Experimental Evaluation of Feature Extraction Schemes for Face Recognition

Shaghayegh Parchami

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
February 2015
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Serhan Çiftçioğlu
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Işık Aybay
Chair, Department of
Computer Engineering Department

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Hakan Altınçay
Supervisor

Examining Committee

1. Prof. Dr. Hakan Altınçay
2. Prof. Dr. Hasan Kömürcügil
3. Asst. Prof. Dr. Ahmet Ünveren

ABSTRACT

In this thesis, we studied the use of Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Gabor wavelets for face recognition. Both PCA and LDA are applied for the extraction of features from the raw pixel values. Then, their use for the extraction of features from the outputs of Gabor wavelets is considered. Lattice-based selection of a subset of Gabor outputs is considered for this purpose. A rectangular grid of various sizes is considered and the Gabor filter outputs extracted from the grid points are employed for feature extraction using PCA and LDA. As an alternative approach, Best Individual Selection (BIS) and Sequential Forward Selection (SFS) are employed for feature subset selection. The k nearest neighbor classifier is employed as the classification scheme. The experiments have been carried out on FERET database. It is observed that the accuracies achieved using Gabor wavelets are superior when compared to the features derived from the raw pixel values. Moreover, superior scores are generally achieved using BIS and SFS approaches when compared to PCA and LDA.

Keywords: Face recognition, sequential feature selection, best individual selection, Gabor wavelets, principal component analysis, linear discriminant analysis

ÖZ

Bu tezde, Ana Bileşenler Analizi (ABA), Doğrusal Ayırtaç Analizi (DAA) ve Gabor dalgacıklarının yüz tanımada kullanımı üzerinde çalışılmıştır. Hem ABA hem de DAA, yüz resimlerindeki ham piksel değerlerinden öznitelik çıkarımı için uygulanmıştır. Daha sonra, Gabor dalgacıklarının çıktılarından öznitelik çıkarımı için kullanımları değerlendirilmiştir. Gabor çıktılarının alt kümelerinin örgü-tabanlı seçimi bu amaçla kullanılmıştır. Değişik boyutlardaki dikdörtgen örgüler kullanılmış ve örgü noktalarında hesaplanan Gabor çıktılarından ABA ve DAA kullanılarak öznitelikler çıkarılmıştır. Alternatif yaklaşım olarak, Eniyi Bireysel Seçimi (EBS) ve Sıradan İleri Seçimi (SİS) de öznitelik altkümesi seçimi için değerlendirilmiştir. k en yakın komşu sınıflandırma yöntemi olarak kullanılmıştır. Deneysel çalışmalar FERET veri kümesinde yapılmıştır. Gabor dalgacıkları kullanıldığında, ham piksel değerleri kullanımına göre daha iyi sonuçlar elde edildiği gözlenmiştir. Ayrıca, EBS ve SİS yaklaşımları ile genelde ABA ve DAA’ya göre daha iyi sonuçlar elde edilmiştir.

Anaytar sözcükler: Yüza tanıma, sıradan ileri seçimi, eniyi bireysel seçimi, Gabor dalgacıkları, doğrusal bileşenler analizi, doğrusal ayırtaç analizi

This master thesis is dedicated to my family with love

ACKNOWLEDGMENT

I would like to express my deep gratitude to my dear supervisor Prof. Dr. Hakan Altınçay for his beneficial guidance and continuous support during the provision of my master dissertation. Without his supervision and guidance this thesis would not have been accomplished.

Worth extremely regard to Prof. Dr. Hasan Kömürcügil and Asst. Prof. Dr. Ahmet Ünveren for serving me as committee members and making my defense become unforgettable for me and also all the staff and and members of computer engineering department without whose collaboration I would not be able to attain the results in this dissertation.

Last but not least important, I owe more than thanks to my parents and two younger brothers who supported me and devoted their love in my whole time. I would like to declare great respect to my dear friend Hamid Mir Mohammad Sadeghi who has shown a tower of patience and endless knowledge. I want to thank him for his supprot and concern throughout my degree. I love them all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Biometric systems

Biometric recognition corresponds to classification of human beings using their physical or behavioral characteristics. In biometric verification, the main aim is to identify whether the input belongs to the target person. In biometric identification, the person to which the given input belongs is computed using a closed-set of people. These systems generally employ one or more measurable characteristics such as facial images, finger prints, iris images, palm prints, voice and hand writing signatures [1]. There are several advantages of using biometric techniques based authentication in practice, some of which are listed below [2]:

- Decreased ID deception and promoted security.
- Automated confirmation.
- No necessity of preserving password.
- No demand of any token to be taken.

1.2 Face Recognition

Face recognition is one of the most important problems in computer vision. It is a challenging pattern classification problem which has attracted the interest of many researchers in recent decades. It has a wide range of applications recognition in practice such as access control, information security, law enforcement and video surveillance. In face recognition, the main purpose is to find best match between the input facial image and the existing images in a given data set [3]. A typical face recognition system is implemented in three major steps as presented in Fig. 1.1. The first step involves detection of the face from a given image.

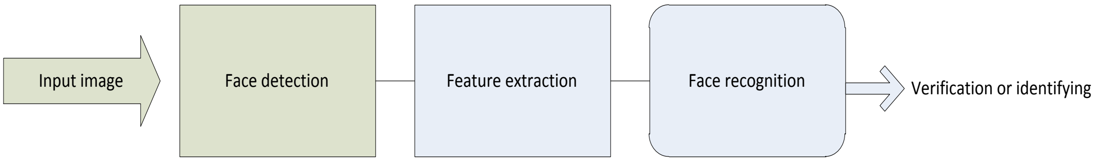

Figure: 1.1: General structure of face recognition system

This step is also essential for some other applications such as pose estimation, face tracking and compression. The following step is the feature extraction where the major concern is to extract coherent information from the facial image. Numerous techniques have been proposed which mainly focus on effective representation of the face so as to extract the most discriminative information from facial images. These efforts can be categorized into two groups as holistic and local features based approaches. Holistic approaches extract features from the whole face. Eigenfaces is an example of the holistic methods. This approach is based on principal component analysis (PCA) which reduces the feature dimensionality while retaining the characteristics of dataset. Local features based methods employ various facial features from more discriminative regions of the faces such as eyebrows, eyes and mouth. A popular local features approach is to use Gabor wavelets [22].

Face recognition is a challenging problem due to several reasons. Changing poses, occlusion of some parts of the faces and the use of glasses may deteriorate the recognition performance. The facial features generally changes due to aging. Illumination and lighting condition can also affect the recognition performance. In practice, numerous techniques are generally employed to detect and, if possible, minimize the distortion to the classification system [4].

1.3 Objectives

As mentioned above, feature extraction has a key (critical) role for face recognition. In this thesis, we studied both holistic and local features based feature extraction techniques. More specifically, we studied the performances of PCA and Linear Discriminant Analysis (LDA) based feature extraction schemes. As the local features approach, we considered Gabor wavelets. The feature vectors extracted by considering all pixels have very large dimensionality. In general, 5 scales and 8 orientations are considered which leads to *40xP* dimensional feature vectors where *P* is the number of pixels. Taking into account the fact that the contributions of different Gabor kernels and pixels to the recognition performance are not equivalent, various techniques are proposed to reduce the feature dimensionality.

In this thesis, transformation of Gabor feature space into a reduced space by exploiting PCA and LDA are firstly addressed. As an alternative approach, lattice-based selection approach is also considered. In this method, a set of points are initially specified by placing a rectangular lattice of size $N \times N$ on the center of the image. Then, a subset of these N^2 points having the most discrimination power is selected. The selection process may be based on individual or joint evaluation. In this thesis, best-individual selection (BIS) where the selection is based on individual performance of the lattice points and sequential forward selection are considered.

1.4 Lay out of the thesis

This thesis consists of four chapters. Second chapter presents a literature review on face recognition techniques. Chapter 3 presents the experimental results obtained using PCA, LDA, and Gabor filter. Chapter 4 is dedicated to conclusions and future work.

Chapter 2

LITERATURE REVIEW

Face recognition has been one of the popular field researches in computer vision over the past several decades. The main objective of face recognition is to compute best match between input image and existing images in a database. In order to achieve this, several intermediate steps such as preprocessing, feature extraction, feature selection and classifier construction are applied. Many uncontrolled conditions such as head orientation and changing in facial expression and so on can have an influence on the performance of face recognition system. Changing lighting conditions is another serious problem that face recognition system designers has to cope with [5]. Preprocessing steps are expected to affect the process of feature extraction and contribute the performance of recognition [6].

This chapter presents an overview of the basic steps of implementing a face recognition system such as feature extraction, feature selection and classifier design. The dataset considered in simulation studies is also presented.

2.1 Preprocessing

The main goal of image preprocessing is to enhance the images so as to raise the discriminative information included and make sure that ambient factors such as lighting conditions cannot negatively influence the process of feature extraction [7]. In this thesis, histogram equalization and illumination normalization are applied during preprocessing.

2.1.1 Histogram Equalization

Histogram equalization is applied for contrast adjustment of the images. As illustrated in Figure 2.1, when histogram equalization is applied, the intensity values are more uniformly distributed in the resultant histogram. Assume that *I(x, y)* is an image with *n* pixels. Let the total number of possible intensity levels in the image and the k^{th} intensity value be represented by L and r_k, respectively. It should be noted that, for 8 bits image, the number of intensity levels is 256. The probability of occurrence of intensity level r_k in the image is defined by

$$P(r_k) = \frac{n_k}{n} \qquad (1)$$

where the number of pixels having the intensity r_k is expressed by n_k. Histogram equalization converts the distribution of pixel intensity values into uniform distribution [7, 8]. This function is defined as follows:

$$S_K = T(r_k) = (L-1)\sum_{j=0}^{k} P(r_j) \qquad (2)$$

where k = 0, 1, 2…, L-1

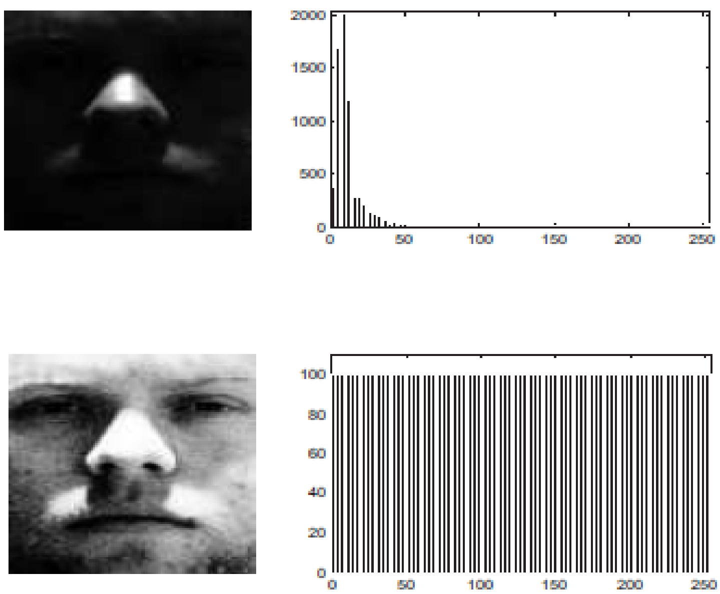

Figure 2.1: A sample image before (upper row) and after (lower row) applying histogram equalization [7]

2.1.2 Illumination Normalization

All images in the dataset should be normalized after the histogram equalization is carried out. The idea of normalization is to standardize images by setting the mean (μ) and standard deviation (σ) of the pixel values of the images to zero and one, respectively. In other words, the intensity value x is modified as $\frac{x-\mu}{\sigma}$. In order to normalize the images, as the first step, mean and standard deviation of all pixels of the image is found. Then, the normalized pixel values are computed. By this method, images become sharp, obvious and noiseless for feature extraction and image analysis [9].

