FUSING LOCAL APPEARANCE MODELS FOR FACE RECOGNITION

by

Nuri Murat Arar

B.S. Computer Engineering, Bilkent University, 2010

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2012

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis supervisors Prof. Lale Akarun and Assist. Prof. Hazım Kemal Ekenel for their insight, guidance and endless support. I would also like to express my deepest gratitude to my mentor, Hua Gao, for his inspiring ideas and endless help.

I am grateful to Prof. Bülent Sankur, Prof. Fikret Gürgen and Assist. Prof. Arzucan Özgür for participating in my thesis jury, their criticism and valuable feedback have greatly improved this thesis.

I would like to thank the members of MediaLab and PILAB for providing an enjoyable working environment, and especially to Neşe Alyüz and Yunus Emre Kara for their support throughout the thesis.

I would especially like to present my gratitude to my friends, Belma Engin, Beyza Ermiş, Burak Şişman, D. Sinem Yıldırım, M. Gülbey Yanar, Mesut Öner, N. Çağrı Kılıboz, N. Kaan Bekmezci, Safiye Çelik, Sevgiye Sönmez and Tunahan Yıldırım for their valuable friendship and continuous support.

I would like to express my deepest gratitude to my parents Emine Arar and H. Ali Arar, and my brother Ş. Şamil Arar for their endless love, support and encouragement throughout my life.

I am specifically grateful to Fatma Güney. I honestly could not bear the burden without her support and motivation.

Lastly, I would like to offer my gratitude to the Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting me with National Scholarship Programme for M.Sc. Students - 2210.

ABSTRACT

FUSING LOCAL APPEARANCE MODELS FOR FACE RECOGNITION

Face recognition is a popular research area due to its scientific challenges and potential applications. Therefore, it attracts attention from both diverse research communities and the industry for several years. Various techniques have been intensively investigated to provide a robust face recognition system. Many of these techniques have already achieved very high recognition accuracies under controlled conditions. On the other hand, face recognition under uncontrolled conditions is still a very hard problem. The difficulty arises from facial appearance variations caused by various factors, such as expression, illumination and partial occlusion, and the time gap between training and testing data capture. Face recognition based on local features usually outperforms holistic approaches because local representation is less sensitive to appearance variations caused by occlusions and facial expressions. In this thesis, a local appearance based face recognition algorithm, which works reliably under real-world conditions, is proposed. The proposed algorithm uses different local appearance models to represent face images. Fundamentally, it exploits Gabor features that have been extensively used for facial image analysis due to their powerful representation capabilities. It utilizes curvature Gabor features in addition to conventional Gabor features. The system focuses on selecting and combining multiple Gabor classifiers. The final Gabor classifier is obtained by LLR-based fusion of classifiers that are selected using SFFS-based classifier selection algorithm. In addition, the system uses DCT features as extra evidence. Finally, classifiers trained on different local representations are combined at score-level by PLSR-based fusion. The system is evaluated on FRGC version 2.0 Experiment 4, and achieves 94.16% verification rate @ 0.1% FAR, which is the highest accuracy reported on this experiment so far in the literature.

ÖZET

YÜZ TANIMA İÇİN YEREL GÖRÜNÜM MODELLERİNİN TÜMLEŞTİRİLMESİ

Yüz tanıma, bilimsel zorlukları ve olası uygulamaları dolayısıyla hem çeşitli araştırma toplulukları hem de piyasa tarafından yıllardır ilgi gören, popüler bir araştırma alanıdır. Gürbüz bir yüz tanıma sistemi sağlamak için çeşitli teknikler yoğun olarak incelenmiştir. Bu tekniklerin bir çoğu, kontrollü koşullar altında hali hazırda çok yüksek tanıma başarısı elde etmiştir. Buna karşılık, kontrolsüz koşullar altında yüz tanıma, hala çok zorlu bir problemdir. Bu zorluk, ifade ve aydınlanma değişimi, kısmi örtüşme, kullanılan eğitim ve test verileri arasındaki zaman farkı gibi etmenlerden kaynaklanmaktadır. Yerel öznitelik tabanlı yüz tanıma, örtüşmelerden ve ifadelerden kaynaklanan görünüm değişimlerine daha az hassas olduğu için bütünsel tabanlı yaklaşımlardan daha iyi sonuçlar göstermiştir. Bu tezde, gerçek-dünya koşulları altında güvenli bir şekilde çalışan yerel görünüm tabanlı bir yüz tanıma algoritması önerilmektedir. Önerilen algoritma, yüz imgelerini temsil etmek için farklı yerel görünüm modelleri kullanmaktadır. Sistem yüz imgelerini güçlü bir biçimde temsil etmeleri dolayısıyla yüzsel imge analizinde yaygın olarak kullanılan Gabor özniteliklerinden faydalanmaktadır. Klasik Gabor özniteliklerinin yanı sıra eğrisel Gabor özniteliklerinden de yararlanılmıştır. Sistem çoklu Gabor sınıflandırıcılarının seçilmesine ve birleştirilmesine odaklanmıştır. Son büyük Gabor sınıflandırıcısı, SFFS tabanlı sınıflandırıcı seçme algoritması kullanılarak seçilen sınıflandırıcıların LLR tabanlı tümleştirme ile birleştirilmesiyle elde edilmiştir. Ek olarak, sistem DCT özniteliklerini ekstra bulgu olarak kullanmaktadır. Son olarak, farklı yerel temsiller üzerinde eğitilen sınıflandırıcılar, PLSR tabanlı tümleştirme ile skor seviyesinde birleştirilmiştir. Sistem FRGC veritabanı 2.0 sürümü Deney 4 üzerinde test edildiğinde %0.1 yanlış kabul oranınında %94.16 doğrulama oranı yakalayarak literatürde şimdiye kadar raporlanan en yüksek başarımı elde etmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

A_1	The scaling component subband image
B_i	Input image block in tree representation of one level, 2D wavelet decomposition
$B_{i,H}$	High frequency content in tree representation of one level, 2D wavelet decomposition
$B_{i,L}$	Low frequency content in tree representation of one level, 2D wavelet decomposition
c	Curvature ratio
$C(u, v)$	2D DCT transform
CG	A CG classifier
CG_i	ith CG classifier
d_{ncc}	Normalized cross correlation as distance metric
D_1	Diagonal subband image
$e^{(ik_{\nu,\mu}\vec{z})}$	Oscillatory wave function
f_0	Central frequency of the sinusoidal signal
f	Spacing factor between kernels in the frequency domain
f_{test}	Test feature vector
$f_{training}$	Training feature vector
$f(x, y)$	Inverse transform of DCT
H_1	Horizontal subband image
H	High-pass filtering
J	Total number of selected CG classifiers
$k_{\nu,\mu}$	Wave vector
k_{max}	Maximum frequency
L	Low-pass filtering
P	The dependent variable in PLSR analysis
V_1	Vertical subband image
$VR_{current}$	Current Verification Rate
VR_{new}	New Verification Rate

X	Input data matrix in PLSR analysis
Y	Set of selected CG classifiers
z	Z-normalization
$\downarrow 2$	Downsampling by a factor of two
$\|\| \cdot \|\|$	Norm operator
α	Sharpness of Gaussian
γ	Constant ratio in normalized Gabor function
γ_n	nth block classifier
Γ	A CG classifier
Γ^+	CG classifier that maximizes the verification rate when combined with the current set of classifiers
Γ^-	CG classifier that maximizes the verification rate when excluded from the current set of classifiers
Γ_{SS}	A CG classifier which obtained by fusion of block classifiers using simple sum
$\Gamma_{SS,j}$	jth CG classifier which obtained by fusion of block classifiers using simple sum
Γ_{LLR}	A CG classifier which obtained by LLR-based fusion of block classifiers
$\Gamma_{LLR,j}$	jth CG classifier which obtained by LLR-based fusion of block classifiers
Γ_{SS+LLR}	A CG classifier, which is obtained by simple sum fusion of block classifiers, after LLR-based score conversion is applied
$\Gamma_{SS+LLR,j}$	jth CG classifier, which is obtained by simple sum fusion of block classifiers, after LLR-based score conversion is applied
Λ_{LLR+SS}	Simple sum fusion of selected CG classifiers that are obtained by LLR-based fusion of block classifiers
Λ_{SS+LLR}	LLR-based fusion of selected CG classifiers that are obtained by simple sum fusion of block classifiers
μ	Orientation of Gabor wavelet
$\mu_{different}$	Mean of Gaussian density $p(\Gamma\|different)$
μ_{match}	Mean of the scores of matched pairs in X
$\mu_{nonmatch}$	Mean of the scores of nonmatched pairs in X

μ_{same}	Mean of Gaussian density $p(\Gamma\|same)$
ν	Scale of Gabor wavelet
σ	Gaussian size
μ_{match}	Variance of the scores of matched pairs in X
$\mu_{nonmatch}$	Variance of the scores of nonmatched pairs in X
$\Sigma_{different}$	Variance of Gaussian density $p(\Gamma\|different)$
Σ_{same}	Variance of Gaussian density $p(\Gamma\|same)$
τ_n	Score matrix of nth block classifier
$\phi(f)$	Fourier transform of Gabor function
$\varphi(t)$	General form of Gabor signals
$\psi(\vec{z};\nu,\mu)$	2D Gabor Wavelet function
$\psi(\vec{\tilde{z}};\nu,\mu)$	Curvature Gabor wavelets
$\Psi(Y)$	Evaluation function for set of classifiers, Y
Ω	Score matrix of CG classifier
Ω^*	Score matrix of CG classifier that maximizes the verification rate
Ω_{SS}	Score matrix of a CG classifier which obtained by fusion of block classifiers using simple sum
$\Omega_{SS,j}$	Score matrix of jth CG classifier which obtained by fusion of block classifiers using simple sum
Ω_{LLR}	Score matrix of a CG classifier which obtained by LLR-based fusion of block classifiers
$\Omega_{LLR,j}$	Score matrix of jth CG classifier which obtained by LLR-based fusion of block classifiers
Ω_{SS+LLR}	Score matrix of a CG classifier, which is obtained by simple sum fusion of block classifiers, after LLR-based score conversion is applied
$\Omega_{SS+LLR,j}$	Score matrix of jth CG classifier, which is obtained by simple sum fusion of block classifiers, after LLR-based score conversion is applied

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AC	Alternating Current (Coefficient)
BEE	Biometric Experimentation Environment
CFA	Class-dependence Feature Analysis
CG	Curvature Gabor
DC	Direct Current (Coefficient)
DCT	Discrete Cosine Transform
EBGM	Elastic Bunch Graph Matching
ECG	Extended Curvature Gabor
FAR	False Acceptance Rate
FLD	Fisher's Linear Discriminant
FRGC	Face Recognition Grand Challenge
FRR	False Rejection Rate
GFC	Gabor-Fisher Classifier
HOG	Histogram of Oriented Gradients
ICA	Independent Component Analysis
KDE	Kernel Density Estimation
KFA	Kernel Fisher Analysis
KLT	Karhunen-Loeve Transform
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LFA	Local Feature Analysis
LFW	Labeled Faces in the Wild
LGFC	Local Gabor Fisher Classifier
LLR	Log-likelihood Ratio
MAP	Maximum a Posteriori
NIST	National Institute of Standards and Technology

PCA	Principal Component Analysis
PCLDA	PCA followed by LDA
PLS	Partial Least Square
PLSR	Partial Least Square Regression
ROC	Receiver Operator Characteristics (Curve)
SFFS	Sequential Forward Floating Selection/Search
SFS	Sequential Forward Selection/Search
XML	Extensible Markup Language

1. INTRODUCTION

Recognition of a human face has received a great deal of attention from a wide variety of fields, from psychology and neuroscience to computer science over the last 25 years [4]. Humans often use faces to recognize individuals, and therefore, the main purpose of face recognition is to identify or to verify humans from their facial data in the same manner as humans do.

The motivation of the common interest in face recognition for computer scientists arises from both the scientific challenges of the problem and the abundance of potential applications. With the advent of video surveillance, smart cards, human computer interfaces and electronic services such as e-banking, e-commerce and e-home, and an increased emphasis on the privacy and security of information stored in various databases, automatic personal identification / verification has become a very important topic. Accurate automatic personal identification / verification is now needed in a wide range of civilian applications.

A perfect identity authentication system needs a biometric component. A biometric is a representation of a unique part or characteristic of an individual which has the potential to distinguish between an authorised person and an impostor. Since biometric characteristics are distinctive, cannot be forgotten or lost, and the person to be authenticated needs to be physically present at the point of identification, biometrics are inherently more reliable and more capable than traditional knowledge-based and token-based techniques. Currently, there are many biometric technologies used for personal authentication such as face, fingerprint, hand geometry, iris, retina, signature and voice. Despite the fact that fingerprint or iris scanning methods can be more accurate, they are not appropriate for natural interactions due to their intrusive nature. Thus, face recognition presents an attractive alternative because the person to be identified does not need to cooperate or take any specific action. This makes face recognition technology a perfect match for natural interaction applications. It can work unobtrusively in the background without disturbing or interrupting the subjects to be

identified. However, there is a major difficulty in face recognition that the performance of most systems is easily affected by many factors because the appearance of a human face has potentially very large variations due to head pose, illumination changes, facial expression changes, partial occlusions and aging.

The major focus of this thesis is to develop a reliable automatic face recognition system, which is robust against real-world conditions. For overcoming some of the mentioned problems caused by uncontrolled conditions, we propose a local appearance models-based face recognition framework because local representations are less sensitive to changes in the appearance of the person and partial occlusions. The proposed framework exploits Gabor features that have been recognized as one of the most successful local feature extraction methods for face representation, which are also interesting due to their biological relevance. As an important contribution, it utilizes curvature Gabor features in addition to conventional Gabor features. In other words, it uses various parameter combinations in order to acquire complementary information from different curvature and scale spaces. This comprehensive feature extraction process provides a well established face representation, and improves the robustness of the proposed approach against the previously mentioned problems. In order to fully take advantage of this feature extraction framework and the resulting classifiers, SFFS-based classifier selection algorithm is employed. Then, selected classifiers are combined by LLR-based fusion to form the final CG classifier. In addition, as extra evidence, the system uses DCT features that have been extensively used together with Gabor features due to the complementariness between these two local representations. Finally, classifiers trained on different local representations are combined at score-level by PLSR-based fusion. The proposed framework is evaluated on FRGC version 2.0 Experiment 4, which is a very challenging experiment that requires matching of images collected under controlled and uncontrolled conditions, and it achieves high verification performance.

1.1. Terminology

This section describes some terminology widely used in the face recognition studies for clarification. Face recognition systems fall into two categories: “verification” and

"identification". The term "recognition" is used to denote both categories. Face verification is a $1:1$ match that compares a face image against a single face image, whose identity is being claimed. On the contrary, face identification is a $1:N$ problem that compares a face image against all images in a face database to determine the identity of the query face. The term "gallery", or sometimes referred to as "target set", is used to denote a set of faces with known identities. The faces that are to be identified against the gallery are referred to as the "probe", "query" or "test". In addition, the terms "impostor" and "genuine" are used to mean non-matching and matching query-target image pairs, respectively. There are also terms extensively used in evaluations. For example, Receiver Operating Characteristics (ROC) curves are commonly mentioned. A ROC curve is a visual characterization of the percentage of falsely rejected queries as a function of the falsely accepted queries. False Acceptance Rate (FAR) is the probability that the system incorrectly matches the query to a non-matching target in the database. It measures the percent of queries which are incorrectly accepted. False Rejection Rate (FRR) is the probability that the system fails to detect a match between the query and a target in the database. It measures the percent of queries which are incorrectly rejected.

1.2. 2D Face Recognition Literature Review

This chapter gives an overview of the related work that has been conducted on 2D face recognition. Recent surveys on 2D face recognition literature are given in [5] by Jafri *et al.*, and in [4] by Zhao *et al.* Following from these surveys, we categorize previous studies into three main types by considering mainly the face representation methodology. In Section 1.2.1, global-based face representation approaches that use holistic characteristics of the face are explained. Then, local-based face representation methods are described in Section 1.2.2. Finally, hybrid approaches that use and combine holistic and local features, are explained in Section 1.2.3.

1.2.1. Holistic Methods

There have been many generic face recognition algorithms proposed. Among these, there are two algorithms that have had a very large impact on the face recognition research community and they have inspired countless studies. These are eigenfaces [1], and Fisherfaces [6].

The eigenfaces approach is the most well-known face recognition algorithm [1]. In the algorithm, first a face subspace is constructed from training face images using the Principal Component Analysis (PCA). The face images are then represented with the eigen vectors, also called eigenfaces, of the covariance matrix of the images. To achieve dimensionality reduction, by considering the proportion of the obtained variance over the total variance, only a subset of the eigenvectors are used to represent the face images. Therefore, the selected eigenvectors are the ones having the highest eigenvalues for the variance to be maximum. This way the face images can be reconstructed with the smallest mean-square error for any given subspace dimensionality. The classification is done by comparing the face images in this subspace.

Figure 1.1. Input face images (top row) and their calculated eigenfaces (bottom row) [1].

Later, some other holistic and local methods were proposed extending the eigenfaces approach. For example, [7] introduces the view-based eigenfaces approach to handle head pose variations. It constructs an eigenspace for each different view of an

individual. When a test image arrives, at first the eigenspace that can best represent the view of the face image is determined and then classification is done in that eigenspace. On the other hand, nonlinear extensions of the eigenfaces algorithm via kernel methods have been also studied in [8–10]. In these approaches, the input face space is first mapped into a higher dimensional space by using nonlinear functions such as a polynomial kernel. The PCA is performed in this higher dimensional new space.

