2012

Design and Analysis of A New Illumination Invariant Human Face Recognition System

Aryaz Baradarani
University of Windsor

DESIGN AND ANALYSIS OF A NEW ILLUMINATION INVARIANT HUMAN FACE RECOGNITION SYSTEM

by

Aryaz Baradarani

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Electrical Engineering at the
University of Windsor

Windsor, Ontario, Canada

2012

Design and Analysis of A New Illumination Invariant Human Face Recognition System

by

Aryaz Baradarani

APPROVED BY:

Dr. Sridhar Lakshmanan
Department of Electrical Engineering
University of Michigan, Dearborn, MI, USA

Dr. Yung H. Tsin
School of Computer Science
University of Windsor

Dr. Maher A. Sid-Ahmed
Department of Electrical Engineering
University of Windsor

Dr. Majid Ahmadi
Department of Electrical Engineering
University of Windsor

Dr. Q. M. Jonathan Wu
Department of Electrical Engineering
University of Windsor

Dr. Kenneth Cramer, Chair of Defense
Faculty of Graduate Studies

October 12, 2012

Declaration of Co-Authorship and Previous Publication

I hereby declare that this dissertation incorporates material that is the outcome of research under the supervision of Prof. Q. M. Jonathan Wu. In Section 3.6.1, Pankaj Mendapara collaborated under the supervision of Prof. Q. M. Jonathan Wu to elaborate an application. I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my dissertation, and have obtained a written permission from each of the co-authors to include the above materials in my dissertation. I certify that, with the above qualification, this dissertation and the research to which it refers is the product of my own work. This dissertation includes the original papers that have been previously submitted to or published in peer reviewed journals and conferences as follows:

Thesis Chapter	Publication Title	Status
Chapter 2	A. Baradarani, Q. M. Jonathan Wu, M. Ahmadi, An efficient illumination invariant face recognition framework via illumination enhancement and DD-DT$\mathbb{C}$WT filtering, Pattern Recognition, 46 (1) (2013) 57–72.	Published
Chapter 3	A. Baradarani, Q. M. Jonathan Wu, M. Ahmadi, P. Mendapara, Tunable halfband-pair wavelet filter banks and application to multifocus image fusion, Pattern Recognition, 45 (2) (2012) 657–671.	Published
Chapter 4 and Chapter 2	A. Baradarani, Q. M. Jonathan Wu, Design of an efficient illumination invariant face recognition system via resonance based image representation.	Submitted
Chapter 2	A. Baradarani, Q. M. Jonathan Wu, M. Ahmadi, Illumination suppression for illumination invariant face recognition.	Accepted
Chapter 4	A. Baradarani, Q. M. Jonathan Wu, Illumination invariant human face recognition: frequency or resonance?	Accepted

I certify that I have obtained a written permission from the copyright owners to include the published materials in my dissertation. I certify that the material describes the work completed during my registration as a graduate student at the University of Windsor. I declare that, to the best of my knowledge, my dissertation neither infringes upon anyone's copyright nor violates any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owners to include such materials in my dissertation.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

In this dissertation we propose the design and analysis of a new illumination invariant face recognition system. We show that the multiscale analysis of facial structure and features of face images leads to superior recognition rates for images under varying illumination. We assume that an image $I(x, y)$ is a black box consisting of a combination of illumination and reflectance. A new approximation is proposed to enhance the illumination removal phase. As illumination resides in the low-frequency part of images, a high-performance multiresolution transformation is employed to accurately separate the frequency contents of input images. The procedure is followed by a fine-tuning process. After extracting a mask, feature vector is formed and the principal component analysis (PCA) is used for dimensionality reduction which is then proceeded by the extreme learning machine (ELM) as a classifier. We then analyze the effect of the frequency selectivity of subbands of the transformation on the performance of the proposed face recognition system. In fact, we first propose a method to tune the characteristics of a multiresolution transformation, and then analyze how these specifications may affect the recognition rate. In addition, we show that the proposed face recognition system can be further improved in terms of the computational time and accuracy. The motivation for this progress is related to the fact that although illumination mostly lies in the low-frequency part of images, these low-frequency components may have low- or high-resonance nature. Therefore, for the first time, we introduce the resonance based analysis of face images rather than the traditional frequency domain approaches. We found that energy selectivity of the

subbands of the resonance based decomposition can lead to superior results with less computational complexity. The method is free of any prior information about the face shape. It is systematic and can be applied separately on each image. Several experiments are performed employing the well known databases such as the Yale B, Extended-Yale B, CMU-PIE, FERET, AT&T, and LFW. Illustrative examples are given and the results confirm the effectiveness of the method compared to the current results in the literature.

Acknowledgements

First of all, I would like to thank my advisor Dr. Q. M. Jonathan Wu; by providing a fruitful ground, encouragement, supervision, and an excellent academic environment, he made my Windsor experience unique. I am sincerely thankful to Dr. Wu for this opportunity.

Also, I would like to thank the remaining members of my Ph.D. committee, Dr. Maher Sid-Ahmed, Dr. Majid Ahmadi, Dr. Sridhar Lakshmanan, and Dr. Yung H. Tsin for their constructive comments, valuable suggestions and great help in preparation of this dissertation.

My appreciation extends to Andria Ballo from the Department of Electrical Engineering for her reminders, advice, and assistance in various administrative tasks. I am also thankful to Enrique Chacon, from the International Student Centre, who has always made time for all students in his busy schedule and has given key feedbacks at several occasions. It is not possible to name all friends and the great time we shared together, however, I would like to thank my colleagues in our Computer Vision and Sensing Systems Lab and the Essex Hall for their friendship in the past four years.

Finally, I would like to thank my mother and my father for their encouragement and unconditional love; without them none of this was possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DNA	Deoxyribonucleic Acid
PCA	Principal Component Analysis
2D-PCA	Two-dimensional Principal Component Analysis
ELM	Extreme Learning Machine
CMU-PIE	Carnegie Mellon University Pose, Illumination and Expression
FERET	Face Recognition Technology
AT&T	American Telephone and Telegraph
ORL	Olivetti Research Lab
LFW	Labeled Faces in the Wild
HE	Histogram Equalization
GIC	Gamma Intensity Correction
QI	Quotient Image
SQI	Self-Quotient Image
DCT	Discrete Cosine Transform
CVT	Curvelet Transform
DWT	Discrete Wavelet Transform
WVS	Weighted Voting Scheme
DVS	Democratic Voting Scheme
S–P	Steerable Pyramid
LBP	Local Binary Patterns
CW	Choi-Williams
THP	Tunable Halfband Pair
NVM	Number of Vanishing Moments
PSD	Positive Semidefinite
MFIF	Multifocus Image Fusion
SFF	Shape-From-Focus
SNR	Signal-to-Noise Ratio
PSNR	Peak Signal-to-Noise Ratio
HMM	Hidden Markov Model
MCA	Morphological Component Analysis

LTP	Local Ternary Patterns
HD-DWT	Higher-Density Discrete Wavelet Transform
pro	Proposed
log	Logarithm
CDF	Cohen-Daubechies-Feauveau
HPFB	Halfband-Pair Filter Bank
QMF	Quadrature Mirror Filter
LDA	Linear Discriminant Analysis
SVM	Support Vector Machines
SLFN	Single-hidden Layer Feedforward Neural Networks
FNN	Feedforward Neural Networks
NTR	Number of Training Samples
NTS	Number of Testing Samples
TIN	Training Image Name
DD	Double-Density
PP	Preprocessing
CPU	Central Processing Unit
RAM	Random Access Memory
MI	Mutual Information
SSIM	Structural Similarity
TQWT	Tunable Q-factor Wavelet Transform
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
RADWT	Rational-dilation Wavelet Transform
BW	Band Width
MCA	Morphological Component Analysis
DD-DWT	Double-Density Discrete Wavelet Transform
DD-DT$\mathbb{C}$WT	Double-Density Dual-tree Complex Wavelet Transform
sft	Soft-thresholding
sgn	Signum Function
ntp	No-tuning-parameters
RES	Resonance
ms	Millisecond
min	Minimum
Tenen	Tenengrade
GLV	Gray Level Variance
SML	Sum of Modified Laplacian
Mexh	Mexican Hat
MSE	Mean Square Error

List of Symbols

$I(x, y)$	Pixel (x,y) of Image I
$R(x, y)$	Reflectance at (x,y)
$L(x, y)$	Illumination at (x,y)
t	Transpose
ξ	Tuning Parameter
ϵ	Tuning Parameter
λ	Tuning Parameter
α	Real Number
β	Real Number
$\mathcal{F}$	Primary Bank
$H(z)$	Digital Filter
$h(t)$	Impulse Response of H(z)
$\phi_{(.)}(t)$	Scaling Function
$\psi_{(.)}(t)$	Wavelet
z	z-transform Parameter
ω	Angular Frequency
j	Square Root of -1
$\Psi_{(.)}(\omega)$	Fourier Transform of $\psi_{(.)}(t)$
ℓ	Shift
f	Filtered
$*$	Convolution
τ	Light Source and Camera Axis Angle
n	Input Neurons
$\widetilde{N}$	Hidden Layer Neurons
k	Output Neurons
N	Integer Number
$\mathbf{x}_i$	ith Input

$\boldsymbol{\gamma}_i$	ith Estimated Output
$\mathfrak{g}(x)$	Activation Function
$f_{(.)}(\boldsymbol{b})$	Function of Polynomial Coefficients
$g_{(.)}(\boldsymbol{a})$	Function of Polynomial Coefficients
$\mathbf{w}_i$	Weight Vectors Input/Hidden-Node
β_i	Weight Vectors Hidden-Node/Output
$\boldsymbol{\Upsilon}$	Hidden-Layer Output Matrix
$\mathbf{T}$	Output Matrix
η	Error Function
$\mathbf{R}$	Real Vectors
$\Lambda(x)$	Filter Kernel
$\Omega(x)$	Filter Kernel
$\boldsymbol{a}$	Polynomial Coefficients
$\boldsymbol{b}$	Polynomial Coefficients
$P(z)$	Polynomial in z-domain
$T(x)$	Polynomial in x-domain
m	Number of Zeros at $\omega{=}\pi$
$\mathcal{L}$	Filter Length
$\mathcal{D}$	Finite and Bounded Interval
$p(x)$	Univariate Polynomial
$s(x)$	Univariate Polynomial
$\mathbb{R}$	Set of Real Numbers
v	Monomial Vector
$\mathbb{S}_+$	Symmetric PSD Square Matrices
δ	Ripple Size
mx	Maximally Flat
s.a.	Stopband Attenuation
$\bar{\omega}$	Normalized Frequency
$D(x, y)$	Distance Function
$x(n)$	Discrete-time Input Signal
$u(n)$	Discrete-time Output Signal
f_s	Sampling Frequency
π	Mathematical Constant 3.1415
Q	Quality Factor
r	Redundancy
J	Number of Stages
$\mathbf{O}$	Objective Function
ς	Real Constant
$\mathbf{w}$	Weight Vector
$\mathcal{S}$	Array of Matrices
$\mathfrak{T}$	Threshold Value
$\mathfrak{c}$	Class Vector
$\mathbf{c}$	Feature Vector

Chapter 1

Introduction

1.1 Face Recognition

In computer vision and pattern recognition the term biometrics is specifically used to deal with human identification and recognition based on human characteristics including but not limited to facial information, fingerprint, iris and retina scanning techniques, voice, ear, signature, or DNA information.

It is evident that in the past two decades or so, face recognition has become one of the important and popular tools in biometrics due to relatively lower computational cost and reliability of the application in real life. For instance, although DNA-based recognition excels in terms of accuracy, it is still an expensive technology for daily usage. Perhaps fingerprint recognition and identification is the most frequently used biometrics in a wide range of applications from public security systems to personal items such as laptops and touch screen games. However, fingerprint analysis needs a physical contact to take a sample from the subject. Keeping these points in mind, face recognition is neither that expensive as the DNA-based approaches nor in charge of a mandatory contact with the subjects to be recognized. Therefore, in the last few years, face recognition techniques have been developed rapidly not only in general aspects of recognition purposes, but also narrowed down to specific categories such

as pose invariant, illumination invariant, and occlusion invariant, few to mention. In order to reach higher accuracy in uncontrolled environment, these performance degrading factors need to be well investigated and developed.

1.2 Performance Degrading Factors

In general, performance of a face recognition system is directly related to the environmental conditions and to the quality of features. By the term environmental conditions, it is mostly refereed to the controlled and uncontrolled conditions under which an image is taken. Nowadays the regular face recognition under controlled environment is a straightforward task and several outstanding techniques are available both in the literature and in basic practical and educational applications. However, in reality and in order to deal with real life situations and applications one needs to come up with several performance degrading factors caused by the uncontrolled environmental and photographic conditions. The main influencing factors that can affect the recognition accuracy and computational complexity of any face recognition system are namely the pose, illumination, occlusion, and blur. While each of these factors has their own nature to degrade the quality of the system, performance of a face recognition technique is significantly changed under various illumination and pose. Fig. 1.1 visually presents the difference between the factors. In this figure, the first picture in (a) shows a sample image without a notable illumination, pose, blur or occlusion effects. Other images (b)–(e), however, show the sample image with illumination effects, pose variation, blur and occlusion from left to right, respectively. Although by looking at figures one may simply recognize the subject in (b)–(e) if compared to (a), this is not an easy task for an artificial intelligence system to show the same performance.

Generally the two factors, illumination and pose, are widely known as the most influencing factors in deteriorating the quality of features. Illumination is in fact

(a) (b) (c) (d) (e)

Figure 1.1 – Performance degrading factors. (a) A sample image with no variation in pose, illumination, blur, and occlusion; The same subject in (a) under: (b) high illumination effects; (c) pose variation; (d) blur; (e) occlusion.

seduces the feature extraction algorithms due to the redundant shapes and shadows generated by knobs and ditches of a typical face under various illuminations. Furthermore, partially too bright and dark images require a preprocessing step even if the lighting angle is not too wide. Similarly, any remarkable variation in pose and position of head can significantly change the recognition rates. Ideally, a frontal view with no up-down or left-right position is considered as the most convenient pose to correctly recognize a face image. However, in real applications, face images contain notable changes to sides in several directions. It can considerably reduce the number of informative and useful feature pixels and eventually leads to lower recognition results. In most cases a 3D face image reconstruction step may be required to map the shape to a fixed surface to obtain a normalized applicable face image. Face image normalization is in fact a process to geometrically and photometrically normalize an image. The geometrical normalization contains techniques such as face cropping and eye-alignment that can provide focused information about important and desired features in a fixed position for all images in a database. Likewise, the photometric normalization mostly deals with the illumination and consists of several techniques to initially enhance and improve the lighting condition effects. In addition to illumination and pose, as it can be seen in Fig. 1.1 (d) and (e), blur and occlusion are two other performance affecting factors need to be taken into consideration. There are plenty of blur invariant and occlusion invariant methods developed for face recognition purposes. For instance, in [66] the authors proposed a weighted voting scheme

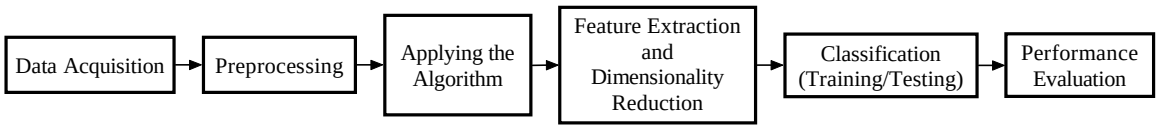


Figure 1.2 – Block diagram and structure of a typical face recognition system.

that takes an image apart into small blocks first. Based on the amount of occlusion on each block a weighted voting procedure is applied to blocks to reduce the effect of unwanted occlusion. Although image deblurring is relatively an old problem in image processing, it has been quite attractive subject for researchers in face recognition, medical imaging, and space image quality enhancement. For the problem of blur invariant face recognition, deblurring techniques or blur invariant feature extraction algorithm are employed. A traditional approach to deblur an image is to find a mathematical model to describe how the image was blurred. Otherwise, one may need to estimate the blur for which stochastic and optimization methods are on the table. While there is no way to get rid of unwanted influencing factors, it is possible to improve the efficiency of face recognition methods via the feature extraction phase and classification step. As it is seen in Fig. 1.2, a face recognition system consists of several sections for which the system works properly if all the involved sections maintain a satisfactory performance. Feature extraction and classification techniques are briefly discussed in the following subsections.

1.3 Feature Extraction

The Merriam-Webster Learner's Dictionary defines the word extraction as "The act or process of getting something by pulling it out, forcing it out, etc". The same dictionary describes the word feature as "An interesting or important part, quality, ability, etc. A part of the face such as the eyes, nose, or mouth". Perhaps the term feature extraction should refer to pulling out the eyes, nose, or mouth from a face

image. Although it might be a sufficient definition at first glance, the key point here is how to pull out the desired parts from an image correctly. On the other hand, selecting and extracting proper features from a given image is considered as an important part of any recognition technique. Technically speaking, a set of variables that contain discriminating and characterizing information about a face image are defined as features. Consequently, a selection from such features to construct an array is called a feature vector. Feature extraction is indeed a mathematical transformation applied on the data to generate the feature vector. Plenty of techniques can be found in the literature and text books. The most widely used transformations are briefly discussed as follows. It should be noted that the linear transformations are frequently used in different applications due to lower computational complexity and adequate performance.

1.3.1 Principal Component Analysis

Principal component analysis (PCA) is one of the oldest methods and at the same time is the most commonly used dimensionality reduction and feature extraction algorithm in the literature. The main idea behind the PCA is to map the data from a higher to a lower dimensional space such that the key information is preserved. PCA takes advantage of a linear transformation to keep the data with significant variance using the least number of elements to construct the low dimensional space vector. It means if a feature possesses higher variation then more information can be obtained from that feature. The data projection is based on the well known least-squares minimization and one can show that the optimal approximation of a sample feature vector is achieved by a linear combination of independent vectors that project the feature vector onto the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of the data. Traditionally, the eigenvectors of the covariance matrix are called as the principal axes and the mapping of the data on these principal axes are named the principal components. These axes are interpreted as dimensions and

thus the dimensionality reduction is usually referred to as the procedure of preserving only the axes that contain most of the high variance information. Implementation and extensive details on PCA can be found in [44] for instance.

1.3.2 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is basically to consider the information that may not be necessarily in the same direction as the one with large variations. In other words, the distances between different clusters are maximized and the sum of distances between objects in the same cluster is minimized. This helps to take the information in directions with small variation in the account. The main drawback of LDA is related to the fact that LDA maps a sample vector from the original source onto a feature vector. The dimension of the vector can be at most close to but not equal to the number of classes. This is independent from the dimensionality of the main data source or pattern. The above mentioned limitation is considered as a serious restriction. Actually, it can lead to lose of important information when a problem contains higher dimensions for which the distribution of desired information may be still meaningful in higher dimensions. In such cases any reduction in the number of dimensions close to but not equal to the number of classes results in missing the information. It is worth pointing out that LDA is less effective for the patterns with overlapped distributions. Linear discriminant analysis or Fisher's linear discriminant is discussed in detail in [33] and [34]. The popularity of PCA is due to the property that if the distinct information remains through the variances then PCA performs better than LDA.

1.3.3 Other Feature Extraction Methods

Principal component analysis and linear discriminant analysis are widely used in the literature, however, performance of a feature extraction method may be changed

from one application to another. On the other hand, depending on the nature of an application and the type of features required to be extracted, several researchers have developed different techniques. Basically, it can be considered as defining a feature vector whose elements are determined based on some strategies. For example stochastic moments, Zernike moments, wavelets and other multiresolution based transformations are some of the examples in which the feature vector is constructed in view of the desired properties of each tool in accordance with the corresponding application that the method is aimed to be applied. In [16] the authors presented a rotation and gray scale transform invariant texture recognition method employing the combination of wavelet subbands of a quadrature mirror filter bank and hidden Markov model (HMM). The subbands obtained by decomposing the texture image are used to derive the gray scale transform invariant features based on the statistics of the first order distribution of gray levels of each subband. The procedure is then followed by modeling a feature vector via a hidden Markov model for each class of textures. Zernike moments is another possible useful feature construction method. A modified direct method has been shown in [36] for the computation of the Zernike moments, that is, a novel factorial-free direct method can be developed based on the use of the Stirlings approximation formula. While the obtained moments are not exactly identical to the ones that can be achieved via the direct method, they are sufficiently acceptable in relevant applications.

It should be noted that dimensionality reduction methods and direct feature vector generating techniques are different in the way they face a pattern. A dimensionality reduction method may be considered as a type of filtering stage as the data are mapped from a higher dimensional space to a lower dimension with less number of features. A feature vector generation algorithm, however, may take advantage of some transformation or statistical approaches to customize and tailor some features which may not necessarily configure a lower dimensional feature vector.

1.4 Classification

In pattern recognition and computer vision the term classification is used to classify and categorize the objects in a database into one of the classes or feature types, namely known as classes. A classifier at first look may seem an intelligent machine that can be trained to learn particular patterns. As it can be seen in Fig. 1.2, classification is usually the last step in a recognition system as the performance evaluation phase has nothing to do with the input data and the information regarding the application. In fact, before being eligible to eventually choose a classifier and its respective training method, one needs to first pass the data acquisition, preprocessing, core processing, and feature extraction and dimensionality reduction steps to sift the most important and key information for better training and classification. In its classic form, classification is actually a mapping and approximation technique in which a wide range of approaches from softcomputing and optimization to mathematical techniques are involved. In the following sections some of the most commonly used classifiers are briefly described. There are many classifiers each of which has its own advantages and disadvantages. Bayes, naive Bayes, k-nearest neighbors, neural networks, support vector machines, hidden Markov model, and extreme learning machine are among the well known and most frequently used classifiers in the literature. In the following section some of the classifiers are discussed briefly to have an idea about the nature of a classifier. The final decision regarding the choice of a proper classifier in an application lies in the type of data and the way a recognition system deals with the information provided to its different parts.

1.4.1 Bayes Classifier

Bayes classifier is a statistical method that works based on allocating certain probabilities to patterns belonging to a particular class. In fact probability distribution of data are considered for class assignment where the distributions must be learned

from the data as it is usually not the case to have a sufficient estimation about these distributions in advance. In a Bayes classifier it is generally assumed that, given the class variable, the presence or absence of a specific feature belonging to a class does not depend on the presence or absence of other features. One of the advantages of the Bayes classifier is that a small training dataset is quite enough to estimate the statistic parameters required for classification. Let us consider a feature vector of dimension N_1 as $\mathbf{c} = \{\mathrm{c}_i \mid i = 1, \ldots, N_1\}$ where c_i denote the features of a feature vector $\mathbf{c}$. A feature vector $\mathbf{c}$ needs to be addressed by a class among the existing N_2 classes $\mathfrak{c}_j$ for $j = 1, \ldots, N_2$. Using the conditional probability of $\mathbf{P}(\mathbf{c}|\mathfrak{c}_j)$, the feature vector $\mathbf{c}$ that belongs to class $\mathfrak{c}_j$ is a representative randomly selected from a conditional probability distribution on class $\mathfrak{c}_j$. This is what referred to as the so-called likelihood in the literature. Using the Bayes theorem, the prior likelihood of each class is related to the conditional probability given the feature vector probability distribution as

$$\mathbf{P}(\mathfrak{c}_j|\mathbf{c}) = \frac{\mathbf{P}(\mathbf{c} \cap \mathfrak{c}_j)}{\mathbf{P}(\mathbf{c})} = \frac{\mathbf{P}(\mathbf{c}|\mathfrak{c}_j).\mathbf{P}(\mathfrak{c}_j)}{\sum_{l=1}^{N_1} \mathbf{P}(\mathbf{c}|\mathfrak{c}_l).\mathbf{P}(\mathfrak{c}_l)}.$$

Although a small set of training data is works out for this type of classification, the main drawback of the Bayes classifier is that it is not always an easy task to determine the likelihood if the original data belong to higher dimensional spaces.

1.4.2 *k*-Nearest Neighbors

Perhaps one of the simplest classifiers for recognition purposes is the so-called k-nearest neighbors (k-NN) method. k-NN actually specifies the decision boundary locally. The notation k is used to denote that each sample is assigned to the majority class of its k closest neighbors. In fact, for a small positive integer k a sample is classified and recognized in one of the classes by a majority vote of its k nearest neighbors where proximity or distance is determined based on Euclidian distance

measure. Detailed explanations are given in [84] for instance. For some given training vectors, k-NN detects k nearest neighbors regardless of the class label and then assigns the feature vector to the class with the maximum number of samples. k-NN is often considered as a lazy learning approach whereas the method only takes the local information into account. On the other hand, given a sample to be classified, the k-NN algorithm figures out the majority class among the k closest points in the training gallery based on the distance calculation. After carrying out the first step, then the new situation classified to a particular class. Although k-NN possess a simple algorithm, easy implementation, remarkably satisfactory performance in several applications, it has some notable limitations that may restrict its application. The main drawbacks with k-NN are related to the necessity of a large memory and execution time for larger training databases as most of the calculations are done at the end of procedure. Furthermore, k-NN is often sensitive to features that are too different from the other features and may seem as redundant data. It may be avoided by voting based feature selection or weighted features that can require more calculations.

1.4.3 Support Vector Machines

Support vector machines known as SVMs [46][98] are one of the mostly used classification in recent years. One of the reasons behind this frequent usage is related to the mechanism of SVMs, that is, a kernel function is used to project the feature space into a higher dimensional space. It is important to note that the classes in the latter space are linearly separable. Therefore, one can find the support vectors in a high dimensional space which is linear. Basically SVMs works based on maximizing the geometric margin of different classes in a database. In its simple form two parallel hyperplanes used to identify the separation boundaries for each class in the feature space. The condition is considered as a deterministic tool that can provide the important and desired vectors, named as support vectors, to find the marginal separation distances. One of the advantages of the SVMs is the ability to separate the classes in

a condition that the boundaries between classes are not linearly separable. Also, unlike some other classifiers, SVMs can perform sufficiently even in higher dimensions. Common assumption regarding the disadvantages of SVMs mostly lies in choice of the kernels. Despite of the inspiring performance of SVMs and the popularity of the classifier, practically, another drawback of the SVMs is the complexity and computational cost of the algorithmic in large scale test data. In the following section a faster and more convenient classifier is introduced.

1.4.4 Extreme Learning Machine

Perhaps among several existing neural networks, feedforward neural networks are one the commonly used ones in various applications. A feedforward neural network includes one input layer that is in contact with external inputs, hidden layers, and an output layer. Feedforward neural networks are mostly preferred in nonlinear mappings that utilize gradient descent approach for weights and bias optimization. In such learning scheme, a small value of learning parameter can simply lead to slow convergence of the learning algorithm. In contrast and vice versa, a higher value of learning parameter may cause instability. In addition, Gradient descent based learning phase of a traditional feedforward neural network is usually a slow and time consuming cycle. Huang et al. proposed an interesting approach in which by using a single hidden layer feedforward neural network the input weights and the hidden layer biases are not necessarily tuned and the hidden layer output matrix can remain unchanged once random values have been assigned to these parameters in the beginning of learning [38][40]. The method which is known as the extreme learning machine (ELM), and works based on feedforward neural networks, performs remarkably faster than other learning techniques. Therefore, ELM has been selected as the classifier used in this dissertation. The required derivations and discussions regarding ELM has been given in Section 2.4.

1.5 Outline of the Dissertation

Despite the fact that illumination is one of the major and crucial performance degrading factors in face recognition and variety of relevant problems, most of the face recognition techniques consider the controlled environmental conditions in general. In the last few years, and due to the fact that traditional face recognition has now reached a satisfactory performance in recognition rate and computational costs, performance degrading issues such as pose, illumination, occlusion, and facial expression changes, have been amongst the hot topics for several authors. The objective and main goal of the research and study in this dissertation is to focus on the design, analyze, and investigation of the illumination invariant face recognition problem to find efficient, reasonable and robust solutions in uncontrolled conditions. The rest of thesis is organized as follows.

In Chapter 2 it is shown that, unlike the commonly used assumption to consider a pure logarithmic based approximation to linearly separate the reflectance and illumination, input images can be assumed as an unknown combination of illumination and reflectance. We then propose a new tunable expression that can be controlled by two parameters to neatly approximate the combination for more flexible separation of the two components. The design procedure proposed in this chapter is based on multiresolution analysis of image components in the frequency domain to efficiently separate the unwanted illumination effects from images under varying illumination. In order to uphold the quality of edges, we have proposed a versatile method, called lifting, which is based on shifting down the raw output image just before mask extraction and subtraction phase. The feature vector construction in our work takes the double-density complex wavelets subbands of the extracted mask. For dimensionality reduction and to prepare training and testing data, principal component analysis (PCA) has been employed. The classification and system performance evaluation cycle of the method is based on the use of the extreme learning machine (ELM). Several

experimental results and simulations are presented in Chapter 2 using the well known databases to elaborate the effectiveness of the proposed approach.

Chapter 3 presents an extensive analysis over the frequency based illumination invariant analysis. Our study and research while working on Chapter 2 and frequency based illumination suppression showed that internal characteristics of a multiresolution transformation may have an impact on the performance of a face recognition system. These intrinsic specifications are basically narrowed down to the factors that can change the quality of the pixels of subbands of a multiresolution transformation which is in fact based on the characteristics of the frequency responses of some digital filters. Frequency selectivity of these filters is mainly related to the transition band width and size of ripples. In Chapter 3 we first propose a general method to deal with the design and control of the above mentioned characteristics. We then design sets of filters with different specifications and frequency selectivity range using the proposed method. We show how frequency selectivity of a multiresolution transformation can affect the performance and recognition rates of the proposed illumination invariant method in Chapter 2.

In Chapter 4, for the first time in this field, we introduce the possibility of investigating the problem of illumination invariant face recognition via the concept of resonance rather than frequency. Frequency domain analysis has been widely used in the literature for different purposes including face recognition. Our main motivation for this research started while facing expressions stating that illumination lies in the low-frequency part of an image. Although we first attacked the illumination invariant human face recognition problem based on the mentioned statement, further investigations regarding this issue left more question at our desk. For instance, if illumination resides in the low-frequency part of images, do these low-frequency contents contain low-resonance or high-resonance components? The concept of resonance is substantially different from the frequency; a low-resonance signal can consists of both low- and high-frequency components. Chapter 4 first draws the reader's attention to the

possibility of resonance based illumination enhancement and then proposes a new and efficient procedure for illumination invariant face recognition via resonance based signal representation and decomposition. Unlike the method in Chapter 2, the resonance based technique is free of tuning parameters, and furthermore, the extra fine-tuning phase has been removed from the procedure. Also, feature vector construction has been reduced to PCA features only.

Summary of contributions and findings have been presented at the end of each chapter, and an overview of the dissertation and future research directions are given in Chapter 5.

Chapter 2

Design of a New Illumination Invariant Face Recognition System

2.1 Literature Review

Several authors have demonstrated a number of techniques for face analysis, recognition and applications [1][10][35][49][66][93][94][97][100][102], few to mention. Illumination is one of the basic characteristics of a visible surface and it provides information for scene interpretation. Recent developments in this field have shown that there is room for improvements. Most of the traditional face recognition algorithms are satisfactory under controlled conditions, however, when dealing with performance degrading issues such as variation in pose, illumination, and facial expression, their accuracy is greatly diminished. As the performance of a face recognition technique is significantly changed under various illumination and lighting effects, illumination is known to be one of the key factors that plays an important role in human face recognition system design.

The approaches proposed for the problem of variable illumination in face recognition are mostly based on illumination modeling, preprocessing, and illumination invariant feature extraction. Face modeling techniques are basically statistical or physical models where linear subspaces with low dimensions form different illumina-

tion conditions to extract the model parameters [8][35][56]. The main limitation of this category is that the images need to construct a linear subspace. In [55], Lee et al. show that there exists configurations of single light source directions so that the obtained images of each object can be directly used as basis images of the linear subspaces effective for face recognition. Another method for face recognition under illumination variation is to simulate the distribution of the images with varying illuminations and generate new training images for face recognition with single frontal views [113]. The approaches in this category may have the downside that several training sample images under varying illumination might be required for training.