2.2 Feature Extraction

Numerous techniques have been proposed which mainly focus on effective representation of the face so as to extract the most discriminative information from facial images. These efforts can be categorized into two groups as holistic and local features based approaches. Holistic approaches such as PCA extract features from the whole face. On the other hand, local approaches extract features from parts of a given image [10, 11, 12].

A popular local features approach is to use Gabor wavelets. However, the feature vectors extracted by considering all pixels have very large dimensionality. Taking into account the fact that the contributions of different Gabor kernels and pixels to the recognition performance are not equivalent, various techniques are proposed to reduce the feature dimensionality. In fact, it is known that smaller number of features on the order of 200 is enough to achieve comparable recognition accuracy to using all features. Transformation of Gabor feature space into a reduced space by exploiting PCA, LDA and general discriminant analysis (GDA) are also studied where it is shown that GDA generally provides higher accuracies compared to PCA and LDA.

Alternatively, salient facial points based local features approaches which aim at computing features from discriminative parts of the images are studied. Experiments have shown that better feature vectors generally involve local features extracted from eyes and mouth regions of the facial images. An important step in local feature extraction is localization of salient points from which discriminative features can be generated. This is also known as landmark-based sampling. Since mouth and eyes regions are known to convey discriminative information, the salient points may be manually placed within these regions. Alternatively, automatic selection of salient facial points can be considered.

In order to speed up automatic learning of discriminative facial locations, the search space may be reduced by using lattice-based approach. In this method, a set of points is initially specified by placing a rectangular lattice on the center of the image. Then, a subset of these points having the most discrimination power is selected. The selection process may be based on individual or joint evaluation. For instance, BIS may be used where the selection is based on individual performance of the lattice points. Computation of the optimal set of facial pints is a challenging problem.

The features extracted from either landmark-based or lattice-based facial points are generally concatenated to form a single feature vector representing the face which can be considered as feature-level combination of information from different pixels. Then, classification is performed using these composite feature vectors. As an alternative approach, the feature vectors can be combined using the model-level fusion approach where a different classifier is implemented for each facial point. Then, the outputs of these classifiers are combined so as to determine the most likely person.

This study will consider PCA, LDA and PCA+LDA as holistic methods and Gabor wavelet and lattice sampling as local-features based methods.

2.2.1 Principal Component Analysis (PCA)

Principal component analysis is a statistical technique to express the given data as a linear combination of principal components. PCA is a useful method to reduce the dimensionality while preserving the variability on the data. The principal components are perpendicular to each other since they are computed as the eigenvectors of the symmetric covariance matrix [13].

Each two dimensional image is expressed as a 1-D vector. This vector is constructed by concatenation each column (or row). Assume that the number of training images is M and each image can be shown as a vector of size N (number of rows x number of columns). Hence, the whole image can be represented by M vectors (X_i) of size N.

$$X_i = [p_1, p_2, \dots, p_N]^T, \; i = 1, \dots, M \qquad (3)$$

where p expresses the pixel values. Let μ represent the average of the training images which is defined by

$$\mu = \frac{1}{M}\sum_{i=1}^{M} X_i \qquad (4)$$

In PCA, the mean vector is then subtracted from each image as

$$r_i = X_i - \mu \qquad (5)$$

In order to find the eigenvalues and eigenvectors, the covariance matrix should be calculated using the following equation:

$C = WW^T$ (6)

where $W = [r_1, r_2, \ldots, r_M]$ and C is a square matrix with dimensionality of $N \times N$.

The eigenvalues and eigenvectors of the covariance matrix should then be computed. However, since the size of C is too large, it is not generally feasible to find eigenvalues and eigenvectors directly. As an alternative approach, the eigenvectors and eigenvalues of matrix C can be obtained from the eigenvectors and eigenvalues of $W^T W$. Suppose that V_i stands for the eigenvectors and λ_i for the eigenvalues of $W^T W$ such that

$W^T W\, V_i = \lambda_i V_i$ (7)

Multiplying both sides by W, we obtain

$W\, W^T (W\, V_i) = \lambda_i W\, V_i$ (8)

This equation implies that $W\, V_i$ and λ_i provide the eigenvectors and eigenvalues of $W\, W^T$, respectively.

Thus, $W^T W$ is employed for computing the eigenvectors of the covariance matrix. The eigenvectors would be sorted from highest to lowest according to their eigenvalues. The top 10% to 15% of the eigenvectors generally contains 90% of total variance in the images and for this reason a subset of the eigenvectors are generally selected [14, 15]. The resultant eigenvectors are computed using

$U_i = W\, V_i$ (9)

U_{i} are generally named as Eigenfaces [16]. Each facial image in the training set is then projected onto a lower space $(M' << M)$ using

$P_k = U^T\, (x_k - \mu)$ (10)

Each test images is also projected onto the Eigenspace. Let a transformed test image be denoted by P. During classification, the minimum distance between P and the training images is computed as follows:

$$\epsilon_k = \|P - P_k\|, k = 1, \dots, M \qquad (11)$$

2.2.2 Linear Discriminant Analysis (LDA)

The main objective of linear discriminant analysis is to reduce the dimensionality of the facial images while preserving the separability of different people. In order achieve this, the projection vectors are computed by employing between-class scatter matrix and within-class scatter matrix [16, 17].

Suppose that training set includes D persons and each person has k_i images $(i = 1,2, \dots, D)$. The total number of training images is equal to $M = \sum_{i=1}^{D} k_i$. Each person corresponds to a different class for face recognition where the i^{th} class is represented by ω_i. Assume that $k_i = k, (i = 1,2, \dots, D)$ and ω_{ij} is the j^{th} image of i^{th} class. For each class, the average image (μ_i) is obtained as

$$\mu_i = \frac{1}{k}\sum_{j=1}^{k} \omega_{ij} \quad , (i = 1,2, \dots, D) \qquad (12)$$

Moreover, for all classes the overall mean can be defined as

$$\mu = \frac{1}{D}\sum_{i=1}^{D} N_i \mu_i \qquad (13)$$

where N_i is the number of samples in class ω_i. S_W is the within-class scatter matrix which can be computed as follows

$$S_W = \sum_{i=1}^{D} \sum_{X_j \in \omega_i} (X_j - \mu_i)(X_j - \mu_i)^T \qquad (14)$$

Additionally, between-class scatter matrix is defined as

$$S_B = \sum_{i=1}^{D} N_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad (15)$$

In order to maximize the separability of different classes, the criterion to be maximized is defined as [18, 19]

$$W_{opt} = argmax_W \frac{|W^T S_B W|}{|W^T S_W W|} = [V_1, V_2, \dots V_m] \qquad (16)$$

where W_{opt} denotes the optimal transformation matrix. The solution of the above problem corresponds to solving the following equation

$$S_W^{-1} S_B V_i = \lambda_i V_i \qquad (17)$$

In other words, the eigenvectors $S_W^{-1} S_B$ corresponds to the candidate projection directions. As the number of classes is equal to D, the projection matrix has at most $D-1$ eigenvectors corresponding to the non-zero eigenvalues.

2.2.3 PCA+LDA Approach

PCA is generally preferable when the number of samples is small and the dimension is high. On the other hand LDA is preferred when we have a large dataset including large number of different classes [16].

Note that LDA has some problems. Firstly, the eigenvectors of $S_W^{-1} S_B$ are not orthogonal since the $S_W^{-1} S_B$ matrix is not generally a symmetric matrix. Hence, LDA is not able to produce an orthonormal projection set. Furthermore, the dimension of S_W and S_B are too large and the processing time of $S_W^{-1} S_B$ is very high. Moreover, the within-class scatter matrix may be singular which means that this matrix may not be invertible. Therefore, $S_W^{-1} S_B$ cannot be computed directly [20]. In order to overcome these drawbacks, PCA+LDA algorithm was proposed. In this approach, PCA performs as an intermediate space. This implies that, before starting LDA computation, the training set is projected onto a reduced space by PCA. Then, LDA uses this new space to calculate the within-class scatter matrix and between-class scatter matrix using equation (14, 15) and hence the eigenvectors of $S_W^{-1}S_B$ [21].

2.2.4 Gabor Wavelet

Two dimensional Gabor wavelet (or filter) function is defined to be in the following form [22]

$$\Psi_j(x,y) = \frac{k_{u,v}^2}{\sigma^2}\left(e^{-\frac{\kappa_{u,v}^2(x^2+y^2)}{2\sigma^2}}\right).(e^{ik_{u,v}(x\cos(\varphi_u)+y\sin\varphi_u))} - e^{-\frac{\sigma^2}{2}}) \quad (18)$$

The filter is defined as the product of a Gaussian envelope and a complex plane wave. $\left(e^{-\frac{\kappa_{u,v}^2(x^2+y^2)}{2\sigma^2}}\right)$ is the Gaussian function which represents optimal localization of Gabor wavelet in both time and frequency domains [23]. σ specifies the width of the Gaussian envelope and it is set to be 2π. The wave vector $(k_{u,v})$ is defined as

$$k_{u,v} = k_v e^{i\varphi_u} \quad (19)$$

where $k_v = 2^{-\frac{v+2}{2}}$ and $\varphi_u = \frac{\pi u}{8}$ (20)

The index can be stated as

$$j = u + 8v \quad (21)$$

Five different scales frequencies $(v = 0,1 \ldots 4)$ and eight different orientations $(u = 0,1 \ldots 7)$ define 40 different Gabor filters.

Real and imaginary parts of Gabor filter can be defined by the following equations, respectively [23, 24, 25].

$$R_e(\Psi) = \frac{\kappa_{u,v}^2}{\sigma^2}\left(e^{-\frac{\kappa_{u,v}^2(x^2+y^2)}{2\sigma^2}}\right).\cos\left(k_{u,v}(x\cos(\varphi_u) + y\sin(\varphi_u))\right) \quad (22)$$

$$I_m(\Psi) = \frac{\kappa_{u,v}^2}{\sigma^2}\left(e^{-\frac{\kappa_{u,v}^2(x^2+y^2)}{2\sigma^2}}\right).\sin\left(k_{u,v}(x\cos(\varphi_u) + y\sin(\varphi_u))\right) \quad (23)$$

The magnitude of complex outputs is defined as

$$O(x,y) = \sqrt{{I_m}^2 + {R_e}^2} \qquad (24)$$

Consider a face image denoted by *I (x, y)*. The convolution of *I(x, y)* and Gabor kernels provides the Gabor wavelet transform which can be written as

$$F(x,y) = I(x,y) * \Psi_j(x,y) \qquad (25)$$

Gabor filters are applied on the images in two different ways to extract facial features. One way is that the whole image is convolved with all Gabor kernels (40 filters). The obtained image has the same size as the original image. Another method is to apply the filter on selected or fiducial points on the face to emphasize significant areas like eyes and mouth. A feature vector is then formed from all complex coefficients which are computed by the convolution of each selected point and all 40 filters. In this thesis, we applied the selected-point method where the Gabor filters will be applied only on a fixed set of points [22]. Figure 2.2 and 2.3 present 40 different Gabor filters and the magnitudes obtained after applying on a facial image.