Fisherfaces is another well-known face recognition approach [6]. This is also a subspace-based algorithm. It uses Linear Discriminant Analysis (LDA) for subspace projection. The aim of the LDA is to extract the projection directions that are effective for discrimination. More specifically, it searches for the best subspace by also utilizing class information. Unlike PCA which maximizes the total scatter across all classes, LDA maximizes the ratio of between-class scatter over within-class scatter, and is thus purportedly better for classification.

Figure 1.2. Sample eigenfaces (top row) showing the tendency of the principal components to capture major variations such as lighting direction and their corresponding fisherfaces (bottom row) showing the ability to discount those factors unrelated to classification [2].

The Fisherfaces approach also has extensions. In [11], a LDA mixture model is used. For each cluster in the face space, a separate LDA is performed. In [10], a nonlinear extension of fisherfaces via kernel methods have been also studied.

The third subspace-based algorithm is Independent Component Analysis (ICA)-based face recognition [12]. Similar to eigenfaces and Fisherfaces algorithms, this algorithm also attempts to represent a face image as a linear combination of the basis images. But ICA aims to find an independent, rather than an uncorrelated, image decomposition and representation by maximizing the statistical independence of the estimated components.

The standard eigenfaces and the Fisherfaces approaches assume the existence of an optimal projection that projects the face images to distinct non-overlapping regions in the reduced subspace where each of these regions corresponds to a unique subject. However, in reality, that assumption may not necessarily be true since images of different people may frequently map to the same region in the face space and, thus, the regions corresponding to different individuals may not always be disjoint. Moreover, many researchers propose alternative approaches that use Bayesian methods or extract facial features by using spatial frequency techniques such as Fourier transform, discrete cosine transform (DCT) and wavelet transforms.

Moghaddam *et al.* [13] proposed a Bayesian face recognition approach. This approach uses a probabilistic measure of similarity, based primarily on a Bayesian Maximum A Posteriori (MAP) analysis of image differences for direct visual matching of images. In other words, the algorithm mainly formulates the multi-class face recognition problem as a two-class classification problem. Intra-personal and extra-personal differences are used to exploit the knowledge of critical variations for discriminating the individuals.

Lai *et al.* [14] combines the wavelet transform and the Fourier transform. Firstly, they extract facial expression insensitive features using wavelet transform because the effect of different facial expressions can be attenuated by removing the high-frequency components and the low-frequency components are only sufficient for recognition. Thus, they propose to use the wavelet transform for face image decomposition. Secondly, they utilize holistic Fourier features in order to obtain translation, scale and on-the-plane rotation invariance with Fourier's basic properties.

Hwang *et al.* [15] propose the hybrid Fourier features extracted from different frequency bands and multiple face models. The hybrid Fourier feature comprises three different Fourier domains; merged real and imaginary components, Fourier spectrum and phase angle. They also define three different frequency bandwidths so that additional complementary features can be obtained. Then, they are individually classified by LDA. This enables analyzing a face image from the various viewpoints. Moreover, they propose multiple face models based on different eye positions with a same image size. In this way, the system achieves 74.33% verification rate in Experiment 4 of FRGC database.

On the other hand, DCT is also a widely used spatial frequency technique that can be performed globally or locally. [16] and [17] are well-known DCT-based methods that perform holistic feature extraction. For instance, [17] uses class dependent kernel DCT features. The system extracts nonlinear kernel features in the discrete transform domain and uses these features in a class dependent feature analysis framework (CFA) to reduce the dimensionality of the data coupled with online discriminant learning using Support Vector Machines. The system obtains 79.33% verification rate in Experiment 4 of FRGC database. Similarly, [18] also presents CFA that reduces the computational complexity of correlation pattern recognition. This algorithm outperforms [17] and results in 87.5% verification rate in the same experiment.

There are also some studies which utilizes Gabor wavelet representation. For example, [19] propose Gabor-Fisher Classifier (GFC). The GFC method applies the Fisher's Linear Discriminant (FLD) model to an augmented Gabor feature vector derived from the Gabor wavelet representation of face images.

In [20], Liu *et al.* presents a pattern recognition framework that capitalizes holistic Gabor image representation with a multi-class Kernel Fisher Analysis (KFA) method. Gabor image representation, which increases dimensionality by incorporating Gabor filters with different scales and orientations, is characterized by spatial frequency, spatial locality, and orientation selectivity for coping with image variabilities such as illumination variations. The KFA method first performs nonlinear mapping from the

input space to a high-dimensional feature space, and then implements the multi-class Fisher discriminant analysis in the feature space. The proposed algorithm achieves 76% verification rate in Experiment 4 of FRGC database.

In addition to all above methods, curvelet-based methods [21], [22], [23], [3] are presented to show that curvelets have better directional decomposition capabilities than wavelets for face images due to the characteristics of facial components. Recently, Hwang *et al.* [3] introduced extended curvature Gabor wavelet (ECG)-based face recognition method on low-resolution images. They extend conventional Gabor wavelet by including smaller frequencies, multiple Gaussian sizes, and different curvature parameters for object representations. They employ AdaBoost for feature selection in order to reduce the dimensionality of the extended features. Then, they learn ECG classifiers by applying LDA on each extended filter. Lastly, they merge a bunch of these classifiers using LLR-based fusion. This approach achieves 90.36% verification rate which is the highest holistic performance in Experiment 4 of FRGC database.

1.2.2. Local Methods

Face recognition based on local facial regions has attracted a significant amount of interest because local features are believed to be more robust to the variations of facial expression, illumination and occlusions. Approaches that utilize local regions either use salient regions or they just partition the face image into rectangular blocks (see Figure 1.3).

Following from the idea in [1], [24] and [25] use the mixtures of eigenfaces approach to represent the variations in the face space more efficiently. Instead of representing the face images with only one subspace, these mixture methods use more than one subspace. The motivation behind these methods is the belief that the face space can possess clusters corresponding to the variations. Therefore, representing each cluster by a local subspace is a more reasonable approach than representing the whole space with a single linear subspace.

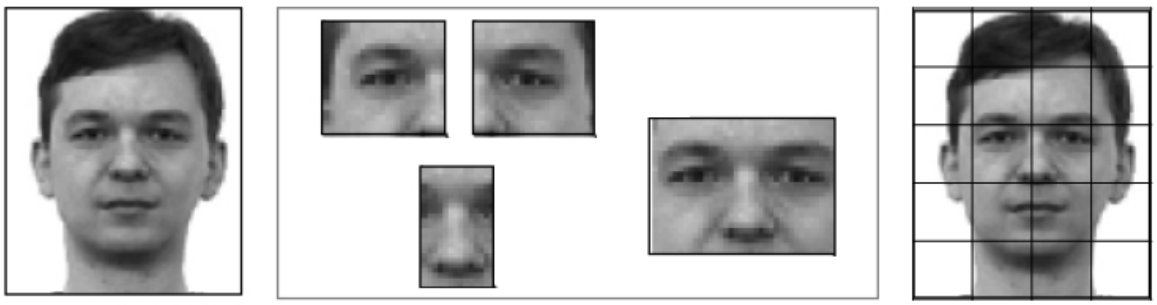

Figure 1.3. Holistic face representation (left), local face representation with salient regions (middle) and with partitioning (right).

In addition, Penev and Atick [26] propose Local Feature Analysis (LFA) to encode the local topological structure of face images. LFA is considered as local method as it utilizes a set of kernels to implicitly detect the local structure such as eyes, nose, and mouth.

Timo *et al.* [27] adopted Local Binary Pattern (LBP), which originated from the area of texture analysis, for face representation. In their method, LBP operator is first applied and then the resulting LBP representation is divided into local regions from which LBP histograms are extracted. The recognition is performed using a nearest neighbor classifier in the computed feature space with Chi square as a dissimilarity measure.

Local Gabor features are also used in facial image analysis. In [28], a facial expression recognition system based on local Gabor filter bank. PCLDA, PCA followed by LDA, is applied to select and compress the Gabor features. Then, minimum distance classifier is employed to recognize facial expression. In addition, Sang *et al.* [29] use Local Gabor Fisher Classifier (LGFC) for face recognition. In LGFC, LFA is exploited to select the most informative Gabor features optimally. The selected low-dimensional local Gabor features are then classified by FLD for final face identification.

Gao et al. [30] introduce multi-resolution local appearance based method. They model the non-overlapping local face blocks with Gabor features and project them

into a discriminant identity space. These local classifiers are fused to combine the image classifier. To acquire complementary information in different scales of face images, they integrate the local decisions from various image resolutions. In other words, they combine local Gabor classifiers obtained from high-resolution, low-resolution and medium-resolution face images. Finally, they combine the multi-resolution Gabor classifier with a local appearance-based DCT classifier as described in [31]. The proposed system achieves 92.5% verification rate which is the highest reported performance in Experiment 4 of FRGC database.

On the other hand, there are many local DCT-based algorithms in face recognition [32], [33], [34], [31] and [35]. [34] show that PCA and LDA can be directly implemented in the DCT domain. First, DCT is applied on the 8×8 pixels resolution blocks. The resulting DCT coefficients are quantized and then ordered according to the zig-zag scan pattern. Only a number of DCT coefficients containing high magnitudes is kept. The obtained DCT coefficients from each block are concatenated. PCA and LDA is applied on this combined vector.

Similar to the idea in [34], Ekenel and Stiefelhagen [35] divide a face image into non-overlapping local blocks of size 8×8, and perform DCT on each block. They perform feature selection by keeping a number of DCT coefficients that provide discriminative information. After applying feature normalization on block features, the feature vectors of blocks are concatenated to form the overall feature vector. Finally, the nearest neighborhood classifier is employed with different distance metrics such as L1 norm, L2 norm, cosine angle, and covariance.

Liu and Liu [36] also have presented a hybrid color space-based face recognition. They utilize three different types of features such as Gabor, LBP and DCT. Similar to the idea in [37], the local classifiers are obtained by the patch-based Gabor features using high-resolution images. They also use LBP feature-based classifiers using multi-resolution and component-based DCT classifiers using low-resolution images. The proposed system achieves 92.4% verification rate in Experiment 4 of FRGC database.

1.2.3. Hybrid Methods

Following from above studies, it has been proven that both holistic representation and local representation are useful for face recognition. Therefore, researchers have had a tendency to combine global and local information extracted using one or more different types of features to improve the recognition performance.

Fang *et al.* [38] propose to combine global PCA features and component-based local features extracted by Haar wavelets. Kim *et al.* [39] propose an effective face descriptor by decomposing a face image into several components, extracting FLD features from each component, and finally combining these component FLD features together with the features extracted by using a holistic FLD. Following a similar idea, [40] show that the combined subspace gives smaller Bayesian error than the subspaces of either the global or local features. On the other hand, Lee *et al.* [41] combine local structures extracted by LFA into composite templates which show compromised aspects between kernels of LFA and eigenfaces.

Elastic Bunch Graph Matching (EBGM) [42] is known as one of the most successful structural matching systems. Wiskott *et al.* model global topological information with graph structure and encode local information as the attribute of the nodes on the graph. To do this, Gabor wavelet transform is used to create dynamic link architecture that project the face onto the elastic grid as shown in Figure 1.4. Recognition is based on the similarity of the Gabor response at each node.

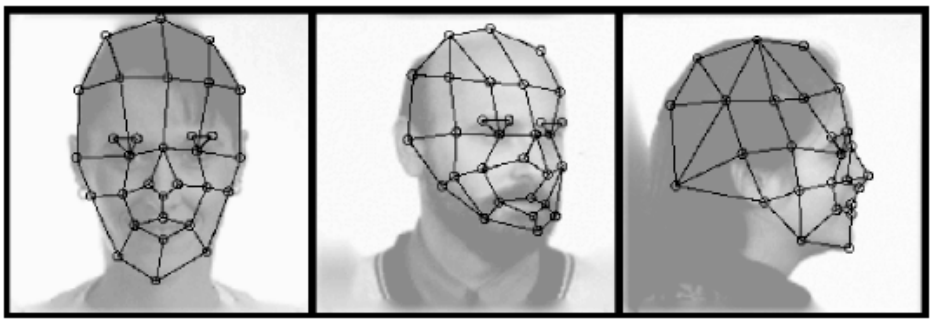

Figure 1.4. Elastic Bunch Map Graphing.

In [43], a normalized face image is divided into four predefined regions such as the

left eye region, the right eye region, the nose region, and inner face region excluding the mouth as shown in Figure 1.3. Similar to the approach in [44], the mouth region is excluded due to its sensitivity to expression variations. DCT and LDA are applied successively both on the entire face image and on predefined facial components. The DCT is applied both on the intensity and edge images. Among DCT coefficients, selected ones are kept and fed into the LDA. The classification is done by finding the minimum weighted Euclidean distance between the feature vectors.

[45] utilize Gabor and LBP for face representation. Instead of directly using the intensity to compute the spatial histogram, multi-scale and multi-orientation Gabor filters are used for the global decomposition of a face image. Then, the resulting Gabor image is further divided into non-overlapping rectangle regions with specific size. On each local region, LBP operator is performed and histograms are computed. Finally, all histograms obtained by local regions are concatenated to form the overall face model. In this way, the combination of Gabor and LBP further enhances the representation power of the spatial histogram. In a very similar manner, Tan and Triggs [46] suggests a heterogeneous method that combine globally extracted Gabor features and locally extracted LBP features. This system achieves 83.6% verification rate in Experiment 4 of FRGC database.

Moreover, Su *et al.* [37] propose a hierarchical framework that combines global and local classifiers. The main difference of this work is that the global classifier is based on Fourier features originating from low-resolution images while the local classifiers are organized by the combination of patch-based Gabor features from high-resolution images. After feature extraction, FLD is separately applied to the global Fourier features and each local Gabor patch. The proposed framework results in 89% verification rate in Experiment 4 of FRGC database.

1.3. Outline of the thesis

Chapter 2 describes the theoretical and mathematical details of the local appearance methods used throughout the thesis.

In Chapter 3, the methodology of the proposed system is explained. The details of face registration, feature extraction, generation of classifiers and approaches for selection and fusion of classifiers are given.

In Chapter 4, first, the details of the dataset used in the evaluation of the thesis are described. Then, the experimental setup for face registration, feature extraction and classifiers is described. Experimental results are then presented at the end of the chapter. Finally, the performance comparisons of the proposed system and the previous works are given.

Chapter 5 concludes the thesis with a summary of the obtained results, followed by possible future work.

2. LOCAL APPEARANCE MODELS FOR 2D FACE RECOGNITION

In this chapter, we describe the details of appearance models used in the proposed face recognition system, namely, curvature Gabor wavelets transform and discrete cosine transform.

2.1. 2D Wavelet Transform

Wavelet transform is a powerful signal analysis tool, widely used for feature extraction, compression, and denoising. It represents the signal with small waves of limited durations, which are called wavelets. This representation provides examination of the signal both in frequency and time domains. Wavelets are frequently used in biometric and other image processing based applications.

The 2D wavelet transform is evaluated by applying the 1D wavelet transform to the rows and columns of the input image block consecutively. Figure 2.1 shows the tree representation of one level, 2D wavelet decomposition. In this figure, H represents high-pass filtering, L represents low-pass filtering, and $\downarrow 2$ represents downsampling by a factor of 2. The input image block B_i of resolution $m \times m$ is first filtered along the rows and downsampled by 2. This produces two $m \times \frac{m}{2}$ resolution images, $B_{i,H}$ and $B_{i,L}$, that have high and low frequency contents, respectively. After this decomposition, the wavelet transform is applied to the columns of these $m \times \frac{m}{2}$ resolution images. In the final stage of the decomposition, there are four $\frac{m}{2} \times \frac{m}{2}$ resolution subband images: A_1, the scaling component containing low-pass global information obtained by low-pass filtering the rows and columns, H_1, the horizontal details obtained by low-pass filtering the rows and high-pass filtering the columns, V_1, the vertical details obtained by high-pass filtering the rows and low-pass filtering the columns, D_1, the diagonal details obtained by high-pass filtering the rows and columns. To obtain higher order wavelet transforms, the scaling component can be further decomposed.

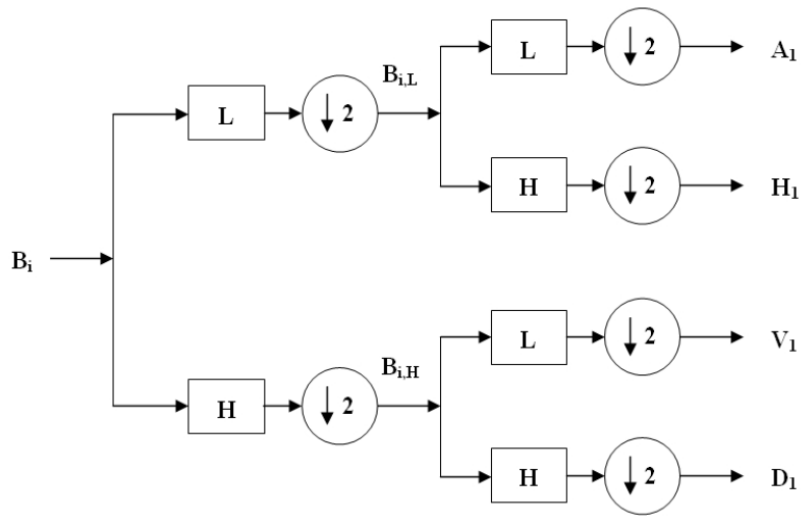

Figure 2.1. Tree representation of one level, 2D wavelet decomposition.