The second category, which is known as preprocessing and normalization approaches, is referred to as improving the illumination and lighting conditions in images where in most of the cases there is no need to face models or surface information. Histogram equalization (HE) [85] and gamma intensity correction (GIC) [85] are very well known techniques in this group. Quotient image (QI) [86] and self-quotient image (SQI) [100] are also considered as preprocessing ideas although they implicitly indicate illumination invariants. The quotient image approach eventually determines the ratio between a given test image and linear combination of three other images of the same face under different illumination effects. This ratio is assumed to be illumination invariant. The difference between QI and SQI is that the illumination invariant ratio for SQI is obtained using a given test image and its smoothed version. SQI, which is in fact a multiscale retinex approach, is simpler than QI as it needs a single image only. In [103], Xie and Lam proposed a preprocessing and illumination normalization algorithm for face recognition, in which the intensity of pixels is locally normalized using a small window centered at the corresponding pixel. For most of the methods in this category there is a compromise between the simplicity and performance.

For the illumination invariant based techniques, illumination invariant feature extraction is the main goal. For instance, Chen et al. [18] proposed a discrete cosine transform (DCT) based method where illumination variations are mainly assumed

to lie in the low-frequency subbands. Local binary pattern (LBP) [67] is another illumination invariant approach in which local contrast based image representation is employed to improve illumination losses. In [1], the authors address that several commonly used image representations, such as edge maps that are often assumed to be insensitive to illumination variations, are insufficient for the problem of illumination invariant face recognition.

Recently, several interesting techniques have been demonstrated for face recognition that take the frequency information of images into consideration. Zhang et al. [110] proposed a discrete wavelet transform (DWT) based algorithm using denoising techniques to detect and eliminate the illumination effects. In [14], it has been shown that wavelet based NeighShrink denoising technique, employing two parameters λ_1 and λ_2, leads to higher recognition rates. El Aroussi et al. [30] introduced the use of steerable pyramid (S–P) to compute the statistics of each block by dividing S–P subbands. Nabatchian et al. [66] showed that a simple lowpass filtering followed by a weighted voting scheme (WVS) can offer better results for illumination invariant recognition purposes. In [61], curvelet transform (CVT) was shown to be useful in recognizing changes in facial expression, however, there is no examples of images with significant variations in illumination such as the ones in the Extended-Yale B and CMU-PIE databases. Although some of the mentioned techniques do not address the problem of illumination invariant face recognition unless a preprocessing step is applied first, they emphasize the role of multiresolution analysis and frequency component filtering.

Multiscale representation of signals has been used in a number of image processing and computer vision applications from multiresolution image representation [48] to our recent work on multifocus image fusion and shape-from-focus [4]. In [13], optimization of filter banks is investigated for invariant supervised texture segmentation. Multiscale directional filter bank is discussed extensively in [19] to suppress the aliasing effect as well as to minimize the reduction in frequency resolution where the

problem of aliasing in decimated bandpass images on directional decomposition has been addressed. Multiresolution analysis or multiscale approximation was basically referred to as the theory and design of discrete wavelet transform (DWT) and filter banks, and later on extended to multiwavelets, dual-tree complex wavelet transform (DT-$\mathbb{C}$[1]WT), curvelets, higher-density discrete wavelet transform (HD-DWT), and double-density dual-tree complex wavelet transform (DD-DT$\mathbb{C}$WT). Multiresolution analysis allows the decomposition of a signal into its frequency components known as approximation and details [23][59][60]. It is basically an analysis-synthesis configuration, where a signal is represented in various frequency subbands via the analysis side of a wavelet filter bank. In a perfect reconstruction filter bank the synthesis side in turn can reconstruct the original signal.

The double-density dual-tree complex wavelet transform is a powerful multiresolution tool which is in fact a notable enhancement over the weakness of DWT in general. As the transformation increases the density (number) of the high-frequency subbands, and due to the fact that we need to separate and remove illumination, which mostly lies in the low-frequency part of images, the DD-DT$\mathbb{C}$WT is a reasonable choice of algorithm to reduce the illumination effects in images.

2.2 Motivation and Problem Statement

In this dissertation we propose a novel multiresolution based method to improve the recognition rates for the problem of illumination invariant face recognition. As mentioned earlier, the proposed technique takes advantage of celebrated properties of the DD-DT$\mathbb{C}$WT; shift-invariance, directionality, and extra wavelets imposition, which yields a systematic and easy to implement high-performance framework that works faster than similar methods. The method does not need any prior information regarding the image contents or face shape, and can be applied separately on each

[1]The complex number symbol $\mathbb{C}$ is used to avoid confusion with the CWT for the continuous wavelet transform [78].

image. Experimental results indicate that the technique does not require a complicated feature extraction phase and can be effectively used as a preprocessing scheme in several applications. It is worth pointing out that, as shown in Section 2.6, the proposed method shows a robust behavior against the reduction in the number of training images.

In view of the success of the presented method in illumination invariant face recognition, and availability of convex optimization techniques and stochastic models, it is reasonable to expect promising results to answer such questions as 1) is there a possibility of an analytical model to represent the reflectance and illumination separately, 2) does an efficient framework exist to deal with shadow and illumination in color images, 3) is there a reasonable and systematic procedure to purely amplify only the desired features, and 4) more importantly, how can the whole procedure be replaced with a system level design concept.

In this chapter we first show that any input image can be considered a black box which contains a combination of illumination and reflectance. To do so, a new approximation has been proposed and the commonly used assumption regarding the linear logarithmic based subtraction has been ignored in our proposition. We then propose a design procedure based on the multiresolution analysis of image components to efficiently suppress and remove the redundant illumination effects. The procedure is further developed via the lifting step and extra fine-filtering to amplify and polish the obtained image to extract an illumination invariant mask. A new feature vector configuration is defined based on the double-density complex wavelets components and scaling functions. Principal component analysis (PCA) and the extreme learning machine (ELM) are used for dimensionality reduction and classification, respectively. Experimental results and detailed explanations about databases and simulation conditions are provided to show the effectiveness and performance of our proposed method.

(a) (b) (c)

Figure 2.1 – A sample image without considerable illumination (a). The same subject partially affected by illumination variation from left (b); right (c).

2.3 Illumination Invariant Representation

2.3.1 Black Box Approximation

Land's experiments show that it is possible to recover surface reflectance information from image brightness measurements [50][51]. In the one-dimensional case, a scheme which is called the retinex works well for certain simple scenes [37]. Now let us consider a condition in which the surface orientation does not play a critical role. In that case, scene radiance may be assumed (approximately) to be proportional to the product of the illumination on the object and the reflectance of the surface as

$$I(x, y) \propto R(x, y).L(x, y) \tag{2.1}$$

where $R(x, y)$ and $L(x, y)$ are the reflectance and illumination at point (x, y) of an image $I(x, y)$, respectively. The challenge presented here is how to efficiently and simply extract or estimate $R(x, y)$, the illumination invariant part of an incoming image. Fig. 2.1 (a) shows a sample image without a considerable illumination variation. The same subject has been affected by nonuniform lighting from left and right, as depicted in Fig. 2.1 (b) and (c), respectively.

Physiological and biological researches show that human visual system and the response of retina cells to illumination variation is a nonlinear function that can be represented and approximated by the logarithm of the intensity of pixels in an image.

Therefore, the following approximate equality, a *common assumption*, has been used in several papers, e.g., in [66], based on the relation in (2.1)

$$I'_{\log}(x,y) = \log(I(x,y)) = \log(R) + \log(L) = R'_{\log}(x,y) + L'_{\log}(x,y) \tag{2.2}$$

where $I'_{\log}$, $R'_{\log}$ and $L'_{\log}$ denote the logarithm of the image, reflectance and illumination, respectively. As the illumination mostly lies in the low-frequency part of the image $I(x,y)$, a lowpass filter can be used to filter the logarithm image $I'_{\log}(x,y)$ to extract the illumination variant part $(L'_{\log}(x,y))$ which is then subtracted from $I'_{\log}(x,y)$ to obtain $R'_{\log}(x,y)$, resulting in an approximation of R. To the best of our knowledge, there is no analytic evidence to realize how such an approximation may be close to the ideal case. In [110], the authors have addressed a key issue stating that "the logarithm of the luminance is a crude approximation to the perceived brightness, hence logarithm transform can partly reduce the effect of lighting... it is worth pointing out that solving Eq. (2.2) is also an ill-posed problem, absolutely separating key facial structure R and L is very difficult even if under the common assumption". This statement motivates the idea that filtering the frequency subbands, to preserve or to eliminate low- or high-frequency information in an image, should not be confused with the simple separation of R and L as in (2.2). Ignoring the common assumption in (2.2), we have assumed a box consisting of a combination of illumination and reflectance. Subtraction of illumination from this box in our research and in this chapter is different from algebraic subtraction in (2.2).

Our experiments with human face images under a wide range of illumination variation show that an expression that may better fit the illumination variation has an offset from being the exact log function. Although one may find a transformation to artfully represent the phenomenon, we propose the use of a normalized version of the log function controlled by two parameters. That is

$$I'_{pro}(x,y) = \xi \frac{\log(I(x,y))}{\sqrt{\log(I(x,y))} - \epsilon} \tag{2.3}$$

where in this notation *pro* stands for the proposed, and parameters ϵ and ξ can offer a fine-tuning opportunity depending on an application. In [110], a parameter λ is introduced to find a threshold (T) values. Empirically, the effective range of λ in [110] was reported to be from 0.01 to 0.30. Likewise, and in [14], higher recognition rates can be obtained for a wide range of a parameter (λ) from 0.9 to 1.2. Cao et al. proposed the use of two parameters λ_1 and λ_2 to be applied to training and testing images independently, where $0.9 \leq \lambda_1 \leq 1.2$ and $\lambda_2 > 2$. It has been illustrated that the values of parameters to reach the highest recognition rates are $\lambda_1 = 0.95$ and $\lambda_2 > 2$ [14].

In Section 2.5, it is shown that the practical range for ϵ and ξ to achieve higher accuracy is $0 < \epsilon \leq 0.1$ and $0.6 \leq \xi \leq 1$. It should be noted that the parameter selection in [110][14] and this chapter is not an automatic process. Depending on the structure of each method, there is an optimum range for parameters which is not a wide range to be a bottleneck. Our early experiments show that local stochastic distribution analysis to evaluate the level of illumination on any input image may lead to a narrower range, however, it computationally expensive to find an optimum value of parameters for each input image independently.

At this point, the main question is how to efficiently extract the illumination invariant part, R'_{pro}, of a given image if the proposed approximation is used. In fact, to separate L'_{pro} from I'_{pro}, an efficient frequency information discriminator is required, that is, multiresolution analysis can be a reasonable solution. As stated in the introduction, multiresolution analysis or multiscale approximation, recalls the theory, design and application of the transformations such as DWT, multiwavelets, DT-$\mathbb{C}$WT, CVT, HD-DWT, and DD-DT$\mathbb{C}$WT. First proposed by Kingsbury [65], the dual-tree complex wavelet transform (DT-$\mathbb{C}$WT) is a recent enhancement to the DWT, with two important additional properties, that is, the transformation is shift-invariant and directionally-selective in two and higher dimensions which have addressed the directionality problem in DWT. The double-density discrete wavelet transform (DD-

DWT) [77] is another improvement upon the critically sampled [21] DWT, whereas the DD-DWT outperforms the standard DWT in several applications such as denoising. The transformation can be significantly improved and upgraded in terms of directionality and shift-invariance. In other words, although DD-DWT takes the advantage of more wavelets, it is not entirely directionally-selective. A solution to this problem is provided by Selesnick, introducing the concept of double-density dual-tree complex wavelet transform (DD-DT$\mathbb{C}$WT) which combines the characteristics of the DD-DWT and DT-$\mathbb{C}$WT.

In the next section, the double-density dual-tree complex wavelet transform is briefly introduced. We then show how the directionally- and frequency-selective subbands of the transformation are used to filter and extract the illumination invariant part of an image.

2.3.2 DD-DT$\mathbb{C}$WT and Illumination Removal via Subband Filtering

Double-density dual-tree complex wavelet transform is derived using two scaling filters and four wavelets at the same time, where one scaling and two wavelets are used for real and imaginary parts of a complex wavelet, respectively. Although each of the two wavelet pairs hold half a delay as an offset, it is unnecessary to specify any explicit constraint to ensure the half sample delay [77]. Let us consider two filter banks $\mathcal{F}$ and $\widetilde{\mathcal{F}}$, the primary and dual filter banks of a complex transformation, respectively [77][78], where the analysis and synthesis lowpass (highpass) filters of $\mathcal{F}$ are denoted by $H_0(z)$ ($H_1(z)$, $H_2(z)$) and $F_0(z)$ ($F_1(z)$, $F_2(z)$), respectively. The filters associated with the dual bank are defined similarly with superscript '$\sim$'. For the double-density discrete wavelet transform (DD-DWT) with a single filter bank, perfect reconstruction condition is derived based on the relation between input and

output that can be written as

$$V(z) = \frac{1}{2}([H_0(z)H_0(\frac{1}{z}) + H_1(z)H_1(\frac{1}{z}) + H_2(z)H_2(\frac{1}{z})]X(z) \tag{2.4}$$

$$+[H_0(z)H_0(\frac{-1}{z}) + H_1(z)H_1(\frac{-1}{z}) + H_2(z)H_2(\frac{-1}{z})]X(-z)) \tag{2.5}$$

In order to satisfy the perfect reconstruction condition, it is required the input $X(z)$ and the output $V(z)$ be identical, thus

$$H_0(z)H_0(\frac{1}{z}) + H_1(z)H_1(\frac{1}{z}) + H_2(z)H_2(\frac{1}{z}) = 1 \tag{2.6}$$

$$H_0(z)H_0(\frac{-1}{z}) + H_1(z)H_1(\frac{-1}{z}) + H_2(z)H_2(\frac{-1}{z}) = 0. \tag{2.7}$$

Apart from the perfect reconstruction, the wavelets and scaling functions are derived using the iterative equations similar to the ones used in traditional DWT, that is, the scaling and wavelet functions associated with the analysis side of $\mathcal{F}$ are defined iteratively via two-scale equations

$$\phi_h(t) = 2\sum_{\kappa} h_0[\kappa]\phi_h(2t-\kappa) \tag{2.8}$$

$$\psi_{h_i}(t) = 2\sum_{\kappa} h_i[\kappa]\phi_h(2t-\kappa) \quad \text{for} \quad i = 1, 2 \tag{2.9}$$

where $h_0[\kappa]$ and $h_i[\kappa]$ are the impulse responses of $H_0(z)$ and $H_i(z)$, respectively, and z refers to the z-transform with $z = e^{j\omega}$. The scaling function ϕ_f and wavelets ψ_{f_i} in the synthesis side of the primary filter bank $\mathcal{F}$, and the scaling function and wavelets of the dual bank $\widetilde{\mathcal{F}}$ in the analysis and synthesis sides are similarly defined. In order for the primary and dual filter banks to constitute a dual-tree complex wavelet structure, $\psi_{\widetilde{h}_i}(t)$ and $\psi_{f_i}(t)$ must be the Hilbert transform of $\psi_{h_i}(t)$ and $\psi_{\widetilde{f}_i}(t)$, respectively. In other words

$$\Psi_{\widetilde{h}_i}(\omega) = \begin{cases} -j\Psi_{h_i}(\omega) & \omega > 0 \\ j\Psi_{h_i}(\omega) & \omega < 0 \end{cases}, \quad \Psi_{f_i}(\omega) = \begin{cases} -j\Psi_{\widetilde{f}_i}(\omega) & \omega > 0 \\ j\Psi_{\widetilde{f}_i}(\omega) & \omega < 0 \end{cases}$$

where $\Psi_{h_i}(\omega)$, $\Psi_{f_i}(\omega)$, $\Psi_{\widetilde{h}_i}(\omega)$, and $\Psi_{\widetilde{f}_i}(\omega)$ are the Fourier transforms of wavelet functions $\psi_{h_i}(t)$, $\psi_{f_i}(t)$, $\psi_{\widetilde{h}_i}(t)$, and $\psi_{\widetilde{f}_i}(t)$ respectively, and $j = \sqrt{-1}$. The Hilbert pair condition is similarly derived and extended for pairs of wavelets for the filters in the synthesis side, and the transformation possesses desired properties. The three key properties of DD-DT$\mathbb{C}$WT, 1) shift-invariance, 2) directional-selectivity, 3) double-density wavelets, play important role in multiresolution based image analysis. The double-density dual-tree complex wavelet transform contains more wavelet subbands than the regular DWT. In addition, wavelets associated with the DD-DT$\mathbb{C}$WT are directional and free of checkerboard effect. This is related to the real and imaginary parts of a dual-tree complex structure, and to the extra wavelets imposed to the double-density transformation. Fig. 2.2 shows the structure of a typical DD-DT$\mathbb{C}$WT. Only the analysis (decomposition) side of the filter banks has been shown in this figure and the synthesis (reconstruction) side can be similarly configured. It is important to note that, due to the above mentioned properties of the DD-DT$\mathbb{C}$WT, the transformation can localize features in several directions [77]. Furthermore, as the transformation contains doubled number of high-frequency subbands in each level, the so-called density of wavelets, it fits the problem of extracting low-frequency information by suppressing more number of high-frequency subbands. As discussed earlier, the latter is used to be subtracted from the input image to obtain the reflectance part of the image.

Now considering the concept of double-density strategy, shown in Fig. 2.2 and discussed in detail in [77] and [82], the two separate filter banks $\mathcal{F}$ (tree 1) and $\widetilde{\mathcal{F}}$ (tree 2) are designed such that the subbands of upper filter bank can be interpreted as the real part of a complex wavelet transform, and subbands of lower filter bank are assumed as the imaginary part of the transformation. More specifically, for the

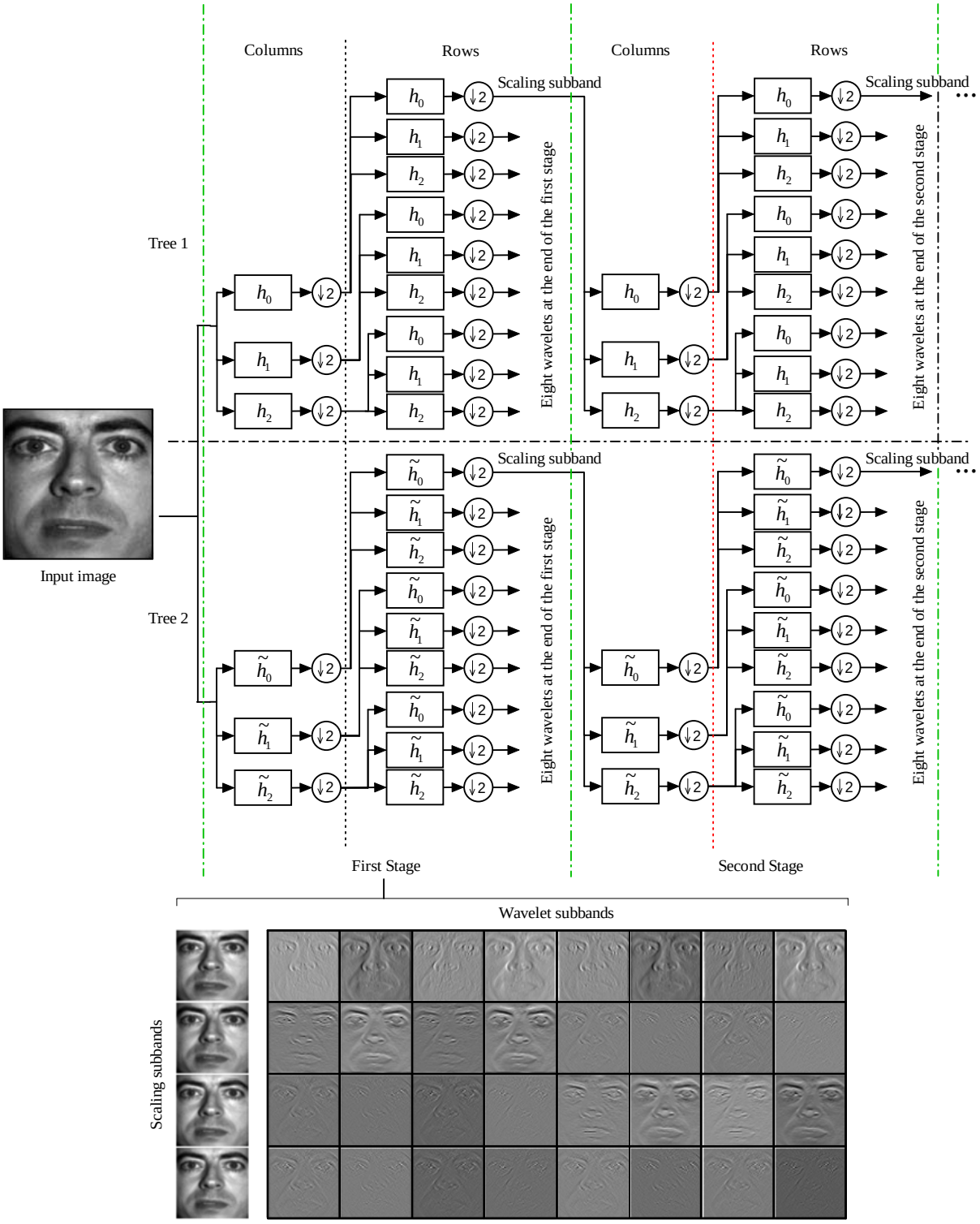


Figure 2.2 – Structure of the DD-DT$\mathbb{C}$WT. Only the analysis (decomposition) side of the transformation is shown in this figure and the synthesis (reconstruction) side can be similarly configured via tree 1 ($\mathcal{F}$) and tree 2 ($\widetilde{\mathcal{F}}$).

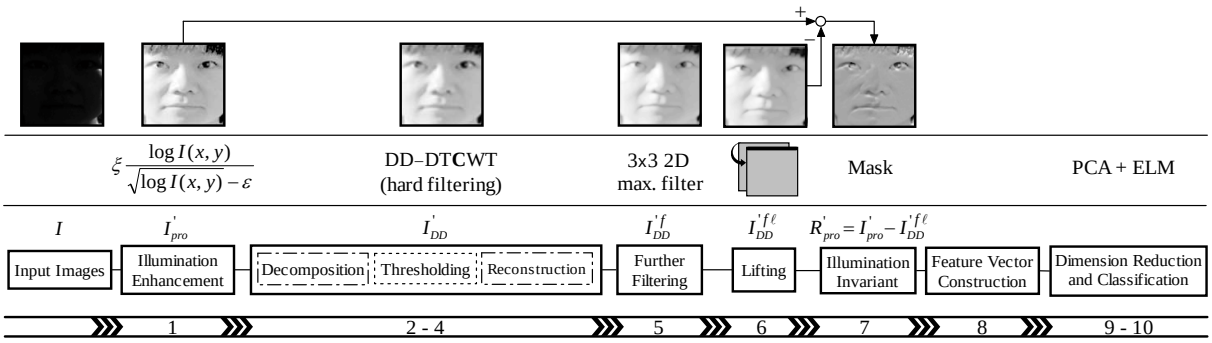


Figure 2.3 – Block diagram of the proposed method.

filters designed for such a configuration, the wavelets associated with the real part must be (approximately) the Hilbert transform of the wavelets associated with the imaginary part of the dual-tree shown in Fig. 2.2. In that case, the DD-DTℂWT is in fact a 2D oriented higher density complex transformation with more directionally-selective wavelets that can be widely used in relevant applications in image processing and pattern recognition. Recall that for both the DWT and DD-DWT some of the wavelets have no dominant orientation, and therefore, DD-DTℂWT can simply outperform the other wavelet family members.

Motivated by these facts, we propose an efficient and customized frequency subband filtering method for the problem of illumination invariant face recognition. The idea is suggested as a means of preserving the illumination invariant information and at the same time suppressing the redundant data. The whole idea and problem statement can be summarized as follows:

1. For a given image, illumination is initially enhanced using (2.3) to obtain I'_{pro}.

2. I'_{pro} is decomposed into frequency subbands using DD-DTℂWT.

3. Thresholding is performed on the high-frequency subbands.

4. Using the inverse DD-DTℂWT, reconstruction is done based on the low-frequency and thresholded high-frequency subbands to obtain I'_{DD}.

5. Maximum filter is applied on the reconstructed image to further enhancing any nonuniform spikes that cannot be considered as illumination.

6. Amplifying the role of edges to lift the quality of the reconstructed image.

7. Extracting a mask to obtain the illumination-invariant image.

8. Decomposing the obtained mask, R'_{pro}, into frequency subbands via DD-DT$\mathbb{C}$WT to form a feature vector.

9. Applying PCA to reduce the dimension of feature space.

10. Performing ELM for classification.

The block diagram of the proposed algorithm is shown in Fig. 2.3, where the discussion in this section refers to the blocks 2 to 4 of the diagram. The input data to this block, which is in the spatial domain, are transformed to the complex wavelet domain. Each image, $I'_{pro}(x, y)$, is first decomposed by DD-DT$\mathbb{C}$WT into frequency subbands. Unlike the traditional DWT with only three wavelets, DD-DT$\mathbb{C}$WT offers 32 directional wavelet subbands. The 32 wavelets for DD-DT$\mathbb{C}$WT and the 3 wavelets associated with DWT has been shown in Fig. 2.4 (a) and (b), respectively. Using a 3D representation, Fig. 2.4 (c) shows the variety and differences of the 32 wavelets of the DD-DT$\mathbb{C}$WT. As can be observed, the frequency subbands of the DD-DT$\mathbb{C}$WT are more discriminative than the ones in DWT.

We then apply thresholding on each high-frequency subband to essentially deactivate the high-frequency information. The value of threshold is fixed to the mean of the minimal coefficients of each row of a subband. The coefficients that are less than threshold remain unchanged and the rest of coefficients are set to zero. In fact, thresholding the wavelet (high-frequency) subbands preserves low-frequency information, which is ideally desired to be subtracted from I'_{pro} to obtain R'_{pro}. The final step, at this point, is to reconstruct the thresholded subbands via an inverse DD-DT$\mathbb{C}$WT.

2.3.3 Smoothing and Lifting

Shown as block 5 in Fig. 2.3, additional filtering is applied for fine-tuning. The superscript f in this figure refers to filtering. Among several existing well known filters such as the Cohen-Daubechies-Feauveau (CDF) filters, halfband-pair filter bank (HPFB),

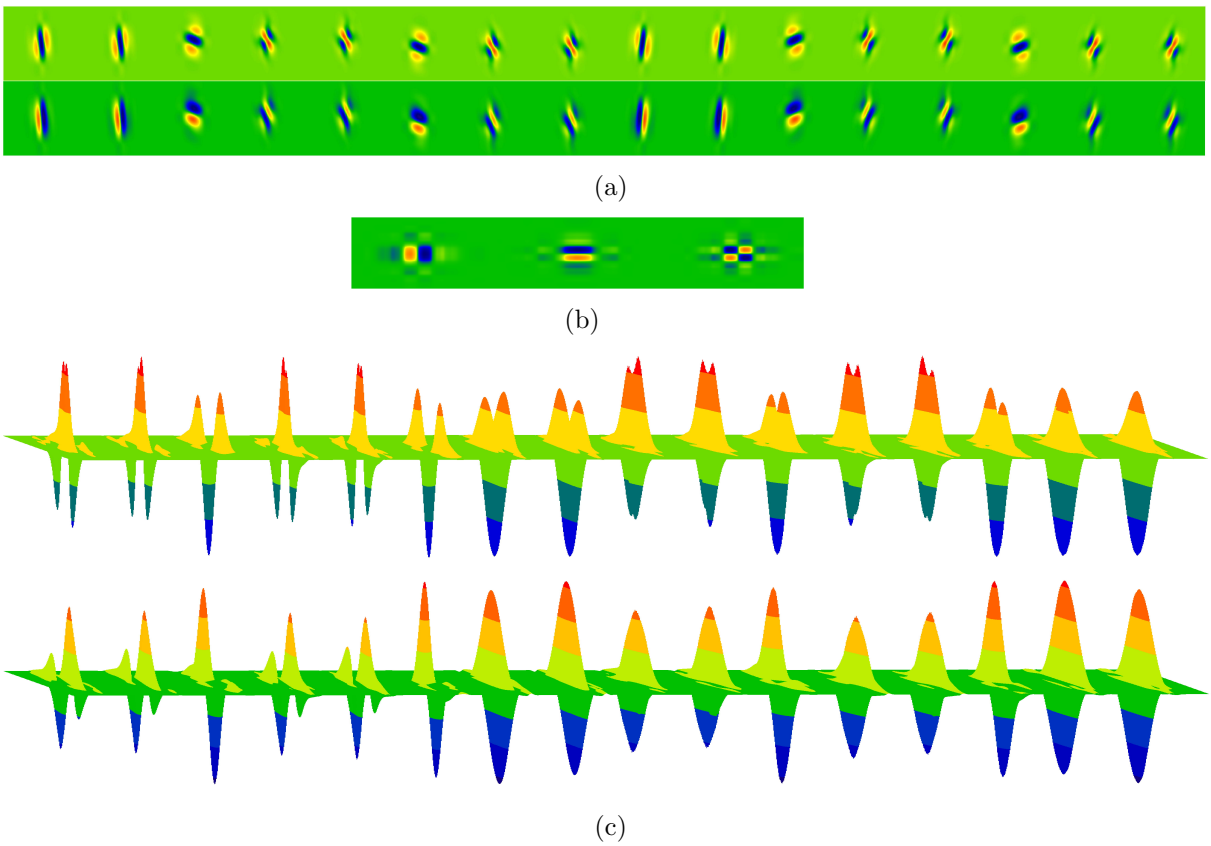


Figure 2.4 – Visual representation of wavelets. (a) 32 wavelets related to the 2D double-density dual-tree complex wavelet transform; (b) Three wavelets in a regular DWT; (c) 3D representation of the 32 wavelets in (a) with horizontal and vertical view of 356.5° and 7°, respectively.

quadrature mirror filter (QMF), median filter, maximum filter, few to mention, we found that both the CDF and maximum filter perform better than others for this application. We have used a 3×3 2D statistic maximum filter before sending the result to block 6, in which the features edges are stimulated and sharpened as is shown in Fig. 2.5. It should be noted that the maximum filter is a nonlinear filter and it does not have a frequency domain equivalent. As the images under this operation lose some sharpness with removed pepper-type noise, it resembles a lowpass filter nature [66] where the filtered image contains less details than the initial one. One should note that a single lowpass filtering, regardless of the type of filter that might be used, cannot compete with the DD-DTℂWT which can accurately discriminate

(a) (b) (c) (d) (e)

Figure 2.5 – Feature stimulation by lifting. (a) A sample image from the Extended-Yale B database; (b) ${I'_{DD}}^{f}$; (c) $I'_{pro} - {I'_{DD}}^{f}$; (d) ${I'_{DD}}^{f\ell}$; (e) $I'_{pro} - {I'_{DD}}^{f\ell}$.

the frequency information in an image. In other words, the main role of filtering procedure relies on the use of the DD-DT$\mathbb{C}$WT and the maximum filter is employed for smoothing and polishing the images.