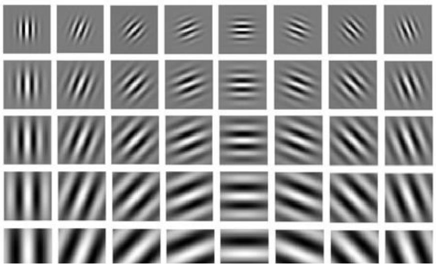

Figure 2.2: 40 different Gabor filters [23].

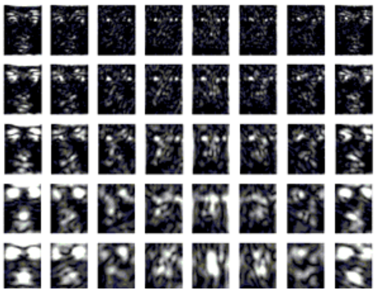

Figure 2.3: The magnitudes of the Gabor feature representation [23].

2.2.5 Lattice and Landmark Sampling

Two methods can be utilized for specifying important facial location: lattice sampling and landmark sampling.

In lattice based approach, a rectangular grid of size $m \times m$ is placed over the face image. The convolution is performed with Gabor wavelet kernels at different frequencies and orientations at each point of this grid and then a feature vector for the entire face is formed by the concatenation of the magnitude of the complex outputs of Gabor wavelet.

In the landmark method, some salient facial points are utilized. Generally, 30 salient points ($S = 30$) over the facial image are employed by the researchers. The goal of these sampling schemes is to define the important location between these points and to test the points that are really discriminative [30].

2.3 Feature Selection

The objective of feature selection is to select an optimal subset of features to minimize classification error and redundancy [10]. Feature selection methods are able to enhance learning performance, degrade computational cost and storage requirement, reduce feature space dimensionality, decrease the redundant and noisy data and construct generalizable models [35]. Feature selection techniques can be categorized into two groups, namely filter methods and wrapper methods.

Filter methods rely on some intrinsic characteristics of training data to choose features individually. However in wrapper methods, learning algorithms are also considered and the features may also be jointly evaluated [31]. It should be note that the criterion function and search strategy are very important in feature selection [34].

In this study, we have used the wrapper methods, namely Best Individual Selection (BIS) and Sequential Forward Selection (SFS)

2.3.1 Best Individual Selection (BIS)

Assume that a feature set has *n* variables, $F = \{f_1, f_2, \ldots f_n\}$. The goal of this method is to find a subset with the best d features ($d<n$).

Define S to be the set of all features. The criterion is denoted by $j(f_i)$ which shows the discrimination performance of f_i for face recognition. This method evaluates $j(f_i)$ for all features and sorts them in decreasing order. The top ranked *d* features are used during classification. As the criterion function, the classification accuracy in face recognition can be considered [31, 32, 33].

2.3.2 Sequential Forward Selection (SFS)

Sequential Forward Selection starts with an empty set of selected features denoted by S. In each step, this algorithm adds one feature to set S as the most effective additional feature. In order to decide on the best additional feature, it evaluates the candidate features together with the already selected ones. The algorithm can be summarized as follows:

1. Choose S as selected features set which is empty , $S = \phi$.
2. Find the best feature f_y: $f_y = argmax_{f_{x \notin S}} j(S \cup f_x)$.
3. Add f_y to the selected features set S: $S = (S \cup f_y)$.
4. Go back to step 2.

This algorithm continues until the candidate features do not add any benefit to the already selected set. Mostly, the effectiveness of SFS is higher than BIS [31, 32, 34].

2.3.3 Sequential Backward Selection (SBS)

This method is similar to SFS; however the procedure is in the exact opposite order. This implies that, instead of adding the most effective feature to the selected features set, it removes the least effective feature from it. This algorithm considers all the features as the selected features set (S) and takes into account the performance of S by absence of one feature (f_y) from S. By removing f_y from set S, more useful features remain in S. This process should be carried out until further improvement is not possible by omitting any of the remaining ones [34, 35]. The steps of this method are summarized below.

1. Choose S as the set of all existing features.
2. Find the most useless feature $f_y : f_y = argmax_{f_{x \in S}} j(S - \{f_x\})$
3. Discard f_y from S: $S = (S - \{f_y\})$.
4. Go back to step 2.

Since the running-time of this method is too long, we have considered the BIS and SFS as feature selection methods for this thesis.

2.4 Classifiers

After the features are selected, the next step is the design of a classification scheme. There are various methodologies that can be used for this purpose. In face recognition, since the number of samples for each class is limited, simpler models are generally preferred [36]. These techniques are mainly based on evaluating the similarity of the samples [10]. In this thesis, k-nearest neighbor classifier is employed.

k-Nearest Neighbor Classifier is one of the oldest and popular scheme. It is based on computing the distances between the test sample and all training samples. The labels of the k nearest samples are considered in making the final decision. In general, voting is applied to decide the most likely class. When $k=1$, the classifier assigns the test sample to the class which has the closest training sample to this test sample [36, 37]. In order to measure the similarity between different samples, the Euclidean distance measure is generally used [38] which is defined as

$$dist(x, y) = \left(\sum_{i=1}^{d} |x_i - y_i|^2\right)^{1/2}. \qquad (25)$$

2.5 Datasets

In this thesis, the experiments are carried out on FERET database. 205 arbitrarily selected subjects are considered, each having four frontal images. The images are firstly cropped to the size of 80×64. Histogram equalization followed by zero-mean unit-variance normalization is then applied. The experiments are repeated for four experimental sessions. In each session, one of the images is left out for testing and the remaining three are used during testing. Then, the classification rates are averaged.

Chapter 3

EXPERIMENTAL RESULTS

In order to evaluate the performance of the feature extraction schemes discussed in Chapter 2, experiments are carried out on FERET database. A subset of the database which includes 820 images that correspond to 205 persons is considered. Each person is represented with 4 different frontal gray scale images which have different illumination conditions and facial expressions. The images are cropped to the size of 80×64 and 8 bits gray level representation is used. Three images of each person are employed for training and one image for testing to obtain the accuracy.

The images are firstly preprocessed. It includes histogram equalization followed by zero mean and unit variance normalization. Consequently, the undesirable effects of variations in lighting conditions are avoided.

3.1 Comparing the performances of PCA and PCA+LDA

As mentioned in Chapter 2, each training image is firstly expressed as a 1-D vector. This corresponds to 1×5120 vectors in raw form. The training data matrix is then constructed whose size is 615×5120. The mean vector (μ) is computed using equation (4) which has the size 615×1. Then, using Equation (5), the mean is subtracted from each image and the covariance matrix is computed by Equation (6). Then, the eigenvalues and eigenvectors are computed using Equation (8) and they are sorted in decreasing order. The eigenvectors corresponding to the largest eigenvalues are employed for computing the feature vectors. Since the aim of PCA is to reduce the dimension by extracting principal features, we did not select all eigenvectors. We utilized different number of eigenvectors to obtain feature vectors of various lengths using Equation (10). By applying the same procedure, the feature vectors are computed for the test images. The classification is then carried out by using nearest neighbor classifier (1-NNC). The classification accuracy is computed as the accuracy on the test images. It is the percentage of the test samples which are classified correctly by nearest neighbor classifier. The effectiveness of PCA with different number of features is expressed in Table 3.1.

Table 3.1: Accuracy (in %) of PCA using 1-NNC classifier

No. of features	**PCA**
10	69.7561
20	77.0732
30	78.5366
40	79.5122
50	79.5122
60	79.5122
70	79.5122
80	79.5122
90	79.5122
100	79.5122

As it can be seen in Table 3.1, the accuracy did not change after selecting more than 40 features.

As it was mentioned in Section 2.2.3, LDA has some drawbacks. Since within-class scatter matrix (S_W) is too large and it may not always be invertible. Therefore, we used the PCA+LDA. The feature vectors computed using PCA are used as the input for LDA. The mean of each class and the overall mean are computed using Equations (12) and (13) respectively. The number of samples in each class is 3 which corresponds to the number of training images in each class. Then, S_W and S_B are calculated using Equations (14) and (15), and then the eigenvalues and eigenvectors of $S_W^{-1}S_B$ were calculated. The aim of LDA is to obtain the optimal projection. It provides the projection matrix by finding eigenvalues and eigenvectors of $S_W^{-1}S_B$. In this thesis, the number of selected features for PCA and PCA+LDA are set to be equal. After the feature vectors are constructed, the classification is done using 1-NNC. The results of this method are shown in Table 3.2.

Table 3.2: Accuracy (in %) of PCA+LDA using 1-NNC classifier

No. of features	**PCA+LDA**
10	72.1951
20	87.3171
30	92.6829
40	94.6341
50	94.6341
60	96.5854
70	94.6341
80	96.0976
90	95.6098
100	94.6341

In this table, the highest accuracy is 96.5854 which shows that this method is more effective than PCA. PCA is not an effective method on its own. We did not use all features in both PCA and PCA+LDA since the computational load was increased. Comparison of the performances of PCA and PCA+LDA is shown in Fig 3.1.