Among different wavelet transforms, Gabor wavelets are of high interest since they were inspired by 2D receptive field profiles of the mammalian cortical simple cells. They form a well established image decomposition because of their spatial locality and orientation selectivity characteristics. Therefore, Gabor wavelets are optimally localized in the space and frequency domains, and can be used in facial image processing for face and facial expression recognition and analysis.

2.1.1. Gabor Function

A signal's specificity simultaneously in time and frequency, is fundamentally limited by a lower bound on the product of its bandwidth and duration. Gabor found the general family of signals that optimize this trade-off and thus achieve the theoretical lower limit of joint uncertainty in time and frequency. Gabor signals take the general form:

$$\varphi(t) = e^{-\alpha^2 t^2} e^{j2\pi f_0 t} \tag{2.1}$$

where α is the sharpness of the Gaussian, and f_0 is the centre frequency of the sinusoidal signal. Figure 2.2 shows Gabor elementary function with different frequencies. The function has a Fourier transform:

$$\phi(f) = \sqrt{\frac{\pi}{\alpha^2}}\ e^{-\frac{\pi^2}{\alpha^2}(f-f_0)^2} \tag{2.2}$$

To make the time duration of function $\varphi(t)$ dependent on the central frequency f_0 a constant ratio, $\gamma = \frac{f_0}{\alpha}$, is defined. Therefore, the function behaves as a scaled version of each other when applied to different frequencies [47]. Both the time duration and frequency bandwidth of the Gabor function are now related with the central frequency: the higher the frequency, the smaller the time duration.

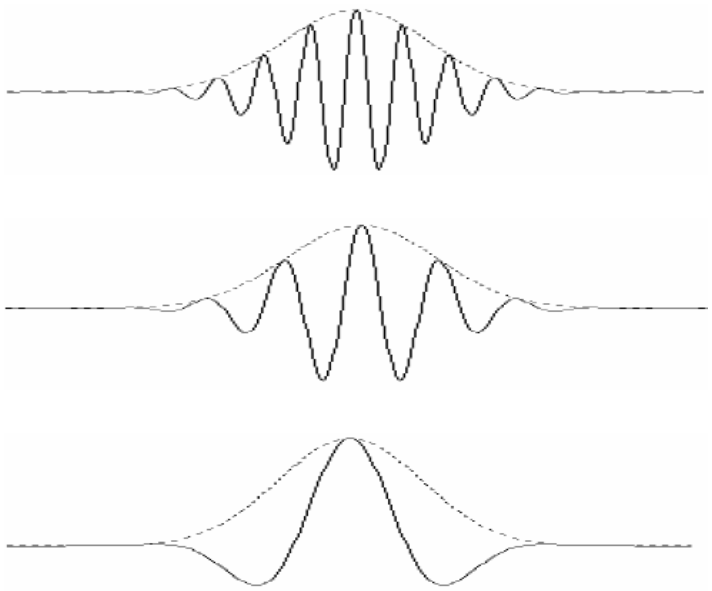

Figure 2.2. Gabor elementary functions with changing frequencies.

The maximum response of the function in the frequency domain can also be normalized to one by multiplying by its inverse $\sqrt{\frac{\alpha^2}{\pi}}$. Consequently the normalized Gabor function is now defined as:

$$\varphi(t) = \frac{|f_0|}{\gamma\sqrt{\pi}}\ e^{\frac{f_0^2}{\gamma}t^2} e^{j2\pi f_0 t} \tag{2.3}$$

2.1.2. Gabor Wavelets

Gabor wavelets have been recognized as one of the most successful local feature extraction methods for face representation. Daugman [48], following from Equation (2.3), generalizes the Gabor function to the following 2D form to model the receptive fields of the orientation-selective simple cells:

$$\psi(\vec{z};\nu,\mu) = \frac{k_{\nu,\mu}^2}{\sigma^2} e^{(-\frac{k_{\nu,\mu}^2||\vec{z}||^2}{2\sigma^2})}[e^{(ik_{\nu,\mu}\vec{z})} - e^{(-\frac{\sigma^2}{2})}] \tag{2.4}$$

$$\vec{z} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} xcos\theta + ysin\theta \\ -xsin\theta + ycos\theta \end{pmatrix} \tag{2.5}$$

where ν controls the scale of the Gabor wavelet, which mainly determines the center of the Gabor filter in the frequency domain; μ controls the orientation of the Gabor filters, $||\cdot||$ denotes the norm operator, and $e^{(ik_{\nu,\mu}\vec{z})}$ is the oscillatory wave function, whose real part and imaginary parts are the cosine and sine functions, respectively. The wave vector $k_{\nu,\mu}$ is defined as:

$$k_{\nu,\mu} = k_\mu \ e^{j\theta_\nu} \tag{2.6}$$

where $\theta_\nu = \frac{\pi\nu}{8}$, $k_\mu = \frac{k_{max}}{f^\mu}$, k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain.

In most cases, Gabor wavelet is used with the following parameters for face recognition: five scales $\nu \in \{0,1,2,3,4\}$ and eight orientations $\mu \in \{0,1,..,7\}$ with Gaussian size $\sigma = 2\pi$, $k_{max} = \frac{\pi}{2}$ and $f = \sqrt{2}$. Figure 2.3 shows the real parts of the Gabor kernels at five scales and eight orientations; while Figure 2.4 illustrates the magnitude coefficients of Gabor kernels at five scales. Figure 2.5 shows the results of convolving a face image with some of the filters in different scales and orientations.

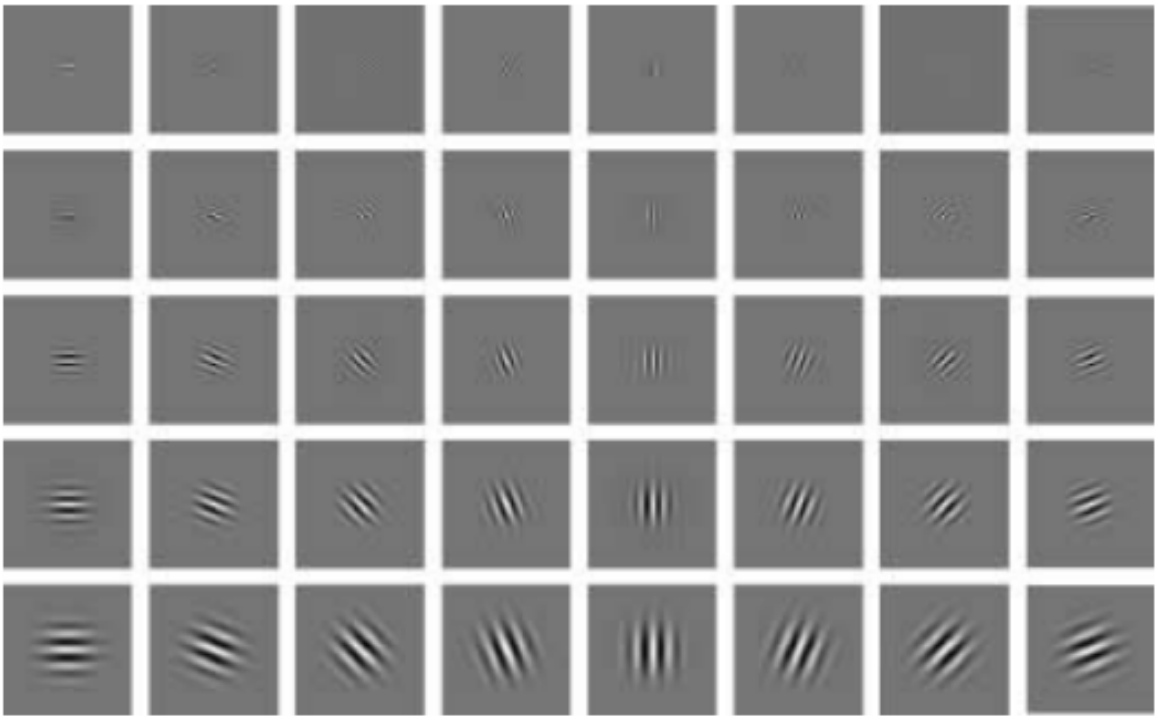

Figure 2.3. The real part of the Gabor kernels at five scales and eight orientations when $\sigma = 2\pi$, $k_{max} = \frac{\pi}{2}$ and $f = \sqrt{2}$.

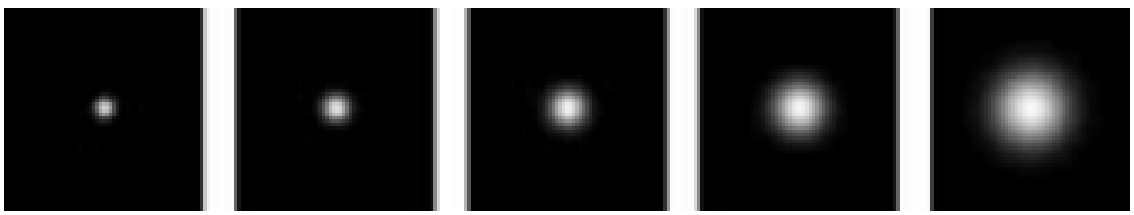

Figure 2.4. The magnitude of the Gabor kernels at five different scales.

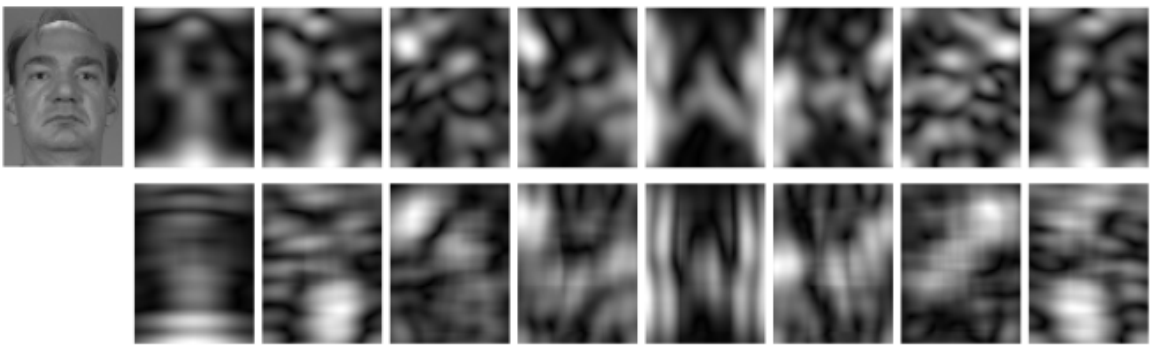

Figure 2.5. Input image, and its Gabor representations at one scale $\nu = 3$ and eight orientations with two different Gaussian sizes; $\sigma = \pi$ (top) and $\sigma = 2\pi$ (bottom).

2.1.3. Curvature Gabor Wavelets

A typical face image contains facial components such as eyes, nose, cheeks, lips, and eyebrows. These components show curved characteristics rather than straight ones. Therefore, it is natural to represent a face image with curvature kernels as well as straight ones as described in Section 2.1.2.

Using curvature Gabor wavelets, which are called as banana wavelets, is one way of modelling curve-like features of a face image. They are obtained by adding a curvature parameter to the conventional Gabor formulation [49]. The curvature Gabor wavelets are defined as follows:

$$\psi(\vec{\acute{z}};\nu,\mu) = \frac{k_{\nu,\mu}^2}{\sigma^2} e^{(-\frac{k_{\nu,\mu}^2 ||\vec{\acute{z}}||^2}{2\sigma^2})} [e^{(ik_{\nu,\mu}\acute{x})} - e^{(-\frac{\sigma^2}{2})}] \tag{2.7}$$

$$\vec{\acute{z}} = \begin{pmatrix} \acute{x} \\ \acute{y} \end{pmatrix} = \begin{pmatrix} xcos\phi + ysin\phi + c(-xsin\phi + ycos\phi)^2) \\ -xsin\phi + ycos\phi \end{pmatrix} \tag{2.8}$$

where the curvature ratio $c \in \{0, 0.05, 0.1, 0.15, 0.2\}$.

When we insert Equation (2.8) in Equation (2.7), we obtain the following:

$$\psi(x,y;\nu,\mu,c) = \frac{k_{\nu,\mu}^2}{\sigma^2} \left(e^{-\frac{k_{\nu,\mu}^2}{2\sigma^2}((xcos\phi_\mu + ysin\phi_\mu + c(-xsin\phi_\mu + ycos\phi_\mu)^2)^2 + (-xsin\phi + ycos\phi)^2)} \right) \left(e^{(ik_{\nu,\mu}(xcos\phi_\mu + ysin\phi_\mu + c(-xsin\phi_\mu + ycos\phi_\mu)^2))} - e^{(-\frac{\sigma^2}{2})} \right) \tag{2.9}$$

Curvature Gabor wavelets do not have the orientation symmetry as in conventional Gabor wavelets as shown in Figure 2.6. Therefore, if the curvature ratio is not zero, the number of orientations used in curvature Gabor wavelets increases to 16. For $c = 0$, it is simply a conventional Gabor wavelet. Some sample results of convolving a

face image with some of the curvature Gabor filters in different scales and orientations are presented in Figure 2.7.

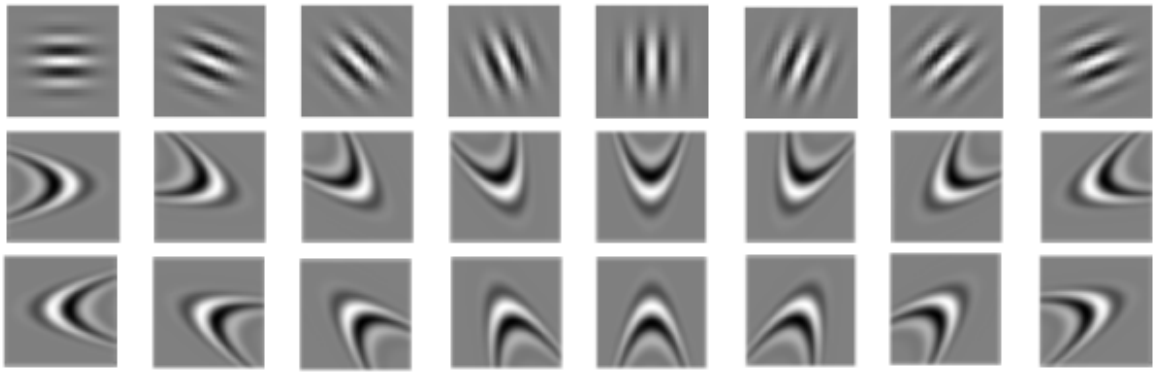

Figure 2.6. Orientation asymmetry in curvature Gabor wavelets $c = 0.1$ (middle and bottom row) unlike conventional Gabor wavelets (top row).

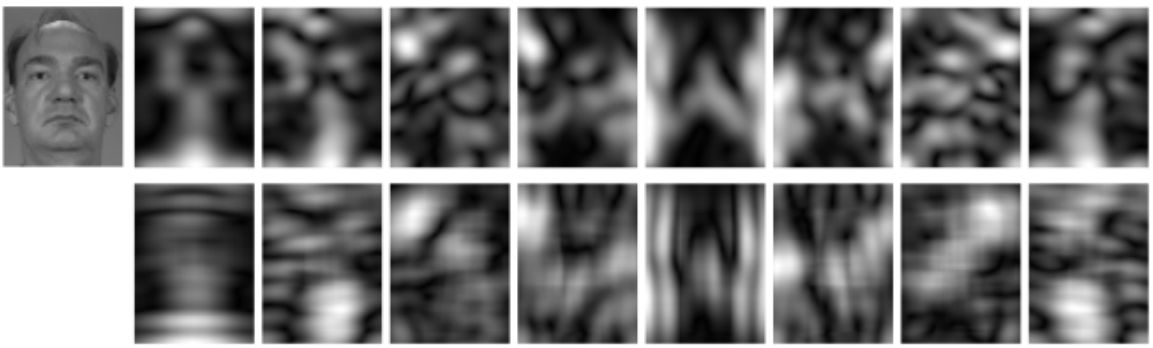

Figure 2.7. Input image, and its curvature Gabor representations at one scale $\nu = 3$ and eight orientations (0-180 degrees) with $c = 0.05$ and two different Gaussian sizes; $\sigma = \pi$ (top) and $\sigma = 2\pi$ (bottom).

2.2. 2D Discrete Cosine Transformation

Discrete cosine transform (DCT) [50] is a well-known signal analysis tool, widely used in feature extraction and compression applications due to its compact representation power. Although Karhunen-Loeve Transform (KLT) is known to be the optimal transform in terms of information packing, its data dependent nature makes it infeasible for use in most practical tasks [51]. Furthermore, DCT closely approximates the compact representation ability of the KLT for a class of signals, which makes it a very useful tool for signal representation both in terms of information packing and in terms

of computational complexity due to its data independent nature. Like other transforms, DCT attempts to decorrelate the image data. After the decorrelation each transform coefficient can be encoded independently without losing compression efficiency. The 2D DCT of an $N \times N$ image is defined as:

$$C(u,v) = \omega(u)\ \omega(v)\ \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\ cos\left[\frac{(2x+1)u\pi}{2N}\right]\ cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{2.10}$$

where $u \in \{0,1,...,N-1\}$, $v \in \{0,1,...,N-1\}$.

