mnl\right)$ [2], where l is the decomposition level. After wavelet decomposition, the resolution of the image becomes $r \times s$, where $r < m$ and $s < n$. The complexity of the dimension-reduction algorithm (PCA) is $O\left(rsd\right)$, where d is the number of dimension in reduced dimension space. After that, distance between the probe image and each subject in the gallery is calculated using the Nearest Linear Combination (NLC) [58] classification approach. During this step, required time complexity is $O\left(rsc\right)$, where c is the total number of subjects to be compared. Thus, the total time complexity in the testing stage is $O\left((mn(log(\frac{n^2}{m}) + l)) + (rs(d + c))\right)$.

The complexity of the training stage is similar to that of the testing stage. Assume that each gallery image of a person is decomposed into q feature images using the TV+L^1 model, where k is the number of images per person in gallery. Therefore, the estimated complexity is $O\left(kq((mn(log(\frac{n^2}{m}) + l) + (rs(d + c)))\right)$.

5.3 Summary

A comprehensive comparison of the proposed approach, with the state-of-the-art techniques, has been presented in this chapter. It is appropriate to indicate that the developed approach has been shown to perform well for the case of illumination variations. It shows quite agreeable results with large databases as well. The chapter also includes a complexity analysis of the proposed approach.

Chapter 6

Conclusions

Performing face recognition under different illumination conditions is a challenging field for researchers. Many approaches have been proposed to tackle the problem of illumination changes, yet the problem is far from being solved. The approach presented in this research attempts to provide an illumination-invariant face recognition framework based on the combination of scale-space decomposition and wavelet decomposition.

Scale-space decomposition methods decompose a facial image into large-scale and small-scale features. One of the popular scale-space decomposition methods is the Total Variation model $(TV+L^1)$, which can be utilized to estimate the illumination component. The advantage of this model is that it requires only one parameter, lambda (λ). Most of the $TV+L^1$ based methods use only illumination-invariant small-scale features for recognition after removing the large-scale features.

Another popular tool for image analysis is Discrete Wavelet Transform (DWT), which decomposes an image into different sub-band images with various frequency ranges. The resultant low frequency sub-band (LL sub-band) image has lower resolu-

59

tion than the original image and contains the most important information.

Unlike other scale-space decomposition based methods, the proposed approach is designed to use both large-scale and small-scale facial features of gallery images. For multiscale decomposition process using the TV+L^1 model, a schedule for the values of λ is provided. To include all possible lighting conditions (from normal to abnormal), features of different scales are extracted using the λ values within a predefined range. Then the LL sub-band images are obtained after applying DWT on the extracted feature images, which can reduce the computational complexity of deriving feature vectors by Principal Component analysis (PCA). For the classification, the Nearest Linear Combination (NLC) approach is used to combine the prototypical vectors of a person to represent that individual. Lastly, the probe image is compared to all the linear combinations to determine the class.

To evaluate the proposed approach, several experiments were conducted on different databases. The proposed approach has shown to be superior than the other two baseline approaches: PCA and NLC (for the Yale, ORL and PIE datasets). It also has a better or fairly comparable result to other TV+L^1 based methods: the TV+L^1&PCA and TV+L^1&NLC. It reveals that the integration of large-scale and small-scale features by the proposed method improves the results of the methods (TV+L^1&PCA and TV+L^1&NLC) that perform recognition based on only small-scale features.

In this research, the proposed approach is validated with different databases. In the future, further research is needed to be done to improve the performance of the system for large databases.

Although the TV+L^1 model is utilized in the proposed framework, the effectiveness of the proposed framework should be applicable when it is in conjunction with

other scale-space decomposition approaches for illumination normalization. Further research on replacing the TV+L^1 with the LTV [24] and TVQI [23] is needed, to validate the generality of the proposed approach. The possibility of embedding other feature extraction methods will also be studied in future works.

Furthermore, the NLC classification approach is utilized in the proposed approach for the recognition task. Additional research could be done to get better performance by adding weights for computing linear combinations. The performance of the system can be tested by using other classification techniques, such as the Nearest Constrained Linear Combination (NCLC) approach and Support Vector Machines (SVM).

The proposed approach outperforms two different holistic approaches, and is also superior or equivalently accurate with other TV+L^1 based methods. The initial studies in this research show that there is still room for improvement, which can increase the effectiveness of the proposed method.

Bibliography

[1] Georgia Tech Face Database. www.anefan.com/research/facereco.htm. Feb.16, 2014.