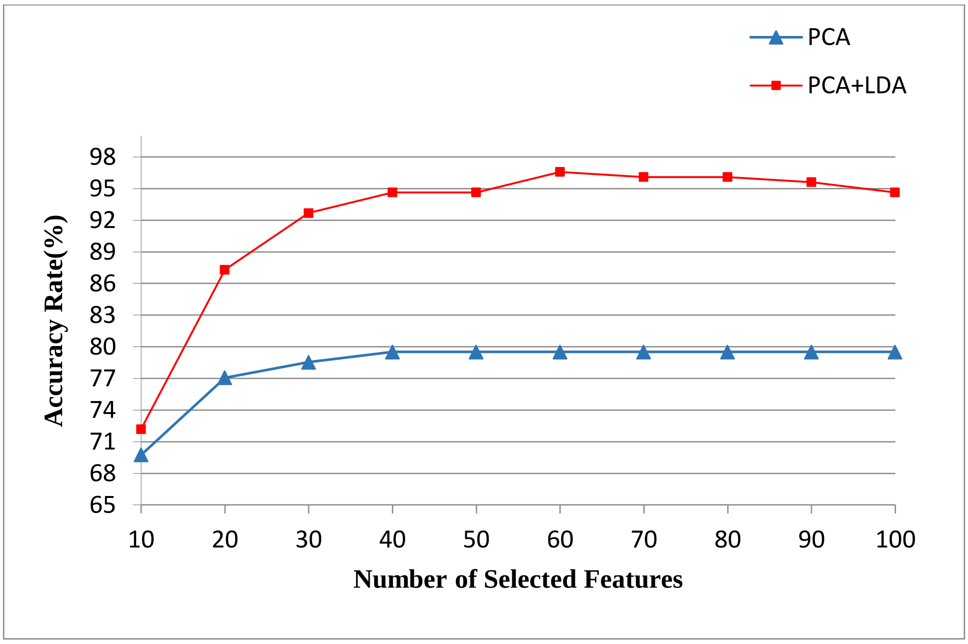

Figure 3.1: The comparative performance of PCA and PCA+LDA

3.2 Evaluation of the performance of Gabor wavelets and Gabor wavelets together with PCA+LDA

In order to evaluate the performance of Gabor filters, further experiments are conducted. Instead of applying the Gabor filters on the entire images, some points of the image are firstly selected by lattice sampling which was explained in section 2.2.5, and then Gabor filters are applied on these points. In fact, we utilized lattice-based sampling for using Gabor filters, which was explained in subsection 2.2.4. The lattice sampling was used with 3 different sizes: 7×7, 15×15 and 21×21. The selected grid was positioned on the centers of the facial images. Each point of this grid was convolved with Gabor kernels, as explained in subsection 2.2.4. A feature vector including 40 real and imaginary entries are extracted for each point due to employing 5 frequencies and 8 orientations. Then, for each point, the magnitude of all complex outputs is computed. For each of the 49 points of an image, we obtained a different magnitude feature vector which is then concatenated to obtain a feature vector of size 1960 (49×40). The classification is done using 1-NNC classifier as before. The same procedure is carried out for grids of size 15×15 and 21×21. The accuracies obtained using this technique are shown in Table 3.3 for 3 different grid sizes.

Table 3.3: Accuracy (in %) of Gabor filters using 1-NNC classifier

Grid $\mathbf{7 \times 7}$	**Grid $\mathbf{15 \times 15}$**	**Grid $\mathbf{21 \times 21}$**
95.6098	96.0976	96.5854

The use of PCA+LDA on the Gabor features is also considered. The scores obtained are represented in Table 3.4.

Table 3.4: Accuracy (in %) of Gabor features using PCA+LDA for 3 different sizes of lattice sampling

No. of Features	**Grid 7×7**	**Grid 15×15**	**Grid 21×21**
25	89.7561	90.7317	93.6585
35	93.6585	95.6098	96.5854
50	95.6098	97.5610	97.5610
60	96.0976	98.0488	98.0488
80	95.6098	98.5366	98.0488
100	95.6098	98.0488	97.0732
120	96.0976	97.0732	97.561
140	96.0976	97.0732	97.561
160	96.0976	97.561	98.0488
180	96.0976	97.561	98.0488
200	96.5854	97.0732	97.561
220	96.5854	97.0732	98.5366
240	97.561	97.561	98.5366
250	96.0976	97.0732	98.0488

Table 3.5: Accuracy (in %) of PCA and PCA+LDA using 1-NNC classifier

No. of features	**PCA**	**PCA+LDA**
10	69.7561	72.1951
20	77.0732	87.3171
30	78.5366	92.6829
40	79.5122	94.6341
50	79.5122	94.6341
60	79.5122	96.5854
70	79.5122	94.6341
80	79.5122	96.0976
90	79.5122	95.6098
100	79.5122	94.6341

It can be seen that the recognition rate for PCA+LDA using Gabor features is higher than PCA or PCA+LDA when the raw pixel values are considered. It can also be seen that, increasing the number of features by using denser lattices helps to acquire more discriminatory features from the images, and consequently provides increased recognition rates. The performance of this technique on the different sizes of lattice sampling is represented in Fig 3.2.

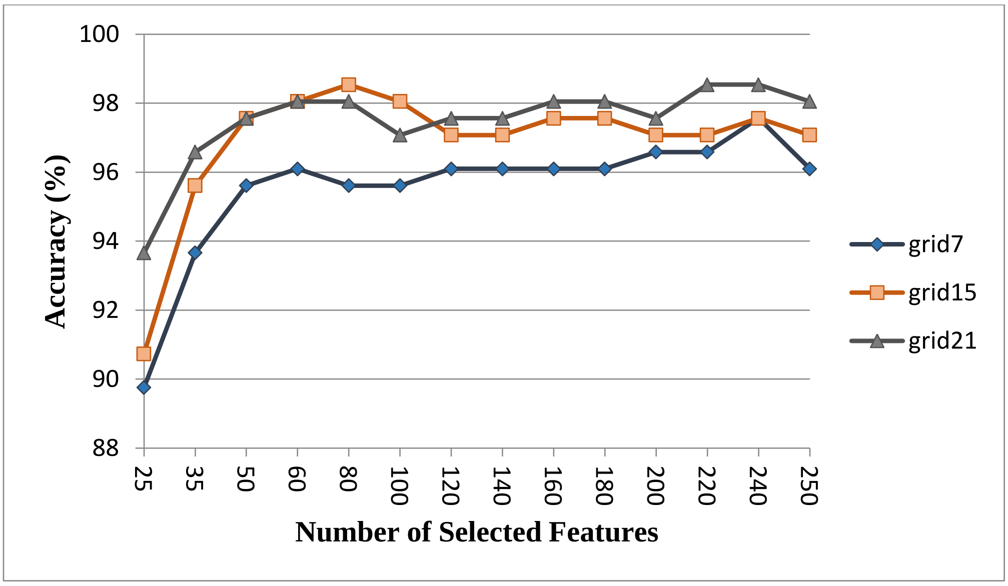

Figure 3.2: The performances achieved using Gabor features and PCA+LDA for 3 different sizes of lattices

3.3 Evaluation of the performance of Best Individual Selection (BIS)

Each feature contains a degree of discrimination ability when considered on its own. Hence, individual evaluation of the features can help to find a subset of individually discriminative features to be employed for recognition. In order to measure the significance of each local feature, the recognition performance of each feature can be considered. As it was explained in subsection 2.3.1, the objective of BIS is to achieve a subset with the best d features by considering the discrimination performance of each local feature when used individually. This method is made up of two parts. In the first part, the discrimination performance of each feature is found individually and top *d* features are selected. Then, the classification is done using these *d* features.

As mentioned earlier, the number of classes is 205 and each class has 4 images. Three images are used for training and the remaining image is used as a test. In order to find the performance of each feature, we considered 3 training images of each class. Two images were used as training images and the remaining image was used as a test. Hence, there are 3 different permutations for each session. In order to determine the performance of each feature, the grid is located over the face images and the performance for each point is computed. For each point, Gabor filter outputs are computed as explained in subsection 2.2.4. This size of the corresponding feature vector is 40. Then, the classification is carried out by 1-NNC and the performance of this point is recorded as the average of 3 possible permutations. This procedure is repeated for all lattice points. Then these lattice points are sorted according to their accuracies. The best *d* lattice points are then selected. Assume that 5 points are selected. After applying these points on all images, the size of final feature vector is computed as $200\ (5 \times 40)$.

Two different schemes are considered for the combination of these lattice points. In feature level approach, as described above, the feature vectors from each sample point are concatenated. Alternatively, model level is studied. In this approach, a classifier is designed for each lattice point and the scores obtained from these points are averaged during testing.

The performance of BIS for 3 different grid sizes is shown in Tables 3.5, 3.6 and 3.7. Also, a comparison of the performances of feature level and model level based BIS for different grid sizes are presented in Figs. 3.3, 3.4, 3.5.

Table 3.6: Accuracy (in %) of BIS for 7 × 7 lattice (49 points)

No. of selected Features	**Feature level**	**Model level**
5	94.1463	88.7805
10	95.1220	93.1707
15	95.6098	95.6098
20	96.0976	96.0976
25	96.5854	96.5854
30	95.6098	96.5854
35	96.0976	96.5854
40	96.5854	97.0732
45	96.5854	96.5854
49	96.0976	97.0732

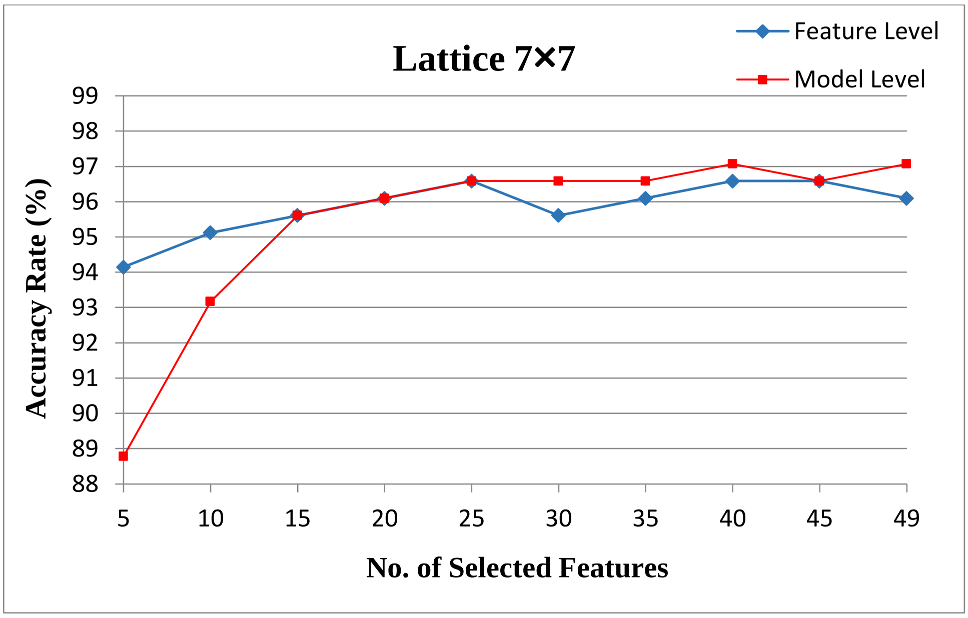

Figure 3.3: The performances of feature level and model level BIS for 7 × 7 lattice

Table 3.7: Accuracy (in %) of BIS for 15 × 15 lattice (225 points)

No. of selected Features	**Feature level**	**Model level**
5	93.6585	92.1951
10	94.6341	94.1463
15	94.1463	95.6098
20	95.1220	95.6098
25	95.6098	95.6098
35	96.5854	96.5854
45	96.0976	97.0732
55	97.0732	97.5610
65	96.0976	97.0732
75	96.0976	97.5610
85	95.6098	97.5610
95	95.6098	98.0488
105	96.0976	98.0488
115	96.0976	98.0488
125	95.6098	98.0488
135	96.0976	98.0488
145	96.0976	98.0488
155	96.0976	98.0488
165	96.0976	98.0488
175	95.6098	98.0488
185	95.6098	98.0488
195	96.0976	97.5610
205	96.0976	97.5610
215	95.6098	97.5610
225	95.6098	97.5610