Our experiments in the design and application of wavelets in multifocus image fusion [4] and the method of Huang et al. in moving object detection introducing the concept of the double-change detection (DCD) [42], motivated us to consider a simple yet versatile approach to stimulate features. Fig. 2.5 (a) shows a randomly selected image from the Extended-Yale B database. The image I is passed through the blocks 1 to 5, depicted in Fig. 2.3, and the obtained image ${I'_{DD}}^{f}$ is then subtracted from the enhanced version of the original image I'_{pro}. The results are given in Fig. 2.5 (b) and (c), respectively. Now let us shift ${I'_{DD}}^{f}$ down by two rows to obtain ${I'_{DD}}^{f\ell}$, followed by the same subtraction from I'_{pro}. This is shown in Fig. 2.5 (d) and (e) where the superscript ℓ denotes the lifting. As can be observed, the fine features, e.g., eyes, eyebrows, nose and mouth in Fig. 2.5 (e) are clearly visible and separable more in details than the ones in Fig. 2.5 (c), resulting in sharper edges and smoother face which improves the recognition rates. It should be pointed out that the amount of shift, in general, depends on the size of images. One should note that most of the recognition systems must resize the input images to a smaller size first, to reduce the computational complexity and the execution time, and therefore, it is not reasonable to work on images with larger sizes. Roughly, for images of size 32×32 to 128×128, the effective value of shift is only one to four rows. Fig. 2.6 (c) and (d)

(a) (b) (c) (d)

Figure 2.6 – The mask achieved using the method in [66] and the one obtained by the proposed method. (a) A sample image from the CMU-PIE database without considerable illumination effect; (b) The same subject with high illumination effects; (c) and (d) the mask used in [66] and the one obtained by using our method, respectively.

show the masks obtained using one of the recent methods in the literature [66] and the proposed method, respectively. There is an obvious difference and improvement between the masks in Fig. 2.6 (c) and (d), where the latter leads to superior results to recognize an image as dark as in (b). In the next section we shall employ the obtained masks to extract the feature vector required for the training and testing phases of the recognition system to evaluate the effectiveness of the proposed framework.

2.4 Feature Extraction and Classification

Principal component analysis (PCA) [34][44], two-dimensional PCA (2D-PCA) [107], and linear discriminant analysis (LDA) [33][34] are some of the well known general algorithms for dimensionality reduction and feature extraction used in the literature. Also, several authors have demonstrated many feature vector construction algorithms based on innovative techniques such as the Zernike moments [36], statistical moments [34], and steerable pyramid [30]. Among the existing techniques, PCA has been widely used for dimensionality reduction in recent relevant papers [30][66][110]. To keep consistency, we have used PCA for dimensionality reduction and feature extraction. The feature vector in our work is formed by the subbands of the DD-DT$\mathbb{C}$WT multiscale representation, that is, the obtained mask $R'_{pro} = I'_{pro} - {I'_{DD}}^{f\ell}$ is convolved with each

of the 16 complex wavelets ($\psi^{\mathbb{C}}_{(.)}$) as well as the four scaling filters. In other words, given a mask R'_{pro}, scaling coefficients ϕ_1 to ϕ_4, and wavelets ψ_i for $i = 1, \ldots, 32$, the feature vector is formed as $R'_{pro} * \{\psi^{\mathbb{C}}_1, \ldots, \psi^{\mathbb{C}}_{16}, \phi_1, \ldots, \phi_4\}$ where notation '$*$' denotes the convolution.

The final step of a recognition system consists of classifiers that can efficiently and easily handle linear as well as nonlinear mapping amongst features and their classes. Although out of numerous intelligent techniques neural networks and support vector machines (SVMs) have had a large impact, they face some challenging issues, e.g., slow learning speed, trivial human intervene, and poor computational scalability. A new learning algorithm, extreme learning machine (ELM), for single-hidden layer feedforward neural networks (SLFNs) can overcome some of the above-mentioned limitations [38][41]. In [40], it has been shown that extreme learning machine generally outperforms SVM in various kinds of cases. Feedforward neural networks (FNN) have been used in several areas because of the ability of approximation for nonlinear mappings. Traditionally, parameter tuning techniques such as gradient descent-based algorithms are used to determine the input weights and hidden layer biases. Gradient descent-based learning methods are generally slow due to inappropriate or iterative learning steps. ELM randomly chooses hidden nodes and analytically determines the output weights of SLFNs, and therefore, the method is extremely fast in learning.

Shown in Fig. 2.7, for N arbitrary distinct samples $(\mathbf{x}_i, \boldsymbol{t}_i)$ a typical ELM with $\widetilde{N}$ hidden nodes and an activation function $g(x)$ is modeled as

$$\sum_{i=1}^{\widetilde{N}} \beta_i g_i(\mathbf{x}_l) = \sum_{i=1}^{\widetilde{N}} \beta_i G(\mathbf{w}_i, b_i, \mathbf{x}_l) = \mathbf{o}_l, \ l = 1, 2, \ldots, N \tag{2.10}$$

where g_i denotes the output function $G(\mathbf{w}_i, b_i, \mathbf{x})$ of the ith hidden node. An activation function for any additive note can be defined as $g_i = G(\mathbf{w}_i, b_i, \mathbf{x}) = g(\mathbf{w}_i.\mathbf{x} + b_i)$ and therefore

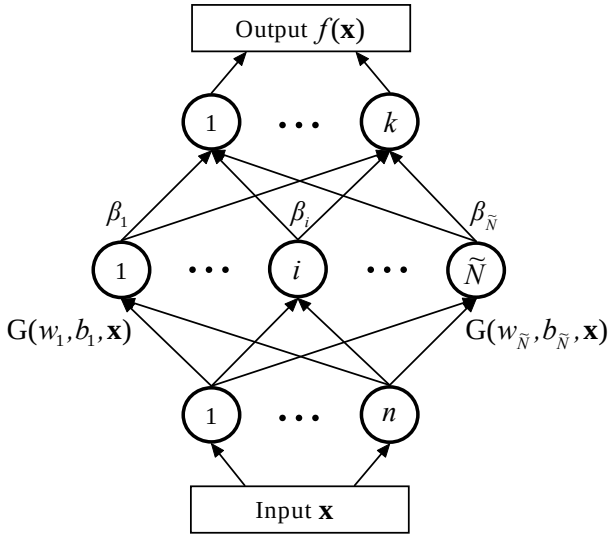


Figure 2.7 – A typical structure of an ELM.

$$\sum_{i=1}^{\widetilde{N}} \beta_i g_i(\mathbf{x}_l) = \sum_{i=1}^{\widetilde{N}} \beta_i g(\mathbf{w}_i.\mathbf{x}_l + b_i) = \mathbf{o}_l,\ l = 1, 2, \ldots, N \tag{2.11}$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^t \in \mathbf{R}^n$, $\boldsymbol{t}_i = [t_{i1}, t_{i2}, \ldots, t_{ik}]^t \in \mathbf{R}^k$, and $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^t$ and $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{ik}]^t$ denote the weight vectors connecting the input nodes to ith hidden node, and from the ith hidden node to the output nodes, respectively. b_i is a threshold for the ith hidden node, $\mathbf{w}_i.\mathbf{x}_l$ indicates the inner product of $\mathbf{w}_i$ and $\mathbf{x}_l$, and superscript 't' stands for the transpose. This model can approximate N samples with zero error $\sum_{l=1}^{\widetilde{N}} \|\mathbf{o}_l - \boldsymbol{t}_l\| = 0$, that is, there exist β_i, $\mathbf{w}_i$, and b_i such that $\sum_{i=1}^{\widetilde{N}} \beta_i g(\mathbf{w}_i.\mathbf{x}_l + b_i) = \boldsymbol{t}_l,\ l = 1, 2, \ldots, N$. These equations can be written as $\boldsymbol{\Upsilon}\beta = \mathbf{T}$ where $\beta = [\beta_1^t, \ldots, \beta_{\widetilde{N}}^t]_{\widetilde{N}\times k}^t$ and $\mathbf{T} = [\boldsymbol{t}_1^t, \ldots, \boldsymbol{t}_N^t]_{N\times k}^t$ [38][41]. With this notation, $\boldsymbol{\Upsilon}$ denotes the hidden layer output matrix of ELM where ith column of $\boldsymbol{\Upsilon}$ is the output of ith hidden node with respect to inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$.

If the activation function g is infinitely differentiable, it is proved that the required number of hidden nodes satisfy $\widetilde{N} \ll N$ [38]. Traditionally, in order to train an ELM, one may need to minimize the error function defined as $\eta = \sum_{l=1}^{N}(\sum_{i=1}^{\widetilde{N}} \beta_i g(\mathbf{w}_i.\mathbf{x}_l + b_i) - \mathbf{T}_l)^2$. Generally, gradient based algorithms are employed to search for the

minimum of $\|\boldsymbol{\Upsilon}\beta - \mathbf{T}\|$. Gradient descent-based learning schemes are generally slow, that is, if the learning rate is too small, the learning algorithm converges very slowly. In contrast, if the learning rate is too large, then the algorithm becomes unstable and may diverge. Huang et al. show that the input weights and the hidden layer biases are not necessarily tuned and the hidden layer output matrix can remain unchanged once random values have been assigned to these parameters in the beginning of learning [38][40][41]. Therefore, for fixed input weights and the hidden layer biases, training an ELM is equivalent to find a least-squares solution $\hat{\beta}$ of the linear system $\boldsymbol{\Upsilon}\beta = \mathbf{T}$, that is

$$\|\boldsymbol{\Upsilon}(\mathbf{w}_1, \ldots, \mathbf{w}_{\widetilde{N}}, b_1, \ldots, b_{\widetilde{N}})\hat{\beta} - \mathbf{T}\| \tag{2.12}$$

$$= \min_{\beta} \|\boldsymbol{\Upsilon}(\mathbf{w}_1, \ldots, \mathbf{w}_{\widetilde{N}}, b_1, \ldots, b_{\widetilde{N}})\beta - \mathbf{T}\|. \tag{2.13}$$

In most cases the number of hidden nodes is much less than the number of distinct training samples, and therefore, $\boldsymbol{\Upsilon}$ is a nonsquare matrix. If the number of hidden nodes is equal to the number of distinct training samples, SLFNs can approximate these training samples with zero error. In case of $\boldsymbol{\Upsilon}$ being nonsquare, the smallest norm least-squares solution of the above linear system is $\hat{\beta} = \boldsymbol{\Upsilon}^*\mathbf{T}$, where $\boldsymbol{\Upsilon}^*$ is the *moore-penrose* generalized inverse of a matrix $\boldsymbol{\Upsilon}$, and the smallest training error can be reached by this special solution as

$$\|\boldsymbol{\Upsilon}\hat{\beta} - \mathbf{T}\| = \|\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^*\mathbf{T} - \mathbf{T}\| \tag{2.14}$$

$$= \min_{\beta} \|\boldsymbol{\Upsilon}\beta - \mathbf{T}\|. \tag{2.15}$$

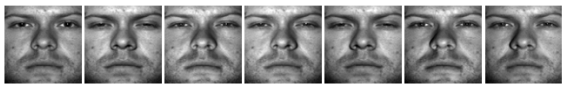
Although almost all learning algorithms wish to approach the minimum training error, most of them cannot reach it because of local minimum or infinite training iteration that is usually not allowed in applications (see Section 3.2 in [38]). Details and discussions on extreme learning machine are given in [38]–[41].

2.5 Experiments and Results

Our experiments can be basically divided into two categories. The first group contains the highly illuminated databases, e.g., the Yale B, Extended-Yale B, and CMU-PIE. The experiments on these databases show the performance of the proposed method to extract illumination invariants under notable illumination variations. The results of these experiments have been compared with the papers in this category in which the problem of illumination invariant face recognition is investigated. In the second group, we use the FERET, AT&T, and LFW databases. These databases contain a low or moderate illumination effects, however, they have notable variations both in pose and facial expressions.

2.5.1 The Yale B and Extended-Yale B Databases

The Yale B database [105] contains 5760 single light source images of size 192×168 for 10 subjects with 576 viewing conditions, i.e., 9 poses × 64 illumination conditions, for each individual. The database was expanded to 38 subjects, known as the Extended-Yale B [106], with 21,888 single light source images each of which is seen under the same 576 viewing conditions as in the Yale B. Similar to previous papers, and to keep consistency for comparison, the same setup has been taken into account for the subsets, that is, the databases are divided into five subsets corresponding to the different light source directions which is in fact the angle (τ) between the light source and camera axis. Subset 1 consists of 70 (266) images with $\tau < 12°$. Subsets 2 to 5 contain 120 (456) $\{20° < \tau < 25°\}$, 120 (456) $\{35° < \tau < 50°\}$, 140 (532) $\{60° < \tau < 77°\}$, and 190 (722) $\{\tau > 78°\}$ images, respectively, where the numbers inside parenthesis denote the number of images for the corresponding subset in the Extended-Yale B database. Thus, for the Yale B and the Extended-Yale B databases, each individual has 7+12+12+14+19=64 images. Detailed information about the two databases, including photographic conditions and subsets, is given in [105][106]. All



(a) Subset 1

(b) Subset 2

(c) Subset 3

(d) Subset 4

(e) Subset 5

Figure 2.8 – Sample representation for individual 1 in the Extended-Yale B database.

(a) Subset 1

(b) Subset 2

(c) Subset 3

(d) Subset 4

(e) Subset 5

Figure 2.9 – The corresponding masks obtained by the proposed method for images in Fig. 2.8.

images are first eye-aligned and then cropped and resized into 128×128 gray level pixels for which we set the lifting value $\ell = 3$ and follow the rest of procedure via a single-stage decomposition to obtain the illumination invariant masks required for recognition. Subset 1 contains seven images per person, taken under low illumination variations, which is used to train the recognition system. The level of illumination is related to the angle between the light source direction and the camera axis and is

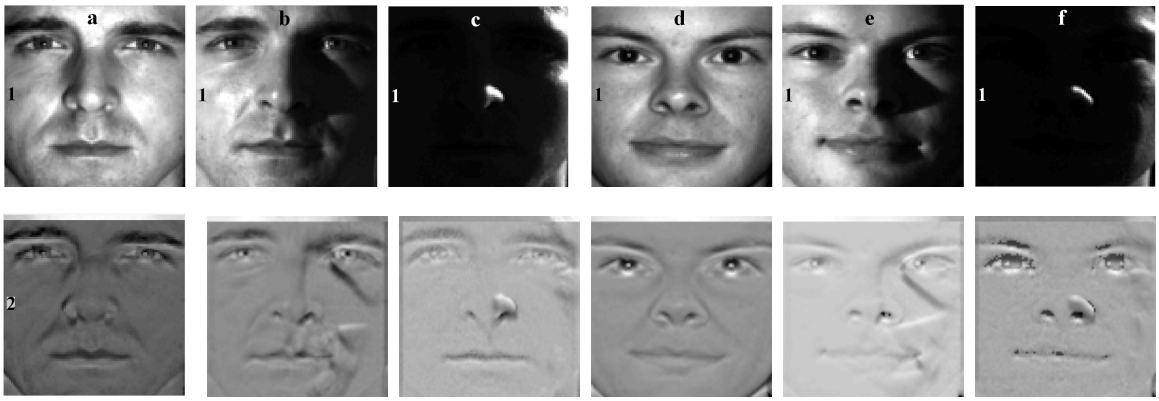
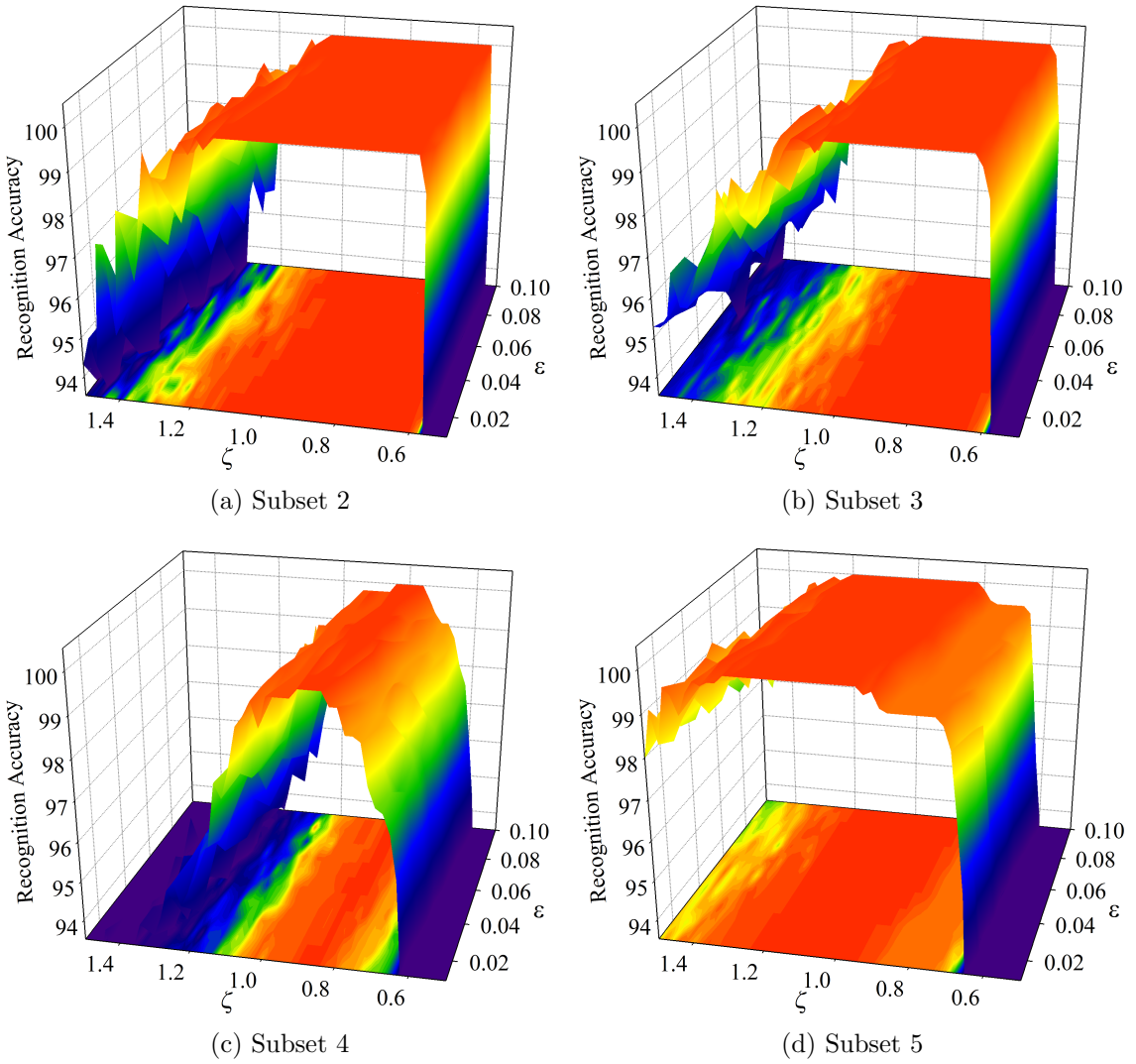


Figure 2.10 – Visual representation of the proposed algorithm for the Yale B and the Extended-Yale B databases, respectively. In the first row, (a) shows a sample image without considerable illumination effect from the Yale B database; (b) and (c) indicate the same subject as in (a) but under an average and high illumination, respectively. The same is shown for a sample image from the Extended-Yale B database in (d)–(f), respectively. The second row shows the obtained mask for each image in the first row, respectively.

increased for the remaining subsets, Subset 2 to Subset 5, which are used for testing purposes. The images of a sample subject are shown in Fig. 2.8 for Subsets 1 to 5, followed by the extracted mask in Fig. 2.9 using the proposed method on each image in subsets, respectively. To have a visual perception, three different samples for a particular individual from each database and the corresponding mask, have been shown in Fig. 2.10 for the Yale B and Extended-Yale B databases, respectively.

The first row shows an individual from the Yale B database under three different illumination effects labeled as (a), (b) and (c) where the same is shown for a randomly selected individual from the Extended-Yale B database depicted in (d), (e) and (f), respectively. The second row shows the obtained mask for each image in the first row, respectively. To maintain consistency with other works, we have used seven images of Subset 1 for training, and the remaining Subsets 2–5 for the testing gallery. The results are given in Tables 2.1 and 2.2 for the Yale B and Extended-Yale B, respectively. The difference between the Yale B and Extended-Yale B is that the former contains 10 individuals while the latter includes 38 subjects, which makes the experiments



(a) Subset 2

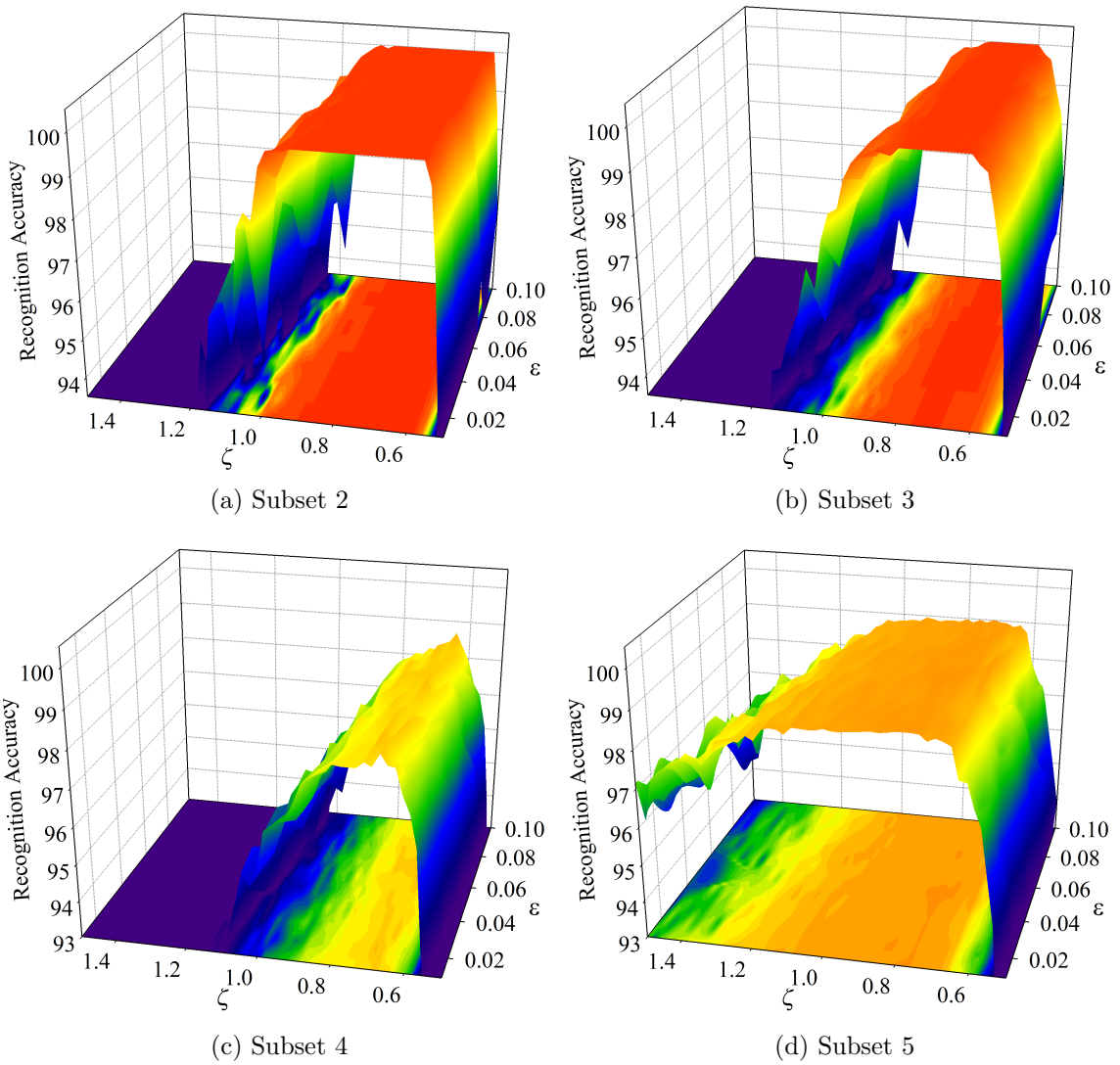
(b) Subset 3

(c) Subset 4

(d) Subset 5

Figure 2.11 – Recognition rate versus ξ and ϵ for the Yale B database.

more difficult. The results indicate that the proposed method can outperform the previous approaches in terms of the recognition accuracy. Some authors reported the results on different training subsets. For example, state-of-the-art method in [102], which is eventually an excellent feature extraction algorithm with robustness against occlusion, uses half of the images ((7+12+12+14+19)/2 = 32 images per individual) in the Extended-Yale B database for training and the other half for the testing phase. Ignoring the commonly used Extended-Yale B settings for the illumination



(a) Subset 2

(b) Subset 3

(c) Subset 4

(d) Subset 5

Figure 2.12 – Recognition rate versus ξ and ϵ for the Extended-Yale B database.

invariant face recognition in the literature, we have performed the same experiment as in [102] on this database. The obtained recognition rate of 99.42%, employing our method, compares favorably with 98.1% in [102]. Recognition rate versus ξ and ϵ has been shown in Fig. 2.11 (a)–(d) for Subsets 2–5 for the Yale B database. It is important to note that the recognition rate varies marginally along the ϵ axis in contrast to the changes in ξ direction. Therefore, the role of ϵ is mostly to calibrate the highest accuracy in applications such as Subset 4 of the Extended-Yale B database.

Table 2.1 – Yale B: Recognition rate for different techniques (%).

Method	Subset 2	Subset 3	Subset 4	Subset 5
PCA [11]	98.33	79.17	30.00	15.79
Cones Attached [56]	100	100	91.40	N/A
QI [86]	99.30	61.90	34.10	23.30
QIR [86]	100	100	90.60	78.80
Cones Cast [56]	100	100	100	N/A
Hist. Equz. [85]	100	89.00	55.10	44.40
Linear Subspace [8]	100	100	85.00	N/A
GIC [85]	100	88.10	39.90	27.50
SQI [100]	100	100	96.40	97.90
LTV+PCA [17]	100	99.17	96.43	92.12
PP+LTP [92]	100	100	99.20	97.20
LBP [92]	100	97.60	65.20	44.40
MQI [112]	100	100	100	98.40
Wavelet+PCA [110]	100	100	100	100
Wavelet+PCA [14]	100	100	100	100
S–P [30]	100	100	87.70	45.68
S–P+PP [30]	100	100	100	100
II+PCA [66]	100	100	98.60	98.90
II+PCA+DVS [66]	100	100	99.29	99.47
II+PCA+WVS [66]	100	100	100	99.47
Proposed [3]	100	100	100	100

In all experiments, except Subset 4 and Subset 5 of the Extended-Yale B database, one may represent the recognition accuracy versus the ξ by setting the variable ϵ to a constant value. The same experiments on the Extended-Yale B database leads to similar results which are depicted in Fig. 2.12 (a)–(d) for Subsets 2–5, respectively. In Section 2.6, we show that the method presents a resistive nature against the reduction in the number of training samples (NTR).

2.5.2 The CMU-PIE Database

This database contains 41,368 images from 68 subjects with different pose, illumination and expression (PIE) conditions. Depending on these factors, the images are classified into different sets. For each subject, 13 poses and 43 illumination conditions

Table 2.2 – Extended-Yale B: Recognition rate for different techniques (%).

Method	Subset 2	Subset 3	Subset 4	Subset 5
PCA[11]	90.16	41.23	6.37	3.24
LDA[10]	100	98.12	38.35	5.13
II+PCA+SVM [66]	100	99.78	95.44	94.68
II+PCA+1NN [66]	100	100	96.01	92.44
II+PCA+DVS [66]	100	100	96.60	95.40
II+PCA+WVS [66]	100	100	97.91	96.54
Proposed [3]	100	100	98.68	99.03

Figure 2.13 – Visual representation of the proposed algorithm for the CMU-PIE database. In the first row (a) shows a sample image without considerable illumination effect from the lights-on setting; (b) and (c) indicate the same subject as in (a) but under an average and high illumination, respectively. The same is shown for a sample image from the lights-off setting in (d)–(f), respectively. The second row shows the obtained mask for each image in the first row, respectively.

are available. All images were taken at the Carnegie Mellon University 3D Room between October and December 2000 [2]. As in [66] and [87], only frontal images have been used in experiments. The images are first eye-aligned and cropped, and then resized into 96×96 pixels. As discussed earlier, we select ℓ =3 and the procedure in Fig. 2.3 is applied to the images to obtain required masks.

For the lights-on setting with 24 images, the images named as {I06, I08, I11 I20} from each of 68 individuals are selected for training due to their better illumination conditions [87]. The rest of images (20 images) are used for testing. This has been

Table 2.3 – CMU-PIE: Recognition rate for different lighting conditions (%). The second row in each experiment shows the results obtained via [66].

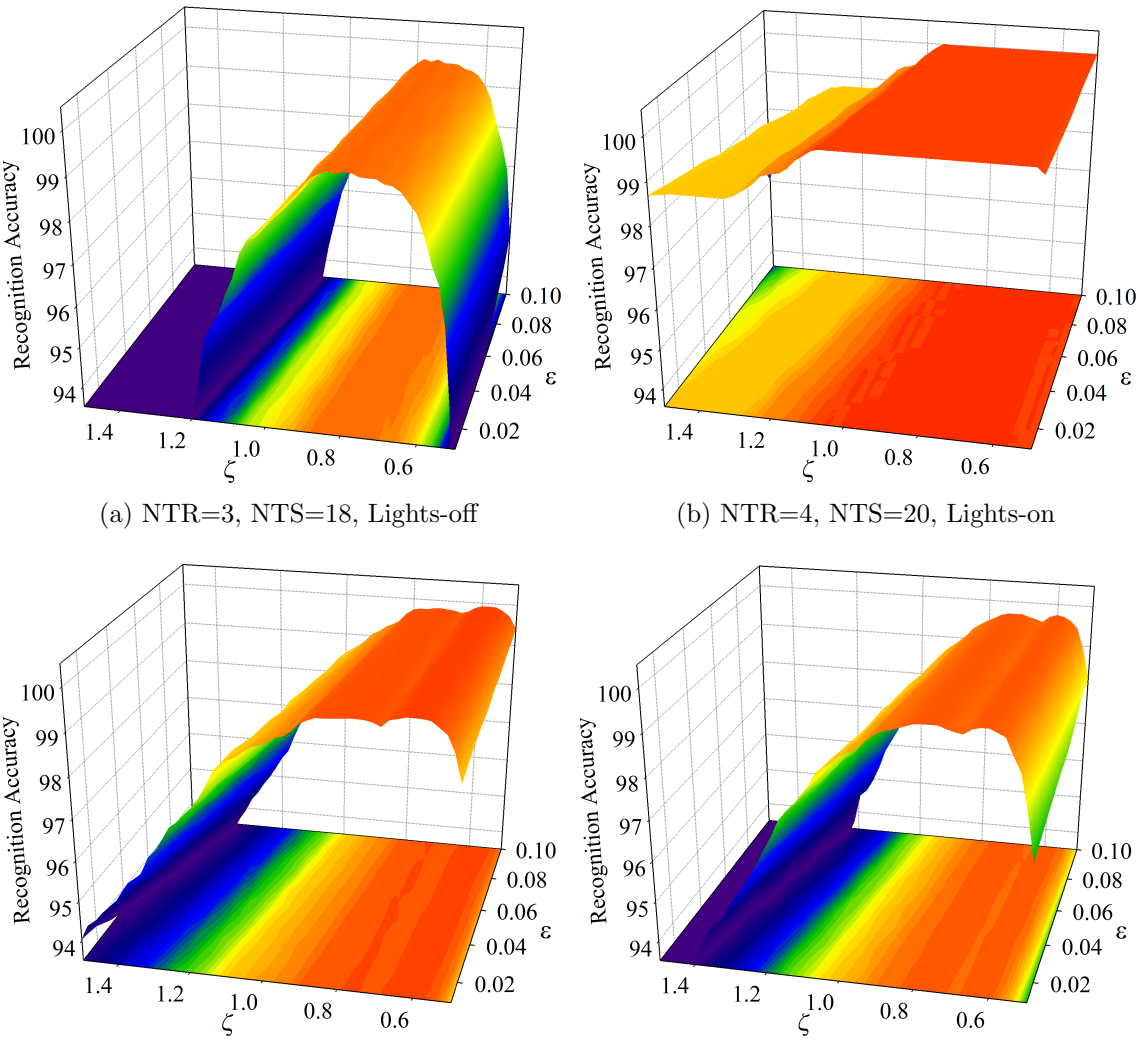
lighting condition	NTR	NTS	Total	%
off	3	18	21	99.59
				94.85
on	4	20	24	100
				100
on/off	7	38	45	99.92
				99.74
	6	39	45	99.77
				99.55
	5	40	45	99.70
				98.99

shown in Table 2.3. Note that the names used for images here are the ones that appear in the corresponding database [2]. For the lights-off setting with 21 images for each of the 68 individuals the images {i08, i11, i20} are used for training and the rest of images are kept for the testing phase. Some samples from different individuals have been shown for the lights-on and lights-off settings in Fig. 2.13, respectively. Recognition rate versus ξ and ϵ has been shown in Fig. 2.14 (a)–(d), for instance, for different lighting settings, respectively.