[2] The Wavelet. http://en.wikipedia.org/wiki/Wavelet. Feb.16, 2014.

[3] P. R. A. J. Noraini and N. Selvanathan. Human face recognition using Zernike moments and Nearest Neighbour Classifier. In *4th student conference on Research and Development*, pages 120–123, 2006.

[4] Y. Adini, Y. Moses, and S. Ullman. Face Recognition: the Problem of Compensating for Changes in Illumination Direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:721–732, 1997.

[5] N. U. Ahmed and K. R. Rao. *Orthogonal Transforms for Digital Signal Processing*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1975.

[6] F. Alizadeh, F. Alizadeh, D. Goldfarb, and D. Goldfarb. Second-Order Cone Programming. *Mathematical Programming*, 95:3–51, 2001.

[7] S. Alliney. A property of the minimum vectors of a regularizing functional defined by means of the absolute norm. *IEEE Transactions on Signal Processing*, 45(4):913–917, 1997.

[8] AT&T. The ORL Dataset. http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html. Feb.16, 2014.

[9] G. A. Bahtivar. Holistic Face Recognition By Dimension Reduction. M.Sc. Thesis, Department of Electrical and Electronics Engineering, Middle East Technical University, Sept. 2003.

[10] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by Independent Component Analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[11] R. Basri and D. W. Jacobs. Lambertian Reflectance and Linear Subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, Feb. 2003.

[12] P. N. Belhumeur, P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 711–720, 1997.

[13] P. N. Belhumeur and D. J. Kriegman. What is the Set of Images of an Object Under All Possible Lighting Conditions? In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, CVPR '96, page 270, Washington, DC, USA, 1996.

[14] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy. Electrophysiological Studies of Face Perception in Humans. *J. Cognitive Neuroscience*, 8(6):551–565, Nov. 1996.

[15] V. Blanz and T. Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.

[16] R. Brunelli and T. Poggio. Face Recognition through Geometrical Features. In *European Conference on Computer Vision (ECCV)*, pages 792–800, 1992.

[17] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[18] D. Cai. Matlab codes for dimensionality reduction (subspace learning). `http://www.cad.zju.edu.cn/home/dengcai/Data/DimensionReduction.html`. Feb.16, 2014.

[19] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[20] T. Chan, S. Esedoglu, F. Park, and A. Yip. Recent Developments in Total Variation Image Restoration. In *In Mathematical Models of Computer Vision*. Springer Verlag, 2005.

[21] T. F. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math*, 2005.

[22] L. Chen and N. Tokuda. A unified framework for improving the accuracy of all holistic face identification algorithms. *Artificial Intelligence Review*, 33(1-2):107–122, 2010.

[23] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Illumination normalization for face recognition and uneven background correction using total variation based image models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pages 532–539, 2005.

[24] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total Variation Models for Variable Lighting Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(9):1519–1524, Sept. 2006.

[25] I. Daubechies. The Wavelet Transform, Time-frequency Localization and Signal Analysis. *IEEE Trans. Inf. Theor.*, 36(5):961–1005, Sept. 2006.

[26] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *Int. J. Imaging Systems and Technology*,

15(5):252–260, 2005.

[27] H. K. Ekenel and R. Stiefelhagen. Local Wavelet Analysis for Face Recognition. In *15th IEEE Conference on Signal Processing and Communications Applications*, pages 1–4, 2007.

[28] M. S. El-Bashir. Face Recognition using Multi-Classifier. *Applied Mathematical Sciences*, 6(45):2235–2244, 2012.

[29] M. J. Er, W. Chen, and S. Wu. High-speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks. *Trans. Neur. Netw.*, 16(3):679–691, May 2005.

[30] G.-C. Feng, P. C. Yuen, and D.-Q. Dai. Human face recognition using PCA on wavelet subband. *J. Electronic Imaging*, 9(2):226–233, 2000.

[31] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(7):179–188, 1936.

[32] W. T. Freeman, J. B. Tenenbaum, W. T. Freeman, and J. B. Tenenbaum. Learning bilinear models for two-factor problems in vision. In *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1997.

[33] J.-Y. W. GAO-YUN AN and Q.-Q. RUAN. Enhanced TV -based Quotient Image Model and its application to Face Recognition with One Sample per Subject Employing Subspace Methods. *Journal of Information Science and Engineering*, 26(1):279–290, 2010.