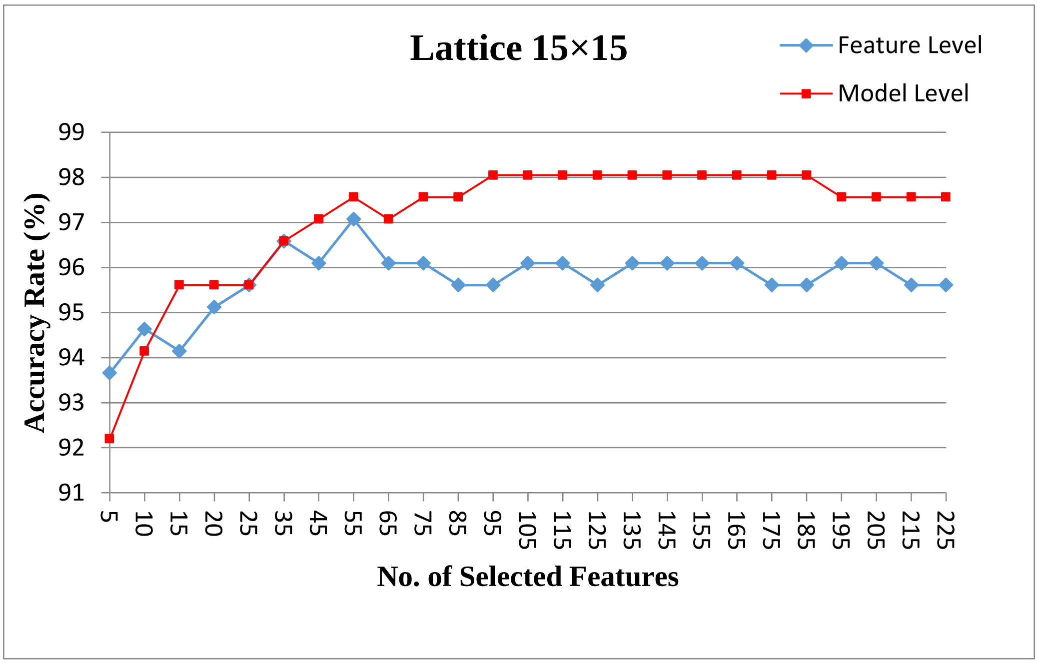

Figure 3.4: The performances of feature and model level BIS for 15 × 15 lattice

Table 3.8: Accuracy (in %) of BIS for 21 × 21 lattice (441 points)

No. of selected Features	**Feature level**	**Model level**
5	94.1463	92.6829
10	93.6585	94.1463
15	94.6341	94.1463
20	94.6341	94.6341
25	94.6341	95.1220
35	95.1220	95.1220
45	94.6341	95.1220
55	93.6585	95.6098
65	93.6585	96.0976
75	94.1463	96.0976
85	94.6341	96.5854
95	96.0976	96.5854
105	96.0976	96.5854
115	96.5854	96.5854
125	96.5854	97.0732
135	96.0976	97.0732
145	96.5854	97.0732
155	96.5854	97.0732
165	96.5854	97.0732
175	97.0732	97.0732
185	97.0732	97.5610
195	97.0732	97.5610
205	96.5854	97.5610
215	96.5854	98.0488
225	96.5854	98.0488
250	96.5854	98.5366
275	96.5854	98.5366
300	96.5854	98.5366
325	96.5854	98.0488
350	96.5854	98.0488
375	96.0976	98.5366
400	96.0976	98.5366
425	96.5854	98.5366
441	96.5854	98.5366

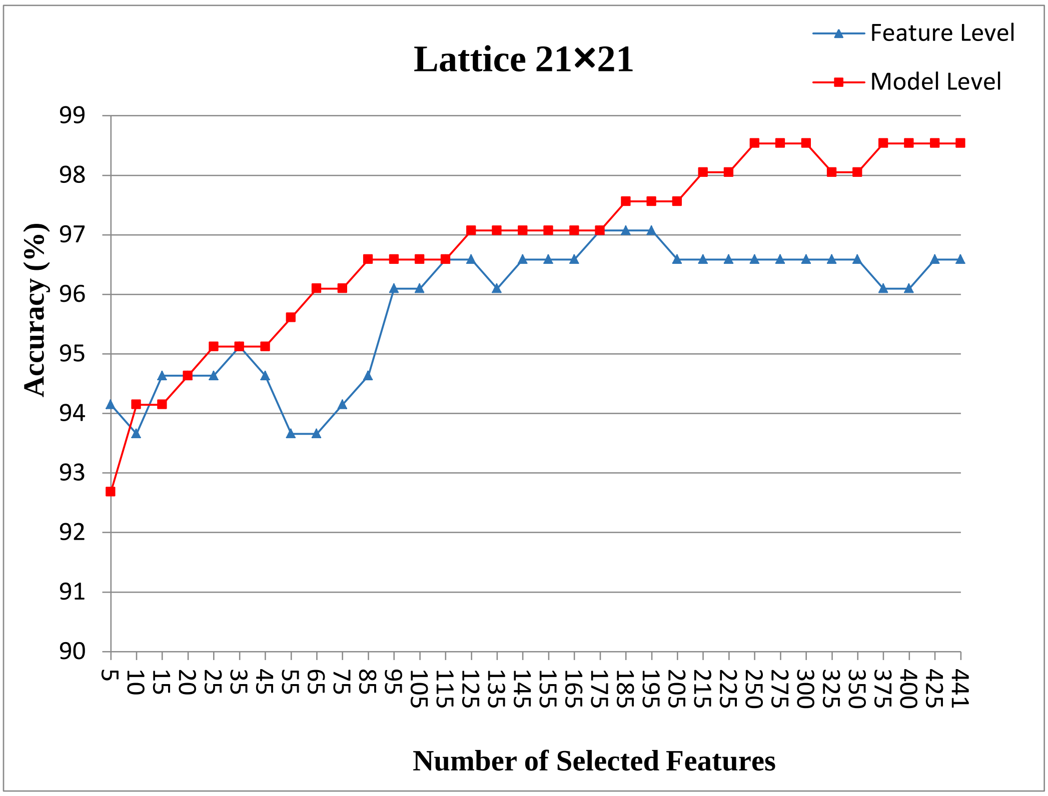

Figure 3.5: The performances of feature and model level BIS for 21 × 21 lattice

3.4 Application of Sequential Forward Selection (SFS)

In most situations, it is better to evaluate each effectiveness of each feature together with the others. Therefore, Sequential Forward Selection is used for this purpose. In this approach, the discrimination performance of each point is evaluated when used together with an existing feature set and the most effective feature was concatenated with the existing set. In order to obtain the best set of features, the first two images of the training set are employed as training images and the third image is used for validation. As explained in subsection 2.3.2, this method started with an empty set. Suppose that we want to choose a good set of 5 features. The first point of the lattice grid is found on all train and test images. Gabor filters are applied on this point and the magnitude of the extracted feature vector is computed for all complex outputs. The size of the obtained feature vector is 40 as before. Then, the classification is accomplished and the accuracy is obtained for each grid point. After finding the performance of all 49 points, these performances were sorted and a feature corresponding to the best performance was added to (S). This procedure was continued for the rest of the features (48). For selecting best performing next feature, we considered the performance of (S) together with each remaining feature. For this purpose, the Gabor filters are applied on these two points and the magnitude of the feature vectors were calculated. The classification is performed and the accuracies obtained are sorted in decreasing order. The best performing pair of points is then selected. This process is continued until 5 grid points are selected. With this selected subset, the classification is accomplished. We considered both feature level and model level combination of features for this method as well. The performance of SFS for 3 different sizes of lattice sampling is shown in Tables 3.8, 3.9 and 3.10. Comparison of the performance of feature level and model level based SFS for different sizes of lattices are presented in Figs. 3.6, 3.7, 3.8.

Table 3.9: Accuracy (in %) of SFS for 7×7 lattice (49 points)

No. of selected Features	**Feature level**	**Model level**
5	95.6098	93.1707
10	95.6098	95.6098
15	95.6098	96.0976
20	95.5854	96.5854
25	96.0976	96.5854
30	95.6098	96.5854
35	95.122	97.561
40	96.0976	97.561
45	96.0976	97.561
49	96.0976	97.0732

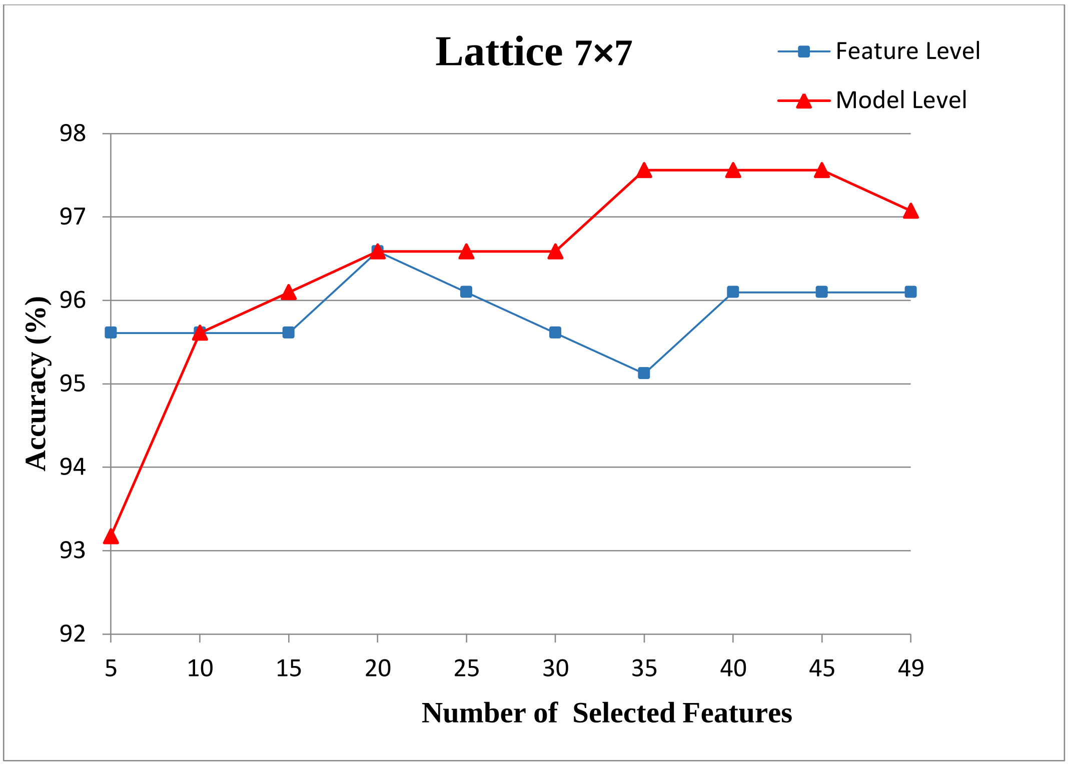

Figure 3.6: The performances of feature level and model level based SFS for 7×7 lattice

Table 3.10: Accuracy (in %) of SFS for 15 × 15 lattice (225 points)