The results are given in Table 2.3 for the CMU-PIE database for the settings of lights-on and lights-off, as well as the results obtained by 4, 5, 6, and 7 number of sample images for training for the case of lights-on/off where the 21+24=45 images are considered as a single set. In this table, NTR and NTS denote the number of training and the number of testing samples, respectively. A minimum accuracy of 99.70% is achieved for the lights-on/off setting with five images for training.

2.5.3 The FERET Database

The Face Recognition Technology (FERET) program database [72] is a large database of facial images, divided into development and sequestered portions. The FERET



(a) NTR=3, NTS=18, Lights-off

(b) NTR=4, NTS=20, Lights-on

(c) NTR=7, NTS=38, Lights-on/off

(d) NTR=6, NTS=39, Lights-on/off

Figure 2.14 – Recognition rate versus ξ and ϵ for the CMU-PIE database.

database consists of 14,051 eight-bit gray scale images of human heads with views ranging from frontal to left and right profiles. The database contains several subsets based on regular facial expression, alternative expression under different illumination, subject faces to his left, subject faces to his right, quarter left and right, half left and right, profile left and right, and random images with angles. For each individual in the database, Subset *ba* [72] indicates regular frontal images. Subset *bj* contains

(a) (b) (c) (d) (e) (f)

Figure 2.15 – In the first row (a)–(c) and (d)–(f) show sample images (32×32 pixels) for two differen individuals from subsets *bj*, *bk*, and *ba* from the FERET database, respectively. Second row shows the obtained mask for each image employing the proposed method.

alternative frontal images corresponding to the *ba* images. Subset *bk* is also a frontal image set that corresponds to *ba* but taken under different lighting. Following the previous approaches, 600 frontal face images from 200 individuals are provided via the Subsets *ba*, *bj*, and *bk* respectively, where Subsets *bj* and *bk* are used to train the recognition system, and Subset *ba* is reserved for the testing phase. Fig. 2.15 shows some randomly selected subjects from the FERET database subsets. The images of size 256×384 in subsets are first cropped and resized to small images of size 32×32 for which ℓ is set to 1 and a three-stage DD-DT$\mathbb{C}$WT decomposition is applied on all images in subsets to obtain the mask required for training and testing. As the size of images is reduced, the decomposition tree is further extended to three stages in this example. As mentioned before, the databases can be divided into two groups. Our experiments show that for the databases with notable variation in pose and facial expressions, the best performance and recognition rate is achieved when smaller images and more number of decomposition stages are used.

For visual representation, two randomly selected individuals have been shown in Fig. 2.15. In this figure, in the first row, (a)–(c) show sample images for an individual

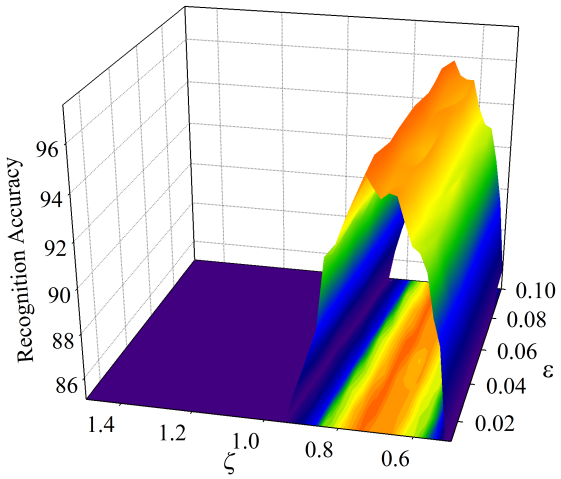


Figure 2.16 – Recognition rate versus ξ and ϵ for the FERET database.

Table 2.4 – FERET: Recognition rate for different techniques (%).

PCA [11]	Gabor	Contourlet	Curvelet	DWT	S–P [30]	Proposed [3]
78.50	89.66	64.14	85.52	82.07	91.72	96.00

from Subsets *bj*, *bk*, and *ba*, respectively. The same is shown for another individual in (d)–(f), respectively. The mask obtained by using the proposed procedure is given in the second row. Similar to previous examples, recognition rate versus ξ and ϵ has been shown in Fig. 2.16 for the FERET database. The result is summarized in Table 2.4. It is seen that the proposed method significantly improves previously reported recognition rates in the literature.

2.5.4 The AT&T Database

The AT&T database, formerly known as the Olivetti Research Lab (ORL) database, contains 10 different images for each of 40 distinct subjects. For some subjects, the images were taken at different times over a period of 2 years, varying the lighting and facial expressions, i.e., open/closed eyes, smiling/not smiling, glasses/no glasses. The images were taken against a dark homogeneous background with the subjects in an

(a) (b) (c) (d) (e) (f)

Figure 2.17 – In the first row (a)–(c) and (d)–(f) show sample images (32×32 pixels) for two differen individuals from the AT&T database, respectively. Second row shows the obtained mask for each image employing the proposed method.

Table 2.5 – AT&T: Recognition rate for different techniques (%).

PCA	Gabor	DWT	CVT+LDA	CVT+PCA	CVT+PCA+LDA [61]	Proposed [3]
91.92	95.50	96.00	95.60	96.60	97.70	97.50

upright frontal position with tolerance for some side movement. The 8-bit gray scale images of size 92×112 pixels have been resized to 32×32 pixels in our experiments. For this small size and similar to the previous example ℓ is set to 1, and using a three-stage decomposition, the proposed algorithm is applied to images. From 10 images per individual, the first five is used for training, that is, 200 images are used for training and the remaining 200 images are reserved for testing phase [30][61].

Figure 2.17 shows three distinct images, (a)–(c) and (d)–(f), from two different individuals randomly selected from the database, respectively. The mask obtained by employing the proposed method is presented in the second row. Recognition rate versus ξ and ϵ has been shown in Fig. 2.18 for the AT&T database. The result is given in Table 2.5. It should be noted that the block-based steerable pyramid method in [30] achieves the 99% accuracy for the best performing subband fusion.

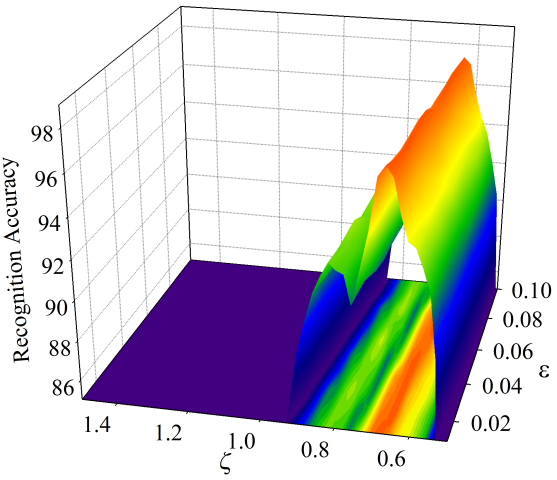


Figure 2.18 – Recognition rate versus ξ and ϵ for the AT&T database.

Figure 2.19 – Sample images from the LFW database for four different individuals. All images, aligned with commercial face alignment software in [91], are available in [43][91]. As it can be seen, the images in LFW are mostly uncontrolled in terms of variation in facial expression, age, race, pose, and occlusion.

2.5.5 The Labeled Faces in the Wild Database

The Labeled Faces in the Wild (LFW) database [43] is a relatively new source of face images designed for studying the problem of unconstrained face recognition. It has been widely used to determine if a pair of face images belong to the same individual or not, where each face image has been labeled with the name of the person pictured. LFW contains 13,233 images of 5749 different individuals collected from the web. Out of 13,233 images in the database, only 1680 of subjects have two or more distinct photos. We use the aligned version of the LFW database [91] available

(a) (b) (c) (d) (e) (f)

Figure 2.20 – Two examples from Subsets 12 and 16 of the LFW database. In the first row, (a)–(f) show sample images (64×64 pixels) for a randomly selected individual from Subset 12. The second row shows the corresponding mask for each image in the first row. The same is represented in rows three and four for another individual from Subset 16, respectively.

online at [43][91]. Fig. 2.19 shows sample images from the LFW database for different individuals. As it can be seen, the images in LFW are mostly uncontrolled in terms of variations in pose, facial expression, lighting, illumination, and occlusion.

To evaluate the performance of the proposed method on this database, we use subsets of the LFW that contain individuals having 10, 11,..., 16 images for instance. Note that the face images with rotation angles more than 30° to left or right, and images with high occlusion have been removed from subsets in our experiments. Subjects with eye-glasses, hat, and the ones with various facial expressions such as closed-eyes remain unchanged. For example, the 19 images of the individual Nicole

Table 2.6 – LFW: Recognition rate for different subsets.

Subset	(%)	NTR	NTS	Subjects	Total images
16	76.25	8	8	10	160
15	89.43	8	7	10	150
14	80.95	7	7	15	210
13	67.78	7	6	15	195
12	72.50	6	6	20	240
11	74.54	6	5	11	121
10	71.42	5	5	21	210

Kidman in the original LFW database has been reduced to 12 images and included in Subset 12 in our experiments. The images and constructed subsets used in our experiments with the LFW database are available from the authors. For the images per individual in each subset, the first half is used for training and the remaining images in the corresponding subset are reserved for testing phase. If the numbers of images are odd, the first half plus one is used for training. All the images were cropped and resized to 64×64 pixels gray level images. We set $\ell = 2$ and the proposed algorithm is applied to all images in subsets, via a three-stage decomposition, to extract the mask required for training and testing. For visual evaluation, some random images have been selected from two different individuals from Subsets 12 and 16, respectively. This has been shown in Fig. 2.20 where in this figure, in the first row, (a)–(f) show sample images for a randomly selected individual from Subset 12. The second row shows the corresponding mask obtained by using the proposed method. The same is shown for another subject from Subset 16, respectively. The recognition results are given in Table 2.6 for different subsets. Similar to previous examples, recognition rate versus ξ and ϵ can be presented for each of the subsets in Table 2.6.

The results indicate that as the total number of images per subset and the number of images used for train and test cycles are different, LFW database contains variety of images in terms of facial expression, age, race, pose, occlusion, and illumination conditions, where the illumination is not the dominant performance degrading fac-

tor. The most influencing characteristics for these images are facial expression and nonuniform photographic style. Subset 15 contains less number of such images, and therefore, a remarkable accuracy difference is observed in contrast with other subsets in this database in Table 2.6.

2.6 Robustness and Training Sample Reduction

Generally, decreasing number of the training samples (NTR) makes the recognition problem more difficult resulting in lower recognition accuracy. Therefore, it is reasonable to reduce the number of images in the training gallery to investigate the effectiveness of the proposed method. Starting from the maximum number of images used for training in each example for the Yale B, Extended-Yale B, and the CMU-PIE databases for instance, we have reduced the training samples until NTR=1. The results of this study have been summarized in Tables 2.7 to 2.9 for the Yale B, Extended-Yale B, and CMU-PIE, respectively. Note that the names of images presented in tables are the same as the ones appear in the corresponding databases in [2][105][106]. This information is given in the last column of the Tables, where TIN denotes the training image name. In each experiment, the remaining images are kept for the testing cycle.

The accuracy obtained by employing our method, even using less number of training samples, compares favorably with recent results in the literature. For example, the results presented in Table 2.1 are obtained using seven images for training (NTR=7). In Table 2.7, we show that the perfect accuracy is guaranteed with only three images for training (NTR=3). Note that the recently proposed method II+PCA+WVS in [66] cannot reach perfect accuracy using seven training samples (Table 2.1). The results in Tables 2.7 to 2.9 indicate that our proposed method is not fully invariant or robust to any reduction in the number of training samples, however, it exhibits a notable robust behavior in the presence of high illumination if compared to similar

Table 2.7 – Yale B: Recognition rate vs. number of training samples.

NTR	Subset 2	Subset 3	Subset 4	Subset 5	TIN
7	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10, +0E+0, +5E+10, +5E-10
6	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10, +0E+0, +5E+10
5	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10, +0E+0
4	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10
3	100	100	100	100	-10E+0, -5E-10, +10E+0
2	100	100	100	100	-10E+0, -5E-10
1	100	100	99.29	100	-10E+0

Table 2.8 – Extended-Yale B: Recognition rate vs. number of training samples.

NTR	Subset 2	Subset 3	Subset 4	Subset 5	TIN
7	100	100	98.68	99.03	-5E-10, +0E+0, -5E+10, -10E+0, +5E-10, +5E+10, +10E+0
6	100	100	98.50	98.89	-5E-10, +0E+0, -5E+10, -10E+0, +5E-10, +5E+10
5	100	100	98.31	98.75	-5E-10, +0E+0, -5E+10, -10E+0, +5E-10
4	100	100	98.12	98.61	-5E-10, +0E+0, -5E+10, -10E+0
3	100	99.78	97.74	98.47	-5E-10, +0E+0, -10E+0
2	100	99.56	96.80	98.20	-5E-10, -10E+0
1	98.27	98.90	95.87	97.37	-5E-10

existing methods. It is worth pointing out that, we have focused only on the use of PCA for feature extraction. In Table 2.1, the results obtained by S–P and S–P+PP (rows 5 and 6 from the bottom in Table 2.1), indicate that how the recognition rate can be significantly improved if a preprocessing step is added to the S–P technique in [30]. Keeping this point in mind and in view of the promising results achieved by the proposed method based on PCA only, it is expected that replacing PCA by a state-of-the-art feature extraction algorithm may improve the results presented in this chapter provided that the computational complexity remains reasonable.

The computational time required for the training is usually an offline process. The execution time, using the proposed method, to process an incoming image of size 128×128 pixels via a three level DD-DTℂWT in MATLAB is 58 milliseconds (ms/image) using a Core 2 Duo 2.40 GHz processor and 2GB of memory. This is the

Table 2.9 – CMU-PIE: Recognition rate vs. number of training samples.

lighting condition	NTR	NTS	Total	%	TIN
lights-off	4	17	21	100	i8, i9, i11, i20
	3	18	21	99.59	i8, i11, i20
	2	19	21	99.42	i11, i20
	1	20	21	98.97	i11
lights-on	4	20	24	100	I8, I9, I11, I20
	3	21	24	100	I8, I11, I20
	2	22	24	100	I8, I11
	1	23	24	100	I8
lights-on/off	7	38	45	99.92	i8, i11, i20, I8, I9, I11, I20
	6	39	45	99.77	i8, i11, i20, I8, I11, I20
	5	40	45	99.70	i8, i11, I8, I11, I20
	4	41	45	99.64	i8, i11, I8, I11
	3	42	45	99.02	i8, I8, I11
	2	43	45	98.70	i8, I11
	1	44	45	84.59	I20

Table 2.10 – Computational time for different methods for an incoming image of size 128×128 pixels.

Method	DWT	Gabor	Curvelet	Contourlet	S–P
ms/image	29	123	59	17	50

time to complete the full procedure from block 1 to 7 in Fig. 2.3 to read an input image and to generate its corresponding illumination invariant mask. The execution time for different algorithms has been shown in Table 2.10 (see Table 10 in [30]) for reference. The experiments in [30] were carried out via a Dual Core processor at 2.00 GHz and 2GB of RAM. Note that feature extraction, dimensionality reduction, and classification are separate and independent procedures where PCA and ELM have been used in our experiments. Compared with other approaches, the proposed method offers promising results for recognition accuracy and at the same time performs reasonably well in terms of computational complexity. It should be noted that for most of the cases there is a trade-off between accuracy and execution time. For example, for the WVS-based method [66] and using the same processor as in our

experiments, it takes seconds to calculate several types of weights for each window in a single image. Recall that, implementation of face recognition algorithms in real applications usually deals with massive databases when several subjects are involved. It is quite computationally expensive, if one may desire to employ a recognition technique with higher accuracy but lower execution timing rate on a large database. It should be noted that the proposed method is not a feature extraction and dimensionality reduction algorithm but an efficient illumination removal technique. Being more specific, most of the feature extraction techniques are eventually used in presence of pose and facial expression variation. However, to apply these algorithms on databases such as Subsets 4 and 5 in the Extended-Yale B, one needs to perform a preprocessing step first. We believe that in case the feature vector formation and dimensionally reduction phase in Fig. 2.3, blocks 8 and 9, is replaced by a specific feature extractor algorithm like [102], the whole procedure can be applied to highly illuminated version of uncontrolled databases such as the Labeled Faces in the Wild to reach higher accuracy.

2.7 Conclusions and Summary

In this chapter we have proposed a new and efficient method for human face recognition under varying illumination. The proposed method is based on enhancing the illumination effects of images first, and then suppressing the illumination via multiresolution decomposition of input images. In view of the directional-selectivity, shift-invariance, and extra number of wavelets of the double-density dual-tree complex wavelet transform, redundant illumination can be reduced by thresholding the high-frequency subbands to construct a raw image. We have also suggested an additional step, the so-called lifting and fine-tuning process, which amplifies the role of fine features and key points in an image. The extracted face image masks significantly facilitate the illumination invariant face recognition task. Principal component anal-

ysis (PCA) is used for dimensionality reduction and the extreme learning machine (ELM) is employed as a classifier, which offers classification at considerably higher speed in comparison with other learning approaches. Several experiments have been performed using available well documented databases used in traditional face recognition problems. The results indicate that the proposed method compares favorably to the recent results in the literature in terms of recognition rate and computational complexity.

Further work in this direction may also include finding 1) an analytical model to represent the reflectance part of an image only, 2) a similar strategy for object recognition (not necessarily face) in color images, 3) a fully automated approach to determine the parameter ξ depending on an input image, 4) embedded solutions, the so-called system level design technique, to combine the algorithm steps as a compact system. The latter is left as an open problem where we expect that the proposed framework can be further studied in embedded-classifier design. We are currently investigating the classifiers attributes and feature selection/extraction algorithms, to be controlled analytically in the wavelet domain, for the design of customized classifiers in complex applications such as military image recognition problems and biomedical instrumentation.

Chapter 3

Frequency Selectivity and Illumination Invariant Analysis

3.1 Literature Review

In Chapter 2, we proposed a novel method for the problem of illumination invariant human face recognition based on illumination suppression via multiresolution analysis. As discussed before, multiresolution analysis allows the decomposition of a signal into its frequency components. It is basically an analysis-synthesis configuration, where a signal is represented in various frequency subbands via the analysis side of a wavelet filter bank. In a perfect reconstruction filter bank the synthesis side in turn can reconstruct the original signal. Now a valid and important question is how can the characteristics of the filters of such a structure affect the recognition results? Being more specific, how the frequency selectivity of the filters of the transformation may change the performance in an application. By the term filter characteristics, it is technically referred to as the transition band and magnitude of ripples of the frequency response of the filters designed and employed in a transformation. This chapter answers to these questions analytically and theoretically, and provides a unique contribution to the relation between the frequency selectivity of the transformation and the recognition accuracy of a face recognition system.

Technically, multiresolution analysis or multiscale approximation is used to deal with the theory and design of discrete wavelet transform (DWT), multiwavelets, dual-tree complex wavelet transform (DT-$\mathbb{C}$WT), higher-density discrete wavelet transform (HD-DWT), curvelets and framelets [23][59][60][73][77][78][89][99]. This family contains several desired properties such as orthogonality, biorthogonality, regularity, and continuity. Although these wavelets cannot be explicitly expressed, they can be written as a function of another wavelet in the family. In addition, there is a small group of wavelets in literature known as crude wavelets, for which the wavelet filters are generated using an explicit mathematical equation. Even with explicit, continuous, and theoretically infinite mathematical equations for these crude wavelets, still one needs to produce discrete finite filters for digital applications. The literature on pulse shaping and crude wavelets is vast and various pulses such as Gaussian, Mexican hat, hyperbolic, Meyer and Morlet, which belong to the crude wavelet group have been studied by researches for decades [52]. An interesting approach to edge detection in hyperbolic- and Gaussian-distributed pixel-intensity images using hyperbolic and Gaussian masks has been presented in [53]. In [54], hyperbolic kernels are further investigated and auto-term functions of the first-order hyperbolic kernel, Choi-Williams (CW) kernel and nth-order hyperbolic kernel have been explicitly derived.

Although many state-of-the-art applications take advantage of existing multiresolution transformations, the theory and design of new multiresolution systems is a difficult and challenging problem whose solution helps to improve relevant applications in image processing and pattern recognition. This chapter deals with the design of a class of wavelet filter banks, halfband pair filter bank (HPFB), which is referred to as tunable halfband pair (THP) throughout the text. The proposed design scheme offers wavelet filters ranging from the lowest frequency selectivity to the sharpest transition band along with a control and tuning opportunity over the filters characteristics required in multiresolution based applications. Traditional maximally flat wavelet filters [23] are highly regular but have poor frequency selectivity. In [76],

Rioul and Duhamel showed that regularity and frequency selectivity are in an inverse relationship. In its basic form a lowpass filter is regular if it has at least one zero at $z = -1$, the so-called number of vanishing moments (NVM). While regularity is an important task in wavelet based compression, *frequency selectivity* is also another significant factor to be considered within applications such as denoising, subband coding, classification and recognition.

In order to analyze the effect of the frequency selectivity of the filters of a multiresolution transformation on the performance of a face recognition system, the challenge here is first to propose a wavelet design method that can provide a tuning opportunity to control the desired specifications. In other words, the problem is defined as how to design wavelet filters with the sharpest transition band for a fixed number of vanishing moments. This chapter is motivated by the possibility, and the need for improvements of wavelet filters characteristics in two-channel perfect reconstruction filter banks. The necessity stems from the lack of a clear control over the frequency response of the filters in terms of passband/stopband edges, number of vanishing moments, and ripples. Various approaches have been considered for the design of perfect reconstruction filter banks, e.g., the Remez exchange algorithm, least squares, and Eigenfilter [73]. In [13], optimization of filter banks is investigated for invariant supervised texture segmentation. Multiscale directional filter bank is discussed extensively in [19] to suppress the aliasing effect, as well as to minimize the reduction in frequency resolution where the problem of aliasing in decimated bandpass images on directional decomposition has been addressed. Dumitrescu formulated an SDP problem [28] that improved the orthogonality error of our earlier orthogonal filter banks. In [29], and based on an SDP framework to guarantee the global optimality, it has been shown that an implicit form of regularity constraint imposition is much more appropriate in terms of numerical accuracy. The idea of using parametric Bernstein polynomial in wavelet design has been investigated by several researchers. In [108], a sum of squares based method was proposed for the design of halfband product filters

for orthogonal wavelets. Zhang indicates that the well-known Remez exchange algorithm is an efficient approach for equiripple design of orthogonal FIR filters [111]. A generalized parametric quadrature mirror filter bank design technique was suggested in [22], where the authors introduced parametric polynomials to approximate the orthogonal wavelet filters. Phoong et al. [73], proposed a design procedure for halfband pair filter bank (HPFB) with the structural perfect reconstruction property, where the lowpass analysis filter, which is assumed to be a halfband filter, was used in a one-stage optimization process employing two kernels where the kernels are mostly assumed to be the same. Tay suggested a two-stage least squares design of HPFB, defining two different polynomial based kernels [95], where the objective function to be minimized to extract free parameters is the energy of ripples. Another method was proposed by Patil et al. [71] to design FIR wavelet filter banks using factorization of a halfband polynomial in the frequency domain.

3.2 Motivation and Problem Statement

Motivated by the capability of parametric polynomials in filter design and in the light of notable progress in semidefinite programming techniques we first derive filter coefficients in the polynomial domain (in the variable x) in terms of the coefficients of the corresponding function in z-domain. We then define a new objective function and problem formulation based on SDP representation of the obtained polynomial. It is pointed out that due to the use of SDP and the way we define the problem, for a fixed filter length and pre-specified filter characteristics, solution to this problem can generate filter pairs ranging from the widest to the sharpest transition band width. The proposed technique offers tuning opportunity on the passband and stopband widths, amplitude of ripples, and number of vanishing moments. Based on these factors, one can tune the specified requirements in an application.

As it was mentioned earlier the main purpose of this chapter is to analyze the effect

of the frequency selectivity of a multiresolution transformation on the performance of a face recognition system which requires to first designing the corresponding filters such that the desired characteristics can be tuned and controlled. While the accuracy of pixel information highly depends on its frequency content in wavelet subbands, it is reasonable to use wavelets with the sharpest transition band width. Knowing that the performance of a face recognition system is mostly related to the quality of the extracted features, informative and healthy feature extraction is an important phase of the recognition problem. On the other hand, several classifiers may be employed for the recognition and classification part, however, the quality of features is the dominant factor. We have implemented and realized several experiments to evaluate and study the effect of the frequency selectivity employing the tunable filters. Illustrative examples are provided and varieties of experiments have been carried out using the well-known databases such as the Yale B and CMU-PIE to support the analysis and results.

Also, to validate the quality filters designed and the effect of the frequency selectivity analysis approach shown in this chapter, a similar analysis has been carried out in multifocus image fusion and shape from focus. The reconstruction of a geometric object and to retrieve spatial information from one or multiple observation is a challenging problem in computer vision. When a 3D scene is projected into a 2D image plane, depth information is lost. Considering the advantage of tunable and frequency selective filter pairs proposed in this chapter the effect of frequency selectivity is inspected in multifocus imaging to select the best fitting feature pixels from the input images. Experimental results and quantitative comparisons show how the performance of a multiresolution based multifocus imaging system can be affected via the use of various filters with different sharpness.

3.3 Problem Formulation and Derivations

Let us consider a two-channel filter bank with the analysis and synthesis lowpass (highpass) filters denoted by $H_0(z)$ ($H_1(z)$) and $F_0(z)$ ($F_1(z)$), respectively. It is now well known that the scaling and wavelet functions associated with the analysis side of the filter bank are defined by the two-scale equations in terms of the impulse response of the filters in the filter bank, h_0 and h_1. The scaling function and wavelets associated with the synthesis side of the filter bank are similarly defined via the corresponding impulse responses f_0 and f_1, respectively [99]. Considering the typical structure of a two-channel filter bank, the goal here is to design $H_0(x)$ and $F_0(x)$, and consequently obtain $H_1(x)$ and $F_1(x)$. Representation of a complicated function (in the variable z for instance) in the polynomial domain (in the variable x) motivates to derive and optimize the filter functions in the variable x. The philosophy of such a strategy is twofold. First, unlike a polynomial in the x-domain, the filter function in the variable z in the frequency domain cannot be simply tailored for a desired response. Secondly, in case one could find a reasonable way to define the problem in the z-domain, the main question is how to easily search or solve for the desired solution. Recent progress in optimization techniques, e.g., Genetic algorithm and convex optimization, provide the opportunity to efficiently search for an optimum solution in the polynomial domain.

Fig. 3.1 (a) shows a typical sample filter pair characteristics in the frequency domain in the variable z where $z = e^{j\omega}$. The same filters are represented in Fig. 3.1 (b) in the variable x in terms of a univariate polynomial. x_{c_p} and x_{c_s} denote the passband and stopband edges, $H_1(x)$ and $F_1(x)$ are the highpass versions of $F_0(x)$ and $H_0(x)$, respectively, and the notations $\Lambda(x)$ and $\Omega(x)$ are used to define filter kernels as is described later. The filter function, which has been defined in the z-domain, is transformed into the x-domain through the change of variable, and thus the problem is converted to a semidefinite programming problem which can be solved

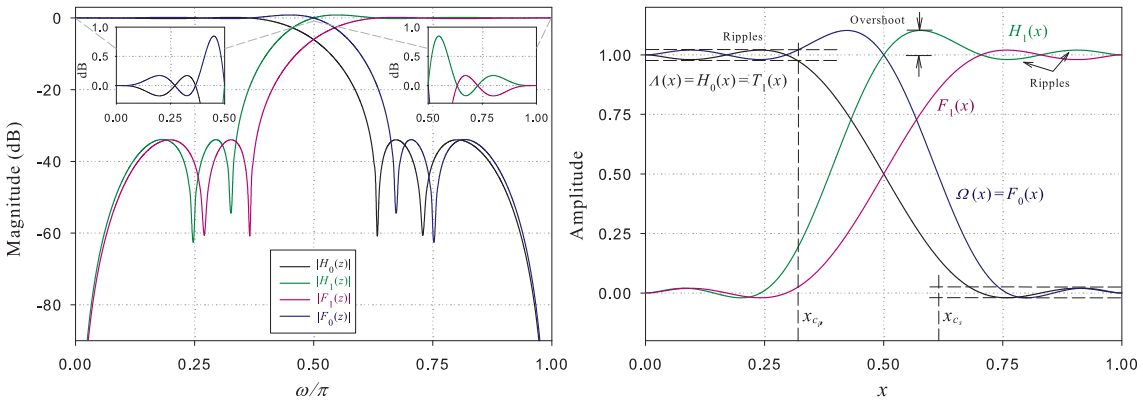


Figure 3.1 – Typical specifications of filters in a sample HPFB pair. (a) in the variable z; (b) in the variable x. Passband and stopband edges, x_{c_p} and x_{c_s} respectively, are shown for H_0 for instance.

efficiently in the variable x to enforce for and to control the desired characteristics. The whole idea and problem statement can be briefly summarized as follows [4].

1. General form of a halfband filter function, $P(z)$, is taken into consideration to be optimized for its coefficients ($\boldsymbol{a}$) such that the desired characteristics are met.

2. Following the discussion above and employing a univariate polynomial, $T(x)$, with the coefficient vector $\boldsymbol{b}$, $T(x)$ is to be written in terms of $\boldsymbol{a}$, the coefficient vector of the frequency domain function $P(z)$.

3. After transformation of $P(z)$ from z-domain into the x-domain, $T(x)$ can be represented via SDP, i.e, the frequency response can be bounded for a given transition band and ripples.

4. Finally, for a given filter length and number of vanishing moments, we define an optimization problem to determine $\boldsymbol{b}$, and consequently $\boldsymbol{a}$, such that the passband and stopband widths be maximized (the sharpest transition band) subject to the desired design characteristics.