[34] C. Garcia, G. Zikos, and G. Tziritas. A Wavelet-based Framework for Face Recognition. In *Workshop on Advances in Facial Image Analysis and Recognition Technology, 5 Th European Conference on Computer Vision*, pages 84–92, 1998.

[35] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From Few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:643–660, 2001.

[36] D. Goldfarb and W. Yin. Second-Order Cone Programming Methods for Total Variation-Based Image Restoration. *SIAM J. Sci. Comput*, 27:622–645, 2004.

[37] D. Goldfarb and W. Yin. Parametric Maximum Flow Algorithms for Fast Total Variation Minimization. *SIAM J. Sci. Comput.*, 31(5):3712–3743, Oct. 2009.

[38] R. Gross, S. Baker, I. Matthews, and T. Kanade. Face Recognition Across Pose and Illumination. In S. Z. Li and A. K. Jain, editors, *Handbook of Face Recognition*, pages 197–221. Springer, 2011.

[39] R. Gross and V. Brajovic. An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, June 2003.

[40] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image Vision Comput.*, 28(5):807–813, May 2010.

[41] M. A. Grudin. On internal representations in face recognition systems. *Pattern Recognition*, 33(7):1161–1177, 2000.

[42] E. Gumus, N. Kilic, A. Sertbas, and O. N. Ucan. Evaluation of Face Recognition Techniques Using PCA, Wavelets and SVM. *Expert Syst. Appl.*, 37(9):6404–6408, Sept. 2010.

[43] G. Guo, S. Z. Li, and K. Chan. Face Recognition by Support Vector Machines. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, FG '00, page 196, Washington, DC, USA, 2000. IEEE Computer Society.

[44] H. Han, S. Shan, X. Chen, S. Lao, and W. Gao. Separability Oriented Prepro-
cessing for Illumination-insensitive Face Recognition. In *Proceedings of the 12th
European Conference on Computer Vision: Part VII*, ECCV'12, pages 307–320,
Berlin, Heidelberg, 2012. Springer-Verlag.

[45] H. Han, S. Shan, L. Qing, X. Chen, and W. Gao. Lighting Aware Preprocessing
for Face Recognition Across Varying Illumination. In *Proceedings of the 11th
European Conference on Computer Vision: Part II*, ECCV'10, pages 308–321,
Berlin, Heidelberg, 2010. Springer-Verlag.

[46] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face Recognition: Component-Based
versus Global Approaches. Computer Vision and Image Understanding, 2003.

[47] Y. Ijiri, M. Sakuragi, and S. Lao. Security Management for Mobile Devices by
Face Recognition. In *Proceedings of the 7th International Conference on Mobile
Data Management*, MDM '06, page 49, Washington, DC, USA, 2006. IEEE
Computer Society.

[48] R. T. Imran Naseem and M. Bennamoun. Linear Regression for Face Recog-
nition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,
32(11):2106–2112, 2010.

[49] M. Ingebretsen. The Brave New Intelligent Interface. *IEEE Intelligent Systems*,
25:4–8, 2010.

[50] R. Jafri and H. R. Arabnia. A Survey of Face Recognition Techniques. *JIPS*,
5(2):41–68, 2009.

[51] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical Pattern Recognition: A re-
view. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(1):4–
37, 2000.

[52] M. K. K. P. Chandar, M. M. Chandra and B. Latha. Multi Scale Feature Extraction and Enhancement using SVD towards Secure Face Recognition System. In *International Conference on Signal Processing, Communication, Computing and Networking Technologies*, pages 64–69, 2011.

[53] S. Kakarwal and R. Deshmukh. Wavelet Transform based Feature Extraction for Face Recognition. *International Journal of Computer Science and Application Issue*, pages 100–104, 2010.

[54] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220:671–680, 1983.

[55] J. H. Kuang-Chin Lee and D. J. Kriegman. Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[56] J. R. B. K. S. Kyungim Baek, Bruce A. Draper. PCA vs. ICA : A comparison on the FERET dataset. In *Joint Conference on Information Sciences*, 2002.

[57] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face Recognition: A Convolutional Neural Network Approach. *IEEE Transactions on Neural Networks*, 8:98–113, 1997.

[58] S. Z. Li. Face Recognition Based on Nearest Linear Combinations. In *IEEE Trans. On Neural Networks*, pages 839–844, 1998.

[59] S. Z. Li and J. Lu. Face Recognition Using the Nearest Feature Line Method. *IEEE Transactions on Neural Networks*, 10:439–443, 1999.