No. of selected Features	**Feature level**	**Model level**
5	95.122	91.2195
10	95.122	94.1463
15	95.122	95.122
20	95.6098	96.0976
25	95.6098	97.0732
35	96.5854	96.5854
45	96.5854	97.0732
55	96.5854	97.561
65	96.0976	97.561
75	95.6098	97.561
85	95.6098	97.561
95	95.6098	97.561
105	95.6098	97.561
115	95.6098	97.561
125	95.6098	97.561
135	95.6098	97.561
145	95.6098	98.0488
155	95.6098	98.0488
165	95.6098	98.0488
175	95.6098	98.0488
185	96.0976	98.0488
195	96.0976	98.0488
205	95.6098	97.561
215	95.6098	97.561
225	95.6098	97.561

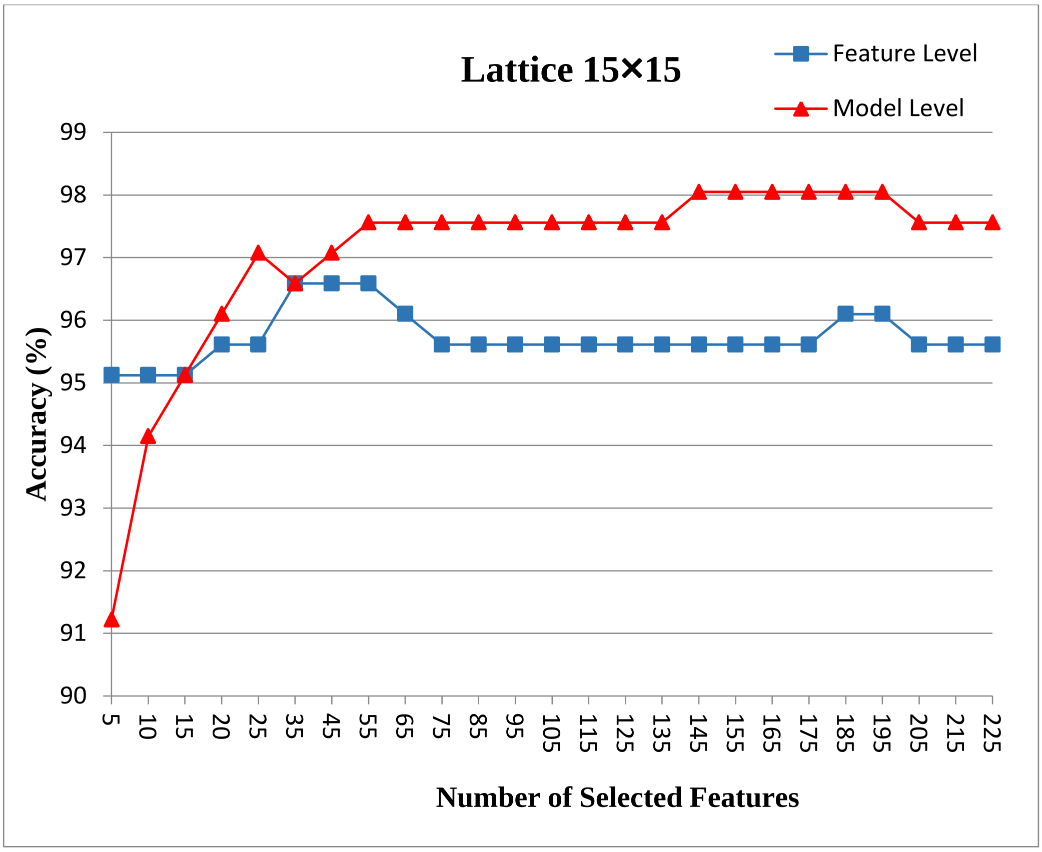

Figure 3.7: The performances of feature level and model level based SFS for 15×15 lattice

Table 3.11: Accuracy (in %) of SFS for 21 × 21 lattice (441 points)

No. of selected Features	**Feature level**	**Model level**
5	9.1463	90.7317
10	95.1219	95.6098
15	95.1219	96.5854
20	95.1219	96.0976
25	95.1219	96.0976
35	95.6098	96.5854
45	96.0976	97.0731
55	96.0976	96.5854
65	96.0976	96.5854
75	96.5854	96.5854
85	96.5854	96.5854
95	95.6098	96.5854
105	96.0976	97.0731
115	96.5854	97.0731
125	96.5854	97.0731
135	96.5854	97.0731
145	96.5854	97.0731
155	96.5854	97.561
165	96.0976	97.561
175	96.0976	97.561
185	96.5854	97.561
195	96.0976	97.561
205	96.0976	97.561
215	96.0976	97.561
225	96.0976	98.0488
250	96.5854	98.0488
275	96.6098	98.0488
300	96.6098	98.5366
325	96.6098	98.0488
350	96.6098	98.0488
375	96.6098	98.5366
400	96.0976	98.5366
425	97.0731	98.5366
441	96.5854	98.5366

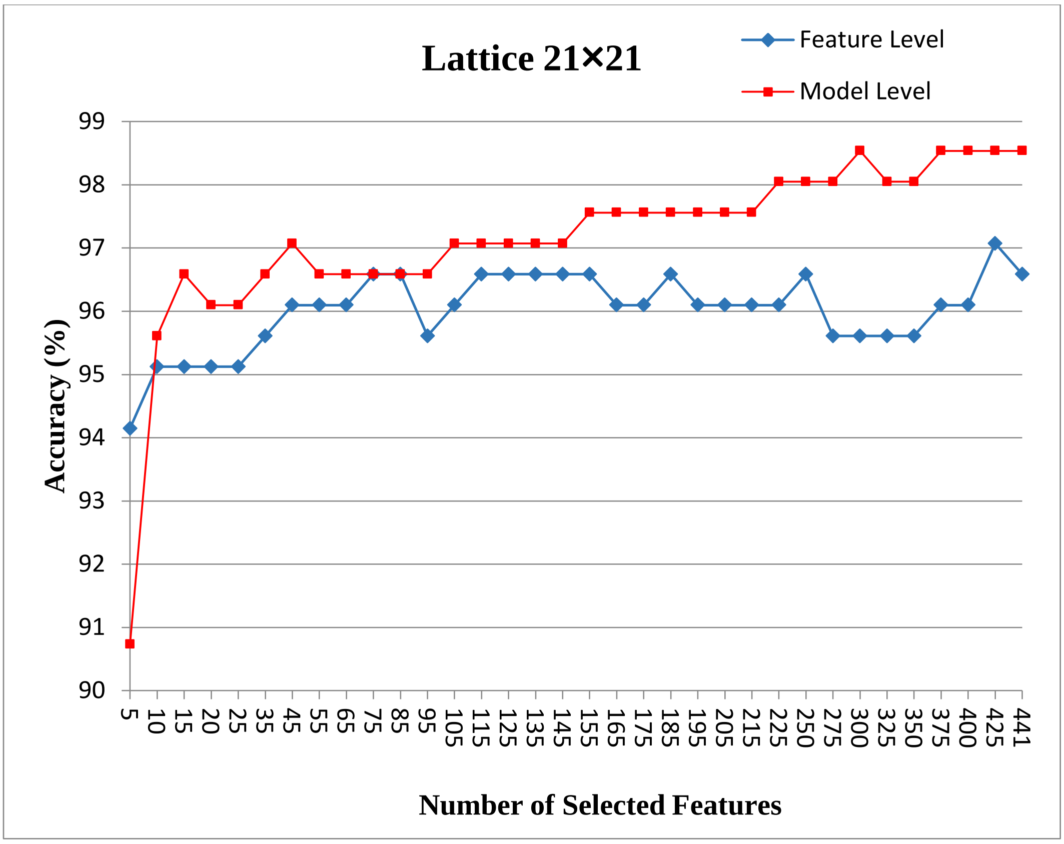

Figure 3.8: The performances of feature level and model level based SFS for 21×21 lattice

The experimental results have shown the model level combination provides better accuracies when large numbers of features are used. Moreover, the selection of a good subset of features is more important in the case of feature level combination since adding more features may lead to reduced accuracies. For instance, in the case of 15 ×15 grid, best accuracy is achieved for 35 features.

Considering Tables 3.5, 3.6 and 3.7, it can be observed that the upper section of the face image such as eyes and eyebrows contain the most discriminative information. Although the lower section of face images also contributes to the performance scores, the upper section is more informative. The comparison of the feature and model level combination schemes for BIS and SFS are presented in Figs. 3.9, 3.10, 3.11, 3.12, 3.13 and 3.14. It can be seen that the performances are comparable in general where SFS can achieve better scores when small number of features are considered.

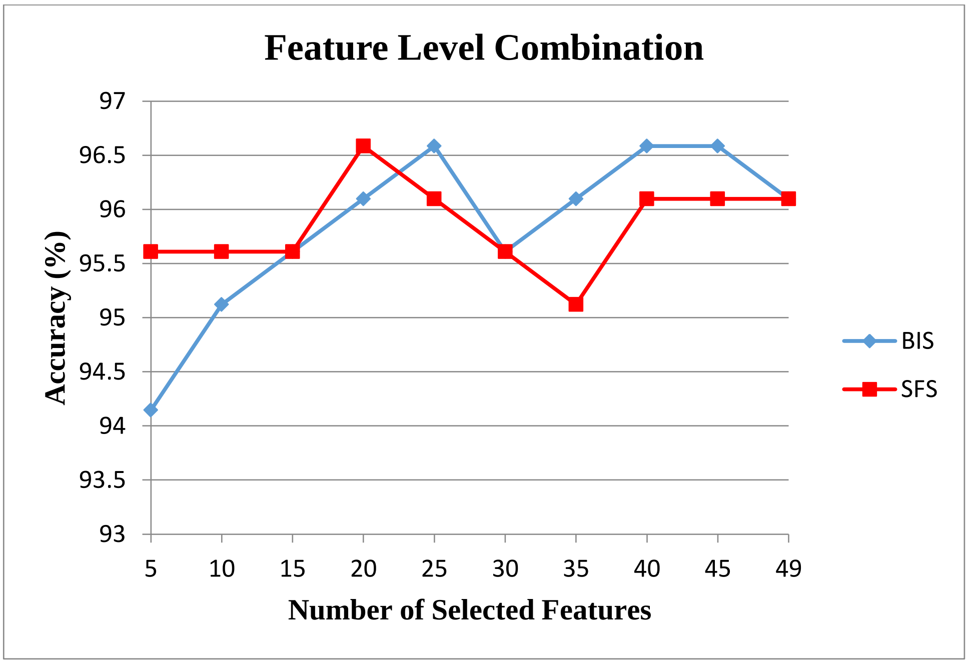

Figure 3.9: The performance of feature level combination for SFS and BIS using 7×7 lattice