The inverse transform is defined as:

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \omega(u)\ \omega(v)\ C(u,v)\ cos\left[\frac{(2x+1)u\pi}{2N}\right]\ cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{2.11}$$

where $\omega(u)$ and $\omega(v)$ is defined as:

$$\omega(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } u = 0 \\ \sqrt{\frac{2}{N}} & \text{for } u = 1,2,...,N-1 \end{cases} \tag{2.12}$$

Obtained basis functions for N = 6 can be seen in Figure 2.8. As can be seen from the top-left part of the basis functions and also from Equation 2.10, the first transform coefficient, the (0,0) component, is the average intensity value of the image. In literature, this value is referred to as the DC Coefficient. All other transform coefficients are called the AC Coefficients. The DC Coefficient can be directly affected by illumination variations, and it is commonly used for normalization. It can be also noted that the basis functions exhibit a progressive increase in frequency both in the vertical and horizontal direction as shown in the (0,1) and (1,0) components.

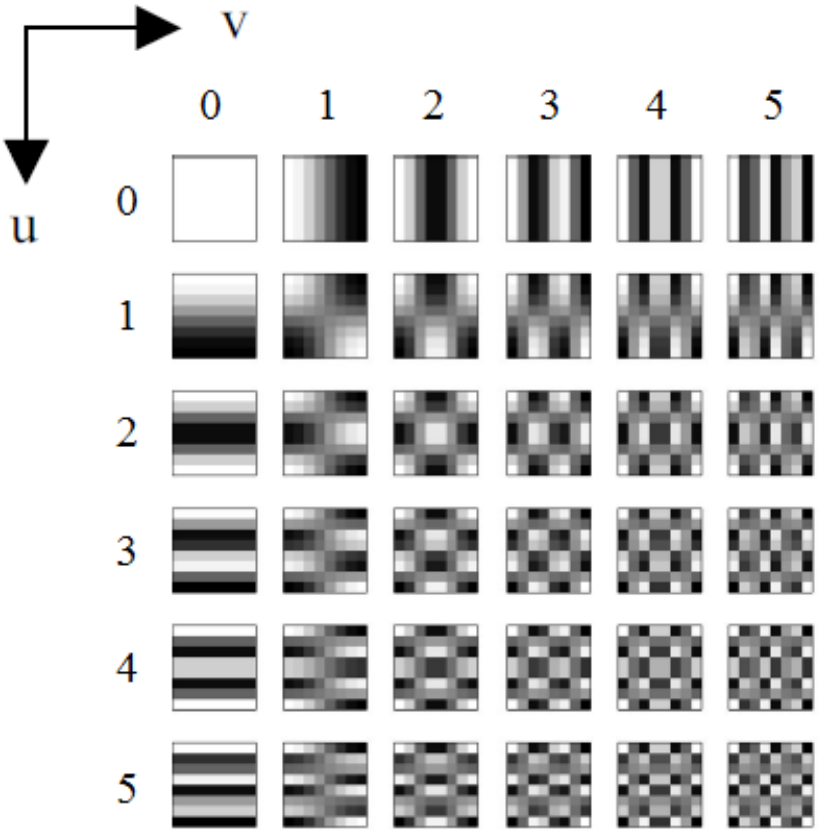

Figure 2.8. DCT basis functions when $N = 6$.

3. FACE RECOGNITION USING LOCAL APPEARANCE MODELS

In this chapter we describe the details of all phases of the proposed face recognition system, namely, face registration, feature extraction, generation of CG and DCT classifiers, selection of CG classifiers, fusion of selected CG classifiers, and fusion of the final CG classifier and DCT classifier.

3.1. Outline of the System

Our system uses a local appearance-based face recognition framework which uses Curvature Gabor (CG) features on spatially partitioned non-overlapping blocks on high-resolution face images. Then, the resulting Gabor images obtained by low pass filtering are downscaled through averaging on grids of the non-overlapping blocks. Various parameter combinations are used in order to acquire complementary information from different curvature and scale spaces. Then, PCLDA, that is, performing PCA followed by LDA, is applied independently on each block to obtain CG classifiers besides further dimensionality reduction. A CG classifier for an entire face image is obtained by combining block decisions with the sum rule. After obtaining CG classifiers, each classifier's performance is improved using Log-likelihood Ratio (LLR) based score conversion. The final classifier is obtained by combining selected classifiers using Sequential Forward Floating Search (SFFS)-based selection mechanism. In addition, DCT-based classifier is combined at score-level by learning the weights with partial least square regression (PLSR). Figure 3.1 shows the outline of our face recognition system.

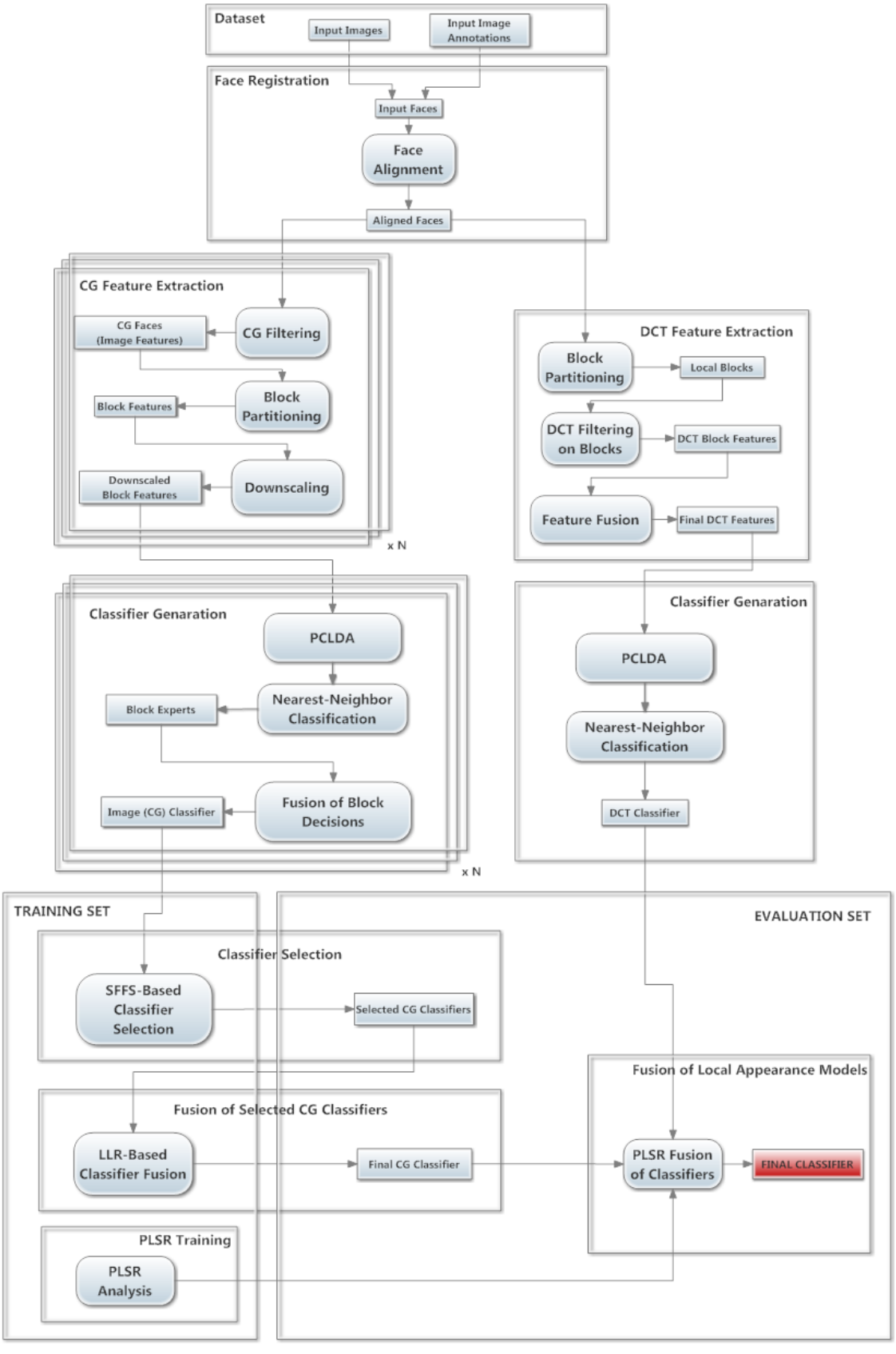

Figure 3.1. Block diagram of the system

3.2. Face Registration

Face registration is the process of defining a face alignment that will closely align all face images. It is also known as facial feature localization since facial features such as eyes, nose and mouth are commonly used as facial landmarks for the alignment procedure. Face registration is a critical first step in many subsequent tasks, such as non-overlapping block partitioning, feature extraction and classification, and therefore, it is crucial for accurate face recognition. For the purpose of alignment, there are some registration parameters which affect both the face region extracted and the proportions of facial components. These parameters can be adjusted to achieve better performance. We use eye centers and inter-ocular distance as registration parameters. We consider the trade-off between information content and noise while determining the alignment parameters. Since we mainly aim to focus on powerful feature extraction, we avoid noisy information like background and hair as much as possible. The result of face registration on a sample input image is shown in Figure 3.2.

Figure 3.2. Face registration: sample input image (left) and extracted face image from input image (right)

3.3. Local Feature Extraction

In this section, we explain the local feature extraction processes for our face recognition system. The underlying idea why we choose the local feature extraction rather than holistic feature extraction is that in holistic feature extraction, a change in a local

region can affect the entire feature representation, whereas in local feature extraction it affects only the features that are extracted from the corresponding block while the features that are extracted from the other blocks remain unaffected. In addition, the local feature extraction can facilitate weighting of local regions. This means that more weight can be assigned to regions which are found to be more discriminative, or less weight can be assigned to the regions where an occlusion is detected.

3.3.1. CG Feature Extraction

As stated in Section 2.1.3, a typical face image consists of facial components which have both linear and curvature characteristics as shown in Figure 3.3. Therefore, it is more reasonable to represent a face image with both curvature and linear Gabor kernels rather than only with linear kernels as in conventional Gabor wavelet representations.

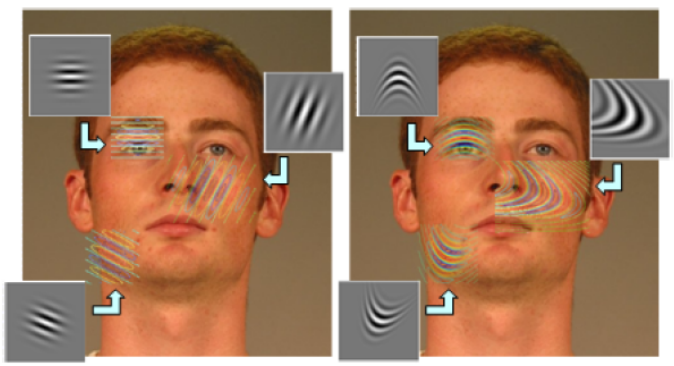

Figure 3.3. Filter response visualization obtained using linear and curvature Gabor wavelets [3].

In this thesis, we use different parameter settings for a comprehensive feature extraction. Unlike conventional Gabor feature extraction, CG feature extraction consists of several kernels with different Gaussian sizes, $\sigma \in \{0.5\pi, \pi, 1.5\pi, 2\pi\}$, for scale space utilization and with different curvature coefficients, $c \in \{0, 0.05, 0.1, 0.15, 0.2\}$, for curvature space utilization. Considering these parameter settings, we have 20 $(= 5 \times 4)$ different CG wavelets (see Figure 3.4) each consisting of 5×16 or 5×8 discrete filters, resulting in 1440 filters $(= 5 \times 16 \times 4 \times 4 + 5 \times 8 \times 4)$.

To sum up, in addition to conventional Gabor wavelets which are good at repre-

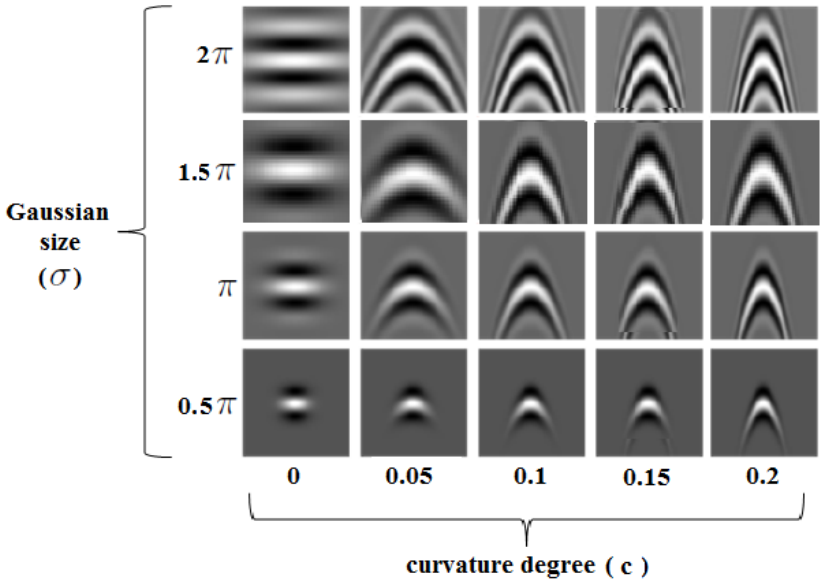

Figure 3.4. CG wavelets' real part representation with changing curvature degrees and Gaussian sizes.

senting coarse and straight structures, CG wavelets provide a way to model finer features with smaller Gaussian sizes: $\sigma = \{0.5\pi, \pi, 1.5\pi\}$, and curvature features with different curvature ratios: $c = \{0.05, 0.1, 1.5, 2\}$. CG wavelet modelling has stronger representation power of modelling facial structures than the conventional Gabor wavelet.

After determining the proper kernels for filtering, we proceed with the local feature extraction using these kernels. In this perspective, one of the critical issues in Gabor feature extraction is the selection of image resolution. Many face characteristics are fine features; and would benefit from high-resolution. However, using high-resolution images brings two disadvantages: First, dimensionality increase brings computational load. Secondly, performance of the system becomes very sensitive to landmark localization and face registration errors. To prevent these setbacks and still benefit from high-resolution, we first perform full convolution of a CG wavelet with the high-resolution face image to obtain the CG magnitude images in different scales and orientations. Afterwards, we spatially partition each CG magnitude image into a number of non-overlapping local blocks of size $m \times m$. The spatial partitioning overcomes the problem of local information loss, makes the system robust to registration errors, and it also provides a relatively lower dimensionality of the overall feature vector. Figure 3.5 shows the construction of overall block feature vectors.

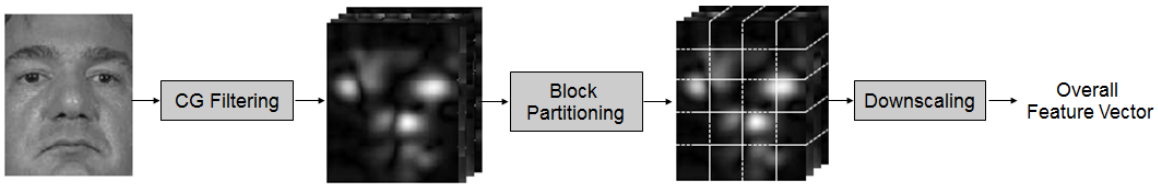

Figure 3.5. Construction of overall CG block feature vectors.

Since we use CG wavelets, the dimensionality of the feature vector of a block is still very high. To overcome the dimensionality problem, we perform downscaling on the block features. In each block, we average the magnitude values in an $p \times p$ grid and use the average value as the new feature. In this way, the dimensionality decreases from D $(= m \times m \times Number\ of\ Filters)$ to $\frac{D}{p^2}$ $(= \frac{m}{p} \times \frac{m}{p} \times Number\ of\ Filters)$. Lastly, the resulting CG block features are Z-normalized before the subspace analysis in order to centralize the data and normalize the variance. Z-normalization, also known as mean-standard deviation normalization, is defined as follows:

$$z = \acute{x} = \frac{x - \mu}{\sigma} \tag{3.1}$$

where x is a feature vector, μ is the mean vector, and σ is the standard deviation vector.

3.3.2. DCT Feature Extraction

As stated in Section 2.2, DCT is a well-known signal analysis tool due to its compact representation power. In the scope of this thesis, we represent faces using DCT in a local appearance based manner as shown in Figure 3.7. More specifically, we first divide an aligned face image into blocks of size $k \times k$. We determine the block size by considering to have small enough blocks in which stationariness is provided and transform complexity is kept simple on one hand, and to have big enough blocks to provide sufficient compression on the other hand. Each block is then represented

by its DCT coefficients. Among these coefficients, the first three DCT coefficients contain general information about the global statistics of the processed block of an image. The first (top left) DCT coefficient only represents the average intensity value of the block while the second and third coefficients represent the average horizontal and vertical intensity change in the image block, respectively. Moreover, the coefficients located on the top left region of the block contain higher representative capability as explained in Section 2.2. Hence, we discard the first DCT coefficient, and then we select M coefficients from the remaining ones in zig-zag scanning order (Figure 3.6) while constructing the block feature vectors. In order to prevent the problems that may occur due to the imbalance between the blocks' impact to the classification, each block feature vector is normalized to unit norm.