[60] K. Ma and X. Tang. Discrete Wavelet Face Graph Matching. In *International Conference on Image Processing*, volume 2, pages 217–220, 2001.

[61] A. Martinez and R. Benavente. The AR Face Database. CVC technical report 24, 1998.

[62] M. Mazloom and S. Kasaei. Combination of Wavelet and PCA for Face Recognition. In *EEE GCC conference*, pages 1–5, 2006.

[63] M. Mazloom, S. Kasaei, and H. A. Neissi. Construction and Application of SVM Model and Wavelet-PCA for Face Recognition. In *Proceedings of the 2009 Second International Conference on Computer and Electrical Engineering - Volume 01*, ICCEE '09, pages 391–398, Washington, DC, USA, 2009. IEEE Computer Society.

[64] J. Meng and W. Zhang. Volume Measure in 2DPCA-based Face Recognition. *Pattern Recogn. Lett.*, 28(10):1203–1208, July 2007.

[65] G. O. J.-M. P. Michel Misiti, Yves Misiti. *Wavelet Toolbox For Use with MAT-LAB*. MathWorks, 1996.

[66] T. Mitchell. *Artificial Neural Networks*. McGraw Hill, 1997.

[67] X. Mu and M. H. Hassoun. Combining Gabor Features: Summing vs. Voting in Human Face Eecognition. In *IEEE International conference on Systems, Man and Cybernetics*, volume 1, pages 737–743, 2003.

[68] C. Nastar and N. Ayache. Frequency-Based Nonrigid Motion Analysis: Application to Four Dimensional Medical Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(11):1067–1079, 1996.

[69] C. Nastar, B. Moghaddam, and A. Pentland. Flexible Images: Matching and Recognition Using Learned Deformations. *Computer Vision and Image Understanding*, 65(2):179–191, 1997.

[70] C. A. Nelson. The Development and Neural Bases of Face Recognition. *Infant and Child Development*, 10:3–18, 2001.

[71] V. D. M. Nhat and S. Lee. An Improvement on PCA Algorithm for Face Recognition. In J. Wang, X. Liao, and Z. Yi, editors, *ISNN (1)*, volume 3496 of *Lecture Notes in Computer Science*, pages 1016–1021. Springer, 2005.

[72] M. Nikolova. Minimizers of Cost-Functions Involving Nonsmooth Data-Fidelity Terms. Application to the Processing of Outliers. *SIAM J. Numerical Analysis*, 40(3):965–994, 2002.

[73] M. Nikolova. A Variational Approach to Remove Outliers and Impulse Noise. *Journal of Mathematical Imaging and Vision*, 20(1-2):99–120, 2004.

[74] M. Nixon. Eye Spacing Measurement for Facial Recognition. In *Application of Digital Image Processing*, pages 279–285, 1985.

[75] S. Osher, A. Sole, and L. Vese. Image Decomposition and Restoration Using Total Variation Minimization and the H^1 Norm. *Simul*, 1:349–370, 2002.

[76] D. K. P. Temdee and K. Chamnongthai. Face Recognition by Using Fractal Encoding and Backpropagation Neural Network. 5th International Symposium on Signal Processing and its Applications, 1999.

[77] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, CVPR '05, pages 947–954, Washington, DC, USA, 2005. IEEE Computer Society.

[78] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, Oct. 2000.

[79] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET Database and Evaluation Procedure for Face Recognition Algorithms. *Image and Vision*

Computing, 16(5):295–306, 1998.

[80] PIE. The CMU PIE database. http://www.cad.zju.edu.cn/home/dengcai/
Data/FaceData.html. Feb.16, 2014.

[81] H. L. Q. Liu and S. Ma. A non-parameter Bayesian classifier for face recognition.
Journal of electronics, 20(5):362–370, 2003.

[82] H. L. Qingshan Liu and S. Ma. Improving Kernel Fisher Discriminant Analysis
for Face Recognition. *IEEE Transactions on Circuits and Systems for Video
Technology*, 14(1):42–49, 2004.

[83] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and
irradiance: determining the illumination from images of a convex Lambertian
object. *Journal of the Optical Society of America A*, 8(10):2448–2459, 2001.

[84] T. P. Ryan. *Modern Regression Methods*. Wiley-Interscience, 1997.

[85] F. S. Samaria. Parameterisation of a Stochastic Model for Human Face Iden-
tification. In *Proceedings of Second IEEE Workshop Applications of Computer
Vision*, pages 138 – 142, Dec 1994.