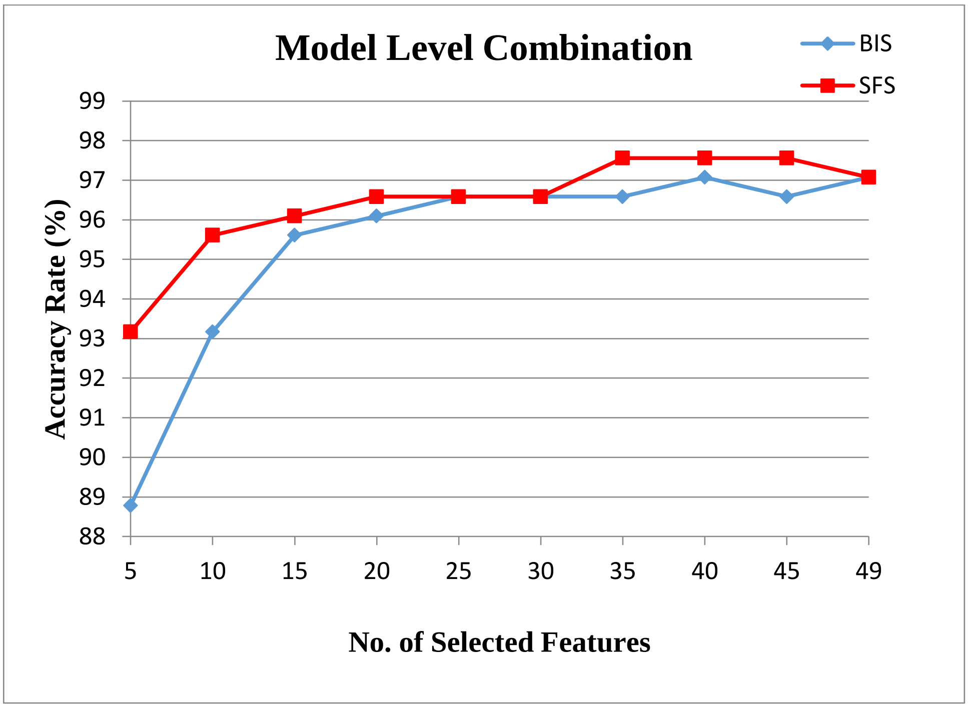

Figure 3.10: The performance of model level combination for SFS and BIS using 7×7 lattice

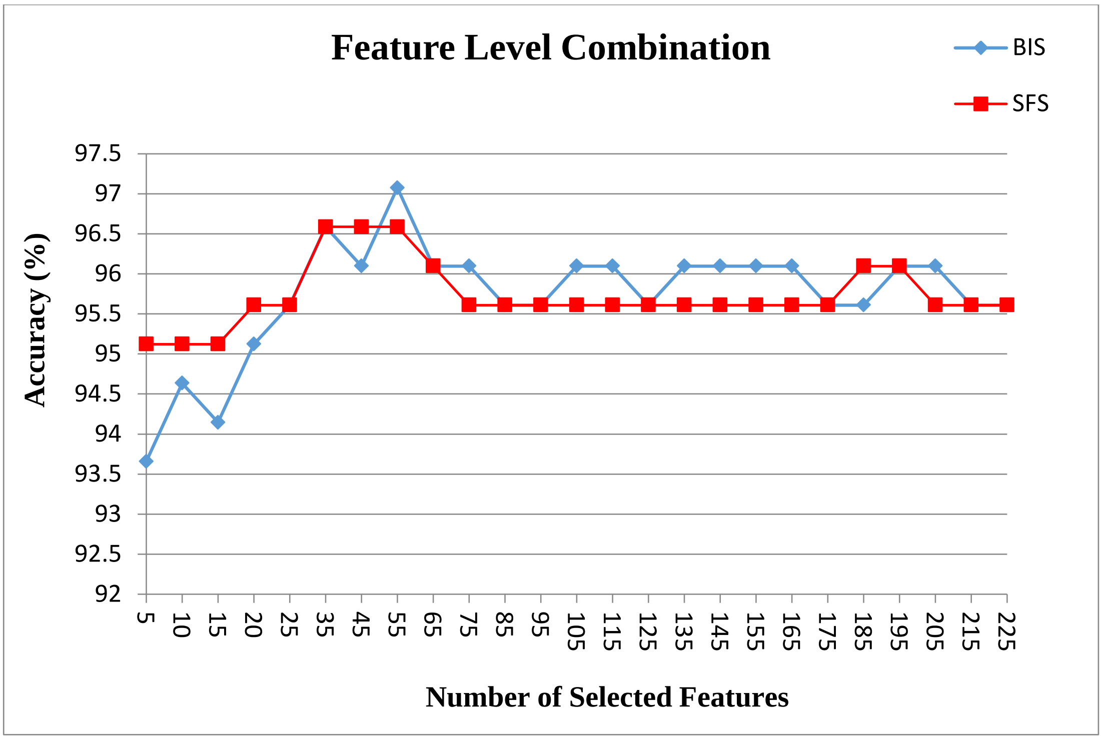

Figure 3.11: The performance of feature level combination for SFS and BIS using 15×15 lattice

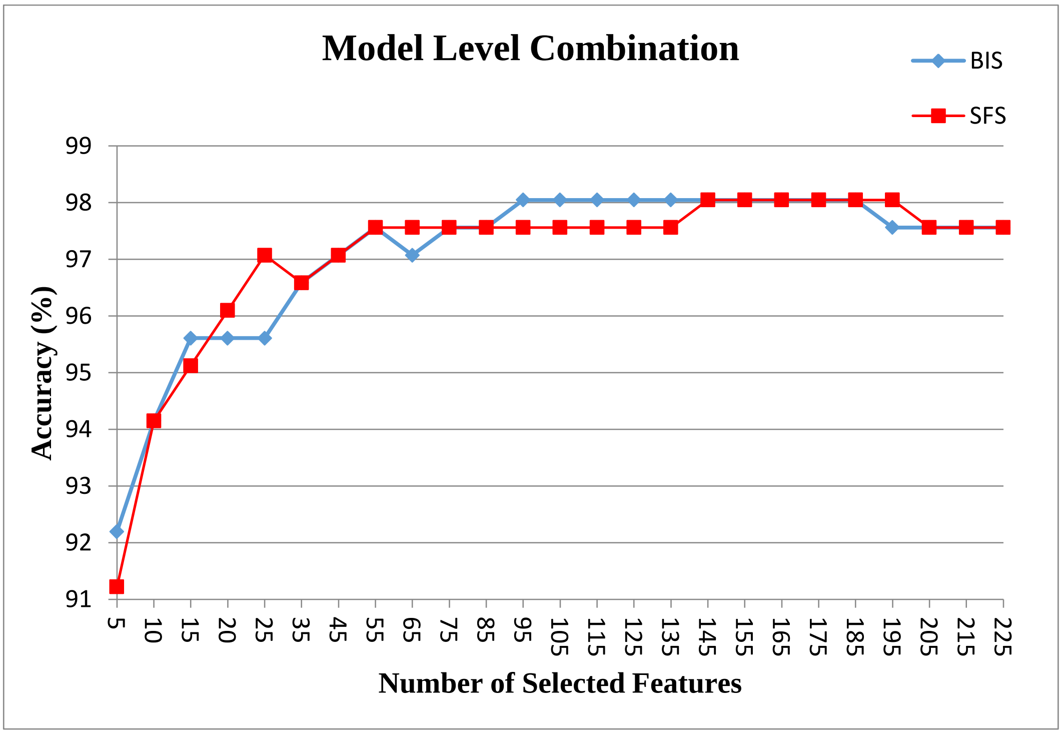

Figure 3.12: The performance of model level combination for SFS and BIS using 15×15 lattice

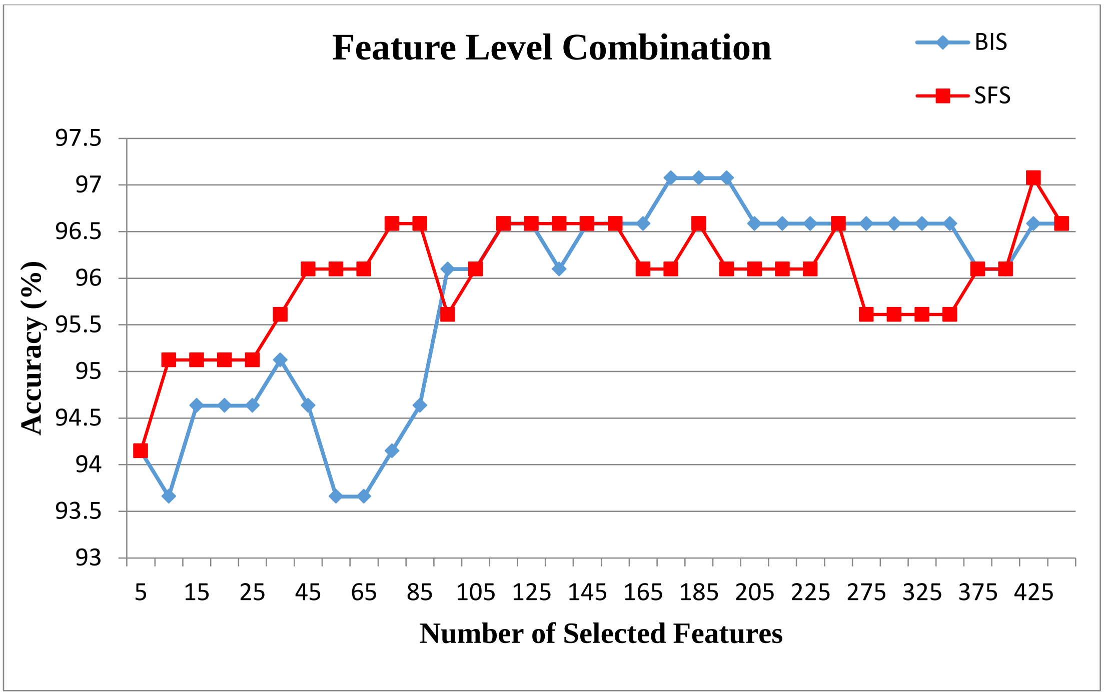

Figure 3.13: The performance of feature level combination for SFS and BIS using 21×21 lattice