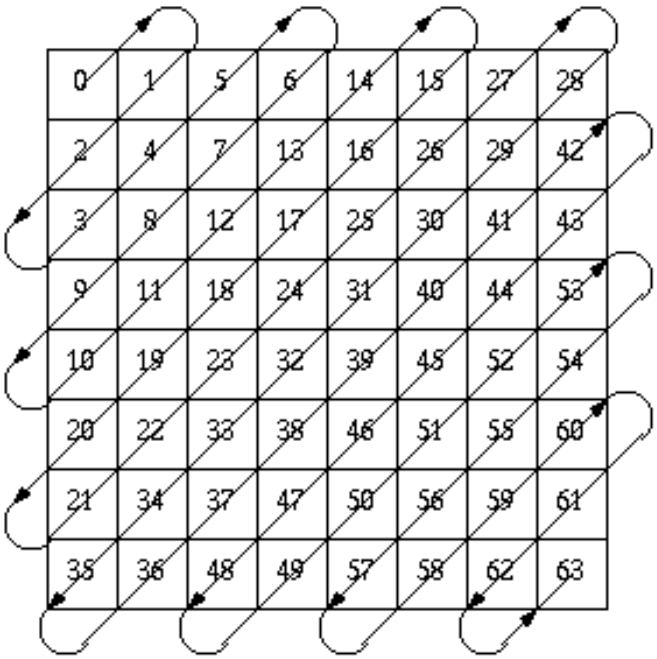

Figure 3.6. The order of DCT coefficients in zig-zag scan pattern.

After obtaining the block features, in feature fusion step as shown in Figure 3.7, the overall feature vector, which is used by the classifier, is obtained by concatenating the block feature vectors. Finally, to balance impacts of coefficients to the classification, the coefficients are divided by their standard deviations that are learned from the training samples. The standard deviation is calculated over all blocks, that is, there

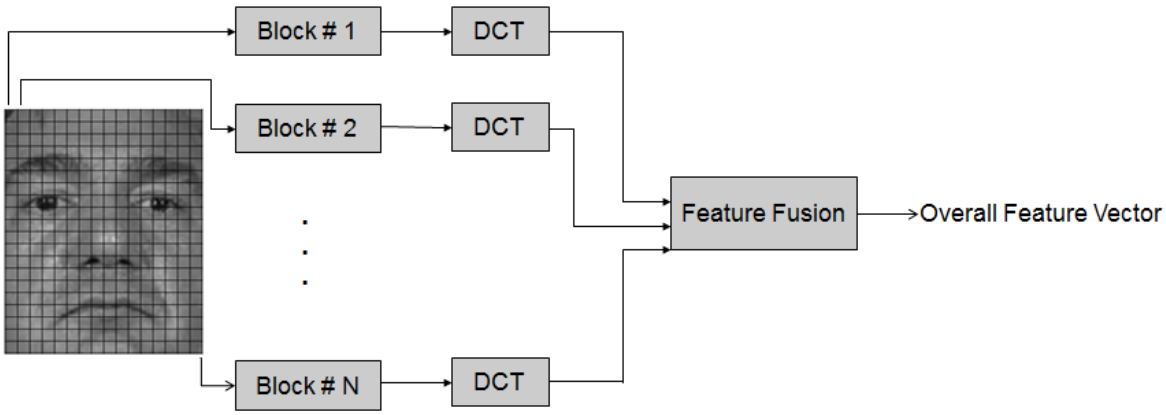

Figure 3.7. Construction of overall DCT feature vector.

are no block specific values for the coefficients.

3.4. Generation of Classifiers

For generation of classifiers, the same classification method and the same decision rule are used for the CG classifiers and the DCT classifier. We use nearest-neighbor classification which is perhaps the simplest of all classification algorithms. It is an easy and efficient classification algorithm, where there is no work done in the training stage, such as density estimation, and all the work is conducted during testing. The classification is done by comparing a query image with all the target images in the database and by finding the training sample that has the closest distance. To compute the score, the similarity or the distance, several metrics can be used for the nearest-neighbor classification, such as L1, L2 norms, and normalized cross correlation. The choice of the metric is very important and the classification performance can change dramatically according to the used metric. In this thesis, we use normalized cross correlation, d, that is as the decision rule by considering the comparative analysis of similarity metrics as stated in [52]. Normalized cross correlation, also known as the cosine similarity, equals to the cosine of the angle between the unit feature vectors.

$$d_{ncc} = \frac{f_{training} \cdot f_{test}}{||f_{training||} * ||f_{test}||} \tag{3.2}$$

where $f_{training}$ and f_{test} denote the training feature vector and the test feature vector, respectively.

3.4.1. CG Classifiers

For the generation of CG classifiers, we perform PCLDA on each of the extracted local block features independently. PCLDA is a widely used method in face recognition because it provides dimensionality reduction while transforming the features to a more discriminative feature space. Applying PCLDA results in N local block classifiers based on nearest neighbor classification with normalized cross correlation as distance metric. Then, we use the scores of block classifiers to form a single image classifier as shown in Figure 3.8.

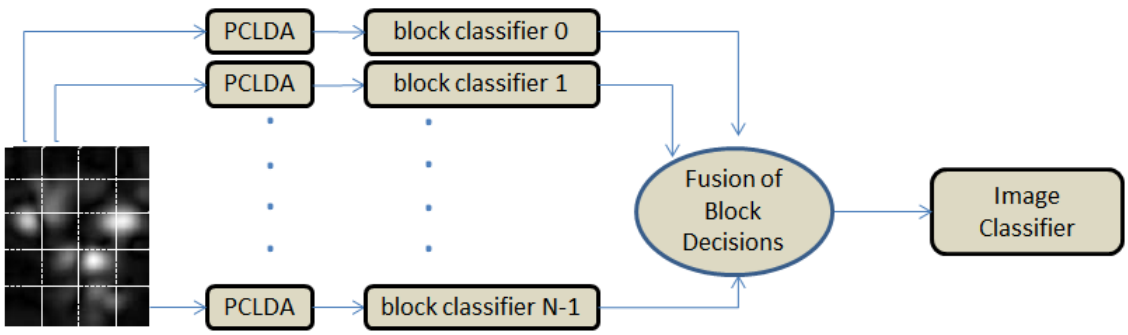

Figure 3.8. Construction of a single image classifier.

In this thesis, we investigate the two approaches, namely, simple sum fusion and LLR-based fusion (please see Section 3.6 for the details of the algorithms), for fusion of block decisions to form a single image classifier.

As stated in Section 3.3.1, there are 20 parameter configurations by using combinations of 5 different curvature degrees, $c \in \{0, 0.05, 0.1, 0.15, 0.2\}$ and 4 different Gaussian sizes, $\sigma \in \{0.5\pi, \pi, 1.5\pi, 2\pi\}$. For each parameter setup, one of the fusion methods is applied to generate 20 CG classifiers. Among these, some classifiers have complementary information when combined with others. In our experiments, we further explore this complementariness by employing a classifier selection algorithm as

described in Section 3.5.

3.4.2. DCT Classifier

For the generation of the DCT classifier, we perform a similar method as in CG classifiers generation. We perform PCLDA on the overall feature vector obtained by concatenating the block features. Applying PCLDA results in a DCT-based classifier, Λ_{DCT}, based on nearest neighbor classification with normalized cross correlation as distance metric.

3.5. Selection of CG Classifiers

Each of 20 CG wavelets is good at representing some particular features; therefore, some of them contain complementary information with other classifiers. This means that fusing some of the CG classifiers can increase classification performance. The optimal combination can be found by exhaustive search, but this requires examination of all subsets of the classifier set (2^{20} trials) which is practically impossible. Another method is to perform empirical selection which is simply fusing the classifiers with the highest verification rates. However, it is important to note that the verification rate of a classifier alone is not the best criterion for determining the classifiers to be fused because the idea is to find the best combination, providing as much complementary information as possible. So we adapt a widely used feature selection algorithm, Sequential Forward Floating Search (SFFS) [53] for selection of classifiers.

Basically, SFFS-based classifier selection algorithm starts from a null classifier set, and for each step, it adds the classifier Γ^+ that maximizes $\Psi(Y_k + \Gamma^+)$, the verification rate on the training set, when combined with the classifiers that have already been selected. The algorithm also verifies the possibility of VR improvement if a classifier is excluded. In such a case, the classifier Γ^- that maximizes $\Psi(Y_k - \Gamma^-)$ is eliminated from the classifier set, that is also known as backtracking. Therefore, the algorithm proceeds dynamically increasing and decreasing the number of classifiers until there is no VR improvement by inclusion or exclusion of any classifier.

The details of SFFS-based classifier selection algorithm is as folllows:

(i) Given the CG classifiers, Γ_i where $i = 0, 1, 2, ..., I-1$ and $I = 20$.

(ii) Initialize the selected classifiers set with a null set, $Y_0 = \{\emptyset\}$

Set the current verification rate to zero, $VR_{current} = 0$

Set the new verification rate to zero, $VR_{new} = 0$

Set the iterator to zero, $k = 0$.

(iii) Apply the below steps until the following two criteria are satisfied:

- The iterator should be less than the number of classifiers, $k < I = 20$
- The new verification rate should be greater then or equal to the current verification rate, $VR_{current} <= VR_{new}$

(a) Set the current verification rate using the current set of selected classifiers, $VR_{current} = \Psi(Y_k)$

(b) Check for inclusion of a classifier:

$\Gamma^+ = \underset{\Gamma \notin Y_k}{\text{argmax}}\ \Psi(Y_k + \Gamma)$

If $\Psi(Y_k + \Gamma^+) > VR_{current}$

- Update $Y_{k+1} = Y_k + \Gamma^+$
- $VR_{new} = \Psi(Y_k + \Gamma^+)$
- $k = k + 1$

(c) Check for exclusion of a classifier (backtrack):

$\Gamma^- = \underset{\Gamma \in Y_k}{\text{argmax}}\ \Psi(Y_k - \Gamma)$

If $\Psi(Y_k - \Gamma^-) > VR_{current}$

- Update $Y_{k-1} = Y_k - \Gamma^-$
- $VR_{new} = \Psi(Y_k - \Gamma^-)$
- $k = k - 1$

3.6. Fusion of CG Classifiers

This section describes the fusion approaches that are used to combine the classifiers at score level, namely, simple sum fusion and LLR-based fusion. We investigate the effect of these two approaches while forming the final CG classifier. We use one of these two approaches according to the fusion method used for the combination of block decisions as stated in Section 3.4.1. In other words, if simple sum is used for the fusion of block classifiers, then the selected CG classifiers are combined with LLR-based fusion, and vice versa. The block diagram for the construction of the final CG classifier using LLR-based fusion is shown in Figure 3.9.

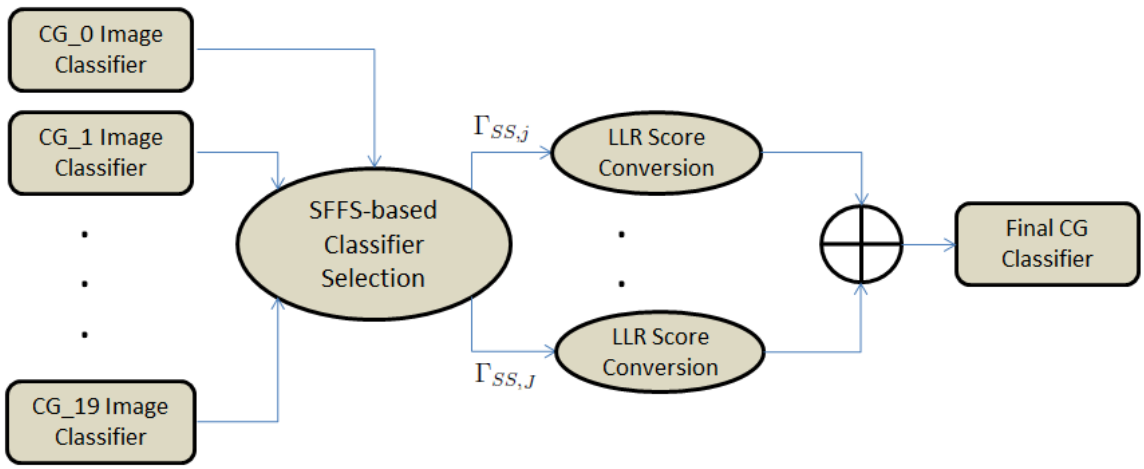

Figure 3.9. Construction of the final CG classifier.

Suppose we have a set of score matrices generated by different classifiers. In these score matrices, each entry corresponds to the distance between a query-target image pair. The distance is computed using nearest-neighbor classification with normalized cross correlation as distance metric as explained in Section 3.4. Then, the details of the fusion methods are as follows:

(i) *Simple sum fusion:*

This is perhaps the simplest of all fusion algorithms. The scores of each classifier to be fused are accumulated in simple sum manner to form the fused classifier. In this thesis, we use this algorithm in either fusion of block classifiers or fusion of selected CG classifiers generated by LLR-based fusion of block classifiers. For

example, fusion of block classifiers using simple sum fusion is defined as follows:

$$\Omega_{SS} = \sum_{n=0}^{N-1} \tau_n \tag{3.3}$$

where τ_n is the score matrix of n^{th} block classifier, and N is the total number of local blocks.

Similarly, fusion of selected CG classifiers that are obtained by LLR-based fusion of block classifiers is defined as follows:

$$\Lambda_{LLR+SS} = \sum_{j=0}^{J-1} \Omega_{LLR,j} \tag{3.4}$$

where $\Omega_{LLR,j}$ is the score matrix of j^{th} selected CG classifier generated by LLR-based fusion of block classifiers, and J is the total number of selected CG classifiers.

(ii) *LLR-based fusion:*

Considering face verification problem, there are two densities, $p(\Omega|same)$ and $p(\Omega|different)$ where Ω is a score matrix generated by a classifier. These densities correspond to the distributions of the scores when the query and target images belong to the same person, and different people, respectively. In this perspective, if these densities are known, then LLR test achieves the highest verification rate for a given false accept rate following from the Neyman-Pearson Lemma [54].

$$log\frac{p(\Omega|same)}{p(\Omega|different)} >< 0 \tag{3.5}$$

Although the real densities are not known, modelling them as two Gaussian distributions with $\mu_{different}$, $\Sigma_{different}$ and μ_{same}, Σ_{same} by observing the scores

computed from query-target image pairs has proven to work successfully while fusing the classifiers in [3]. In this thesis, we employ the same idea for conversion of scores as follows:

$$\Omega_{LLR} = log \frac{\mathcal{N}(\Omega; \mu_{same}, \Sigma_{same})}{\mathcal{N}(\Omega; \mu_{different}, \Sigma_{different})} \tag{3.6}$$

where $\mathcal{N}(\Omega; \mu, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}} e^{-\frac{(\Omega-\mu)^2}{2\Sigma}}$ is the Gaussian density function. The parameters, μ_{same} and Σ_{same} are estimated from the scores obtained by the classifier corresponding to matched target-query pairs generated from the training set. Similarly, $\mu_{different}$ and $\Sigma_{different}$ are computed from the scores obtained by the classifier corresponding to nonmatched target-query pairs generated from the training set.

Moreover, in LLR-based fusion of classifiers, the score matrices of classifiers are first converted using Equation 3.8 and this achieves a higher separability between matching and non-matching classes in individual classifiers. Then, the converted score matrices are fused in simple sum manner. In this thesis, we use LLR-based fusion algorithm for fusion of block classifiers as well as fusion of selected CG classifiers generated only by simple sum fusion of block classifiers. For example, fusion of block classifiers using LLR-based fusion is defined as follows:

$$\Omega_{LLR} = \sum_{n=0}^{N-1} log \frac{\mathcal{N}(\gamma_n; \mu_{same,n}, \Sigma_{same,n})}{\mathcal{N}(\gamma_n; \mu_{different,n}, \Sigma_{different,n})} \tag{3.7}$$

where τ_n is the score matrix of n^{th} block classifier, and N is the total number of local blocks.

Similarly, fusion of selected CG classifiers is defined as follows:

$$\Omega_{SS+LLR,j} = log\frac{\mathcal{N}(\Omega_{SS,j}; \mu_{same,j}, \Sigma_{same,j})}{\mathcal{N}(\Omega_{SS,j}; \mu_{different,j}, \Sigma_{different,j})} \tag{3.8}$$

$$\Lambda_{SS+LLR} = \sum_{j=0}^{J-1} \Omega_{SS+LLR,j} = \sum_{j=0}^{J-1} log\frac{\mathcal{N}(\Omega_{SS+LLR,j}; \mu_{same,j}, \Sigma_{same,j})}{\mathcal{N}(\Omega_{SS+LLR,j}; \mu_{different,j}, \Sigma_{different,j})} \tag{3.9}$$

where $\Omega_{SS,j}$ is the score matrix of a j^{th} CG classifier generated by simple sum fusion of block classifiers. J is the total number of selected CG classifiers.

3.7. Fusion of Local Appearance Classifiers

This section describes the fusion approaches that are used to combine the classifiers which are trained on different local representations at score-level. In the scope of this thesis, we combine the final CG classifier with the DCT classifier. We perform two different methods for the fusion, namely, simple sum fusion and partial least square regression (PLSR) fusion. Simple sum fusion of local appearance classifiers is exactly the same method described in Section 3.4.1 and Section 3.6.

The second method is based on PLSR [55], which is a statistical technique that generalizes and combines features from the PCA and multiple regression. In PLSR fusion, different classifiers are combined at score level by learning the weights with PLSR.