The design procedure starts by defining a filter function in the frequency domain. Daubechies wavelet filters contain the maximum possible number of vanishing moments for a fixed filter length, however, there is no free parameter to have some degree of freedom to control the frequency response of the wavelet filters to be able to in-

crease the frequency selectivity of the filters obtained. We want to incorporate free parameters to control the design specifications, and therefore, we first map a general halfband polynomial from z- into the x-domain by a variable transformation. While it is possible to directly optimize the desired specifications in x-domain, it cannot be easily performed in the variable z. Knowing that a halfband filter function can be defined by the following expression [59], [89]

$$\begin{aligned} P(z) &= a_0 + a_2 z^{-2} + \ldots + a_{K/2-1} z^{-K/2+1} + a_{K/2} z^{-K/2} \qquad (3.1) \\ &+ a_{K/2-1} z^{-K/2-1} + \ldots + a_2 z^{-K+2} + a_0 z^{-K} \\ &= R(z)(z+1)^{-2m} \qquad (3.2) \end{aligned}$$

where $R(z)$ is the reminder term with symmetric coefficients, our aim is to obtain the optimal values for the coefficients $\boldsymbol{a} = \{a_0, a_2, \ldots, a_{K/2-1}, a_{K/2}\}$ such that for a fixed filter length, pre-specified NVM and ripples, the corresponding filter has the sharpest transition band, leading to the best frequency selectivity in its class. In this notation, K is an even integer, $K/2$ must be odd, and m is the representative for the number of required zeros at $z = -1$. Let us assume $T(x)$ be a univariate polynomial in the variable x defined by

$$\begin{aligned} T(x) &= (b_M x^M + b_{M-1} x^{M-1} + \ldots + b_1 x + b_0)(x-1)^m \qquad (3.3) \\ &= f_N(\boldsymbol{b}) x^N + f_{N-1}(\boldsymbol{b}) x^{N-1} + \ldots + f_1(\boldsymbol{b}) x + (-1)^m b_0 \qquad (3.4) \end{aligned}$$

where $M + m = N$, m is the number of zeros at $x = 1$, $N = K/2$, $f_{(.)}(\boldsymbol{b})$ refers to a function of b_is with $i = 0, \ldots, M$ to denote coefficients of the polynomial in (3.4), and we define $\boldsymbol{b} = \{b_M, \ldots, b_1, b_0\}$. The polynomial in (3.4), which represents filter function in the variable x, can be transformed into the z-domain through a change of variable, that is, $x = \frac{1}{2}(1 - \cos\omega) = \frac{1}{2}(1 - \frac{e^{jw}+e^{-jw}}{2}) = \frac{1}{2}(1 - \frac{z+z^{-1}}{2}) = -\frac{1}{4} z (1 - z^{-1})^2$ with $z = e^{j\omega}$. It must satisfy the halfband condition $T(x) + T(1-x) = 1$ required for perfect reconstruction. Assuming that the final goal is to find optimal values for $\boldsymbol{a}$, we shall find the relationship between $\{a_0, a_2, \ldots, a_{K/2-1}, a_{K/2}\}$ and $\{b_M, \ldots, b_1, b_0\}$

before performing an optimization cycle in the variable x. For the class of filters discussed here although $H_0(x)$ is an antisymmetric function about the point ($x = 0.5, H_0(x) = 0.5$), $F_0(x)$ is not necessarily an antisymmetric function and thus, similar to [95] the following kernels are used to design H_0 and F_0.

$$\Lambda(x) \triangleq T_1(x) \tag{3.5}$$

$$\Omega(x) \triangleq T_1(x) + 2T_2(x) - 2T_1(x)T_2(x) \tag{3.6}$$

where T_1 and T_2 are derived from (3.4) with different free parameter vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, and the lowpass filters in the analysis and synthesis sides of the filter bank are determined by

$$H_0(z) = \Lambda(\frac{-z(1-z^{-1})^2}{4}) \tag{3.7}$$

$$F_0(z) = \Omega(\frac{-z(1-z^{-1})^2}{4}) \tag{3.8}$$

with the corresponding highpass filters given by $H_1(z) = z^{-1}F_0(-z)$ and $F_1(z) = zH_0(-z)$. Note that H_1 and F_1 are the highpass versions of F_0 and H_0 respectively, and therefore, it is sufficient to design either H_0 and F_0, or H_1 and F_1. Selection in (3.5) and (3.6) ensures perfect reconstruction which can be verified using the property $T(x) + T(1-x) = 1$, that is

$$H_0(-z) \equiv T_1(1-x) = 1 - T_1(x) \tag{3.9}$$

$$F_0(-z) \equiv T_1(1-x) + 2T_2(1-x) \tag{3.10}$$

$$- \quad 2T_1(1-x)T_2(1-x) \tag{3.11}$$

$$= \quad 1 + T_1(x) - 2T_1(x)T_2(x). \tag{3.12}$$

From (3.5), (3.6), (3.9), and (3.12) it is verified that perfect reconstruction is achieved and $H_0(z)F_0(z) + H_0(-z)F_0(-z) = 1$ [89]. It is important to note that the condition holds if the halfband property of $T(x)$ is assured. In the following section it is shown that how this property is imposed within the design procedure. The obtained lowpass filters, $H_0(z)$ and $F_0(z)$, structurally hold the desired number of vanishing moments,

i.e., $\mathrm{NVM}_\Lambda = 2m_{T_1}$, $\mathcal{L}_\Lambda = 2N_{T_1}+1$, $\mathrm{NVM}_\Omega = 2\mathtt{min}\{m_{T_1}, m_{T_2}\}$, and $\mathcal{L}_\Omega = 4N_{T_2}+1$ where $\mathcal{L}$ refers to filter length. Relations (3.5) and (3.6) employ two different polynomials T_1 and T_2 with m_{T_1} and m_{T_2} zeros at $x=1$, respectively. NVM_Λ and NVM_Ω denote the number zeros at $z=-1$ after the variable transformation from x to z. Performing a change in the variable in (3.4), $T(x)$ is then written in terms of z, where the equivalency of $T(z) \equiv P(z)$ results in

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ (-1)^1\binom{K}{1} & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ (-1)^N\binom{K}{N} & (-1)^{N-1}\binom{K-2}{N-1} & \dots & 1 \end{bmatrix} \times \begin{bmatrix} (-\frac{1}{4})^N f_N(\boldsymbol{b}) \\ (-\frac{1}{4})^{N-1} f_{N-1}(\boldsymbol{b}) \\ \vdots \\ (-\frac{1}{4})^1 f_1(\boldsymbol{b}) \\ (-\frac{1}{4})^0 (-1)^m b_0 \end{bmatrix} = \begin{bmatrix} a_0 \\ 0 \\ a_2 \\ 0 \\ \vdots \\ a_{K/2-1} \\ a_{K/2} \end{bmatrix} \tag{3.13}$$

and therefore

$$\begin{bmatrix} f_N(\boldsymbol{b}) \\ f_{N-1}(\boldsymbol{b}) \\ \vdots \\ f_1(\boldsymbol{b}) \\ (-1)^m b_0 \end{bmatrix} = \Bigg(\begin{bmatrix} 1 & 0 & \dots & 0 \\ (-1)^1\binom{K}{1} & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ (-1)^N\binom{K}{N} & (-1)^{N-1}\binom{K-2}{N-1} & \dots & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} (-\frac{1}{4})^N & 0 & \dots & 0 \\ 0 & (-\frac{1}{4})^{N-1} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & (-\frac{1}{4})^0 \end{bmatrix} \Bigg)^{-1} \begin{bmatrix} a_0 \\ 0 \\ a_2 \\ 0 \\ \vdots \\ a_{K/2-1} \\ a_{K/2} \end{bmatrix} = \begin{bmatrix} g_N(\boldsymbol{a}) \\ g_{N-1}(\boldsymbol{a}) \\ \vdots \\ g_1(\boldsymbol{a}) \\ g_0(\boldsymbol{a}) \end{bmatrix} \tag{3.14}$$

where the right hand side of the latter equality indicates how $f_{(.)}(\boldsymbol{b})$s are related to functions of $\boldsymbol{a}$, denoted by $g_{(.)}(\boldsymbol{a})$. It should be noted that $f_N(\boldsymbol{b}) = b_M$, and thus, b_M is written based on $\boldsymbol{a}$. $f_{N-1}(\boldsymbol{b})$ depends on $f_N(\boldsymbol{b})$, and therefore, $f_{N-1}(\boldsymbol{b})$ is written based on b_M and consequently in terms of $\boldsymbol{a}$. Similarly, $f_{N-2}(\boldsymbol{b})$ needs $f_{N-1}(\boldsymbol{b})$ and $f_N(\boldsymbol{b})$ and so on. Therefore, starting from top to bottom and substituting the values, all $f_{(.)}(\boldsymbol{b})$s are derived based on $g_{(.)}(\boldsymbol{a})$s. In order to structurally impose the desired number of vanishing moments, one may factorize either $P(z)$ as in (3.2) to extract $R(z)$, or $T(x)$ as in (3.3). Recall that $x = 1$ in $x = -\frac{1}{4}z(1 - z^{-1})^2$ yields $z = -1$. In practice, the expression in (3.3) is more convenient than extracting $R(z)$ in (3.2). It is worth pointing out that the matrix equality in (3.14) represents the direct relation between the coefficients of $T(x)$ and $P(z)$ in general taking into account that the number of vanishing moments have been structurally imposed. In addition to this relationship, $T(x)$ must be halfband, i.e., $T(x)$ must satisfy $T(1) = 0$, $T(0) = 1$, $T(0.5) = 0.5$, and $T(x)$ must be an antisymmetric function about the point $(x = 0.5, T(x) = 0.5)$. $T(1) = 0$ is automatically satisfied due to the existence of the term $(x - 1)^m$. $T(0) = 1$ results in $b_0 = (\pm 1)^m$, and $T(0.5) = 0.5$ presents another relationship between the coefficients. The antisymmetry of $T(x)$ is achieved by replacing x and T with $x + 0.5$ and $T + 0.5$, respectively, and then applying the antisymmetric condition at origin, that is, $-T(x) = T(-x)$. Now the coefficients of $T(x)$, $\boldsymbol{b}$, are tightly related to the coefficients of $P(z)$, $\boldsymbol{a}$, holding the required conditions for perfect reconstruction and number of vanishing moments.

In Section 3.4 it is shown that $T(x)$ is bounded in desired intervals depending on the size of ripples and the transition band width. In other words, $T(x)$, which is the filter function in the variable x, is bounded to be optimized for its coefficients, $\boldsymbol{b}$, where they are already written in terms of the coefficient vector $\boldsymbol{a}$. The following two examples give a preview regarding the idea, whereas the details are extensively discussed in Section 3.4.3.

Let us start with the maximally flat filter for the case $N = 7$. We must set $m = 4$

and thus

$$\begin{aligned} P(z) &= a_0 + a_2 z^{-2} + a_4 z^{-4} + a_6 z^{-6} + a_7 z^{-7} + a_6 z^{-8} + a_4 z^{-10} + a_2 z^{-12} + a_0 z^{-14} \\ &= R(z)(z+1)^{-8} \end{aligned}$$

where the factor $R(z)$ is of degree -6 and $\boldsymbol{a} = \{a_0, a_2, a_4, a_6, a_7\}$. The equivalent polynomial in the variable x is then written as

$$T(x) = (b_3 x^3 + b_2 x^2 + b_1 x + b_0)(x-1)^4$$

where $\boldsymbol{b} = \{b_3, b_2, b_1, b_0\}$, and consequently

$$\begin{bmatrix} 1b_3 \\ 1b_2 \\ 1b_1 \\ 1b_0 \\ -4b_0 \\ 6b_0 \\ -4b_0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -4b_3 \\ -4b_2 + 6b_3 \\ -4b_1 + 6b_2 - 4b_3 \\ 6b_1 - 4b_2 + 1b_3 \\ -4b_1 + 1b_2 \\ +1b_1 \\ 0 \end{bmatrix} = \begin{bmatrix} -16384 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 57344 & 4096 & 0 & 0 & 0 & 0 & 0 & 0 \\ -78848 & -12288 & -1024 & 0 & 0 & 0 & 0 & 0 \\ 53760 & 13824 & 2560 & 256 & 0 & 0 & 0 & 0 \\ -18816 & -7168 & -2240 & -512 & -64 & 0 & 0 & 0 \\ 3136 & 1680 & 800 & 320 & 96 & 16 & 0 & 0 \\ -196 & -144 & -100 & -64 & -36 & -16 & -4 & 0 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} a_0 \\ 0 \\ a_2 \\ 0 \\ a_4 \\ 0 \\ a_6 \\ a_7 \end{bmatrix} \tag{3.15}$$

Starting from the first row, b_3 is obtained based on $\boldsymbol{a}$, and by substituting it in the second row, b_2 is also written in terms of $\boldsymbol{a}$ and the procedure is stopped in the fourth row. That is, $b_3 = -16384a_0$, $b_2 = -8192a_0$, $b_1 = -13312a_0 - 1024a_2$. As $T(0) = 1$, then $b_0 = 1$ and from the fourth row $-15872a_0 - 1536a_2 - 1 = 0$. Also $T(0.5) = 0.5$ results in $-672a_0 - 32a_2 - \frac{7}{16} = 0$. Therefore, $a_0 = -0.001220703125$, $a_2 = 0.011962890625$, and consequently $b_3 = 20$, $b_2 = 10$, $b_1 = 4$, and from the other rows in (3.15) we have $a_4 = -0.059814453125$, $a_6 = 0.299072265625$, and $a_7 = 0.5$. While the maximally flat can be obtained without any optimization process, that is not the case if one demands a sharper transition band at the price of lower number of vanishing moments. This is shown through the following example.

For $N = 7$, and in case of $m = 2$ instead of maximally flat ($m = 4$), still we have

$$P(z) = a_0 + a_2 z^{-2} + a_4 z^{-4} + a_6 z^{-6} + a_7 z^{-7} + a_6 z^{-8} + a_4 z^{-10} + a_2 z^{-12} + a_0 z^{-14}$$

however, $P(z) = R(z)(z+1)^{-4}$, where $R(z)$ is of degree -10, $\boldsymbol{a} = \{a_0, a_2, a_4, a_6, a_7\}$, and

$$T(x) = (b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0)(x-1)^2$$

where $\boldsymbol{b} = \{b_5, b_4, b_3, b_2, b_1, b_0\}$ and therefore

$$\begin{bmatrix} 1b_5 \\ 1b_4 \\ 1b_3 \\ 1b_2 \\ 1b_1 \\ 1b_0 \\ -2b_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} 0 \\ -2b_5 \\ -2b_4 + 1b_5 \\ -2b_3 + 1b_4 \\ -2b_2 + 1b_3 \\ -2b_1 + 1b_2 \\ +1b_1 \\ 0 \end{bmatrix} = \begin{bmatrix} -16384 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 57344 & 4096 & 0 & 0 & 0 & 0 & 0 & 0 \\ -78848 & -12288 & -1024 & 0 & 0 & 0 & 0 & 0 \\ 53760 & 13824 & 2560 & 256 & 0 & 0 & 0 & 0 \\ -18816 & -7168 & -2240 & -512 & -64 & 0 & 0 & 0 \\ 3136 & 1680 & 800 & 320 & 96 & 16 & 0 & 0 \\ -196 & -144 & -100 & -64 & -36 & -16 & -4 & 0 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} a_0 \\ 0 \\ a_2 \\ 0 \\ a_4 \\ 0 \\ a_6 \\ a_7 \end{bmatrix} \tag{3.16}$$

Following the same procedure as in the previous example, we have $b_4 = -3b_3 - 6b_2$, $b_5 = 2b_3 + 4b_2$, and therefore, $b_5 = -16384a_0$, $b_4 = 24576a_0$, $b_3 = -13312a_0 - 1024a_2$, $b_2 = 2560a_0 + 512a_2$, $b_1 = 2$, and $b_0 = 1$. Now depending on the desired ripple size and the transition band width, the optimum values for coefficients must be determined via an optimization cycle which has been shown in Case 1 in Section 3.4.3. This is the technical elaboration of the development procedure summarized at the beginning of Section 3.3. In the next section, we shall apply the bound conditions on $T(x)$ to incorporate filters characteristics in the formulation in the variable x. Our aim here is to find filter coefficients $\boldsymbol{a}$, which are directly defined as a semi-definite programming problem variables. Note that coefficients of $T(x)$, $\boldsymbol{b}$, have been written in terms of $\boldsymbol{a}$.

3.4 Kernels and the Design of THPs

The nonnegativity of the frequency response of the product filter on the unit circle is a critical condition that shall be taken into consideration in orthogonal design [25][64][108]. Since the nonnegativity of the frequency response of the product filter is not a needed condition to be satisfied in biorthogonal design, the strict nonnegativity

is relaxed to have some negative ripples. Therefore, we first study the nonnegativity of polynomials in intervals and then show that $T(x)$ can be represented via SDP decompositions. The latter leads to the point that the frequency response can be bounded for a given tolerance error.

3.4.1 Nonnegativity of Polynomials on Domains

For a univariate polynomial $p(x)$, nonnegativity for a finite and bounded interval denoted by $\mathcal{D}$, is utilized via sums of squares. In other words, a necessary and sufficient condition for a univariate polynomial to be positive semidefinite (PSD) is to write it in a sum of squares form [12]. Note that if $p(x) \geq 0$ for all $x \in \mathbb{R}$, then the polynomial is said to be PSD.

Theorem 1: Let $s(x)$ be a univariate polynomial of degree $2k$ with monomials of degree less than or equal to k, i.e., $v = [1,\ x,\ x^2, \ldots,\ x^k]$. If $s(x)$ can be expressed as a positive semidefinite quadratic form $s = v^t \Delta v$, where $\Delta \in \mathbb{S}_+$[1], then s is a sum of squares. Conversely, s can be expressed as a positive semidefinite quadratic form in the monomials, that is, $s = v^t \Delta v$ for some $\Delta \in \mathbb{S}_+$.

Proof. See Chapter 4 in [12].

Theorem 2: Let $p(x)$ be a univariate polynomial of odd degree $2k+1$. If $p(x)$ is nonnegative on a domain $\mathcal{D} = \{x|\ x \in [\alpha\ \beta],\ \alpha, \beta \in \mathbb{R}\}$, that is, $p(x) \geq 0$ for all $x \in \mathcal{D}$, then there exist sum of squares $s_\mu(x)$ and $s_\eta(x)$ of degree not more than $2k$ such that

$$p(x) = (x - \alpha) s_\mu(x) + (\beta - x) s_\eta(x). \tag{3.17}$$

Likewise, if polynomials of even degree are to be used, i.e., $p(x)$ of degree $2k$, then $p(x) \geq 0$ on $\mathcal{D} = \{x|\ x \in [\alpha\ \beta],\ \alpha, \beta \in \mathbb{R}\}$, if and only if there exist sum of squares $s_\mu(x)$ and $s_\eta(x)$ of degree not more than $2k-2$ such that

$$p(x) = s_\mu(x) + (x - \alpha)(\beta - x) s_\eta(x). \tag{3.18}$$

Proof. See [74], or Section III in [25].

[1] $\mathbb{S}_+$ denotes the set of symmetric positive semidefinite square matrices and t stands for transpose.

It is pointed out that not all positive polynomials are necessarily sum of squares. For real polynomials the condition holds for 1) univariate polynomials of any degree, 2) quadratic polynomials of any number of variables, and 3) quartic polynomials of two variables. In general, the condition holds for trigonometric polynomials, where all variables are bounded. The condition that a polynomial be written as sum of squares and viewed as a constraint on its coefficients turns out to be equivalent to a linear programming problem and therefore a variety of optimization problems with such constraints can be posed as SDPs [12]. Strictly speaking, it is shown that the condition $s = v^t \Delta v$ is a set of linear equality constraints relating the coefficients of s and the matrix Δ. Combined with Theorem 1 and 2, this shows that the condition is then equivalent to a set of linear equalities relating Δ and the coefficients of the polynomial, and the matrix inequality $\Delta \geq 0$. See Chapter 17 in [64] and Problem 4.45 in [12].

3.4.2 Representation of Kernels

Noting that N is an odd integer [22, 89], the representation in (3.17) is used to define sets of inequalities required in biorthogonal design scheme illustrated in Fig. 3.1, where the strict positivity in (3.17) is relaxed to have some negative ripples. Thus, $\Lambda(x)$ has passband and stopband edges x_{c_p} and x_{c_s} respectively, if the following inequalities and consequently the corresponding equality decompositions are satisfied

$$1 - \delta \leq \Lambda(x) \leq 1 + \delta, \ x \in [0, x_{c_p}] \tag{3.19}$$

$$\Lambda(x) - 1 + \delta = (x - 0)s_{\mu_1}(x) + (x_{c_p} - x)s_{\eta_1}(x) \tag{3.20}$$

$$1 + \delta - \Lambda(x) = (x - 0)s_{\mu_2}(x) + (x_{c_p} - x)s_{\eta_2}(x) \tag{3.21}$$

$$0.5 \leq \Lambda(x) \leq 1 - \delta, \ x \in (x_{c_p}, 0.5] \tag{3.22}$$

$$\Lambda(x) - 0.5 = (x - x_{c_p})s_{\mu_3}(x) + (0.5 - x)s_{\eta_3}(x) \tag{3.23}$$

$$1 - \delta - \Lambda(x) = (x - x_{c_p})s_{\mu_4}(x) + (0.5 - x)s_{\eta_4}(x) \tag{3.24}$$

where δ denotes ripples, and polynomials $s_{\mu_r}(x)$, $s_{\eta_r}(x)$, for $r = 1, \ldots, 4$, have degrees equal to or less than k. Note that similar derivations, required for the stopband region with $x_{c_s} = 1 - x_{c_p}$, are not necessary due to antisymmetry of $\Lambda(x)$. Likewise, and noting to the structure of $\Omega(x)$ as in (3.6) and shown in Fig. 3.1, we can write

$$1 - \delta \leq \Omega(x) \leq 1 + \delta, \ x \in [0, \ x_{c_p}] \tag{3.25}$$

$$\Omega(x) - 1 + \delta = (x - 0)s_{\mu_1}(x) + (x_{c_p} - x)s_{\eta_1}(x) \tag{3.26}$$

$$1 + \delta - \Omega(x) = (x - 0)s_{\mu_2}(x) + (x_{c_p} - x)s_{\eta_2}(x) \tag{3.27}$$

$$1 - \delta \leq \Omega(x) \leq \delta_{mx}, \ x \in (x_{c_p}, \ 0.5] \tag{3.28}$$

$$\Omega(x) + 1 - \delta = (x - x_{c_p})s_{\mu_3}(x) + (0.5 - x)s_{\eta_3}(x) \tag{3.29}$$

$$\delta_{mx} - \Omega(x) = (x - x_{c_p})s_{\mu_4}(x) + (0.5 - x)s_{\eta_4}(x) \tag{3.30}$$

$$\delta \leq \Omega(x) \leq 1 - \delta, \ x \in (0.5, \ x_{c_s}] \tag{3.31}$$

$$\Omega(x) - \delta = (x - 0.5)s_{\mu_5}(x) + (x_{c_s} - x)s_{\eta_5}(x) \tag{3.32}$$

$$1 - \delta - \Omega(x) = (x - 0.5)s_{\mu_6}(x) + (x_{c_s} - x)s_{\eta_6}(x) \tag{3.33}$$

$$-\delta \leq \Omega(x) \leq \delta, \ x \in (x_{c_s}, \ 1] \tag{3.34}$$

$$\Omega(x) + \delta = (x - x_{c_s})s_{\mu_7}(x) + (1 - x)s_{\eta_7}(x) \tag{3.35}$$

$$\delta - \Omega(x) = (x - x_{c_s})s_{\mu_8}(x) + (1 - x)s_{\eta_8}(x) \tag{3.36}$$

where polynomials $s_{\mu_r}(x)$, $s_{\eta_r}(x)$, for $r = 1, \ldots, 8$, have degrees equal to or less than k, and δ_{mx} in (3.28) refers to the amount of overshoot of the corresponding maximally flat filter which can be determined via the maximally flat function. In fact, we want to control and maintain the overshoot to be less than the corresponding maximally flat in each design. In general, given the ideal H_0 and using the perfect reconstruction equation $H_0(z)F_0(z) + H_0(-z)F_0(-z) = 1$, one can obtain the ideal F_0 defined by 1

for $x \in [0, 0.5)$ and 0 for $x \in (0.5, 1]$. The problem is that $F_0(x) = 1$ at $x = 0.5$ but $H_0(0.5) = 0.5$. The overshoot is a consequence of the perfect reconstruction condition; when H_0 is small in the passband region, F_0 needs to compensate to achieve perfect reconstruction, and therefore, by increasing the passband/stopband width of H_0 the amount of overshoot must be decreased. In a recent paper [47] it has been shown that triplet halfband filter banks can improve the image compression due to better frequency selectivity. Though the overshoot is one of the limitations of HPFB class, this does not seem to pose problems in certain applications, e.g., the 5/3 JPEG2000 filter pair belongs to this class. The class of triplet halfband filter banks does not have this limitation, however, the design procedure is computationally expensive. It is expected that a similar approach as presented in this chapter may lead to a simpler design for the class of triplet halfband filters.

It should be noted that the design procedure contains two cycles. First the coefficients of $T_1(x)$ ($\Lambda(x)$), $\boldsymbol{b}_1$ written based on $\boldsymbol{a}_1$, are obtained. Then a new set of coefficients $\boldsymbol{b}_2$ (in terms of $\boldsymbol{a}_2$), which are the coefficients of $T_2(x)$ are determined via optimization for bounding $\Omega(x)$. Note that $\Lambda(x)$ depends on the coefficients of $T_1(x)$ only, however, $\Omega(x)$ requires $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. Substituting $T_1(x)$ and $T_2(x)$ in (3.5) and (3.6) respectively, $\Lambda(x)$ and $\Omega(x)$ in (3.19)–(3.33) are derived based on polynomials coefficients. Now let us define the following problems.

Problem 1. Let $\Lambda(x)$ be in the form of (3.5) with pre-specified N, m, and δ. We want to determine $\boldsymbol{b}_1$ such that the passband and stopband widths, defined using x_{c_p} and x_{c_s}, be maximized subject to inequalities in (3.20) and (3.22).

Problem 2. Let $\Omega(x)$ be in the form of (3.6) with pre-specified N, m, δ, δ_{mx}, and the coefficients $\boldsymbol{b}_1$ obtained in Problem 1. We want to determine $\boldsymbol{b}_2$ such that the passband and stopband widths of $\Omega(x)$ be maximized subject to inequalities in (3.25), (3.28), (3.31), and (3.34).

Solution to these nonconvex problems cannot be determined if the passband or stopband edges are considered as unknown variables. A traditional approach is to

Table 3.1 – Performance analysis of the filters designed in Case 1 for $N = 7$.

m	$\bar{\omega}_{c_p}^{H_0}$	$\bar{\omega}_{c_s}^{H_0}$	$\bar{\omega}_{c_p}^{H_1}$	$\bar{\omega}_{c_s}^{H_1}$	$\delta_p^{H_0}$	$\delta_s^{H_0}$	$\delta_p^{H_1}$	$\delta_s^{H_1}$	Overshoot	s.a.^{H_0}	s.a.^{H_1}
4	0.2728	0.7272	0.7249	0.2318	0	0	0	0	1.1250	-	-
2	0.3801	0.6199	0.6173	0.3375	0.02	0.02	0.02	0.02	1.1024	-34	-34

solve the problem by changing the role of variables [12] first; ripples and edges, and then iteratively find the optimal solution. However, in order to efficiently search for the solution and to reduce the computational complexity, we have used the Genetic Algorithm. Furthermore, we have supplemented the procedure with Bisection [12] due to the existence of two edges to be optimized. The Bisection step is quite fast as the bisecting interval is not a wide range. It is usually less than half of the passband/stopband width. It is worth pointing out that although we do not solve the problem directly for ripples, the exact desired ripple size is obtained at the end of the cycle. We use a MATLAB toolbox for optimization over symmetric cones [90] and the Genetic Algorithm Toolbox of MATLAB as the solvers. In the following section, the proposed method is used to design sets of new wavelets with various frequency response characteristics, which will eventually be employed along with the proposed face recognition system in Chapter 2 to analyze the performance of a face recognition system under different frequency response characteristics.

3.4.3 Design Examples and Discussions

Various combinations may be considered, depending on filter length, number of vanishing moments, passband and stopband widths, and ripples. For a given filter length and specified number of vanishing moments and ripples, we can obtain a set of THP wavelet filter pairs ranging from the maximally flat to the sharpest transition band.

Case 1: For $N = 7$, $m = 2$ and $\delta = 0.02$, we consider the possible maximal passband/stopband width to determine coefficients of the polynomials T_1 and T_2 required to implement the corresponding wavelet filter kernels $\Lambda(x)$ and $\Omega(x)$. Performance

Table 3.2 – Coefficients to implement $H_0(z)$ and $F_0(z)$ in Case 1 for $N = 7$ and $m = 2$. Note that $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ result in $T_1(x)$ and $T_2(x)$, respectively, where $\Lambda(x) = T_1(x)$, however, $\Omega(x) = T_1(x) + 2T_2(x) - 2T_1(x)T_2(x)$.

$\boldsymbol{a}$	$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	$\boldsymbol{b}$	$\boldsymbol{b}_1$	$\boldsymbol{b}_2$
a_0	-0.0109558471679687	-0.0044677734375000	b_5	179.5006	73.2000
a_2	0.0414515014648589	0.0246503906250143	b_4	-269.2509	-109.8000
a_4	-0.0898694213864434	-0.0783945312497395	b_3	103.3979	34.2330
a_6	0.3093737670909270	0.3082119140635260	b_2	-6.8238	1.1835
a_7	0.5	0.5	b_1	2.0000	2.0000
			b_0	1.0000	1.0000

of the designed filters is presented in Table 3.1 and compared to maximally flat filters with the same length. In this table, $\bar{\omega}$ refers to the normalized frequency and s.a. denotes the stopband attenuation in decibel (dB). The optimization variables $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are obtained and given in Table 3.2. Recall that since the biorthogonal filter banks designed in this chapter belong to the class of HPFB [73], the analysis lowpass and synthesis highpass filters H_0 and F_1 must be halfband, that is, H_0 (and F_1) is an antisymmetric function with respect to the point $(x = 0.5, H_0(x) = 0.5)$. In other words, passband and stopband widths are exactly identical, and equiripple behavior is preserved. However, F_0 (and H_1) is not necessarily an antisymmetric function and the stopband and passband characteristics are quite different, knowing that $F_0(x = 0.5) = 1$ but $H_0(x = 0.5) = 0.5$ [95] (see Fig. 3.1). One should note that the two functions $T_1(x)$ and $T_2(x)$, used to define kernels in (3.5) and (3.6), are antisymmetric functions. As $\Lambda(x) = T_1(x)$, this property is observed for $H_0(x)$, however, $\Omega(x) \neq T_2(x)$ (see (3.6) and (3.8)) and therefore $\Omega(x)$ and consequently $F_0(x)$ and $H_1(x)$ cannot be antisymmetric functions. The proposed formulation in polynomial domain can control the design specifications and provide a trade-off between regularity and frequency selectivity. The increased sharpness of the transition band over the maximally flat, and consequently the better frequency selectivity of the filters has been given in Table 3.1.

Case 2: For the case $N = 13$ and $m = 1$, we set $\delta = 0.02$ both for H_0 and F_0.

Table 3.3 – Performance analysis of the filters designed in Case 2 for $N = 13$ with $m = 7, 5, 3, 1$.