[86] M. P. Satone and G. K. Kharate. Face Recognition Based on PCA on Wavelet
Subband of Average-Half-Face. *JIPS*, 8(3):483–494, 2012.

[87] R. J. Schalko. *Pattern Recognition*. John Wiley & Sons, 1992.

[88] G. A. F. Seber. *Linear Regression Analysis*. Wiley-Interscience, 2003.

[89] G. Shakhnarovich and B. Moghaddam. Face Recognition in Subspaces. In *S.Z.
Li, A.K. Jain (EDS.), Handbook of Face Recognition*, pages 141–168. Springer,
2004.

[90] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination Normalization for Robust
Face Recognition Against Varying Lighting Conditions. In *Proceedings of the*

IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG '03, page 157, Washington, DC, USA, 2003. IEEE Computer Society.

[91] A. Shashua and T. Riklin-Raviv. The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2):129–139, 2001.

[92] S.-W. L. Song-Hyang Moon and S.-W. Lee. Illumination Invariant Face Recognition Using Linear Combination of Face Examplars. *Lecture Notes in Computer Science, Springer*, 3546:112–121, 2005.

[93] H. W. Stan, H. Wang, S. Z. Li, and Y. Wang. Generalized Quotient Image. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 498–505, 2004.

[94] D. M. Strong and T. F. Chan. Edge-preserving and Scale-dependent Properties of Total Variation Regularization. *Inverse Problems*, 19(6):165–187, 2003.

[95] D. Swets and J. juyang Weng. Using Discriminant Eigen features for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:831–836, 1996.

[96] S. B. Terence Sim and M. Bsat. The CMU Pose, Illumination, and Expression (PIE) Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1615–1618, 2003.

[97] M. P. Timo Ojala and T. Maenpaa. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, July 2002.

[98] M. Turk and A. Pentland. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, 3(1):71–86, Jan. 1991.

[99] R. University. The TV+L^1 model. `http://www.caam.rice.edu/~wy1/`
`ParaMaxFlow/`. Feb.16, 2014.

[100] P. D. Wadkar and M. Wakhande. Face Recognition using Discrete Wavelet
Transforms. *International Journal of Advanced Engineering Technology*,
3(1):230–242, 2012.

[101] H.-Y. Wang and X.-J. Wu. Weighted PCA space and its application in face
recognition. *Machine Learning and Cybernetics*, 7:4522–4527, 2005.

[102] L. Wiskott, J.-M. Fellous, N. Krüger, and C. V. D. Malsburg. Face Recognition
By Elastic Bunch Graph Matching. *IEEE Trans. Pattern Analysis and Machine
Intelligence*, 19:775–779, 1997.

[103] D. G. Wotao Yin and S. Osher. The Total Variation Regularized L^1 Model
for Multiscale Decomposition. *Multiscale Modeling & Simulation*, 6(1):190–211,
2007.

[104] X. S. Z. Wotao Yin, Terrence Chen and A. Chakraborty. Background correc-
tion for cDNA microarray images using the TV+L^1 model. *Bioinformatics*,
21(10):2410–2416, 2005.

[105] X. Xie, W. Zheng, J.-H. Lai, and P. C. Yuen. Face illumination normalization
on large and small scale features. In *CVPR*. IEEE Computer Society, 2008.

[106] X. Xie, W.-S. Zheng, J.-H. Lai, P. C. Yuen, and C. Y. Suen. Normalization of
Face Illumination Based on Large-and Small-Scale Features. *IEEE Transactions
on Image Processing*, 20(7):1807–1821, 2011.

[107] Yale. The YALE Dataset. `http://www.cad.zju.edu.cn/home/dengcai/Data/`
`FaceData.html`. Feb.16, 2014.

[108] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):131–137, Jan. 2004.

[109] W. Yin, D. Goldfarb, and S. Osher. Image Cartoon-Texture Decomposition and Feature Selection Using the Total Variation Regularized L^1 Functional. *Lecture Notes in Computer Science, Springer*, 3752:73–84, 2005.

[110] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature Extraction from Faces Using Deformable Templates. *Int. J. Comput. Vision*, 8(2):99–111, Aug. 1992.

[111] W. Zhao and R. Chellappa. Robust Face Recognition Using Symmetric Shape-from-Shading. Technical Report, Center for Automation Research, University of Maryland, 1999.

[112] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Comput. Surv.*, 35(4):399–458, Dec. 2003.