Partial least squares regression (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. It is based on PLS which is used to find the fundamental relations between two matrices (X and P), i.e. a latent variable approach to modelling the covariance structures in these two spaces. A PLS model tries to find the multidimensional direction

in the X space that explains the maximum multidimensional variance direction in the P space. PLS regression is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values.

Following from the idea of PLSR, in PLSR fusion, the goal is to optimally predict a set of dependent variables from a set of predictors. In the scope of this thesis, there is only one dependent variable corresponding to the matching score of a target-query image pair, and there are two predictors, which correspond to the final CG classifier and the DCT classifier. PLSR searches for a set of components, called latent vectors, that performs a simultaneous decomposition of X and P with the constraint that these components explain as much as possible of the covariance between X and P. This process is followed by a regression step where the decomposition of X is used to predict P.

For training of the PLSR, we do not use the entire training set because PLSR analysis is a complicated task. Instead we select a subset which would best represent the training set. More specifically, we include some of the "difficult face pairs" from the training set. Difficult match face pairs ($n = 400$) are sampled randomly from match pairs that had distance scores greater than 0.4 standard deviations ($\sigma_{match} = 0.351$) below the match mean ($\mu_{match} = 16.349$). Difficult nonmatch face pairs ($n = 400$) are sampled randomly from nonmatch pairs that had distance scores less than 3 standard deviations ($\sigma_{nonmatch} = 2.64$) above the nonmatch mean ($\mu_{nonmatch} = 19.999$). After difficult face pairs are sampled, corresponding distance scores from different classifiers are combined in a columnwise matrix to form the observations matrix, X. The dependent variable is P, a $2n$-element vector containing the match status ($+1$ for match and -1 for nonmatch) for each face pair. PLSR is then trained on these matrices.

4. EXPERIMENTS AND RESULTS

4.1. Face Recognition Grand Challenge (FRGC) Database

We evaluated our system on the Face Recognition Grand Challenge (FRGC) [56]. The FRGC database was collected at the University of Notre Dame during 2002-2004 academic years. It is the database that has been used in the face recognition grand challenge experiments which has been conducted by the National Institute of Standards and Technology (NIST). The database contains high resolution still images, multi-images of a person and 3D face scans. The still images are collected both in a controlled and uncontrolled way. The images collected in a controlled way contain frontal faces that are captured in a studio setting. These images have two different lighting conditions caused by the use of two or three studio lights, and two different facial expressions which are neutral and smiling. The images that are collected in an uncontrolled way also contain frontal face images, but this time instead of having a studio setting, the data is captured under changing illumination conditions in hallways, atria, or outdoors. They also contain facial expression changes, out of focus exposure and partial occlusions (please see [56] for more detailed information).

The FRGC is divided in two challenges; version 1 and version 2. Version 1 is designed to introduce the participant to the FRGC challenge problem format and its supporting infrastructure provided by the Biometric Experimentation Environment (BEE). This is an XML based framework for describing and documenting computational experiments. The BEE provides a framework that makes it possible to describe the experiment, record the raw results and provide the analysis, presentation and documentation of the experiment in a common format.

Version 2 is designed to challenge researchers to meet the FRGC performance goal. The FRGC version 2 consists of six experiments:

(i) *Experiment 1:* It is a controlled experiment where the gallery consists of a single

controlled still image of a person, and each probe consists of a single controlled still image.

(ii) *Experiment 2:* This experiment studies the effect using multiple still images of a person has on performance. Each biometric sample consists of the four controlled images of a person taken in a subject session. For example, the gallery is composed of four images of each person where all the images are taken in the same subject session. Likewise, a probe now consists of four images of a person.

(iii) *Experiment 3:* It measures the performance of 3D face recognition. The gallery and probe set consist here of 3D images of a person.

(iv) *Experiment 4:* It measures recognition performance from uncontrolled images. The gallery consists of a single controlled still image, and the probe consists of a single uncontrolled still image.

(v) *Experiment 5:* It examines and compares 3D and 2D images. The gallery consists of 3D images and the probe consists of a single controlled still image.

(vi) *Experiment 6:* This also examines and compares 3D and 2D images. The gallery consists of 3D images, but in contrast to Experiment 5 the probe here consists of a single uncontrolled still image.

In this thesis, we aim to develop a robust face recognition system which works successfully under uncontrolled conditions. Therefore, we use standard Experiment 4 protocol for the system evaluation. For this experiment, the training set consists of 12776 images from 222 individuals. There are 8014 uncontrolled query images of size 2272×1704 pixels and 16028 controlled target images of size 1704×2272 pixels from 466 subjects. The performance measure is the verification rate at 0.1% false acceptance rate (FAR). The performances are reported by three Receiving Operator Characteristic (ROC) curves which correspond to three different time gaps. ROC I corresponds to a subset that consists of the images collected within a semester, ROC II within a year, and ROC III between semesters and years. As ROC III evaluates the matching with large time gap, we compare the performances mostly in terms of the obtained verification rates on ROC III subset in our experiments. Figure 4.1 shows sample images from the dataset.

Figure 4.1. Sample images from the FRGC database: images collected under controlled conditions from Fall 2003 (top) and images collected under uncontrolled conditions from Spring 2004 (bottom).

4.2. Experimental Setup

This section describes the numerical details of the proposed system such as the parameters, the dimensionality of the feature vectors and the variable ranges.

4.2.1. Face Registration Setup

The FRGC dataset provides labels of salient facial features such as eye centers and mouth corners. We use the provided eye labels to align the face images. Since many face characteristics are fine features and would benefit from high resolution we align the face images into 128×160 pixels. While aligning the face images, we use 72 pixels for the registration parameter, which is the eye distance. Figure 4.2 shows sample registered images from the dataset.

Figure 4.2. Sample input images (top) and their corresponding registered images (bottom) from the FRGC database.

4.2.2. Feature Extraction Setup

We use exactly the same face registration parameters, and therefore, aligned images for extracting features from different local appearance representations. The experimental setups for the extraction processes are explained as follows:

(i) *Extraction of CG Features:* We first perform 2D CG filtering, full convolution of a CG wavelet with the aligned face image, to extract the holistic CG features. Then, the resulting features are spatially partitioned into 20 non-overlapping patches of 32×32 pixels. Since we use 5 scales and 16 orientations, the dimensionality of each local block becomes $32 \times 32 \times 5 \times 16 = 81,920$ (if $c \neq 0$), which is very high compared to the number of subjects (222) in the training set. So each local block is downscaled by averaging the magnitude values in an 8×8 grid, and the dimensionality is reduced to 1280 ($= 4 \times 4 \times 5 \times 16$). Lastly, the resulting CG block features are Z-normalized before the subspace analysis in order to centralize the data and normalize the variance before the subspace analysis.

(ii) *Extraction of DCT Features:* We first perform non-overlapping block partitioning on the aligned face images and we obtain $16 \times 20 = 320$ local blocks of size 8×8.

Then, we perform 2D DCT on each of the local blocks. The obtained DCT coefficients are ordered using zig-zag scanning. From the ordered coefficients, we discard the first coefficient and we select the following 3 coefficient to form the local feature vector. Then, the DCT coefficients extracted from each block are concatenated to construct the overall feature vector of size $320 \times 3 = 960$. Finally, in order to balance the contributions of both the blocks and their coefficients to the classification at the same time, a combined normalization is performed by dividing the coefficients by their standard deviations and normalizing the local feature vector to unit norm.

4.2.3. Classifier Setup

We perform PCLDA on the local appearance based features. PCLDA process provides a more discriminative feature space in a lower dimensional space. Applying PCLDA results in a classifier based on nearest neighbour classification with normalized cross correlation as distance metric. In other words, the classification is done by comparing a query image with all the target images in the database in terms of the distance between their feature vectors. In our system, we have $20 + 1$ classifiers: 20 of them are generated using different CG wavelet parameter setups as shown in Table 4.1, and the other one is obtained from DCT features.

Table 4.1. The parameter setups of 20 CG classifiers.

Classifiers	**Scales**	**Orientations**	**Curvature Ratio**	**Gaussian Size**	**Orientation Range**
$CG_0 = \Gamma_0$	5	8	0.0	2π	π
$CG_1 = \Gamma_1$	5	16	0.05	π	2π
$CG_2 = \Gamma_2$	5	16	0.1	π	2π
$CG_3 = \Gamma_3$	5	16	0.2	π	2π
$CG_4 = \Gamma_4$	5	8	0.0	π	π
$CG_5 = \Gamma_5$	5	8	0.0	0.5π	π
$CG_6 = \Gamma_6$	5	16	0.05	2π	2π
$CG_7 = \Gamma_7$	5	16	0.05	0.5π	2π
$CG_8 = \Gamma_8$	5	16	0.1	2π	2π
$CG_9 = \Gamma_9$	5	16	0.1	0.5π	2π
$CG_{10} = \Gamma_{10}$	5	16	0.2	2π	2π
$CG_{11} = \Gamma_{11}$	5	16	0.2	0.5π	2π
$CG_{12} = \Gamma_{12}$	5	8	0.0	1.5π	π
$CG_{13} = \Gamma_{13}$	5	16	0.05	1.5π	2π
$CG_{14} = \Gamma_{14}$	5	16	0.1	1.5π	2π
$CG_{15} = \Gamma_{15}$	5	16	0.15	1.5π	2π
$CG_{16} = \Gamma_{16}$	5	16	0.2	1.5π	2π
$CG_{17} = \Gamma_{17}$	5	16	0.15	0.5π	2π
$CG_{18} = \Gamma_{18}$	5	16	0.15	π	2π
$CG_{19} = \Gamma_{19}$	5	16	0.15	2π	2π

4.3. Results on Individual Block Classifiers

There are 20 local block classifiers that are obtained by applying PCLDA on the local block features. Each block classifier is the expert of the corresponding local region and affects the overall verification performance. Figure 4.3 shows the distribution of the local block classifiers over the face.

Figure 4.3. Indices of the local block classifiers.

Among these block classifiers, some of them shows significantly better performances than the others due to the amount of discriminative information extracted. Figure 4.4 shows the ROC III performances of each local block that belong to a CG wavelet setup (please see Table 4.4 for more detailed results on local block classifiers). This figure indicates that the best performances are obtained by the classifiers generated from the eye-nose region as it is expected.

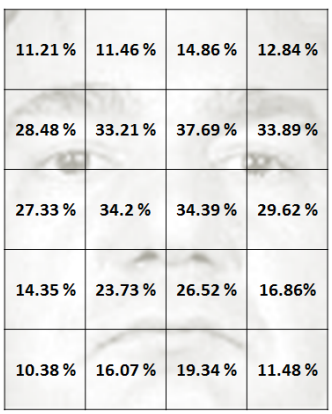

Figure 4.4. The ROC III performances of individual block classifiers generated by CG_0 parameter setup.

Table 4.2. The performances of individual block classifiers obtained using CG_0 parameter setup.

Block Classifiers	**ROC I**	**ROC II**	**ROC III**
γ_0	11.62%	11.57%	11.21%
γ_1	12.94%	12.3%	11.46%
γ_2	14.68%	14.78%	14.86%
γ_3	13.08%	12.97%	12.84%
γ_4	28.57%	28.96%	28.48%
γ_5	34.26%	33.82%	33.21%
γ_6	37.1%	37.39%	37.69%
γ_7	33.09%	33.44%	33.89%
γ_8	27.56%	27.49%	27.33%
γ_9	35.72%	35.04%	34.2%
γ_{10}	36.67%	35.63%	34.39%
γ_{11}	27.76%	28.65%	29.62%
γ_{12}	16.26%	15.37%	14.35%
γ_{13}	26.35%	25.19%	23.73%
γ_{14}	29.31%	28.03%	26.52%
γ_{15}	17.04%	16.95%	16.86%
γ_{16}	10.38%	10.21%	10.03%
γ_{17}	16.33%	16.22%	16.07%
γ_{18}	20.05%	19.69%	19.34%
γ_{19}	11.86%	11.72%	11.48%

4.4. Results on Individual CG Classifiers

As stated in Section 3.4.1, there are two different fusion methods for block classifiers that are used to generate CG classifiers, namely, simple sum fusion and LLR-based fusion. When the block classifiers are combined by performing simple sum fusion, the best verification rate on ROC III, 91.05%, is achieved by a linear Gabor wavelet, $\Gamma_{SS,4}$ while the worst verification rate, 81.92%, is achieved by a curvature Gabor wavelet, $\Gamma_{SS,16}$. On the other hand, LLR-based fusion for combining the block classifiers results in 90.87% for the best, and 81.36% for the worst by $\Gamma_{LLR,4}$ and $\Gamma_{LLR,16}$, respectively. Table 4.3 shows the performances of all CG classifiers using these two methods. As it is seen, only a few of the Γ_{LLR} classifiers have higher verification rates than Γ_{SS} classifiers, but mostly there occurs a decrease on the verification performance by $0.1\% - 0.6\%$. The main reason for this decrease is the sampling quality of the training of LLR.

Table 4.3. The performances of CG classifiers generated by simple sum fusion and LLR-based fusion of block classifiers.

Γ_{SS}	**ROC I**	**ROC II**	**ROC III**	Γ_{LLR}	**ROC I**	**ROC II**	**ROC III**
$\Gamma_{SS,0}$	87.57%	87.61%	87.58%	$\Gamma_{LLR,0}$	88.08%	87.9%	87.7%
$\Gamma_{SS,1}$	89.99%	90.04%	90.02%	$\Gamma_{LLR,1}$	90.12%	90.01%	89.9%
$\Gamma_{SS,2}$	89.1%	89.13%	89.11%	$\Gamma_{LLR,2}$	89.21%	89.06%	88.88%
$\Gamma_{SS,3}$	88.25%	88.34%	88.41%	$\Gamma_{LLR,3}$	88.26%	88.19%	88.11%
$\Gamma_{SS,4}$	91.28%	91.18%	91.05%	$\Gamma_{LLR,4}$	91.52%	91.23%	90.87%
$\Gamma_{SS,5}$	87.39%	86.84%	86.19%	$\Gamma_{LLR,5}$	87.37%	86.61%	85.73%
$\Gamma_{SS,6}$	87.27%	87.4%	87.49%	$\Gamma_{LLR,6}$	87.69%	87.67%	87.57%
$\Gamma_{SS,7}$	88.42%	88.1%	87.72%	$\Gamma_{LLR,7}$	88.37%	87.79%	87.13%
$\Gamma_{SS,8}$	86.13%	86.24%	86.35%	$\Gamma_{LLR,8}$	86.53%	86.51%	86.46%
$\Gamma_{SS,9}$	88.33%	88.03%	87.62%	$\Gamma_{LLR,9}$	88.45%	87.86%	87.17%
$\Gamma_{SS,10}$	85.33%	85.57%	85.71%	$\Gamma_{LLR,10}$	85.76%	85.83%	85.85%
$\Gamma_{SS,11}$	88.21%	87.97%	87.66%	$\Gamma_{LLR,11}$	88.16%	87.75%	87.25%

In contrast to LLR-based fusion of block classifiers, LLR-based fusion of image (CG) classifiers improves the verification performances with respect to simple sum fusion of image classifiers. The reason is that individual CG classifiers are more reliable than the individual block classifiers are not reliable. So, there occurs less outliers among training samples of LLR because genuine and impostor distributions are computed more accurately. The performances of CG classifiers obtained by simple sum fusion of block classifiers and the performances of CG classifiers after LLR-based score conversion are shown in Figure 4.5 and Table 4.4. It can be realized from the figure that LLR-based score conversion mostly increase the performances of CG classifiers.

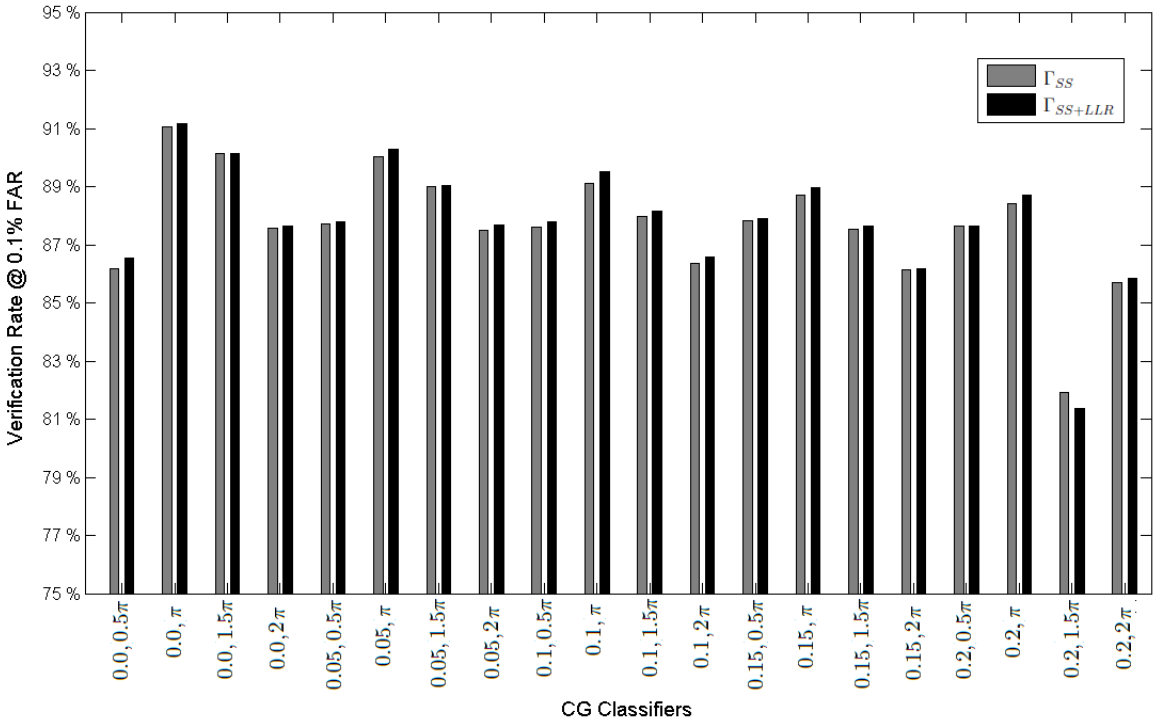

Figure 4.5. The ROC III performance comparison of the individual CG classifiers obtained by simple sum fusion of block classifiers, Γ_{SS}, and the performances of these classifiers after LLR-based score conversion is performed, Γ_{SS+LLR}.