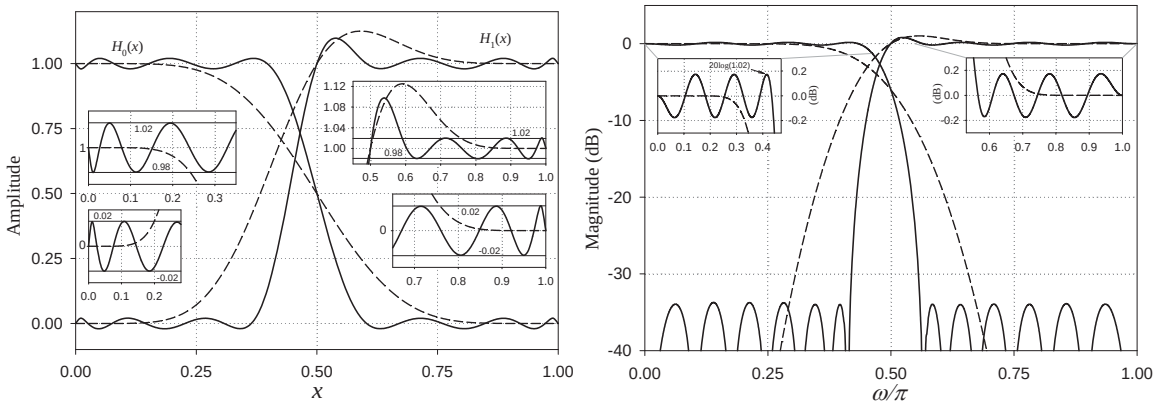
m	$\bar{\omega}_{c_p}^{H_0}$	$\bar{\omega}_{c_s}^{H_0}$	$\bar{\omega}_{c_p}^{H_1}$	$\bar{\omega}_{c_s}^{H_1}$	$\delta_p^{H_0}$	$\delta_s^{H_0}$	$\delta_p^{H_1}$	$\delta_s^{H_1}$	Overshoot	s.a.^{H_0} (dB)	s.a.^{H_1} (dB)
7	0.3268	0.6732	0.6713	0.2934	0	0	0	0	1.1250	-	-
5	0.4161	0.5839	0.5817	0.3853	0.02	0.02	0.02	0.02	1.1018	-34	-34
3	0.4350	0.5650	0.5630	0.4109	0.02	0.02	0.02	0.02	1.0996	-34	-34
1	0.4408	0.5592	0.5573	0.4174	0.02	0.02	0.02	0.02	1.0976	-34	-34

It should be noted that although equiripple must be preserved for halfband filter H_0, this is not a necessity for F_0, and therefore, one may choose a different ripple size for F_0. The filter characteristics for this case are given in Table 3.3, where $\bar{\omega}$ refers to the normalized frequency and s.a. denotes the stopband attenuation in dB. Coefficients of the polynomials, $T_1(x)$ and $T_2(x)$, required to implement the corresponding wavelet filter bank are given in Table 3.4. We have also shown the results for the maximally flat with $m = 7$, and also for the cases with $m = 3$ and $m = 5$ in the same table. Similar to Case 1, Fig. 3.2 (a) and (b) show the filters responses in the variable x and z, respectively, where the sharpness of transition band is clearly observed in Fig. 3.2 in contrast to poor frequency selectivity and wide transition band of the maximally flat filter pairs.

In terms of computational code and programming, a technical question is how can polynomials s_μ and s_η be determined? Although working with univariate polynomials is the simplest case of the problem of positivity of polynomials, in general, it is difficult to find an analytic solution to this question. The authors are aware of some particular techniques such as Sturm sequences and similar ideas based on root counting algorithms which contain some restrictions. Besides its obvious theoretical interest, this is also a relevant question in applied mathematics and engineering. While some algebraic techniques have been proposed to address the problem analytically [75], recent research and results in literature suggest that this problem can be solved more efficiently using numerical techniques such as SDP. Using convex optimization and SDP, the problem is then converted to a standard question of computing eigenval-

Table 3.4 – Coefficients to implement $H_0(z)$ and $F_0(z)$ in Case 2 for $N = 13$ and $m = 1$. Note that $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ result in $T_1(x)$ and $T_2(x)$, respectively, where $\Lambda(x) = T_1(x)$, however, $\Omega(x) = T_1(x) + 2T_2(x) - 2T_1(x)T_2(x)$.

$\boldsymbol{a}$	$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	$\boldsymbol{b}$	$\boldsymbol{b}_1$	$\boldsymbol{b}_2$
a_0	0.0148422480940819	0.00252524529099464	b_{12}	-996046.4088	-169466.3428
a_2	-0.0171503426671027	-0.00825489069819463	b_{11}	5478255.2484	932064.8854
a_4	0.0220590517280641	0.01623159040212130	b_{10}	-13063428.2652	-2200213.8746
a_6	-0.0373170116231449	-0.02908116489754780	b_9	17698512.8304	2910475.7952
a_8	0.0551050808024911	0.05110726157896290	b_8	-15003655.5256	-2364308.4338
a_{10}	-0.1030607602008620	-0.09815436501173560	b_7	8267521.2032	1217313.3385
a_{12}	0.3155217328204610	0.31562632333562400	b_6	-2982873.0376	-397017.7799
a_{13}	0.5	0.5	b_5	693633.2816	80046.8653
			b_4	-99811.4848	-9469.1327
			b_3	8218.8662	593.6343
			b_2	-333.9638	-17.1759
			b_1	3.6280	-0.8895
			b_0	-1.0000	-1.0000



(a) (b)

Figure 3.2 – Filter functions and responses for the case $N = 13$, $m = 1$, and $\delta = 0.02$ using the proposed method (solid-curve) compared with the corresponding maximally flat (dashed-curve) for the same example. Both the lowpass and highpass filters, H_0 and H_1, are shown (H_1 is the highpass version of F_0). (a) in the variable x; (b) in the variable z. Insets show magnification of some portions.

ues. The whole idea with SDP representation of sum of squares decomposition is to generate the positive semidefinite quadratic forms [58][70].

For filters with larger degree (e.g. length of 27 and 53 for $H_0(z)$ and $H_1(z)$ respectively), the SDP problem, including both optimization cycles, is solved in less than a second for each round and less than a minute for the entire problem using a Core 2 Duo PC at 2.80 GHz.

3.5 Frequency Selectivity and the Performance of a Face Recognition System

In order to analyze the effect of frequency selectivity of a multiresolution transformation on the performance of a face recognition system we first need to design the corresponding filters such that the frequency response characteristics could be controlled accordingly. In view of the tuning opportunity of the wavelets proposed and designed in this chapter, and based on the proposed multiresolution based face recognition method in Chapter 2 from the other hand, we are now able to analyze and find the relation between the frequency selectivity and the recognition rate. The proposed face recognition system in Chapter 2 is modified in terms of replacing the DD-DT$\mathbb{C}$WT with the frequency selective filters designed in this chapter, and also, removing the maximum filter block. The maximum filter is removed to avoid any outer affect on the pure performance of the proposed filter pairs. It is worth pointing out that, feature vector construction has been removed as well. The generated mask is directly sent to the PCA for dimensionality reduction. The block diagram of the new proposition has been shown in Fig 3.3 and the procedure is discussed in details as follows.

3.5.1 Illumination Invariant Face Recognition via THPs

Although for images with low variation in illumination the performance of a face recognition system is mostly related to the quality of the features and the feature

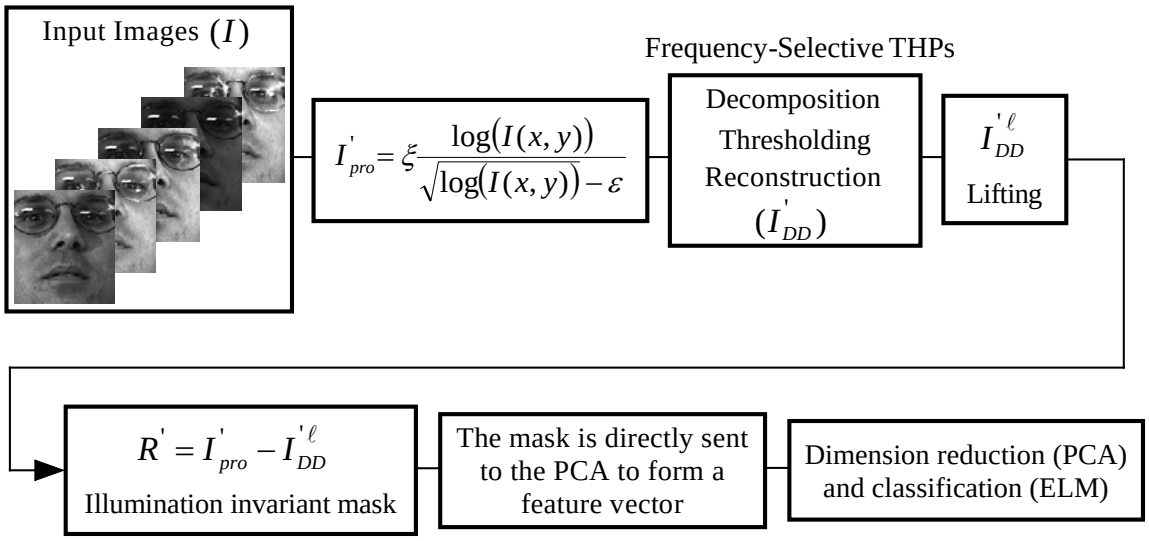


Figure 3.3 – Block diagram of the proposed method in Chapter 2 with modification in terms of the thresholding via THPs, removing the maximum filter, and direct feature extraction.

extraction strategy, that is not the case for the images which are too dark or too bright. One should note that fine features such as eyes, eyebrows, nose and mouth are distinguished by edges which are in fact mostly detectable in high frequency subbands. Increasing or decreasing the sharpness of the filters associated with a multiresolution transformation eventually enables one to keep or discard specific types of information. For the problem of illumination invariant human face recognition, as discussed extensively in Chapter 2, illumination is mostly assumed to lie in the low-frequency part of an image. Therefore, an intelligent approach is to keep healthy high-frequency information as much as possible away from the image. Then the rest could be interpreted as the illumination that can be subtracted from the image to reduce the effect of illumination on recognition rates. Taking the advantage of multiresolution analysis as shown in Chapter 2, and the tunable biorthogonal filter pairs with various frequency selective characteristics designed and proposed in this chapter, one can evaluate and investigate the effect of the frequency selectivity of the transformation on performance of a recognition system. In Chapter 2, we proposed

Table 3.5 – Characteristics of the filters designed with different sharpness and ripples for $N = 7$.

Case	m	$\bar{\omega}_{c_p}^{H_0}$	$\bar{\omega}_{c_s}^{H_0}$	$\bar{\omega}_{c_p}^{H_1}$	$\bar{\omega}_{c_s}^{H_1}$	Ripple	Overshoot	s.a.^{H_0}	s.a.^{H_1}
filt1	4	0.2728	0.7272	0.7249	0.2318	0.00	1.1250	-	-
filt2	2	0.3801	0.6199	0.6173	0.3375	0.02	1.1024	-34	-34
filt3	2	0.4069	0.5931	0.5877	0.3639	0.04	1.0956	-34	-34
filt4	2	0.4228	0.5772	0.5671	0.3784	0.06	1.0898	-34	-34
filt5	2	0.4340	0.5660	0.5468	0.3892	0.08	1.0855	-34	-34

to form the feature vector using the DD-DTℂW subbands, and then obtained vector was dimensionally reduced via PCA. In this chapter we would like to specifically investigate the effect of frequency selectivity of a multiresolution transformation on the performance of a face recognition system. To reach this aim, and using the proposed method in this chapter, we first design and extract the filter pairs with the desired characteristics. For $N = 7$, the ripple size is increased from 0 to 0.08 with the step size of 0.02. Consequently, the transition band is narrowed down resulting in a sharper filter. The frequency response information of the designed filters is given in Table 3.5. As it has been shown in Fig. 3.3, a multiresolution based face recognition system can be developed using the filter pairs introduced in this chapter. In fact, the filter pairs of a traditional DWT are replaced by filt1 to filt5. The face images in databases are now decomposed, thresholded and reconstructed with these customized filters. At this point, we can answer to the main question regarding the effect of the frequency response characteristics to the performance of a multiresolution based face recognition system. In other words, how do ripples, transition band, number of vanishing moments may affect the recognition rates? Ideally, the narrower the transition band, the better the frequency selectivity, however, there is a trade-off between the transition band width and the amplitude of ripples. In fact, the filter length, number of moments, ripples and sharpness of the transition band width are tightly related to each other and interpret the same concept.

3.5.2 Experiments and Results

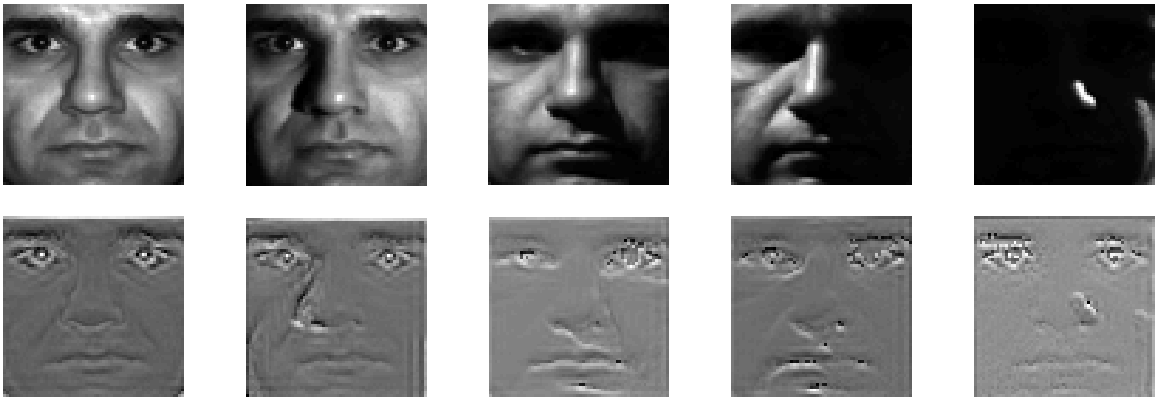
To verify the discussions above regarding how the frequency selectivity of the subbands of a multiresolution transformation may affect the recognition accuracy, several comprehensive experiments have been performed using the well known databases including the Yale B and the CMU-PIE. These databases have been already introduced in Section 2.5 of Chapter 2.

For the Yale B database, each individual has 64 images. All images are first eye-aligned and cropped into 64×64 gray level pixels The seven images of Subset 1 for each individual is used for training. It should be noted that Subset 1 contains the images with low illumination variations. The remaining subsets, Subsets 2–5, are kept for testing phase of the procedure. As mentioned before, and to analyze the frequency selectivity versus recognition accuracy, the filters associated with a typical DWT are replaced with new sets of filters as listed in Table 3.5. Fig. 3.4 shows sample images of an individual from Subset 1 to Subset 5 of the Yale B database. For this experiment and employing filt1 the obtained masks are presented in the second row of the figure, respectively. We then replace the filters of the DWT with filt2 to filt5 (See details in Table 3.5). The results are given in Table 3.6 for the Yale B database. As it was expected, the sharper the filters, the higher the recognition accuracy. However, as the ripple size is increased, the recognition rate is reduced. It can be seen in Fig. 3.5 (d) and (e) for instance. Also, Table 3.6 indicates that for Subsets 2 and 3, which contain images with low variation in illumination, all filters perform satisfactory as the perfect accuracy is achieved. It is then concluded that the role of frequency selectivity is more evident for the subsets consist of darker and highly illuminated images difficult to recognize.

Recognition rate versus ξ and ϵ has been shown in Fig. 3.5 and Fig. 3.6 for Subsets 4 and 5 of the Yale B database, respectively. Note that only the most difficult experiments have been visualized, that is, the representations for Subsets 2 and 3 are excluded. The effect of filters and frequency selectivity can be elaborated via

Table 3.6 – Recognition rate for the Yale B and the CMU-PIE databases employing the designed filters (%).

	Yale B				CMU-PIE	
Case	Subset 2	Subset 3	Subset 4	Subset 5	Lights-off	Lights-on
filt1	100	100	98.67	98.68	98.43	100
filt2	100	100	98.77	98.70	98.47	100
filt3	100	100	98.88	98.71	98.64	100
filt4	100	100	98.97	98.87	98.58	100
filt5	100	100	98.75	98.77	98.32	100

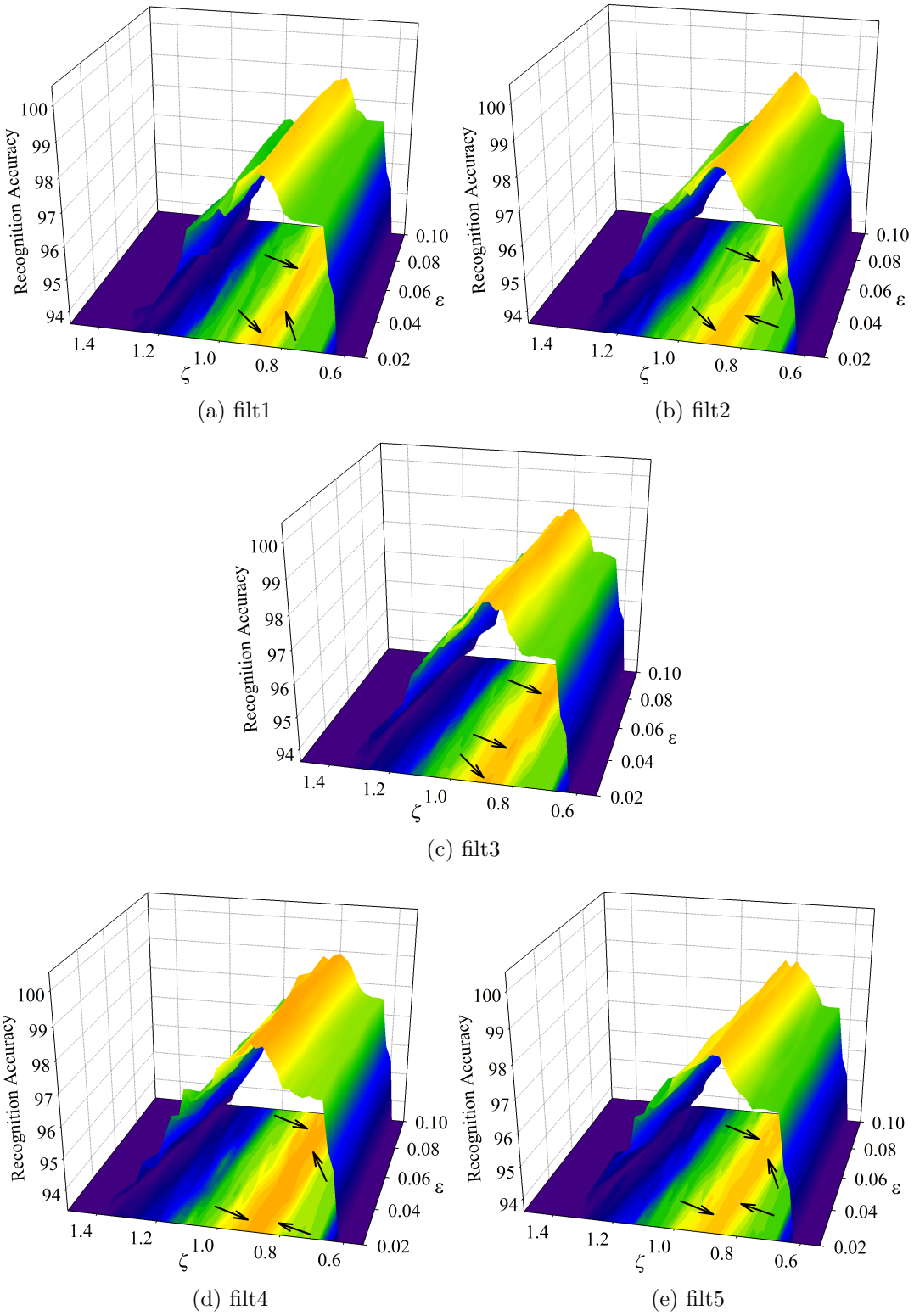


(a) (b) (c) (d) (e)

Figure 3.4 – The first row (a)–(e) show sample images of an individual from Subsets 1–5 of the Yale B database. The second row shows the obtained mask by using filt1 for each image in the first row, respectively.

Fig. 3.5 and Fig. 3.6 (a)–(e). In these figures, it can be seen that sharper filters with shorter transition bands, such as filt4, perform reasonably better for the illumination invariant recognition task. In contrast, extra reduction in the transition band width which is along with the mandatory increment in the ripple size, e.g. filt5, can reduce the recognition accuracy. This phenomenon and the range of remarkable performance of the filters have been shown and highlighted by arrows in Fig. 3.5 and Fig. 3.6 (a)–(e) for Subsets 4 and 5 of the Yale B database, respectively.

For the CMU-PIE database there are two lighting conditions known as the lights-off and lights-on settings. All images are cropped, eye-aligned and resized into 64×64

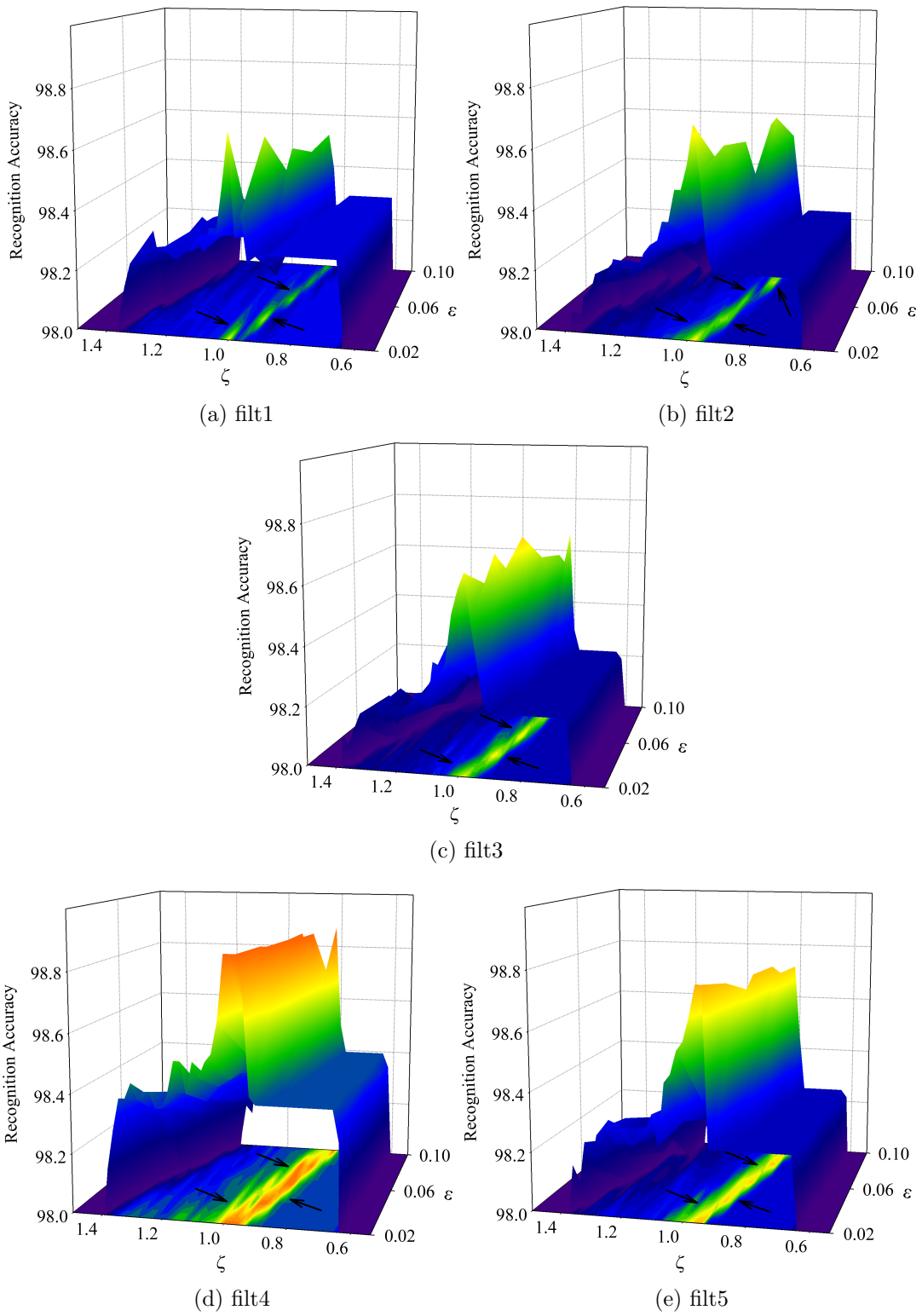


(a) filt1 (b) filt2

(c) filt3

(d) filt4 (e) filt5

Figure 3.5 – Recognition rate versus ξ and ϵ for Subset 4 of the Yale B database employing the designed filters.



(a) filt1

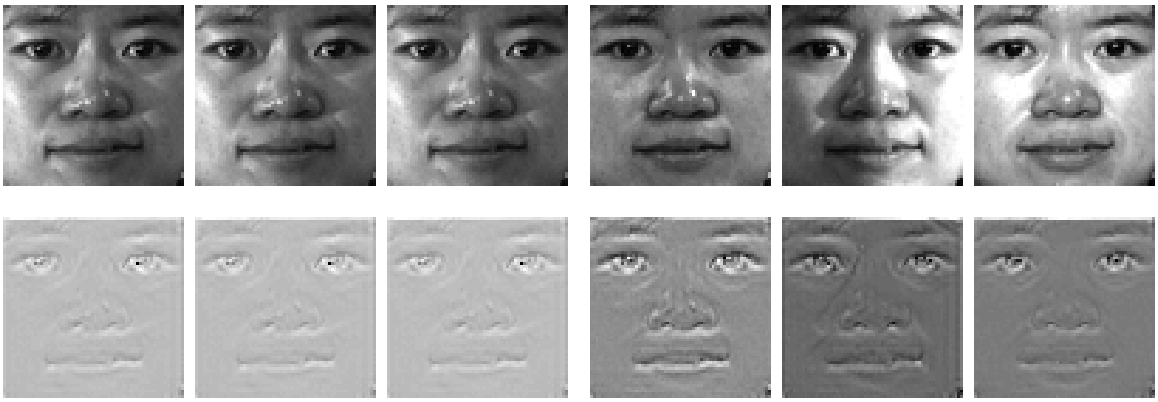
(b) filt2

(c) filt3

(d) filt4

(e) filt5

Figure 3.6 – Recognition rate versus ξ and ϵ for Subset 5 of the Yale B database employing the designed filters.



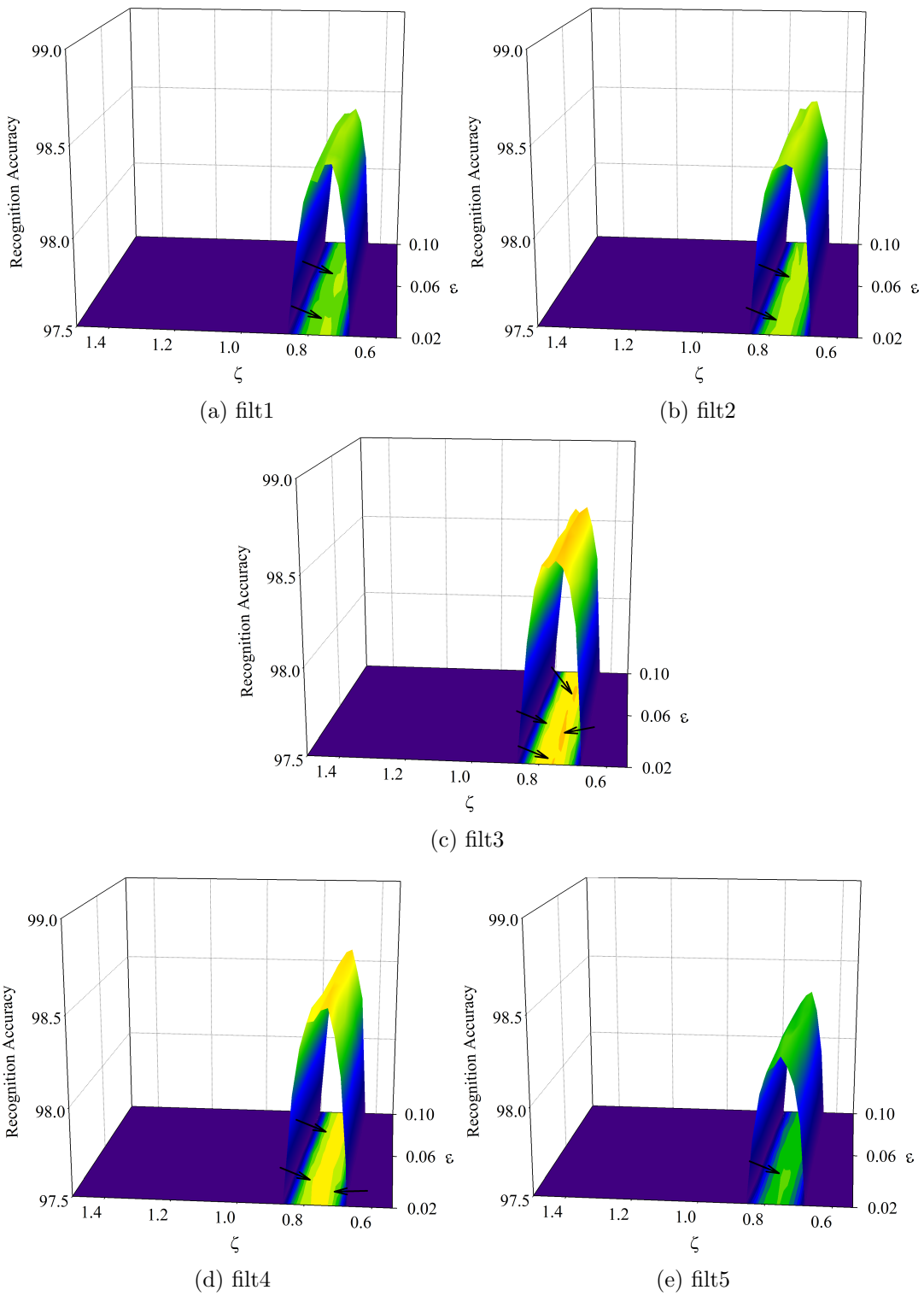
(a) (b) (c) (d) (e) (f)

Figure 3.7 – The first row (a)–(c) and (d)–(f) show sample images of two individuals from the lights-off and lights-on settings of the CMU-PIE database, respectively. The second row shows the obtained mask by using filt1 for each image in the first row, respectively.

pixels. For the lights-off setting with 21 images, three images from each of the 68 individuals are chosen to make a training gallery. The remaining 18 images are reserved for the testing phase of the experiments. Similar to the experiments with the Yale B database, the filters of the DWT decomposition bank are replaced with the ones given in Table 3.5 from filt1 to filt5. Some sample images from the lights-off and lights-on settings have been shown in the first row of Fig. 3.7. The corresponding masks obtained using the proposed filters are presented in the second row of the figure, respectively. The recognition rates are given in Table 3.7 and the visual representation of the findings are illustrated as the recognition accuracy versus ξ and ϵ in Fig. 3.8 and Fig. 3.9 for the lights-off and lights-on settings, respectively. Similar discussion, as elaborated above for the Yale B database, is valid for the CUM-PIE images. The lights-off setting contains images with notable variation in illumination. Therefore, compared to the Yale B database and in contrast to the lights-on setting, it resembles the level of difficulty between Subsets 4,5 and Subsets 2,3 of the Yale B database. While filt3 and filt4, that possess wider tuning for higher recognition rates due to the sharpness of filters, clearly perform better than the other filters for the lights-off

setting it cannot be seen for the lights-on setting. In fact, and similar to the discussion regarding the Subsets 2 and 3 of the Yale B database, the images in subset lights-on contain a low or moderate illumination variation and it is easy to reach the perfect accuracy employing any of the filters given in Table 3.5. The discussion has been visualized by arrows in Fig. 3.8 and Fig. 3.9 for the lights-on and lights-off settings, respectively.

In reality, there is a trade-off between the sharpness of the frequency response, amplitude of ripples, filter length, smoothness of the frequency response at vanishing points, difficulty of design, and the recognition rate. It is too complicated if one may desire to find the finest and the optimum value for each and every parameter to reach the best performance. However, generally speaking, the shorter the filter length, the lower the complexity and realization. In terms of number of vanishing moments, although higher NVM results in a smoother wavelet, it significantly reduces the frequency selectivity by reducing the sharpness of the transition band. A systematic comparison shows that for a fixed filter length, frequency selectivity plays the main role in multiresolution based imaging. Reducing the transition band by increasing the amplitude of ripples leads to better results for recognition, if the ripples are not too large. Our experiments show that $N = 5, 7, \ldots, 15, 17$ is the suitable range to design the filters for such applications. In general, very short filters are simple to implement but suffer from a wide transition band. Also, very long filters may not perform well in real applications. In addition, longer filters are computationally expensive and not practical to implement as common time invariant FIR filters. In terms of oscillations, ripples larger than $\delta = 0.2$ may not lead to remarkable results, although the transition band width can be significantly decreased.