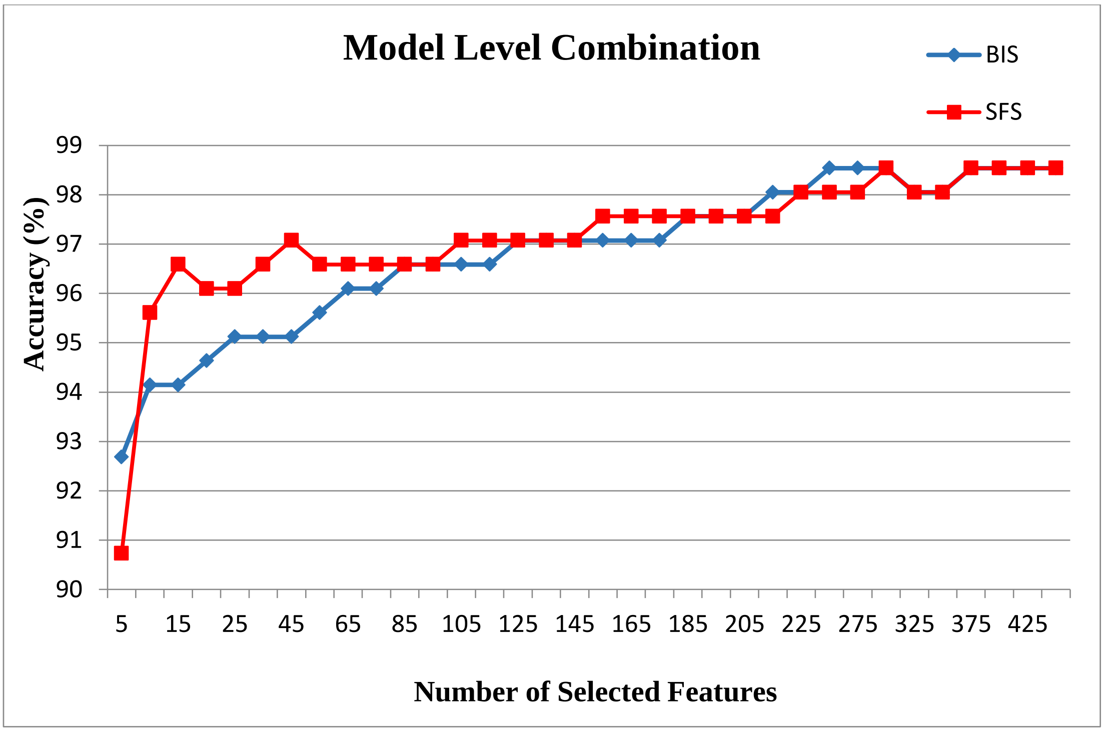

Figure 3.14: The performance of model level combination for SFS and BIS using 21×21 lattice

Chapter 4

CONCLUSION AND FUTURE WORK

In this thesis, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Gabor wavelets are employed for the extraction of features from the facial images. Due to the huge dimensionality of the Gabor feature space, lattice-based selection of a subset of Gabor outputs is considered. A rectangular grid of various sizes is considered and the Gabor filter outputs extracted from the grid points are employed for feature extraction using PCA and LDA. Best Individual Selection (BIS) and Sequential Forward Selection (SFS) are employed for the selection of subsets of features having arbitrary sizes. The combination of features obtained from different grid points are done in both model and feature level. In all simulations, k nearest neighbor classifier is employed as the classification scheme, where $k=1$.

The experiments have been carried out on a subset of 205 people from FERET database. It is observed that the accuracies achieved using the model level combination provides better accuracies than feature level combination when large numbers of features are used. When the best scores are considered, the model level combination scheme leads to better scores for all sizes of grids. Increasing the density of the lattice points is also observed to provide higher accuracies. The performances of feature and model level combination schemes for both BIS and SFS are also compared. It is observed that the performances are comparable in general where SFS can achieve better scores when small number of features is considered.

Larger number of lattice points provides higher scores. It can be argued that this is mainly due to extracting more information, especially from discriminative regions. As an alternative approach, the use of dense sampling only at a priori defined landmark points should be considered. This will help to avoid employing redundant features, leading to decreased computational complexity.

Since the use of more features generally improves the accuracy, the use of backward selection should also be considered for model based combination. It should be noted that, in this thesis, the accuracies are reported for the test samples. In practice, choosing the best number of features using the training data is necessary. This requires cross-validation on the training data. This task should also be considered as a future work.

REFERENCES

[1] I.S. Virk & R. Maini. (2012). Biometric Authentication System: Tools and Techniques. *International Journal of Computer Application*, vol.2, no.2, pp. 150-163.

[2] K. Dharavath, F.A. Talukdar, & R.H. Laskar. (2013). Study on Biometric Authentication Systems, Challenges and Future Trends: A Review. *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-7.*

[3] V. Arulalan, G. Balamurugan, & V. Premanand. (2014). A Survey on Biometric Recognition Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, vol.3, no.2, pp. 5708-5711.

[4] R. Jafri & H.R. Arabnia. (2009). A Survey of Face Recognition Techniques. *Journal of Information Processing Systems*, vol.5, no.2, pp. 41-67.

[5] S. Anila & N. Devarajan. (2012). Preprocessing Technique for Face Recognition Applications Under Varying Illumination Conditions. *Global Journal of Computer Science and Technology Graphics & Vision*, vol.12, no.11, pp. 12-18.

[6] S. Shan, W. Gao, B. Cao, & D. Zhao. (2003). Illumination Normalization for Robust Face Recognition Against Varying Lighting Conditions. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures,* pp. 157-164.

[7] V. Struc, J. Zibert, & N. Pavesic. (2009). Histogram Remapping as a Preprocessing Step for Robust Face Recognition. *Waves transaction on information science and applications, vol.6, no.3,* pp. 520-529.

[8] B. Du, Sh. Shan, L. Qing, & W. Gao.(2005). Empirical Comparisons of Several Preprocessing Methods for Illumination Insensitive Face Recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings. (ICASSP '05),* pp. ii/981 - ii/984.

[9] M.V. Santamarıa & R.P. Palacios. (2004). Comparison of Illumination Normalization Methods for Face Recognition. pp. 27-30.

[10] A.K. Jain, R.P.W. Duin, & J. Mao. (2000). Statistical Pattern Recognition:A Review.*IEEE Trans.Pattern Analysis and Machine Intelligence,* vol.22, no.1, pp. 4-37.

[11] K.M. Lam & H. Yan. (1998). an Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 673-686.

[12] M. Bicego, A.A. Salah, E. Grosso, M. Tistarelli, & L.Akarun. (2007). *Generalization in Holistic versus Analytic Processing of Faces. 14th International Conference on Image Analysis and Processing (ICIAP),* pp. 235-240.

[13] R. Upadhayay & R.K. Yadav. (2013). Kernel Principle Component Analysis in Face Recognition System: A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 6, pp. 348-353.

[14] M. Turk & A. Pentland. (1991). Face Recognition Using Eigen faces. Proc. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586-591.

[15] X. Wang & X. Tang. (2003). Unified Subspace Analysis for Face Recognition. *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, pp. 679-686.

[16] T. Verma & R.K. Sahu. (2013). PCA-LDA Based Face Recognition System & Results Comparison by Various Classification Techniques. *Proceedings of 2013 International Conference on Green High Performance Computing*, pp. 1-7.

[17] P.N. Belhumeur, J.P. Hespanha, & D.J. Kriegman. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720.

[18] Z. Lai, C. Zhao, & M. Wan. (2012). Fisher Difference Discriminant Analysis: Determining the Effective Discriminant Subspace Dimensions for Face Recognition. *Neural Processing Letters*, vol.35, no.1, pp. 203-220.

[19] M.Visani, C.Garcia, & J.M.Jolion. (2006). Tow-Dimensional-Oriented Linear Discriminant Analysis for Face Recognition. *Proceedings International Conference on Computer Vision and Graphics (ICCVG)*, pp. 1008-1017.

[20] P. Navarrte & J. Ruiz-del-Solar. (2001).Eigenspace-based Recognition of Faces: Comparisons and a new Approach. *Proceedings.11th IEEE International Conference on Image Analysis and Processing (ICIAP)*, pp. 42-47.

[21] H.B. Deng, L.W. Jin, L.X. Zhen, & J.C. Huang. (2005). A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA. *International Journal of Information Technology*, vol. 11, no. 11, pp. 86-96.

[22] M. Meade, S.C. Sivakumar, & W.J. Phillips. (2005).Comparative Performance of Principal Component Analysis, Gabor wavelets and Discrete wavelet transforms for Face Recognition. *Canadian Journal of Electrical and Computer Engineering*, Vol.30, No.2, pp. 93-102.

[23] Y.Ch. Lee & C.H. Chen. (2008). Face Recognition Based on Gabor Features and Two-Dimensional PCA. *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 572-576.

[24] T. Barbu. (2010). Gabor Filter-Based Face Recognition Technique. *Proceedings of the Romanian Academy*, Series A, vol.11, no.3, pp. 277–283.

[25] J.Z. Mang, M.I. Vai, & P.U. Mak. (2004). Gabor Wavelets Transform and Extended Nearest Feature Space Classifier for Face Recognition. *Proceedings of the Third International Conference on Image and Graphics (ICIG'04)*, pp. 246-249.

[26] E. Naz, U. Farooq, & T. Naz. (2006). Analysis of Principal Component Analysis-Based and Fisher Discriminant Analysis-Based Face Recognition Algorithms. *Second International Conference on Emerging Technologies*, pp. 121-127.

[27] W. Li & W. Cheng. (2008). Face Recognition Based on Adaptively Weighted Gabor-LDA. *Fourth International Conference on Natural Computation*, pp. 130-134.

[28] S. Shan, W. Gao, Y. Chang, B. Cao, & P. Yang. (2004). Review the Strength of Gabor Features for Face Recognition from the Angle of its Robustness to Mis-alignment. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, pp. 338-341.

[29] C. MageshKumar, R. Thiyagarajan, S.P. Natarajan, S. Arulselvi, & G. Sainarayanan. (2011). Gabor features and LDA based Face Recognition with ANN Classifier. *International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT)*, pp.831-836.

[30] B. Gokberk, M.O. Irfanoglu, L. Akarun, & E. Alpaydın. (2007). Learning the Best of Local Features for Face Recognition. *The Journal of the Pattern Recognition Society*, vol.40, no.1, pp. 1520-1532.

[31] W. Dai, Y. Fang, & B. Hu. (2011). Feature Selection in Interactive Face Retrieval. *4th International Congress on Image and Signal Processing (CISP)*, pp. 1358-1362.

[32] B. Gokberk, M.O. Irfanoglu, L. Akarun, & E. Alpaydın. (2003). Optimal Gabor Kernel Location Selection for Face Recognition. *Proceedings International Conference on Image Processing (ICIP)*, pp. 77-80.

[33] A. Jain and D. Zongker. (1997). Feature Selection: Evaluation, Application, & Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no.2, pp. 153-158.

[34] M.Kudo, J.Sklansky. (2000). Comparison of Algorithms that Select Features for Pattern Classifiers. *The Journal of the Pattern Recognition Society*, vol.33, no.1, pp. 25-41.

[35] L. Ladha & T. Deepa. (2011). Feature Selection Methods and Algorithms. *International Journal on Computer Science and Engineering*, vol.3, no.5, pp. 1787-1797.

[36] P. Viswanath & T.H. Sarma. (2011). An Improvement to k-Nearest Neighbor Classifier. *Recent Advances in Intelligent Computational Systems (RAICS)*, pp. 227-231.

[37] R. Souza, R. Lotufo, & L. Rittner. (2012). A Comparison between Optimum-Path Forest and k-Nearest Neighbors Classifiers. 25th Conference on Graphics, *Patterns and Images (SIBGRAPI)*, pp. 260-267.

[38] X. Wang, Z. Chen, & Z. Lin. (2013). Class-nearest Neighbor Classifier for Face Recognition. *International Conference on Computer Sciences and Applications*, pp. 325-328.

[39] P.J. Phillips, H. Moon, S.A. Rizvi, & P.J. Rauss. (2000). The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans. *Pattern Analysis and Machine Intelligence*, vol.22, no.10, pp. 1090-1104.