The verification rate increase on the individual CG classifiers by LLR-score conversion directly affects the final CG classifiers obtained by fusion of CG classifiers. Hence, we use simple sum fusion of block classifiers to generate the individual CG classifiers and then we use LLR-based fusion of CG classifiers to form the final CG classifier.

Table 4.4. The ROC III performances of the individual CG classifiers obtained by simple sum fusion of block classifiers, Γ_{SS}, and the performances of these classifiers after LLR-based score conversion is performed, Γ_{SS+LLR}.

Image Classifiers	**ROC III**	**Image Classifiers**	**ROC III**
$\Gamma_{SS,0}$	87.58%	$\Gamma_{SS+LLR,0}$	87.65%
$\Gamma_{SS,1}$	90.02%	$\Gamma_{SS+LLR,1}$	90.29%
$\Gamma_{SS,2}$	89.11%	$\Gamma_{SS+LLR,2}$	89.5%
$\Gamma_{SS,3}$	88.41%	$\Gamma_{SS+LLR,3}$	88.7%
$\Gamma_{SS,4}$	91.05%	$\Gamma_{SS+LLR,4}$	91.18%
$\Gamma_{SS,5}$	86.19%	$\Gamma_{SS+LLR,5}$	86.55%
$\Gamma_{SS,6}$	87.49%	$\Gamma_{SS+LLR,6}$	87.69%
$\Gamma_{SS,7}$	87.72%	$\Gamma_{SS+LLR,7}$	87.8%
$\Gamma_{SS,8}$	86.35%	$\Gamma_{SS+LLR,8}$	86.57%
$\Gamma_{SS,9}$	87.62%	$\Gamma_{SS+LLR,9}$	87.78%
$\Gamma_{SS,10}$	85.71%	$\Gamma_{SS+LLR,10}$	85.84%
$\Gamma_{SS,11}$	87.66%	$\Gamma_{SS+LLR,11}$	87.66%
$\Gamma_{SS,12}$	90.13%	$\Gamma_{SS+LLR,12}$	90.12%
$\Gamma_{SS,13}$	89.01%	$\Gamma_{SS+LLR,13}$	89.05%
$\Gamma_{SS,14}$	87.98%	$\Gamma_{SS+LLR,14}$	88.15%
$\Gamma_{SS,15}$	87.53%	$\Gamma_{SS+LLR,15}$	87.65%
$\Gamma_{SS,16}$	81.92%	$\Gamma_{SS+LLR,16}$	81.36%
$\Gamma_{SS,17}$	87.82%	$\Gamma_{SS+LLR,17}$	87.9%
$\Gamma_{SS,18}$	88.69%	$\Gamma_{SS+LLR,18}$	88.97%
$\Gamma_{SS,19}$	86.15%	$\Gamma_{SS+LLR,19}$	86.19%

4.5. Results for Selection and Fusion of CG Classifiers

After LLR-based score conversion, which can improve the performances of individual CG classifiers up to 0.4%, is applied on each CG classifier, we perform fusion of classifiers at score level to form the final CG classifier. However, fusion of all CG classifiers does not result in the best performance because the complementary information extracted from some of the classifiers are cancelled out by some others. Figure 4.6 illustrates the effects of classifier selection and fusion methods to show that classifier selection and fusion play a crucial role for the overall performance.

In our experiments, when there is no selection, that is fusion of all 20 CG classifiers, we achieved 92.85% with LLR fusion, 92.52% with simple sum fusion. On the other hand, we perform two classifier selection methods. The first approach, called as empirical selection, selects the classifiers according to their verification rates, in other words, K of 20 CG classifiers having higher verification rates than others are selected to form the final CG classifiers. This approach results in a better performance than no classifier selection. The best verification rates are 93.15% with LLR fusion of 15 classifiers and 92.87% simple sum fusion of 4 classifiers.

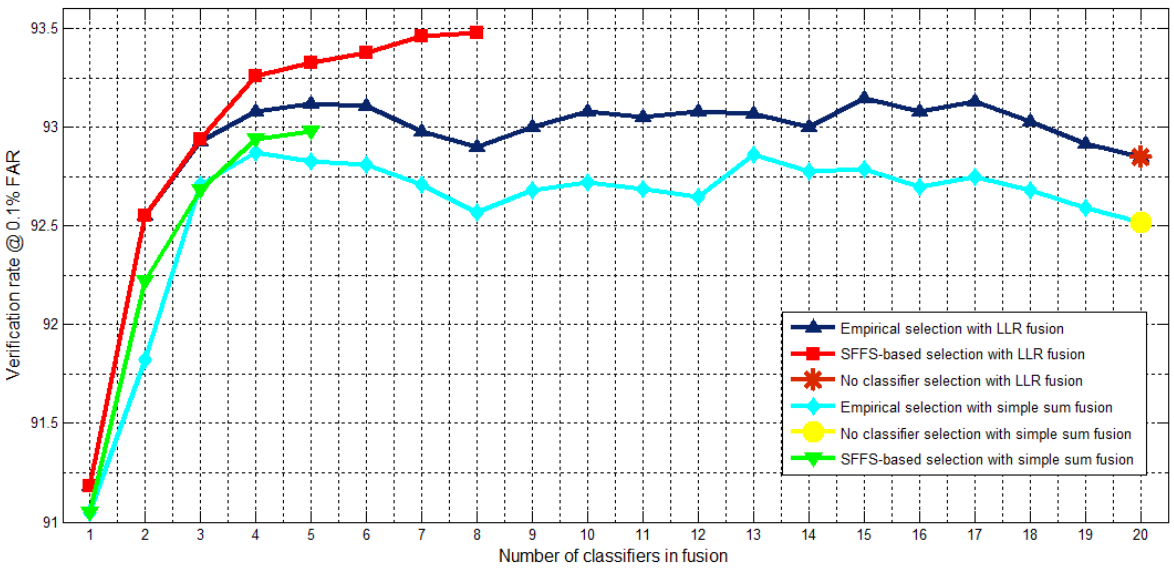

Figure 4.6. Comparison of different classifier selection and fusion methods with changing number of selected classifiers for the fusion.

Another method is SFFS-based classifier selection. This approach uses greedy hill climbing, which iteratively evaluates a candidate subset of classifiers, then modifies the subset and evaluates if the new subset is an improvement over the old. This method achieves the best fusion performances among all approaches. We achieved 93.48% by LLR fusion of 8 classifiers and 92.95% by simple sum fusion of 5 classifiers. Table 4.5 shows the selected CG classifiers and obtained verification rates for each selection approach with LLR fusion of classifiers.

Table 4.5. Comparison of different classifier selection strategies. The best performance, which is used as the final CG classifier, is obtained by SFFS-based classifier selection.

Selection Method	**No of Class.**	**Fusion** (Λ_{SS+LLR})	**ROC III**
No selection	20	$\{\Gamma_0, \Gamma_1, ..., \Gamma_{19}\}$	92.85%
Empirical selection	15	$\{\Gamma_4, \Gamma_1, \Gamma_{12}, \Gamma_2, \Gamma_{13}, \Gamma_{18}, \Gamma_3, \Gamma_{14}, \Gamma_{17}, \Gamma_7, \Gamma_9, \Gamma_6, \Gamma_{11}, \Gamma_{15}, \Gamma_0\}$	93.15%
SFFS-based selection	8	$\{\Gamma_4, \Gamma_1, \Gamma_0, \Gamma_{11}, \Gamma_6, \Gamma_5, \Gamma_{12}, \Gamma_2\}$	93.46%

Figure 4.7 shows the selected classifiers' corresponding wavelet setups when SFFS-based classifier selection and LLR classifier fusion is used. In addition, Table 4.6 shows the classifiers selected in each iteration, and their effects on the fusion performance.

As seen in Figure 4.6 and Table 4.5, there is no classifier exclusion in SFFS algorithm even though we allow backtracking. The main reason for this is that our algorithm deals with classifiers rather than features as in the original SFFS algorithm. In feature selection problem, the number of candidates are much higher compare to classifier selection. This decreases the possibility of backtracking in classifier selection. In this thesis, the algorithm never excludes a classifier since removal of any classifier in any iteration does not improve the verification performance. Therefore, the findings of SFFS in our experiments are exactly the same with the ones of Sequential Forward Selection (SFS) which only considers classifier inclusion. Consequently, the best verifi-

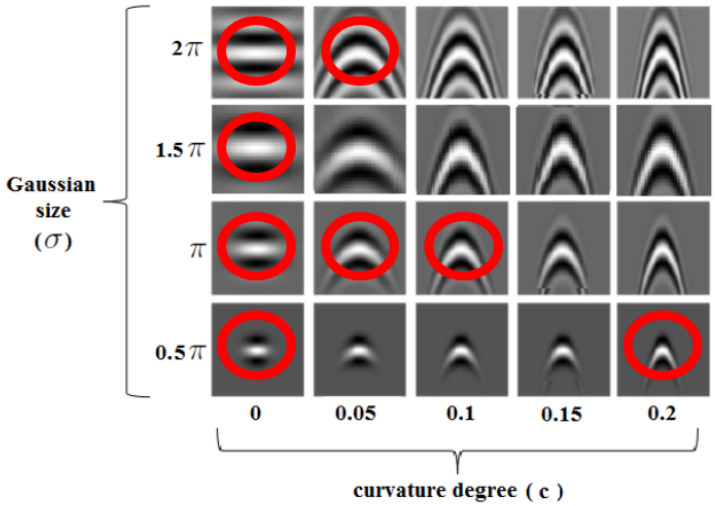

Figure 4.7. Selected wavelet setups that correspond to the classifiers selected by SFFS-based algorithm.

Table 4.6. Selected classifiers and their effects on the fusion performance during the iterations of SFFS classifier selection algorithm.

Iterations	**Selected Classifier**	**Fusion** (Λ_{SS+LLR})	**ROC III**
0	$\Gamma_{SS+LLR,4}$	$\{\Gamma_4\}$	91.18%
1	$\Gamma_{SS+LLR,1}$	$\{\Gamma_4\} \cup \Gamma_1$	92.55%
2	$\Gamma_{SS+LLR,0}$	$\{\Gamma_4, \Gamma_1\} \cup \Gamma_0$	92.94%
3	$\Gamma_{SS+LLR,11}$	$\{\Gamma_4, \Gamma_1, \Gamma_0\} \cup \Gamma_{11}$	93.26%
4	$\Gamma_{SS+LLR,6}$	$\{\Gamma_4, \Gamma_1, \Gamma_0, \Gamma_{11}\} \cup \Gamma_6$	93.33%
5	$\Gamma_{SS+LLR,5}$	$\{\Gamma_4, \Gamma_1, \Gamma_0, \Gamma_{11}, \Gamma_6\} \cup \Gamma_5$	93.38%
6	$\Gamma_{SS+LLR,12}$	$\{\Gamma_4, \Gamma_1, \Gamma_0, \Gamma_{11}, \Gamma_6, \Gamma_5\} \cup \Gamma_{12}$	93.44%
7	$\Gamma_{SS+LLR,2}$	$\{\Gamma_4, \Gamma_1, \Gamma_0, \Gamma_{11}, \Gamma_6, \Gamma_5, \Gamma_{12}\} \cup \Gamma_2$	93.46%

cation performance is achieved using SFFS-based selection mechanism by only fusing 8 classifiers out of 20. In order to form the final CG classifier, we combined the selected classifiers, which are generated using simple sum fusion of block classifiers, by using LLR-based fusion at score level.

4.6. Fusion of Local Appearance Models

In addition to the final CG classifier, we performed additional experiments to see whether a different face representation provides complementary information for verifying faces. We have used another local face representations based on DCT and Histogram of Oriented Gradient (HOG) [57] as extra evidences. Table 4.7 shows the individual performances of different local appearance models.

Table 4.7. Performances of different local appearance models.

Selection Method	**ROC I**	**ROC II**	**ROC III**
Final CG Classifier (Λ_{SS+LLR})	94.34%	93.93%	93.46%
DCT-based Classifier (Λ_{DCT})	67.7%	67.41%	67.01%
HOG-based Classifier (Λ_{HOG})	70.08%	69.82%	68.9%

We combined these local appearance based classifiers at score-level to form the final global classifier. Firstly, we used simple sum fusion of the classifiers. Table 4.8 shows the verification rates obtained by different combinations of classifiers. It indicates that DCT-based classifier provides complementary information to the final CG classifier, on the other hand, HOG-based classifier decreases the performance when it is fused with the final CG classifier.

Table 4.8. Performances of local appearance fusions by simple sum rule at score level.

Fused Classifiers	**No of Classifiers**	**ROC III**
$\Lambda_{SS+LLR} \cup \Lambda_{HOG}$	2	93.27%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	2	94.12%
$\Lambda_{HOG} \cup \Lambda_{DCT}$	2	80.26%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT} \cup \Lambda_{HOG}$	3	93.7%

Another method for the fusion would be the weighted sum of classifiers. In order

to determine the weights, PLSR analysis is performed on a subset of the training dataset as explained in Section 3.7. Table 4.9 shows the verification rates obtained by PLSR fusion.

Table 4.9. Performances of local appearance fusions by PLSR fusion at score level.

Fused Classifiers	**PLSR Kernel**	**No of Latent Vectors**	**ROC III**
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	LINEAR	1	93.13%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	LINEAR	2	94.16%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT} \cup \Lambda_{HOG}$	LINEAR	1	93.64%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT} \cup \Lambda_{HOG}$	LINEAR	2	92.91%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	GAUSSIAN (100)	2	93.91%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	GAUSSIAN (500)	2	94.13%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	GAUSSIAN (1024)	2	94.14%
$\Lambda_{SS+LLR} \cup \Lambda_{DCT}$	GAUSSIAN (1500)	2	94.14%

Our system's performance is further improved using PLSR fusion of different local appearance models. As the best result, our proposed system is achieved 94.16% verification rate at 0.1% FAR on ROC III of Experiment 4.

4.7. Comparison with Previous Work

Table 4.10 summarizes the performances of the previous studies which are also evaluated on Experiment 4 of the FRGC dataset. Earlier, holistic and homogeneous approaches were proposed. [20] presents a framework that capitalizes holistic Gabor image representation with a multi-class Kernel Fisher Analysis (KFA) method. [15] proposes the hybrid Fourier features extracted from different frequency bands and multiple face models. [17] extracts class dependent kernel DCT features and uses them in Class-dependent Feature Analysis (CFA) framework. Similarly, [18] also presents CFA that reduces the computational complexity of correlation pattern recognition.

Later, researchers have tended to combine hybrid methods which combine different types of features in order to improve the results. For example, [46] introduces a method which utilizes Gabor and LBP features. [37] combines Global Fourier features and local Gabor features. The best performances in the literature, 92.4%-92.5%, were achieved by [36] and [30]. [36] combines 17 classifiers whose features were extracted by DCT, LBP, and Gabor, separately. On the other hand, [30] combines local Gabor classifiers obtained from different image resolutions with a DCT classifier.

In addition all above methods, Hwang *et al.* [3] propose to use extended CG (ECG) features similar to our work. However, they extract ECG features globally, and then, they perform feature selection using Adaboost. They learn ECG classifiers by applying LDA. Lastly, they merge a bunch of these classifiers using LLR-based fusion.

As it is seen in Table 4.10, our system outperforms all of the previous studies on the same dataset reported in the literature.

Table 4.10. Performance comparisons with the published methods on Experiment 4 of the FRGC version 2 dataset. The evaluation measure is the verification rate at 0.1% FAR for ROC III.

Methods	**Description and Feature Type(s)**	**No of Cls.**	**ROC III**
[20] - 2006	Holistic Gabor with KFA	1	76%
[15] - 2006	Holistic Hybrid Fourier	3	74.33%
[17] - 2006	Holistic DCT with CFA	1	79.33%
[18] - 2006	Correlation filters with CFA	1	87.5%
[46] - 2007	Holistic Gabor and LBP	2	83.6%
[58] - 2008	Extended Color Image Discriminant Model	3	78.26%
[37] - 2009	Holistic Fourier and Local Gabor	2	89%
[36] - 2009	Local Gabor and LBP and DCT	17	92.4%
[30] - 2010	Multi-resolution Local Gabor and Local DCT	4	92.5%
[3] - 2011	Holistic ECG with Adaboost	6	90.36%
Proposed	**Local CG**	**8**	**93.46%**
Proposed	**Local CG and Local DCT**	**9**	**94.16%**

5. CONCLUSIONS

In this thesis, we present a new local appearance based face recognition system. The proposed system runs reliably under uncontrolled conditions such as illumination and facial expression changes, partial occlusions and out of focus exposure. First, it focuses on robust face representation based on appearances of local facial regions that are presented with curvature Gabor wavelets and DCT. Secondly, it focuses on selecting and combining multiple local appearance based classifiers.