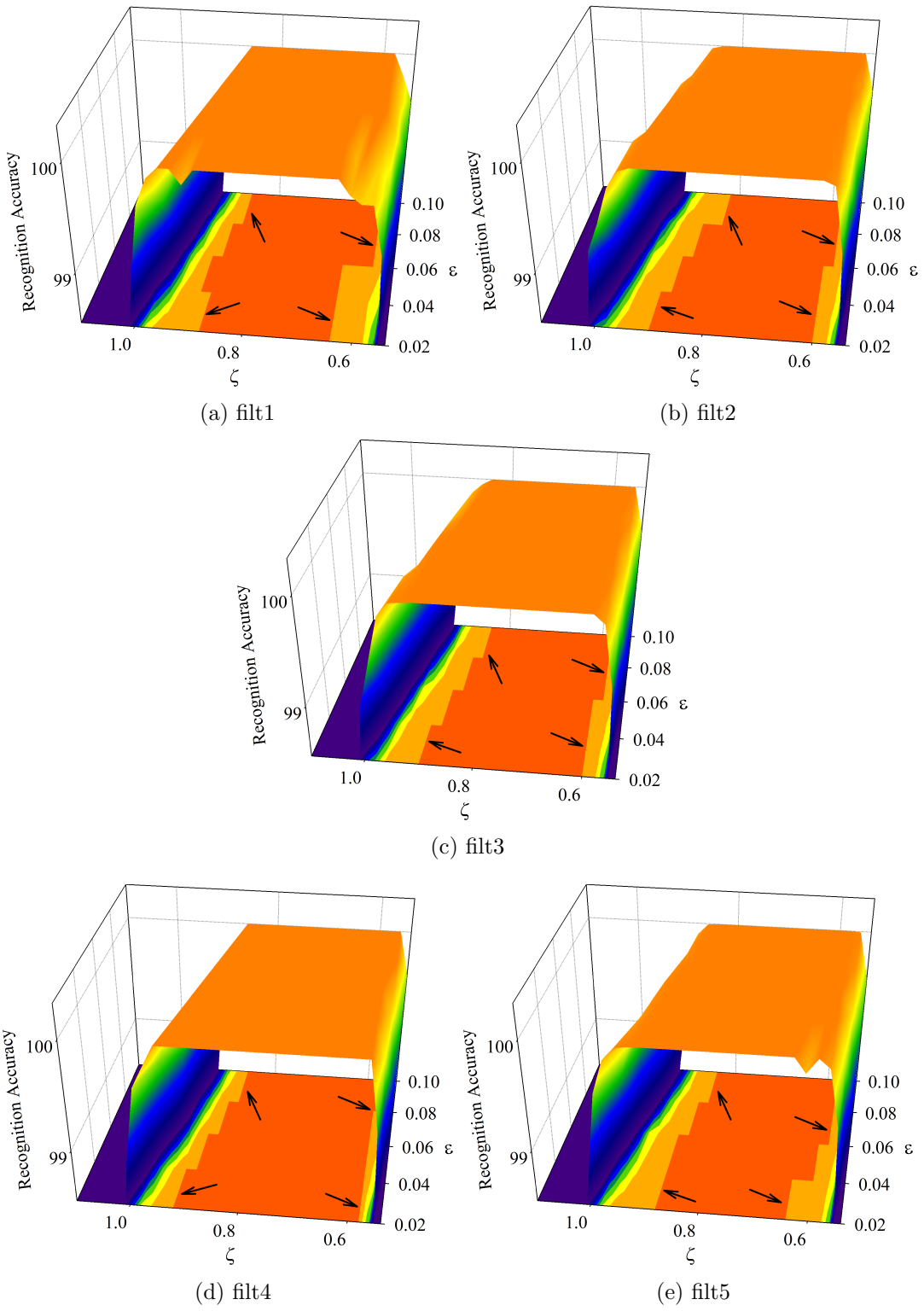


(a) filt1 (b) filt2

(c) filt3

(d) filt4 (e) filt5

Figure 3.8 – Recognition rate versus ξ and ϵ for the CMU-PIE database under the lights-off setting employing the designed filters.



(a) filt1 (b) filt2

(c) filt3

(d) filt4 (e) filt5

Figure 3.9 – Recognition rate versus ξ and ϵ for the CMU-PIE database under the lights-on setting employing the designed filters.

3.6 Frequency Selectivity and Multifocus Fusion

While performing the experiments and investigation for the relation between the frequency selectivity and performance of a face recognition system and in view of our experience in multifocus imaging and shape from focus we are motivated to employ the wavelet filters designed in this chapter to further analyze the effect of frequency selectivity of the filters of a multiresolution filter bank in multifocus image fusion (MFIF). The term focus is used to measure the degree of blurring in an image which is increased with the distance of imaging system from the focus plane. To retrieve depth information, and subsequently to reconstruct a geometric object from one or multiple observations, is a challenging problem in computer vision and sensing systems. The shape-from-focus (SFF) and depth map estimation is a technique to determine the depth of every image point from the camera lens and then to reconstruct the 3D shape of the object. Multifocus image fusion is therefore a process of obtaining a fully-focused image from a set of registered input images. The main goal of multifocus image fusion is to transfer the most relevant information found in source images into a fused image. In the following section, we show how the performance of an MFIF may be affected under frequency response characteristic variations.

3.6.1 Multiresolution Based MFIF Employing THPs

In general, the ability of any focus measure technique to calculate the sharpness value of each image pixel shows the success of a method. To calculate the sharpness value of pixels and subsequently to find the depth of each pixel, we proposed the use of the wavelet transform along with the so-called distance transformation employing an exponentially decaying function [62]. As it has been shown in [62], a multiresolution based MFIF can be developed as is described as follows and shown in Fig. 3.10. The entire multifocus image sequence is decomposed via wavelet transform into scaling and wavelet subbands using the filters designed in Chapter 3.

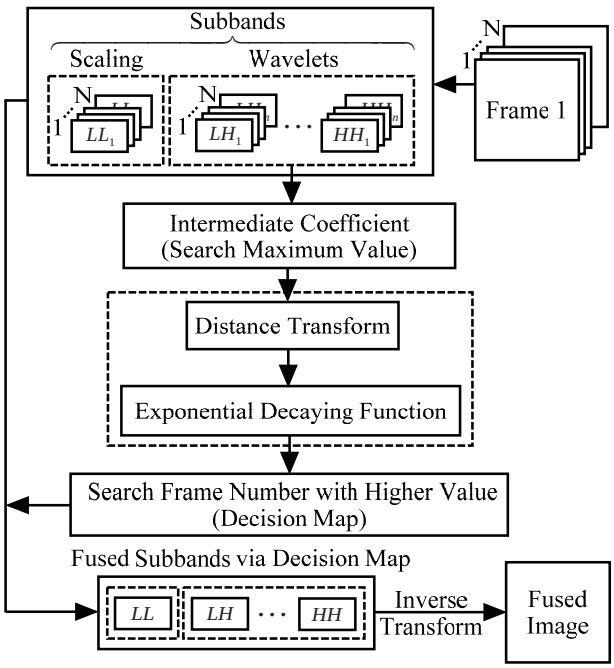


Figure 3.10 – Depth map estimation and image fusion procedure using the proposed wavelet based algorithm employing THPs.

It is technically an important task to know which pixel comes from which image. Also, in order to avoid pixel index complexity, we do not perform downsampling between stages while decomposing data by wavelet transform, and thus, each subband has the same size as of the input image. The larger values in wavelet subbands correspond to sharper brightness changes, and therefore, a good integration rule is to choose the maximum scheme known as the maximum selection rule [57] to pick the coefficient with the larger activity level [68] and discard the others. The intermediate subband with higher pixel value is computed applying maximum selection rule on subbands of each image in the wavelet domain. Finally, a decision map is determined employing the distance transform with the exponentially decaying function. Being more specific, a focus measure which is characterized by the exponentially decaying function is employed to compute the sharpness of each pixel in an image. Such decaying function uses neighborhood information of extracted feature points assuming that the pixel value far from a feature point is equal to 1 and it approaches this limiting

value exponentially. Given an arbitrary point ρ and the set of feature points, the focus measure (FM) is estimated as

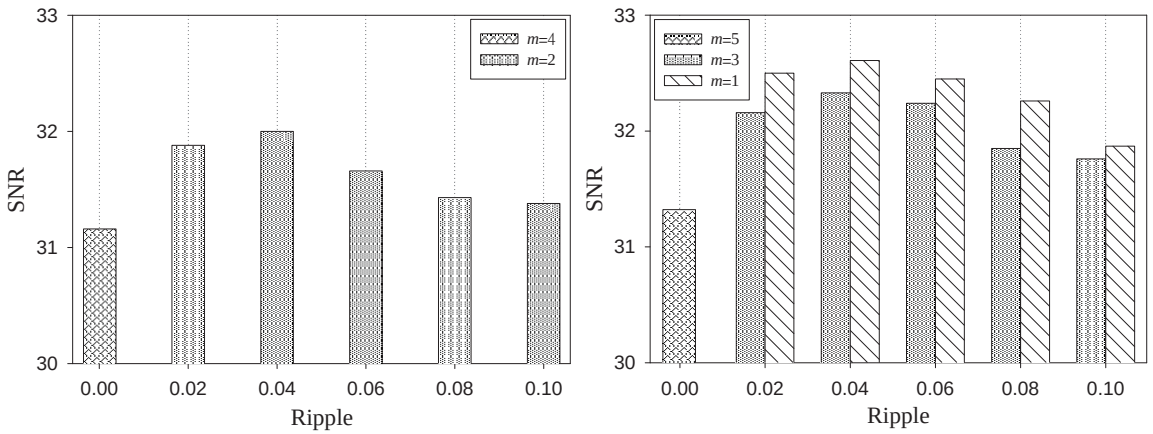
$$FM(x,y) = e^{\frac{-D(x,y)}{2}} I(x,y) + (1 - e^{\frac{-D(x,y)}{2}}) \tag{3.37}$$

where D is the distance between point ρ and the nearest feature pixel of the intermediate subband. The decision map for fusion is constructed by comparing the value of subbands; the frame number with higher value is mapped onto the decision map. Using the decision map, pixels are extracted from corresponding subbands of the image sequence and an inverse wavelet transform yields the final fused image.

Fig. 3.11 (a) and (b) show this trade-off in terms of the signal-to-noise ratio (SNR) for the case with $N = 7$ and $N = 9$, respectively, for the Simulated Cone dataset. For $N = 7$, the number of vanishing moments must be 2, where the maximally flat (with $m = 4$ and $\delta = 0$) is also shown for reference. For the case $N = 9$, the possible moments are 1 and 3, with $m = 5$ and $\delta = 0$ for the corresponding maximally flat filter. Similar SNR plots, as is shown in Fig. 3.11, are obtained for the values of N up to 17 with a slight modification in SNR values. Increasing the length of filters, although offering sharper filters in terms of transition band width, the SNR begins to be reduced considerably. Among the possible designs, we have selected the case with $N = 7$, $m = 2$, and $\delta = 0.02$ for instance.

3.6.2 Experiments and Discussions

The proposed filter pair with sharper transition band and better frequency selectivity is compared to other traditional focus measure systems in the literature, namely Tenengrade (Tenen), gray level variance (GLV) [20] and sum of modified Laplacian (SML) [63], and to the Daubechies orthogonal filter bank (db4), Cohen-Daubechies-Feauveau (CDF) 9/7 biorthogonal filter bank used in JPEG2000, Mexican hat (Mexh), Morlet, and Meyer. Several available multifocus image sequence datasets, e.g., Simulated Cone {72 frames, 318×318}, Chess {29, 800×600}, Lab {6, 512×512}, Clock



(a) $N = 7$, $m = 4, 2$ (b) $N = 9$, $m = 5, 3, 1$

Figure 3.11 – Trade-off between the SNR and ripples.

{2, 512×512}, FlyEye {37, 325×217}, and Rifle {24, 200×150}, have been used in experiments. Fig. 3.12, (a1–a2) and (b1–b2) show some randomly selected sample frames at different focus from the Simulated Cone and Rifle datasets, respectively. It is followed by the corresponding depth map, depth map under Gaussian noise with zero mean and variance of 0.001, fused image, 3D shape reconstruction employing the depth map under noisy condition, in Figs. 3.13 and 3.14, respectively. For the depth maps, shown in Figs. 3.13 and 3.14 (a1)–(a9) and (b1)–(b9), the contour plane axes denote the image size and the vertical axis is obviously the number of frames in each dataset, indicating which pixel comes from which image. Also, the fused images are given in Figs. 3.13 and 3.14 (c1) to (c9) for each example set, respectively. The proposed filters with better frequency selectivity provide promising results when compared to other approaches as well as the crude wavelets family. For the crude wavelets, the worst results belong to the Mexican hat. Using the fused image and the depth map, it is straightforward to extract the 3D shape as is shown in Figs. 3.13 and 3.14 (d1) to (d9) for the Simulated Cone and Rifle datasets, respectively.

To assess the fusion performance statistical measurements such as signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), and mean square error (MSE) may be used if the reference image, fully-focused image in all depths, is available. How-

(a1) (a2) (b1) (b2)

Figure 3.12 – (a1) and (a2) are randomly selected samples from the Simulated Cone dataset. (b1) and (b2) show some sample images from the Rifle dataset. Note to the position of the focus plane in each image.

ever, in practice it can be achieved for simulated and synthetic images such as the Simulated Cone dataset, as the reference image is rarely known for real images. Recent image fusion assessment methods can evaluate the fusion technique without any reference image. These methods assess the fusion on input-output relationship. In [109] a mutual information (MI) based principle has been used to evaluate the fusion technique. MI calculates the quantity of information transferred from source images (input) to a fused image (output). Xydeas and Petrovic [104] proposed a fusion assessing technique based on pixel level (Qp) analysis, in which visual information or perceptual information is directly associated with the edge information while region information is ignored. Among several available quality assessment methods structural similarity (SSIM) index [101] has been widely used in imaging applications in literature when the reference images are not available. SSIM image quality index in fact measures three elements of image patches; the similarity of local brightness, contrast, and structures. The SSIM comparison without noise and in presence of a Gaussian noise are given in Tables 3.7 and 3.8, confirming the experimental results and visual improvements obtained. In these tables, db4, pro, and 9/7 stand for the Daubechies, the proposed, and the CDF 9/7 filter banks. It is seen that for the proposed wavelet based technique the results are independent of the images and datasets, in contrast, the results obtained by other approaches may be affected by the nature of images in each dataset, i.e., a uniform priority cannot be assigned. It can be seen in Tables 3.7 and 3.8 for Tenen, GLV, and SML.

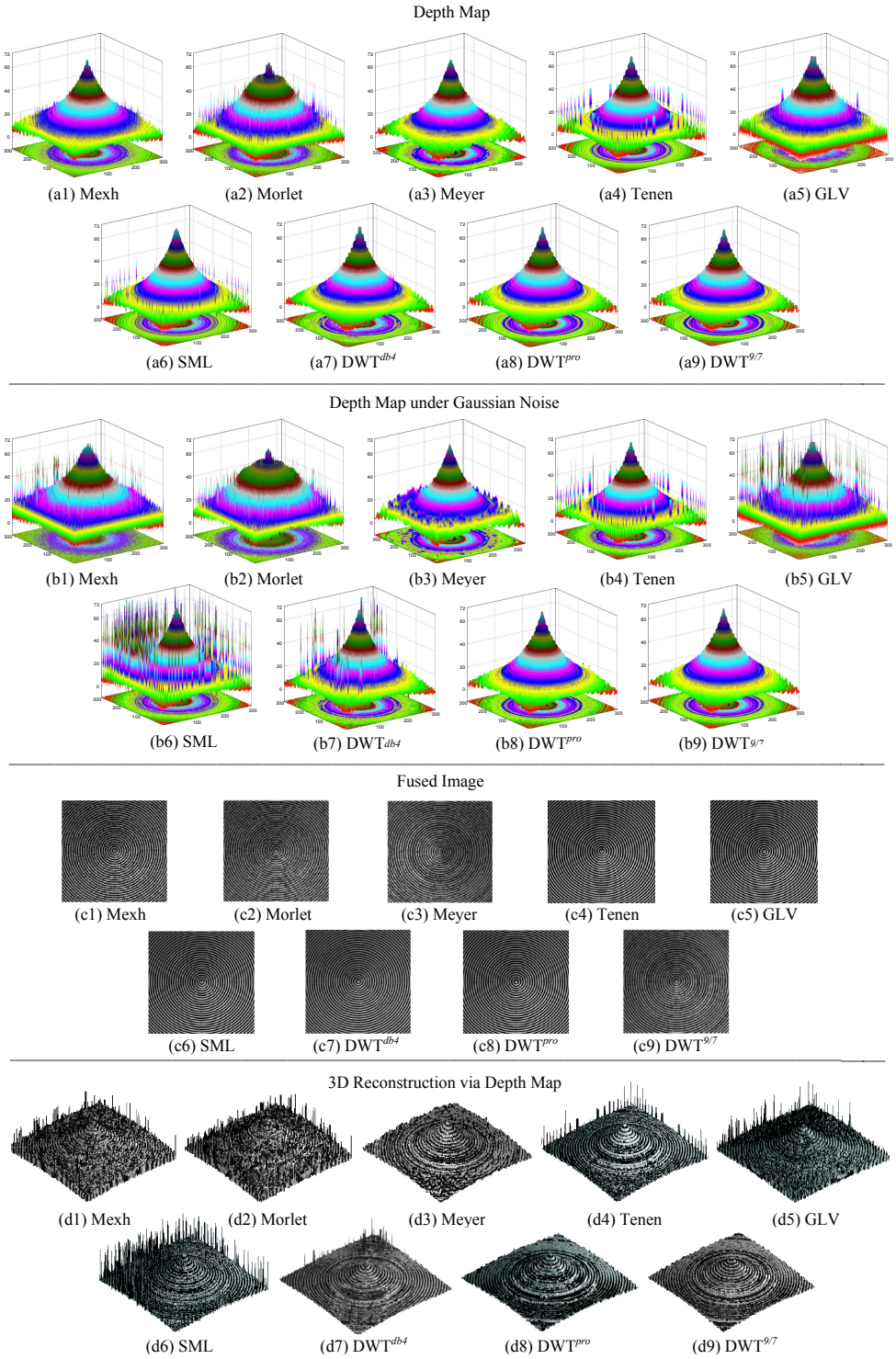


Figure 3.13 – Cone: (a1–a9) depth map without noise; (b1–b9) depth map under Gaussian noise; (c1–c9) fused image; (d1–d9) reconstructed shape using fused image and depth map under Gaussian noise.

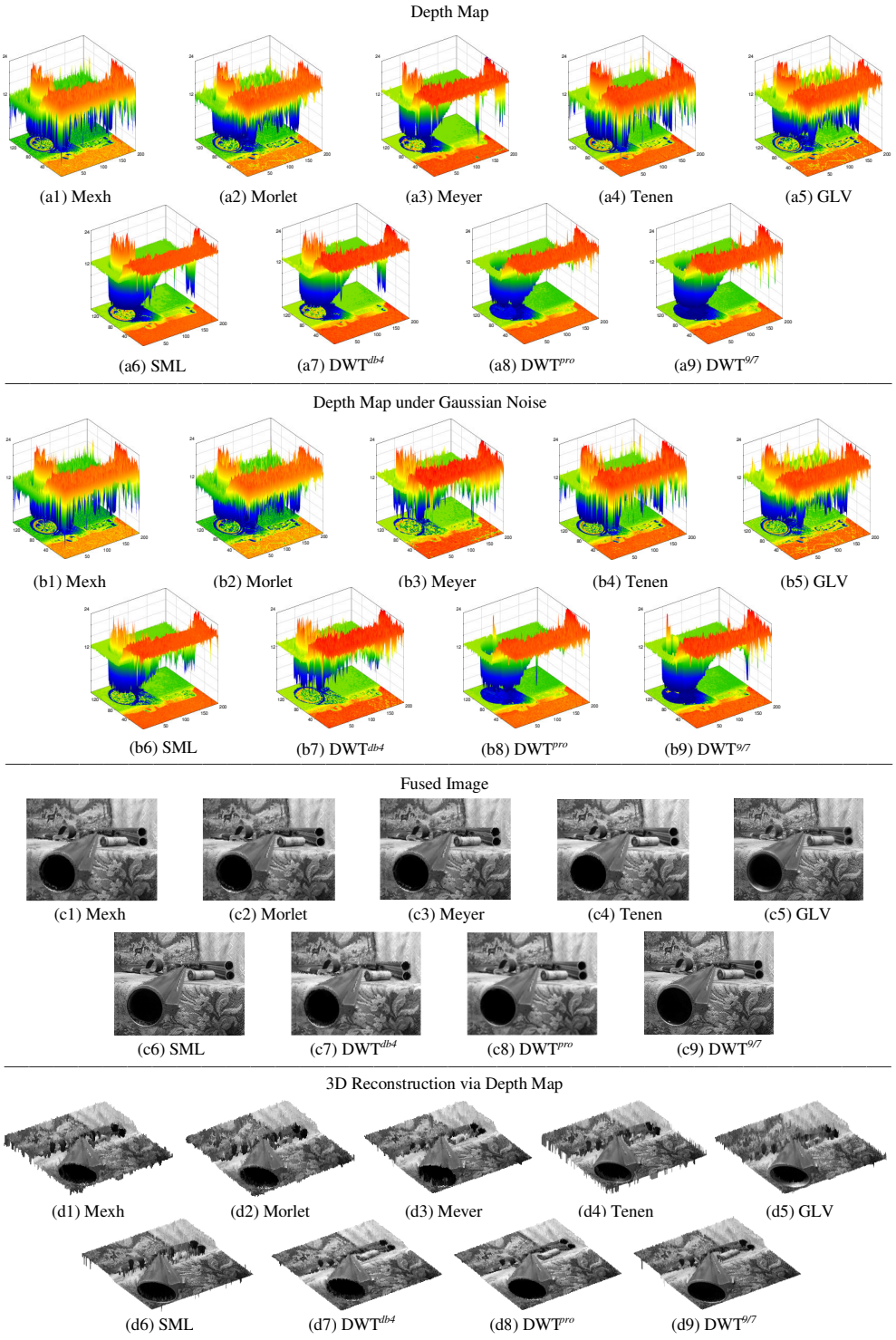


Figure 3.14 – Rifle. (a1–a9) depth map without noise; (b1–b9) depth map under Gaussian noise; (c1–c9) fused image; (d1–d9) reconstructed shape using fused image and depth map under Gaussian noise.

Table 3.7 – SSIM for different datasets using various methods (No noise).

Type	Sim.Cone	Chess	Lab	Clock	FlyEye	Rifle
Mexh	0.2216	0.6602	0.8961	0.8473	0.5753	0.4654
Morlet	0.2268	0.6643	0.9015	0.8505	0.5811	0.4723
Meyer	0.2391	0.7278	0.9254	0.8823	0.6153	0.4965
Tenen	0.2432	0.6757	0.9212	0.8702	0.5929	0.4875
GLV	0.2391	0.7111	0.9193	0.8617	0.6041	0.4997
SML	0.2400	0.7082	0.9240	0.8797	0.6004	0.4884
DWT^{db4}	0.2468	0.7215	0.9253	0.8835	0.6091	0.4941
DWT^{pro} [4]	0.2498	0.7749	0.9339	0.8812	0.6112	0.5170
$DWT^{9/7}$	0.2511	0.7765	0.9369	0.8821	0.6148	0.5204

Table 3.8 – SSIM for different datasets using various methods ($\sigma = 0.001$).

Type	Sim.Cone	Chess	Lab	Clock	FlyEye	Rifle
Mexh	0.1998	0.2912	0.7203	0.6572	0.3004	0.2959
Morlet	0.2033	0.2992	0.7281	0.6624	0.3076	0.3016
Meyer	0.2195	0.3341	0.7315	0.6728	0.3203	0.3323
Tenen	0.2335	0.3104	0.7386	0.6815	0.3173	0.3271
GLV	0.2142	0.3156	0.7373	0.6707	0.3253	0.3153
SML	0.2319	0.3083	0.7401	0.6802	0.3214	0.3128
DWT^{db4}	0.2212	0.3382	0.7282	0.6721	0.3157	0.3305
DWT^{pro} [4]	0.2349	0.3447	0.7429	0.6829	0.3309	0.3363
$DWT^{9/7}$	0.2387	0.3491	0.7493	0.6882	0.3376	0.3398

3.7 Conclusions and Summary

In this chapter, we have analyzed and investigated the effect of the frequency selectivity of filters associated with a multiresolution transformation on the performance and accuracy of a face recognition system. In order for this analysis being carried out, we first propose a general method to the design of the wavelets used in biorthogonal filter banks. The method structurally incorporates the desired number of moments and perfect reconstruction. The proposed formulation is, in general, a systematic

representation of a parametric polynomial. The rationale of the idea is to prove and show that the filter coefficients in the polynomial domain can be written in terms of the coefficients of the corresponding function in the frequency domain in general. The proposed technique offers tuning opportunity on the passband and stopband width and ripples, that is, we can incorporate free parameters to control the transition band, amplitude of ripples and number of moments. Depending on an application and the required trade-off between the filter pair characteristics, one can select different number of free parameters and tuning terms to have several alternatives to control the desired specifications. Based on the properties of the proposed method, and noting that the traditional maximally flat wavelet filters have poor frequency selectivity due to their wide transition band, we can then analyze the effect of the frequency response characteristics of the filters in a multiresolution transformation on the performance and accuracy of a face recognition system.

To this end, a multiresolution based face recognition system can be developed using the filter pairs introduced and designed in this chapter. In other words, the filter pairs of a typical discrete wavelet transform is replaced by our proposed tunable biorthogonal filters. The face images in databases are then decomposed to frequency subbands and high-frequency subbands are thresholded. The reconstruction is performed employing the customized tunable filter pairs. Therefore, the effect of the frequency selectivity of subbands of the multiresolution transformation on the performance of a face recognition system can be studied in details. We found that there is a relation between the sharpness of the frequency response of the filters and the recognition rate of a face recognition system. The amplitude of ripples was also found to be another factor that can influence the recognition accuracy. In general, it is concluded that sharper filters with possibly smaller ripples lead to higher recognition rates. It is interesting that the sharpness of the transition band of the filters are more important than the amplitude of ripples although unreasonably larger ripples degrade the results as in that case the nature of the filters, to act as a filter, may be lost.

Chapter 4

Resonance Based Image Analysis for Illumination Suppression

4.1 Literature Review

The literature on resonance based signal decomposition and oscillatory behavior of resonance components is not that vast as the idea has been just recently introduced in [79]. Therefore, it is expected that, similar to widely used transformations such as wavelets, curvelets, and Fourier transform, resonance based design and analysis will grow up rapidly in the near future. Basically, the representation of signals via traditional frequency based approaches like Fourier transform have been used for many years, where the respective techniques are mostly appropriate for signals with finite duration. In reality, the signals are not necessarily well expressed based on frequency components as they may consist of oscillatory components that cannot be elaborated using the traditional frequency based approaches. In [79], Selesnick shows that processing and study of non-stationary signals and pulses with oscillations can be well described based on a new nonlinear signal decomposition method in view of the concept of resonance rather than the frequency. The method in [79] describes how a signal can be considered as a combination of high-resonance and low-resonance components,

where the terms high- and low-resonance components refer to a signal with sustained and transient oscillations, respectively. Examples of such signals are widely found in medical systems such as electroencephalography, and in speech and voice processing field. In [81], it has been shown that using the rational-dilation wavelet transform (RDWT) and proposes that the high-resonance component can represented by a high Q-factor wavelet transform. Similarly, and as expected, the low-resonance component can be represented via a low Q-factor wavelet transforms. In order to simultaneously decompose a signal into its components, morphological component analysis (MCA) has been shown to be effective to separate the two components. In addition, iterated soft-thresholding approach and split augmented Lagrangian shrinkage have been investigated for the same purpose [81]. Resonance based signal analysis has been studied from a different point of view in [83] in which rational-dilation wavelet transform is used for the sparse representation of resonance components. It has been also shown that the split augmented Lagrangian shrinkage algorithm can be employed instead of MCA to speed up the optimization cycle of the procedure for the problem investigated in [83].

As mentions earlier, the resonance based signal representation is a very new topic, and therefore, the literature and research based on this concept is relatively seldom. In this chapter we show that by transferring a face image into signal domain and employing the resonance based analysis the oscillatory and transient components, the so-called high- and low-resonance components, promising results are achieved for the illumination invariant analysis of face images to the design of an efficient illumination invariant human face recognition system. The idea of designing a resonance based face recognition system in our research is motivated by the possibility of analyzing the nature of unwanted illumination effects in terms of resonance components. In fact, necessity of such investigation is twofold; the recognition accuracy can be further improved, and more importantly the number of tuning parameters can be reduced to lower the computational complexity.

Although illumination is mostly considered as the low-frequency part of images, these low-frequency contents may possess low- and/or high-resonance nature. In this chapter we first assume that an input image can be considered as a combination of illumination and reflectance. We then decompose the images into low- and high-resonance components simultaneously. Because the energy distributions of subbands via resonance decomposition are different for an image with good illumination effects and an image with high illumination variations, the energy of subbands of the two components can be thresholded to deactivate the subbands with unwanted energy distribution created by illumination effects. For dimensionality reduction and classification the principal component analysis and the extreme learning machine have been used, respectively. Experiments and comparisons illustrate the effectiveness of the proposed resonance-based method in illumination invariant face recognition.

4.2 Motivation and Problem Statement

In Chapter 2 we proposed a new method for the problem of illumination invariant human face recognition. The approach is essentially based on the frequency component analysis for which frequency subbands of a discriminator multiresolution transformation the so-called double-density dual-tree complex wavelet transform are taken into account. In Chapter 3, we go deeper and investigate how the frequency selectivity of a transformation may influence the performance of a recognition system. Accordingly a general solution was proposed to the design of tunable biorthogonal filter pairs for which we can control and tune the desired characteristics. We then analyzed how the recognition rate can be affected via these characteristics. While the approaches in Chapter 2 and Chapter 3 are actually based on the density of wavelets and the frequency selectivity of the wavelet subbands, they are in fact expressed in terms of the resolution of frequency subbands. In other words, low- and high-frequency components and overall frequency information is used to design and analyze the recognition

system. In this chapter concept of resonance is admitted to the problem of illumination invariant face recognition [7]. Although frequency based analysis is perhaps the most widely used approach in signal processing, it is indeed suitable for signals that are periodic and possess an oscillatory behavior. Nevertheless, many applications in reality, such as speech, biomedical, and communication signals, as well as physiological phenomena like human vision system, are a mixture of frequency and resonance. The nature of such signals makes it difficult to be studied by linear and frequency based methods. These non-stationary, nonlinear, and oscillatory signals are hard to be analyzed by linear, stationary, and frequency based approaches. In most of the cases, solutions to these problems are a linear approximation in the frequency domain. Recently it has been shown that nonlinear signal analysis based on signal resonance can open doors to investigate and answer to the weakness of the Fourier and wavelet transforms to some extent. In fact, resonance based analysis represents a signal as the sum of high-resonance and low-resonance components rather than traditional high-frequency and low-frequency subbands. In the next section, first the concept of resonance is briefly reviewed in contrast to the frequency based filtering. We then, for the first time, propose a new method for the problem of illumination invariant face recognition based on the resonance components of images.