The system consists of five main steps, namely, face registration, feature extraction, generation of classifiers, selection and fusion of CG classifiers, and fusion of local appearance models.

The proposed system starts with face registration. In this step, face images in the database are aligned in order to localize facial features such as eyes, nose and mouth. It is a critical step because the subsequent tasks are directly affected by the alignment procedure. The main concern in face registration is to determine the registration parameters by considering the trade-off between discriminative information content and noise. For this reason, we adjust the registration parameters so that image regions that contain noisy information such as background and hair are not included in the aligned face image.

The second step, feature extraction, aims to construct a face representation which is robust against uncontrolled conditions. For this reason, we use local face representation rather than holistic approach because local feature extraction is less sensitive to variations on facial appearance. For local face representations, high-resolution face images are spatially partitioned into non-overlapping blocks. For the first local representation, we utilize curvature Gabor wavelets as well as conventional Gabor wavelets in order to acquire complementary information in different scales, orientations and curvatures. For the second local face representation, we exploit local blocks-based DCT as extra evidence because DCT features provide complementary information to curvature

Gabor features.

In the third step, local appearance classifiers are generated from PCLDA which is a widely used method in face recognition. PCLDA transforms the features to a more discriminative feature space besides providing dimensionality reduction. After that, we simply employ nearest-neighbor classifier for recognition. We use normalized cross correlation distance as decision rule to compute the classification scores of target-query pairs.

After obtaining the classifiers, the aim is to determine how to select and fuse the classifiers in order to achieve the best performance. For this purpose, following from the idea in SFFS feature selection, which is known as one of the most powerful feature selection strategies, we employ SFFS-based classifier selection algorithm to determine the best combination of classifiers with the highest discrimination power. Then, we combine the selected CG classifiers to form the final CG classifier by using LLR-based fusion which results in further increase in the verification performance.

The proposed system lastly fuses different local appearance models generated by CG and DCT representations. Since these models are not expected to contribute equally, we perform weighted fusion of these classifiers by learning the weights with PLSR analysis.

From the experiments conducted on the standard Experiment 4 protocol of FRGC version 2.0 database, we have the following observations. The proposed system achieves 94.16% verification rate which is the highest verification performance reported so far in the literature for this experiment.

The experimental results on individual local block classifiers show that high performances of almost all individual classifiers indicate the robustness of the proposed system. The best individual classifier, $\Gamma_{SS+LLR,4}$, has 91.18% verification rate, which is higher than all the holistic methods in the literature. In addition, this classifier shows a better performance than the "T-shaped ECG Classifier Bunch" (90.36%) [3] which

also utilizes CG features with holistic representation. Our approach avoids local information loss which is the case in holistic representation. Moreover, downscaling process provides robustness against individual pixel noises. Therefore, local face modelling that we follow outperforms the holistic one.

The experimental results on individual local block classifiers show that the best performances are obtained by the ones generated from the eye-nose-mouth region. This proves that eye-nose-mouth region is the most discriminative part of face images. As expected, classifiers corresponding to upper and lower parts of the face image show the worst performance since forehead and chin does not provide discriminative information and also these regions are exposed to relatively more noise.

The experimental results on individual CG classifiers show that the best performance is obtained by a conventional Gabor wavelet, $\Gamma_{SS+LLR,4}$. This indicates that a typical face image gives better responses to linear (or straight) Gabor filters than curvature Gabor filters. Even though eyes, nose and mouth as the most discriminative parts of a face show curvature characteristics, a typical face in overall shows more linear characteristics. This confirms why most of the previous studies have used the conventional Gabor wavelets. On the other hand, the fusion of linear and curvature Gabor classifiers improves the results dramatically. This proves that curvature Gabor features provide significantly complementary information to conventional Gabor features.

The experimental results on selection of CG classifiers show that classifier selection is crucial, because the redundant classifiers decrease the overall performance. About the classifier selection criterion, the verification rate of a classifier alone is not the best measure to determine the classifiers to be fused because the idea is to find the best combination providing as much complementary information as possible. The set of classifiers having higher verification rates may not satisfy this constraint.

The experimental results on individual CG classifiers and on fusion of selected CG classifiers indicate that LLR-based fusion is a powerful classifier fusion technique when it is performed for fusion of image classifiers. This means that if the classifiers

are not successful enough, then the LLR score conversion does not provide performance improvement because LLR uses directly the decisions of classifiers to increase the separability between different classes. Consequently, since the image classifiers are much more reliable than the local block classifiers, LLR-based fusion of image classifiers improves the overall performance.

The experimental results on fusion of local appearance models show that Histogram of Gradient-based classifier improves the verification performance when it is fused with DCT-based classifier. On the other hand, when HOG is fused with the final CG classifier, it decreases the performance, which means it does not provide complementary information. As opposed to HOG, DCT-based classifier provides extra evidence when it is fused with the final CG classifier. This proves that CG features and DCT features are complementary.

6. FUTURE WORK

As future directions, we first plan to run our system on a database different than FRGC to show the robustness of the proposed approach. More specifically, we plan to run our system on Labeled Faces in the Wild (LFW) database [59] because there is no system which achieves high performance on both FRGC and LFW, and if we succeed, we will prove that the proposed approach is a generic face recognition algorithm.

In addition, we aim to improve our classifier fusion algorithms. An alternative would be to look for other sampling techniques for the training of LLR and PLSR in order to avoid possible over fitting, e.g., random forests. Another alternative would be to use Gaussian mixture modelling instead of using single Gaussian modelling for estimating the likelihood in LLR-based fusion. Lastly, we also think about using advanced likelihood estimation methods like Kernel Density Estimation (KDE).

In order to make use of CG features more efficiently and effectively, we plan to conduct a comprehensive analysis on the signal processing properties of CG wavelets. For example, we could consider using Gabor phases in addition to / instead of Gabor magnitudes. Besides, we aim to analyze the effects of different block sizes and shapes. For instance, we could use smaller square blocks on the eyes and a greater rectangular block on the mouth by considering the shapes of facial components.

Moreover, we intent to make the system more robust by combining holistic classifiers such as Fourier features-based or DCT features-based holistic classifiers with the existing local classifiers as in [37]. We could also try including a graph matching-based classifier similar to the one mentioned in [42].

We also plan to examine our system further on its sensitivity to pose and illumination by running the system on a more specific database like Yale Face Database B [60].

REFERENCES

1. Turk, M. and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.

2. Swets, D. L. and J. J. Weng, ""Using Discriminant Eigenfeatures for Image Retrieval", *IEEE Transactions on Pattern Analysis And Machine Intelligence*, Vol. 18, pp. 831–836, 1996.

3. Hwang, W., X. Huang, K. Noh and J. Kim, "Face Recognition System Using Extended Curvature Gabor Classifier Bunch for Low-Resolution Face Image", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 15–22, 2011.

4. Zhao, W., R. Chellappa, A. Rosenfeld and P. J. Phillips, "Face Recognition: A Literature Survey", *ACM Computing Surveys (CSUR)*, Vol. 35, pp. 399–458, 2003.

5. Jafri, R. and H. R. Arabnia, "A Survey of Face Recognition Techniques", *Journal of Information Processing Systems*, Vol. 5, No. 2, pp. 41–68, 2009.

6. Belhumeur, P., J. Hespanha and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711–720, 1997.

7. Pentland, A., B. Moghaddam and T. Starner, "View-based and Modular Eigenspaces for Face Recognition", *Proceedings IEEE International Conference Computer Vision and Pattern Recognition*, pp. 84–91, 1994.

8. Kim, K. I., K. Jung and H. J. Kim, "Face Recognition Using Kernel Principal Component Analysis", *IEEE Signal Processing Letters*, Vol. 9, No. 2, pp. 40–32, 2002.

9. Yang, M. H., N. Ahuja and D. Kriegman, "Face Recognition Using Kernel Eigenfaces", *Proceedings of IEEE International Conference on Image Processing*, pp. 37–40, 2000.

10. Yang, M. H., N. Ahuja and D. Kriegman, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods", *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215–221, 2002.

11. Kim, H. C., D. Kim and S. Y. Bang, "Face Recognition Using LDA Mixture Model", *Pattern Recognition Letters*, Vol. 24, No. 15, pp. 2815–2821, 2003.

12. Bartlett, M., J. Movellan and T. Sejnovski, "Face Recognition by Independent Component Analysis", *IEEE Transactions on Neural Networks*, Vol. 13, No. 6, pp. 1450–1464, 2002.

13. Moghaddam, B., T. Jebara and A. Pentland, "Bayesian Face Recognition", *Pattern Recognition*, Vol. 33, No. 11, pp. 1771–1782, 2000.

14. Lai, J., P. Yuen and G. Feng, "Face Recognition Using Holistic Fourier Invariant Features", *Pattern Recognition Letters*, Vol. 34, No. 1, pp. 95–109, 2001.

15. Hwang, W., G. Park, J. Lee and S. Kee, "Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set", *Proceedings IEEE International Conference Computer Vision and Pattern Recognition*, pp. 1574–1581, 2006.

16. Hafed, Z. and M. Levine, "Face Recognition Using the Discrete Cosine Transform", *International Journal of Computer Vision*, Vol. 43, No. 3, pp. 167–188, 2001.

17. Savvides, M., J. Heo, R. Abiantun, C. Xie and B. V. K. V. Kumar, "Class Dependent Kernel Discrete Cosine Transform Features for Enhanced Holistic Face Recognition in FRGC-II", *Proceedings International Conference Acoustics, Speech and Signal Processing*, Vol. 2, pp. 185–188, 2006.

18. Kumar, V., M. Savvides and C. Xie, "Correlation Pattern Recognition for Face Recognition", *Proceedings of the IEEE Transactions on Neural Networks*, Vol. 94, No. 11, pp. 1963–1976, 2006.

19. Liu, C. and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", *IEEE Transactions on Image Processing*, Vol. 11, pp. 467–476, 2002.

20. Kumar, V., M. Savvides and C. Xie, "Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 5, pp. 725–737, 2006.

21. Mandal, T., A. Majumdar and Q. M. J. Wu, "Face Recognition by Curvelet Based Feature Extraction", *International Conference on Image Analysis and Recognition*, Vol. 4633, pp. 806–817, 2007.

22. Mandal, T. and Q. M. J. Wu, "Face Recognition Using Curvelet-based PCA", *International Conference on Pattern Recognition*, 2008.

23. Meynet, J., V. Popovici and J.-P. Thiran, "Face Detection with Boosted Gaussian Features", *Pattern Recognition*, Vol. 40, pp. 2283–2291, 2007.

24. Frey, B. J., A. Colmenarez and T. S. Huang, "Mixtures of Local Linear Subspaces for Face Recognition", *Proceedings IEEE International Conference Computer Vision and Pattern Recognition*, 1998.

25. Kim, H. C., D. Kim and S. Y. Bang, "Face Recognition Using the Mixture-of-eigenfaces Method", *Pattern Recognition Letters*, Vol. 23, No. 13, pp. 1549–1558, 2002.

26. Penev, A. and J. Atick, "Local Feature Analysis: A General Statistical Theory for Object Representation", *Network: Computation in Neural Systems*, Vol. 7, pp.

477–500, 1996.

27. Timo, A., H. Abdenour and P. Matti, "Face Recognition with Local Binary Patterns", *In Proceedings European Conference on Computer Vision*, pp. 469–481, 2004.

28. Deng, H.-B., L.-W. Jin, L.-X. Zhen and J.-C. Huang, "A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA", *IEEE Transactions on Image Processing*, Vol. 11, pp. 467–476, 2002.

29. Sang, N., J. Wu and K. Yu, "Local Gabor Fisher Classifier for Face Recognition", *In Proceedings of Fourth International Conference on Image and Graphics*, pp. 620–626, 2007.

30. Gao, H., H. K. Ekenel, M. Fischer and R. Stiefelhagen, "Multi-resolution Local Appearance-Based Face Verification", *Proceedings International Conference on Pattern Recognition*, pp. 1501–1504, 2010.

31. Ekenel, H. K. and R. Stiefelhagen, "Local Appearance-based Face Recognition Using Discrete Cosine Transform", *In Proceedings of Thirteenth European Signal Processing Conference*, 2005.

32. Zhang, X. and C. Podilchuk, "Face Recognition Using DCT-based Feature Vectors", *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4, pp. 2144–2147, 1996.

33. Nefian, A., *A Hidden Markov Model-based Approach for Face Detection and Recognition*, Ph.D. Thesis, Georgia Institute of Technology, 1999.

34. Chen, W., M. J. Er and S. Wu, "PCA and LDA in DCT Domain", *Pattern Recognition Letters*, Vol. 26, No. 15, pp. 2474–2482, 2005.

35. Ekenel, H. K. and R. Stiefelhagen, "Analysis of Local Appearance-based Face

Recognition: Effects of Feature Selection and Feature Normalization", *Proceedings of CVPR Biometrics Workshop*, 2006.

36. Liu, Z. and C. Liu, "Robust Face Recognition Using Color Information", *Advances in Biometrics*, Vol. 5558, pp. 122–131, 2009.

37. Su, Y., S. Shan, X. Chen and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition", *IEEE Trans. on Image Processing*, Vol. 18, No. 8, pp. 1885–1896, 2009.

38. Fang, Y., T. Tan and Y. Wang, "Fusion of Global and Local Features for Face Verification", *In Proceedings of IEEE International Conference on Pattern Recognition*, pp. 382–385, 2002.

39. Kim, T., H. Kim, W. Hwang and J. Kittler, "Component-based LDA Face Description for Image Retrieval and MPEG-7 Standardisation", *Image Vision Computing*, Vol. 23, No. 7, pp. 631–642, 2005.

40. Kim, C., J. Oh and C. Choi, "Combined Subspace Method Using Global and Local Features for Face Recognition", *In Proceedings of IEEE International Joint Conference on Neural Networks*, Vol. 4, pp. 2030–2035, 2005.

41. Lee, Y., K. Lee and S. Pan, "Local and Global Feature Extraction for Face Recognition", *In Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication*, pp. 219–228, 2005.

42. Wiskott, L., J.-M. Fellous, N. Kruger and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 775–779, 1997.

43. Lee, H. J., H. J. Kim and W. Y. Kim, "Face Recognition Using Component-based DCT/LDA", *Proceedings of IWAIT*, 2005.

44. Pentland, A., B. Moghaddam and T. Starner, "View-based and Modular Eigenspaces for Face Recognition", *Proceedings IEEE International Conference Computer Vision and Pattern Recognition*, pp. 84–91, 1994.

45. Zhang, W., S. Shan, W. Gao and X. Chen, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-statistical Model for Face Representation and Recognition", *In Proceedings of IEEE International Conference on Computer Vision*, pp. 786–791, 2005.

46. Tan, X. and B. Triggs, "Fusing Gabor and LBP Feature Set for Kernel-based Face Recognition", *IEEE International Workshop on Analysis and Modeling of Face and Gestures*, pp. 235–249, 2007.

47. Shen, L., *Recognizing Faces - An Approach Based on Gabor Wavelets*, Ph.D. Thesis, University of Nottingham, 2005.

48. Daugman, J., "Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-dimensional Visual Cortex Filters", *Journal of Opt. Soc. Amer.*, Vol. 2, No. 7, pp. 1160–1169, 1985.

49. Peters, G., N. Kruger and C. von der Malsburg, "Learning Object Representations by Clustering Banana Wavelet Responses", *Proceedings 1st International Workshop Statistical Techniques in Pattern Recognition*, pp. 113–118, 1997.

50. Ahmed, N., T. Natarajan and K. R. Rao, "Discrete Cosine Transform", *IEEE Transactions on Computers*, Vol. 23, No. 1, pp. 90–93, 1974.

51. Stark, H. and J. W. Woods, "Probability, Random Processes, and Estimation Theory for Engineers", *Prentice-Hall*, 1994.

52. Ekenel, H. K., *A Robust Face Recognition Algorithm for Real-World Applications*, Ph.D. Thesis, Karlsruhe Institute of Technology, 2009.

53. Pudil, P., J. Novovicová and J. Kittler, "Floating Search Methods in Feature Selection", *Pattern Recognition Letters*, Vol. 15, pp. 1119–1125, 1994.

54. Cover, T. M. and T. A. Joy, "Elements of Information Theory", *New York: Wiley*, 1991.

55. Abdi, H., "Partial Least Squares Regression (PLS-regression)", *Encyclopedia for Research Methods for the Social Sciences. Thousand Oaks (CA): Sage*, pp. 725–728, 2003.

56. Phillips, P. J., P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge", *Proceedings IEEE International Conference Computer Vision and Pattern Recognition*, Vol. 1, pp. 947–954, 2005.

57. Dalal, N., B. Triggs and C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance", *In Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, pp. 886–893, 2005.

58. Yang, J. and C. Liu, "Color Image Discriminant Models and Algorithms for Face Recognition", *IEEE Transactions on Neural Networks*, Vol. 19, pp. 2088–2098, 2008.

59. Huang, G. B., M. Ramesh, T. Berg and E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, Tech. rep., University of Massachusetts, Amherst, 2007.

60. Georghiades, A., P. Belhumeur and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", *IEEE Trans. Pattern Anal. Mach. Intelligence*, Vol. 23, No. 6, pp. 643–660, 2001.