4.3 Resonance versus Frequency

The concept of resonance based signal decomposition has been introduced and discussed in couple of excellent papers by Ivan W. Selesnick [79]–[83]. Fig. 4.1 shows a sample signal along with its basic and constructive elements. As it can be seen, the elements can have different natures, that is, Low-resonance Low-frequency, Low-resonance High-frequency, High-resonance Low-frequency and High-resonance High-frequency. In other words, and for example, Fig. 4.1 illustrates low-resonance signals where there is no persistency in oscillation. Shown in this figure, a low-resonance

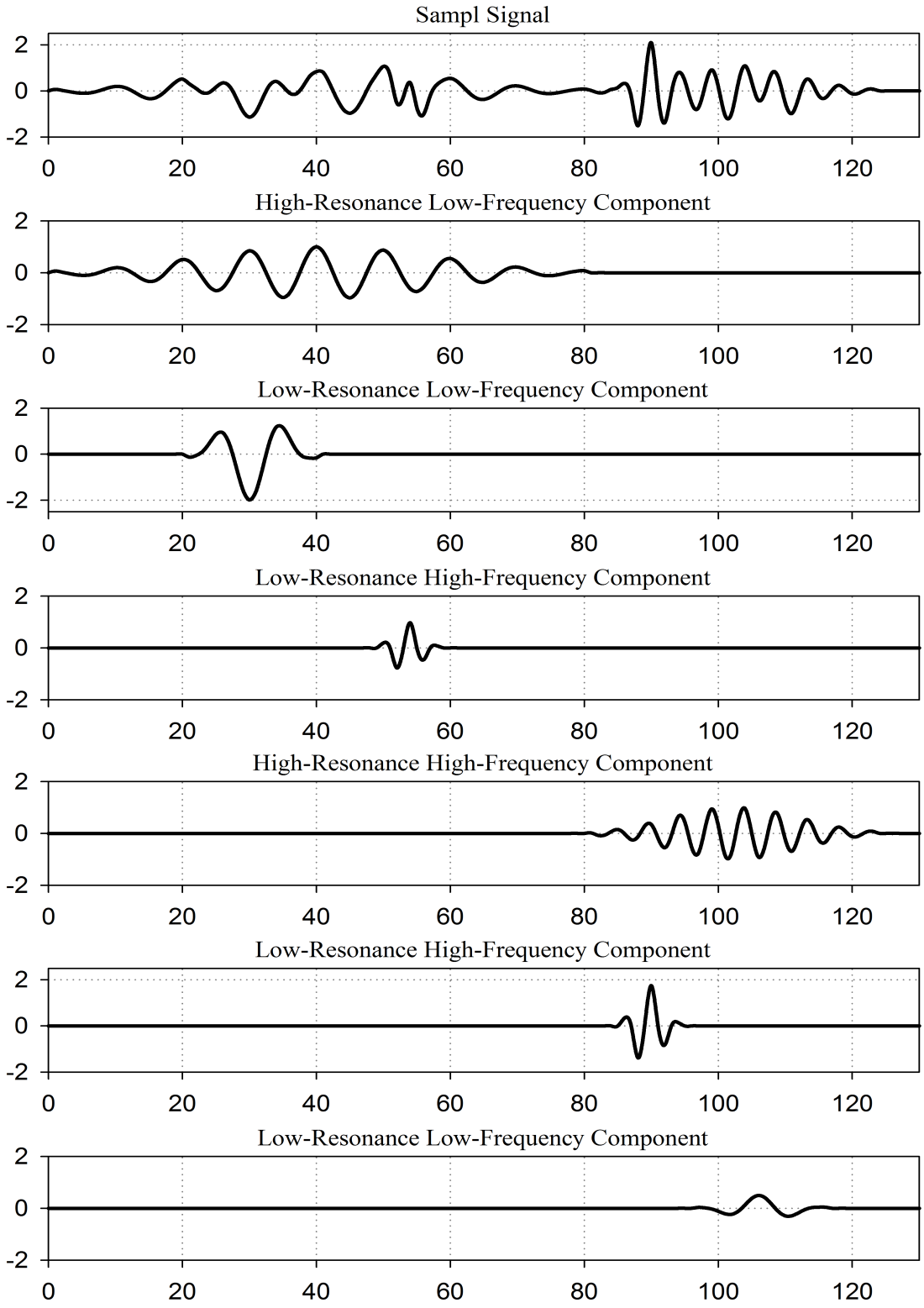


Figure 4.1 – Resonance based signal decomposition.

signal may contain either a high- or low-frequency pulse. The same explanation is valid for other components. On the other hand, Fig. 4.1 also shows examples of high-resonance signals with sustained oscillation behavior. Similar to discussion above, a high-resonance signal may contain either high- or low-frequency pulses. It is worth pointing out that the pulses shown in Fig. 4.1 are indeed time scaled representation of one another, that is, the degree of resonance is not affected by time scaling.

In general, and as it is shown in Fig. 4.2, there is a fundamental difference between frequency and resonance representation. For further explanation, a signal of six pulses with different frequencies and two resonance bands is depicted in Figure 4.2 for instance. It has been clearly observed that the information and content of resonance based and frequency based representations are substantially different. To tell the end of story at the beginning; as the illumination is believed to lie in low-frequency resolutions, we would like to put a question mark as if the low-frequency components may need to have a low- or high-resonance nature as well before being considered as illumination.

In this figure, the sample signal shown in (a) has been decomposed to its components first using a resonance based decomposition and then via the frequency based filtering. It has been shown in Fig. 4.2 (b) and (c), respectively. As it is seen, the beauty of the resonance based representation is that the signal information is not necessarily bounded in frequency resolution subbands. An immediate contribution of such decomposition lies in the fact that linear frequency domain filtering cannot be used to discriminate the low- or high-resonance components of a signal. The separation in Fig. 4.2 indicates that the sample signal is the summation of the high- and low-resonance components, where it is assumed that the reconstruction error is almost zero. Clearly, traditional linear time-invariant filters cannot offer such separation shown in Fig. 4.2 (b) as the frequencies specified in the high-resonance component are exactly the same as the frequencies involved in the low-resonance component. In other words, some parts of the low- and high-frequency components, shown in

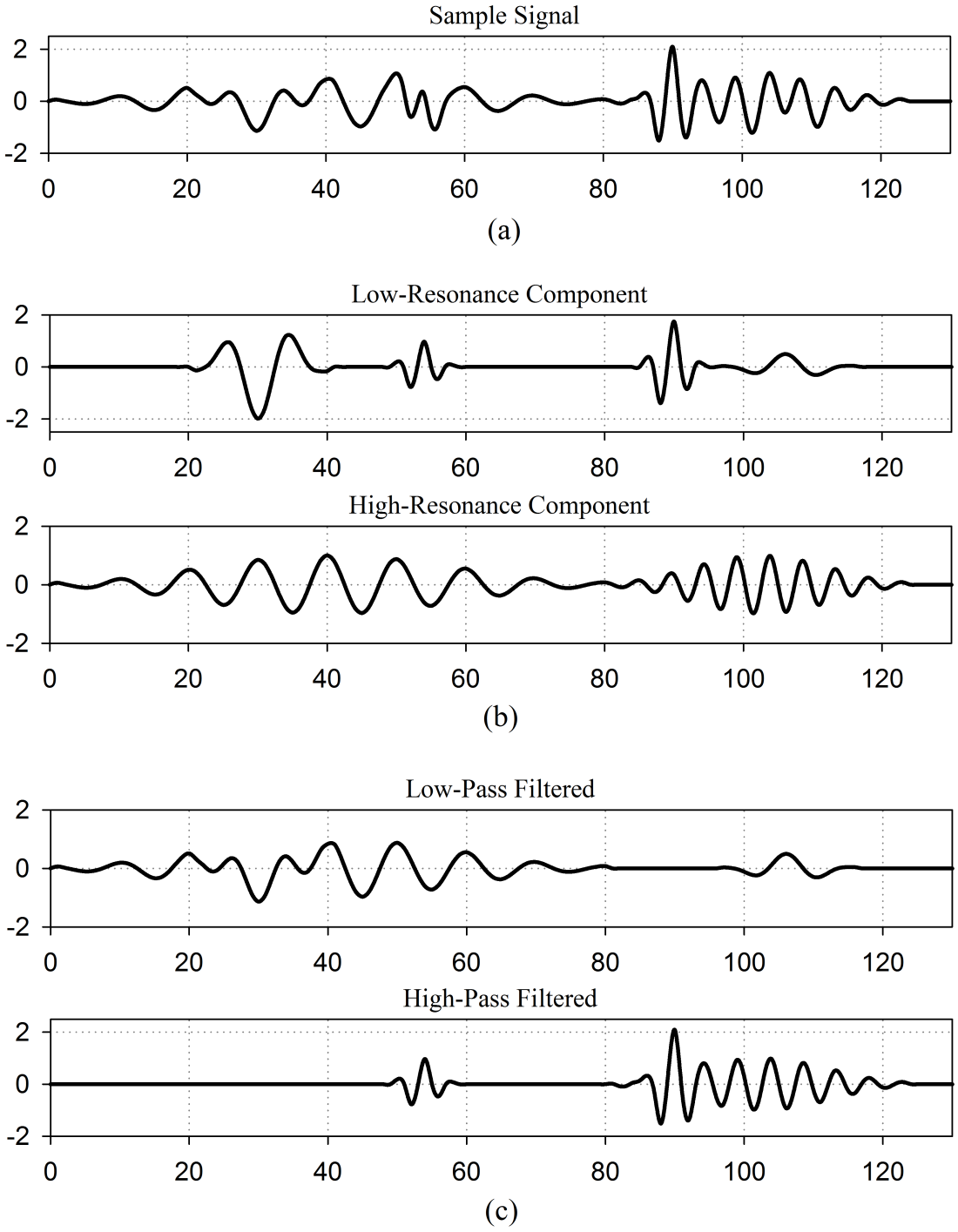


Figure 4.2 – Resonance based decomposition and frequency based filtering of a sample signal. (a) Sample signal; (b) Resonance based decomposition of the signal in (a); (c) Frequency based filtering of the signal in (a).

Fig. 4.2 (c), may appear in low- and high-resonance components in Fig. 4.2 (b) with different degrees of resonance. In [79], Selesnick draws the reader's attention to an interesting observation, that is, what if a sample signal of one-cycle low-resonance and five-cycle high-resonance contain a three-cycle resonance based component? On the other hand, how and under what category such a component should be classified? In general, and to the best of our knowledge, while it is not clear how the resonance of a signal should be defined, a generic signal can be indeed separated into low- and high-resonance components. It is eventually in contrast to the frequency based analysis where the frequency response along with the linearity of the filtering operation, convolution, can simply determine the formal relation between input and output of a system, the so-called transfer function. Although at first glance the resonance based decomposition and representation may seem complicated and elusive, the way it deals with signal analysis is unique and outstanding. One should note that frequency based multiresolution filtering can be written analytically via the use of convolution. In contrast, resonance based decomposition is definitely a nonlinear procedure for which numerical iterative techniques must be employed.

It should be noted that, in general, there is no priority or intention to show whether the resonance based representation or the frequency based filtering may outperform each other. One should note that while they are fundamentally different, each of which may have its own advantages and drawbacks. For instance, applications such as communication, image processing, power systems, few to mention, may demand a specific representation of signals either in terms of frequency or resonance or a combination of both. In this dissertation, we first introduced a new face recognition system. We then analyzed how the frequency filtering and frequency selectivity of subbands may affect the performance of the proposed system. Consequently, and in view of the resonance based representation of signals, a novel illumination invariant face recognition system is proposed in this chapter. The latter is then compared to the recent results in the literature including our frequency based approach.

4.4 Tunable Q-Factor Wavelet Transform

The tunable Q[1]-factor wavelet transform (TQWT) is a new and interesting transformation that has been recently proposed as an efficient and powerful tool to look at the multiresolution analysis from a different point of view [80].

The necessity and importance of the Q-factor of a wavelet becomes more obvious when the problem of signal analysis needs to be tailored application wise. In other words, it can attack a problem based on the oscillatory nature of the signal rather than linear frequency based analysis. For instance, imagine one may face two different problems, that is, processing of an oscillatory signal and processing a signal with less or no oscillation. While the first issue can be addressed with the wavelet transform of relatively high Q-factor, the second problem is better analyzed if a low Q-factor transformation is used. In terms of design, the TQWT possesses technically remarkable structure such as the perfect reconstruction property, reasonably overcomplete rate, and discrete Fourier transform (DFT) based implementation. It can be also implemented more efficiently using radix-2 fast Fourier transform (FFT) [79]. It should be noted that unlike other similar transforms, such as the rational-dilation wavelet transform (RADWT) [9], its parameters can be directly specified by the Q-factor and oversampling rate of the TQWT.

The tunable Q-factor wavelet transform is defined and derived based on two main parameters, that is, the Q-factor and its oversampling rate also known as the redundancy. The TQWT is derived based on two additional blocks, known as the lowpass and highpass scaling blocks, to a regular wavelet filter bank. Shown in Fig. 4.3, the sampling rate for the analysis side subband signals $x_1(n)$ and $x_2(n)$ are αf_s and βf_s, respectively and $0 < \beta \leq 1$, $0 < \alpha < 1$, and f_s is the sampling rate of the input signal. Note that for the strict perfect reconstruction $\beta + \alpha > 1$. For perfect reconstruction it is required to have $x(n)$ and $u(n)$ identical. Let us consider the Fourier transforms

[1]The notation Q in Q-factor stands for the term Quality.

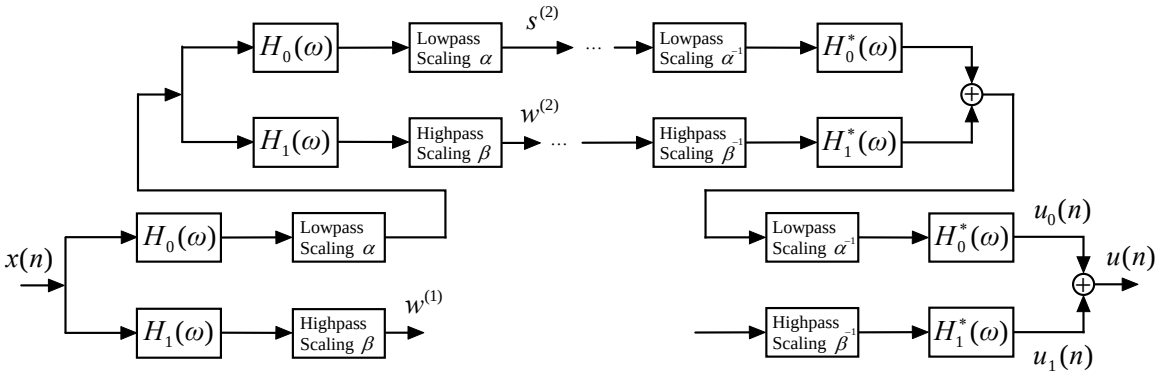


Figure 4.3 – Typical structure of a tunable Q-factor wavelet transform.

of $u_0(n)$ and $u_1(n)$ as

$$U_0(\omega)=\begin{cases}|H_0(\omega)|^2X(\omega) & |\omega|\leq\alpha\pi\\ 0 & \alpha<|\omega|\leq\pi\end{cases},\quad U_1(\omega)=\begin{cases}0 & |\omega|<(1-\beta)\pi\\ |H_1(\omega)|^2X(\omega) & (1-\beta)\pi\leq|\omega|\leq\pi\end{cases} \tag{4.1}$$

and therefore the Fourier transform of the output $u(n)$ can be written as

$$U(\omega)=\begin{cases}|H_0(\omega)|^2X(\omega) & |\omega|<(1-\beta)\pi\\ (|H_0(\omega)|^2+|H_1(\omega)|^2)X(\omega) & (1-\beta)\pi\leq|\omega|<\alpha\pi\\ |H_1(\omega)|^2X(\omega) & \alpha\pi\leq|\omega|\leq\pi.\end{cases} \tag{4.2}$$

Recall that for perfect reconstruction the followings must be held

$$U_0(\omega)=\begin{cases}|H_0(\omega)|=1 & |\omega|\leq(1-\beta)\pi\\ H_0(\omega)=0 & \alpha\pi\leq|\omega|\leq\pi\end{cases},\quad U_1(\omega)=\begin{cases}H_1(\omega)=0 & |\omega|\leq(1-\beta)\pi\\ |H_1(\omega)|=1 & \alpha\pi\leq|\omega|\leq\pi\end{cases} \tag{4.3}$$

also, the transition band needs to satisfy

$$|H_0(\omega)|^2+|H_1(\omega)|^2=1,\quad \omega\in(1-\beta)\pi\leq|\omega|<\alpha\pi. \tag{4.4}$$

The tunable Q-factor wavelet transform is designed based on the iterative procedure in a two-channel filter bank shown in Fig. 4.3. Note that only two stages of the transformation, in the analysis and synthesis sides of the filter bank, have been

shown in this figure and the number of stages (J) can be iteratively increased. The configuration and setup in Fig. 4.3 yields

$$H_{0\atop \alpha\leq 1}^{(J)}(\omega)=\begin{cases}\prod_{i=0}^{J-1}H_0(\omega\alpha^{-i}) & |\omega|\leq\alpha^J\pi\\ 0 & \alpha^J\pi<|\omega|\leq\pi\end{cases} \tag{4.5}$$

and

$$H_{1\atop \alpha\leq 1,\beta\leq 1}^{(J)}(\omega)=\begin{cases}H_1(\omega\alpha^{1-J})\prod_{i=0}^{J-2}H_0(\omega\alpha^{-i}) & (1-\beta)\alpha^{(J-1)}\pi\leq|\omega|\leq\alpha^{(J-1)}\pi\\ 0 & \omega\in[-\pi,\pi]\,.\end{cases} \tag{4.6}$$

It should be noted that although the two frequency responses $H_0(\omega)$ and $H_1(\omega)$ are identically unity and zero in their passbands and stopbands, respectively, they are definitely not ideal lowpass and highpass filters in reality. Therefore, there is a transition band in each case in which the magnitude of the respective frequency response may vary from zero to one; the interval that is often called as the 'don't care' area. In fact, the transition band width is equal to $(\alpha+\beta-1)\pi$ which is actually the same as the overpassed critical sampling rate of the corresponding filter bank where for $\alpha+\beta=1$ the filter bank is technically is in the so-called critically sampled condition for which a zero-width transition band is assumed to be the case. If it happens, theoretically the time domain responses are not adequately localized and that would not be a desired situation in most applications. To avoid this situation it is required to keep $\alpha+\beta>1$. Detailed explanations are given in [79][80]. Now at this point we need to consider the parameters defined and used in a tunable Q-factor wavelet transform. Basically, the main concern here is how parameters such as the Q-factor, bandwidth of the transformation, and redundancy may be related to each other. As already discussed above, for a two-channel configuration shown in Fig. 4.3, the rate of oversampling or the redundancy is defined as $\alpha+\beta$. As the sampling rate at subband J is given by $\alpha^{(J-1)}\beta f_s$, iteration on lowpass output to reach the wavelet transform redundancy is given by $r=\beta(1-\alpha)^{-1}$. Before introducing

the meaning of Q-factor, first it is required to define the center frequency and the bandwidth of the transformation. From the signal processing point of view, the frequency response of $H_1(\omega)$ at stage J, $H_1^{(J)}(\omega)$ is nonzero in $(\omega_1,\ \omega_2)$ where we have $\omega_1 = (1-\beta)\alpha^{(J-1)}\pi$ and $\omega_2 = \alpha^{(J-1)}\pi$ [80]. With this notation the center frequency at stage J is defined as $\omega_c = \frac{\omega_1+\omega_2}{2} = \frac{(2-\beta)\alpha^{(J-1)}\pi}{2}$. An immediate outcome of the discussion is that clearly the band width of the transformation is then defined as BW$= \frac{\omega_2-\omega_1}{2} = \frac{\beta\alpha^{J-1}\pi}{2}$. In physics and communication, quality factor or the so-called Q-factor, is defined as the ratio of the center frequency to the band width, that is, $Q = \frac{\omega_c}{\text{BW}} = \frac{2-\beta}{\beta}$. Interestingly, and due to the independency on the recent equality to the number of stages, the tunable Q-factor wavelet transform is a constant-Q transformation. Furthermore, noting that $r = \beta(1-\alpha)^{-1}$, one can write $\beta = \frac{2}{Q+1}$. Increasing the value of Q makes the wavelets to oscillate more. Technically speaking, the higher the Q-factor, the higher the resonance of the signals in subbands of the transformation. Noting that there is no restriction for Q and r to be integers, it is not desired to set the value of redundancy to unity as it leads to the sinc wavelet for which the transition band for $H_0(\omega)$ and $H_1(\omega)$ will be too narrow to perform the required time-domain response localization. Higher and lower Q-factor representation provides an outstanding opportunity to study and analyze the behavior of a given signal in terms of its resonance components. This is not exactly in contrast to, but very different from the traditional Fourier based or multiresolution based frequency component analysis. In the next section it is shown that how a sample signal can be decomposed into its high- or low-resonance components. We then go one step further and show that how a signal can be decomposed into its high- and low-resonance components simultaneously. The latter is used to design a resonance based human face recognition system that can significantly suppress unwanted illumination effects to notably improve the recognition rates for the problem of illumination invariant face recognition.

4.4.1 Higher and Lower Q-Factor Representations

The tunable Q-factor wavelet transform lets specifying the amount of oscillation for windows to view a given signal in terms of the degree of resonance. It has been shown in Figs. 4.4 and 4.5 in which an image from Subset 5 of the Extended-Yale B database has been transformed into the signal domain first. Fig. 4.4 shows the decomposition of the signal into high-resonance components where the energy of subbands have been depicted in the right corner of the figure for each subband individually. For this decomposition the value of Q-factor, redundancy, and the number of stages are $Q = 4$, $r = 3$, and $J = 31$, respectively. Fig. 4.5 illustrates the same representation for the low-resonance components of the signal via $Q = 1$, $r = 3$, $J = 10$ followed by the energy of each subband.

It can be seen that by increasing the value of Q the wavelets perform more oscillations which means the frequency responses are more narrow. One should note that by increasing the oscillations the number of stages is required to be increased too to be able to span the respective frequency range as the frequency responses become narrow. The signal representation based on the concept of resonance is essentially different from the frequency based signal filtering and this has been shown in these figures. For instance the signals in subbands in Fig. 4.4 contain high-resonance elements although they might be either high- or low-frequency information. The same discussion is valid for the subbands in Fig. 4.5 in which the low-resonance elements are shown regardless of being high- or low-frequency contents. Now one may come up with a question that if a signal could be presented in terms of its resonance components simultaneously, that is, to separate the signal into two visible parts rather than several basic elements to clearly see the high- and low-resonance portions at the same time. In [79] it has been shown that the dual Q-factor signal decomposition can lead to this observation. It is discussed in the next section. In Section 4.5 we propose a new method for the problem of the illumination invariant face recognition that is based on the above mentioned simultaneous resonance based signal decompo-

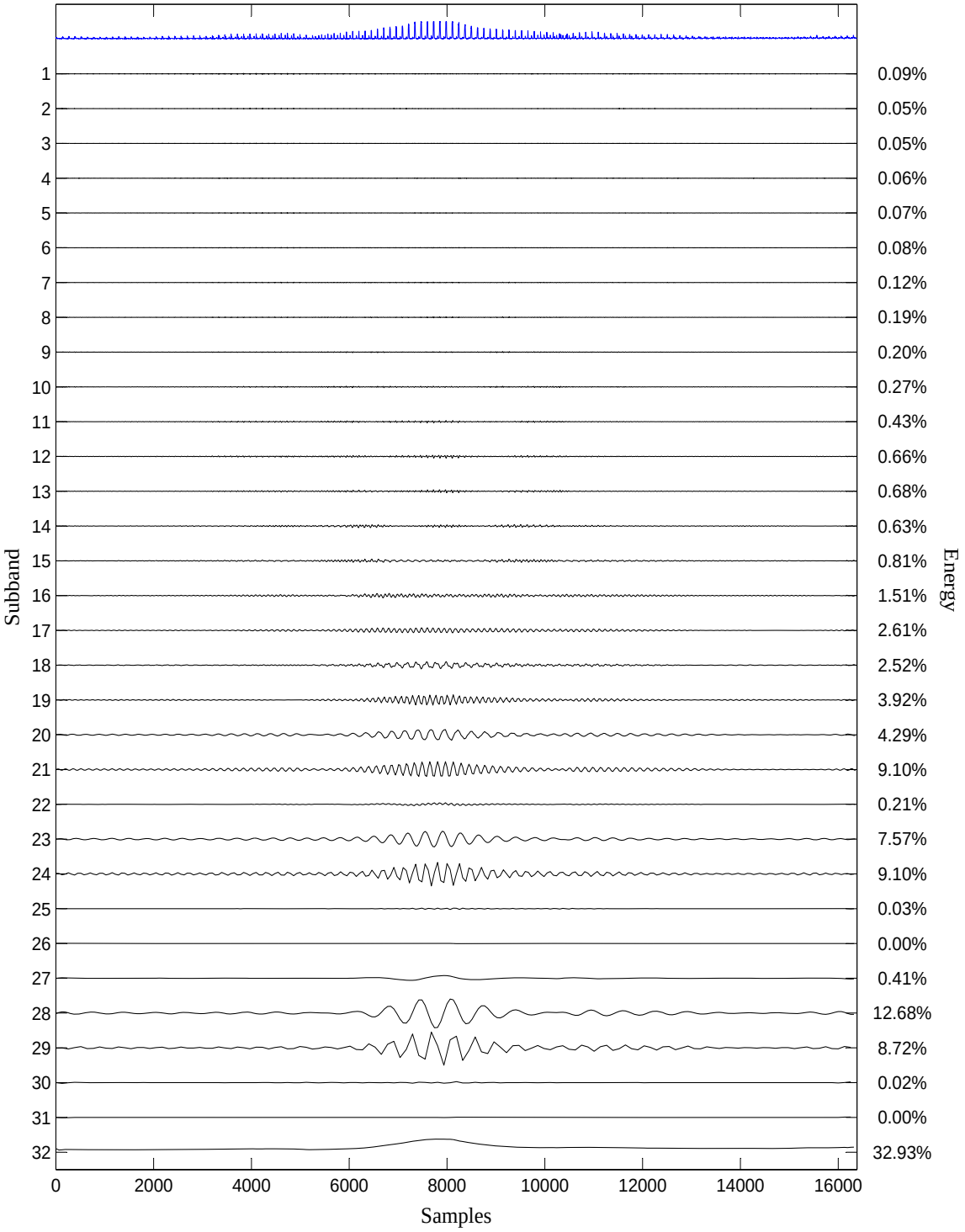


Figure 4.4 – Resonance based signal decomposition to separate the high-resonance components. A sample image is shown as a signal at top. High-resonance components have been presented via $Q = 4$, $r = 3$, $J = 31$.

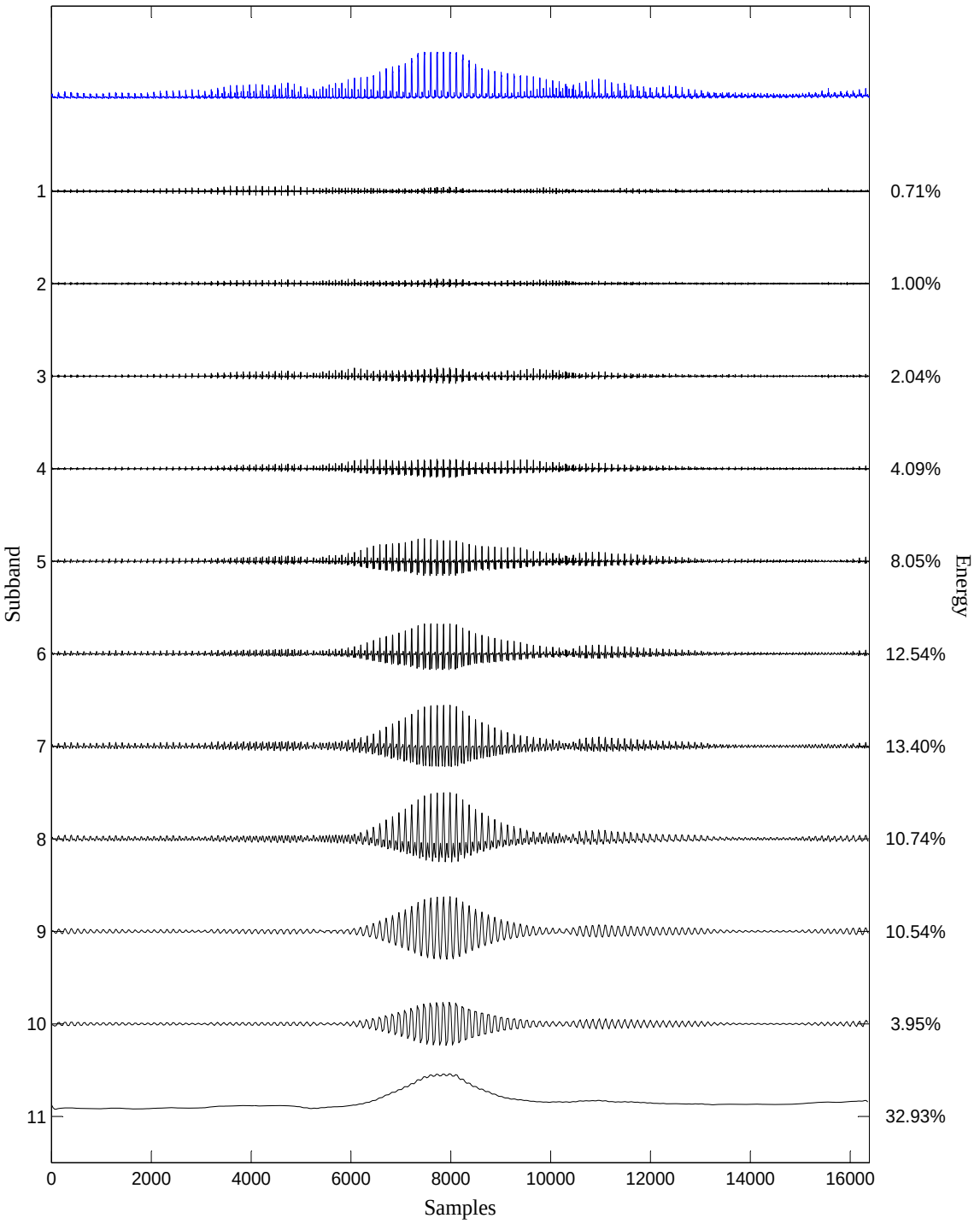


Figure 4.5 – Resonance based signal decomposition to separate the low-resonance components. A sample image is shown as a signal at top (the same image in Fig. 4.4). Low-resonance components have been presented via $Q = 1$, $r = 3$, $J = 10$.

sition and energy-thresholding of the transformation subbands to reduce the effect of unwanted illumination.

4.4.2 Dual Q-Factor Signal Decomposition

In previous section it was shown that an input signal can be illustrated in terms of its high- or low-resonance components. It has been shown in Fig. 4.4 and Fig. 4.5, for a sample image from Subset 5 of the Extended-Yale B database, respectively. Although the single representation, in terms of only the high- *or* low-resonance components, has its simplicity to be used in several relevant applications, one may come up with a valid and important question as how to possibly describe the representation of a signal using two Q-factors simultaneously. On the other hand, how to decompose a signal into high- *and* low-resonance components. This can be seen as higher and lower Q-factors Q_1 and Q_2 shown in Fig. 4.6 (d) and (e) for a sample image given in (a). In this figure, (b) presents the image in (a) after applying (4.19).

Now let us consider a given signal $\mathbf{x}$ to be decomposed into an oscillatory signal and a non-oscillatory signal, $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, defined by $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ accordingly.

Note that technically speaking, the oscillatory and non-oscillatory terms may be frequently used in physics, however, we rely on the signal processing equivalent phrases as the signal $\mathbf{x}$ is desired to be represented by its high- and low-resonance components. According to the discussion in Section 4.3, the procedure to simultaneously represent a signal via its resonance components requires a nonlinear approach where the frequency based techniques such as filtering fail. It is worth pointing out that a high resonance signal cannot be efficiently represented with a low Q-factor transform and similarly a low resonance input cannot not be well represented by using a high Q-factor transform. This is why resonance based signal decomposition is considered as an efficient transformation for which the way that the transformation looks at the problem is entirely different from the frequency based representation of signals. Feasibility of the discrimination of $\mathbf{x}_1$ and $\mathbf{x}_2$ is an optimization problem in

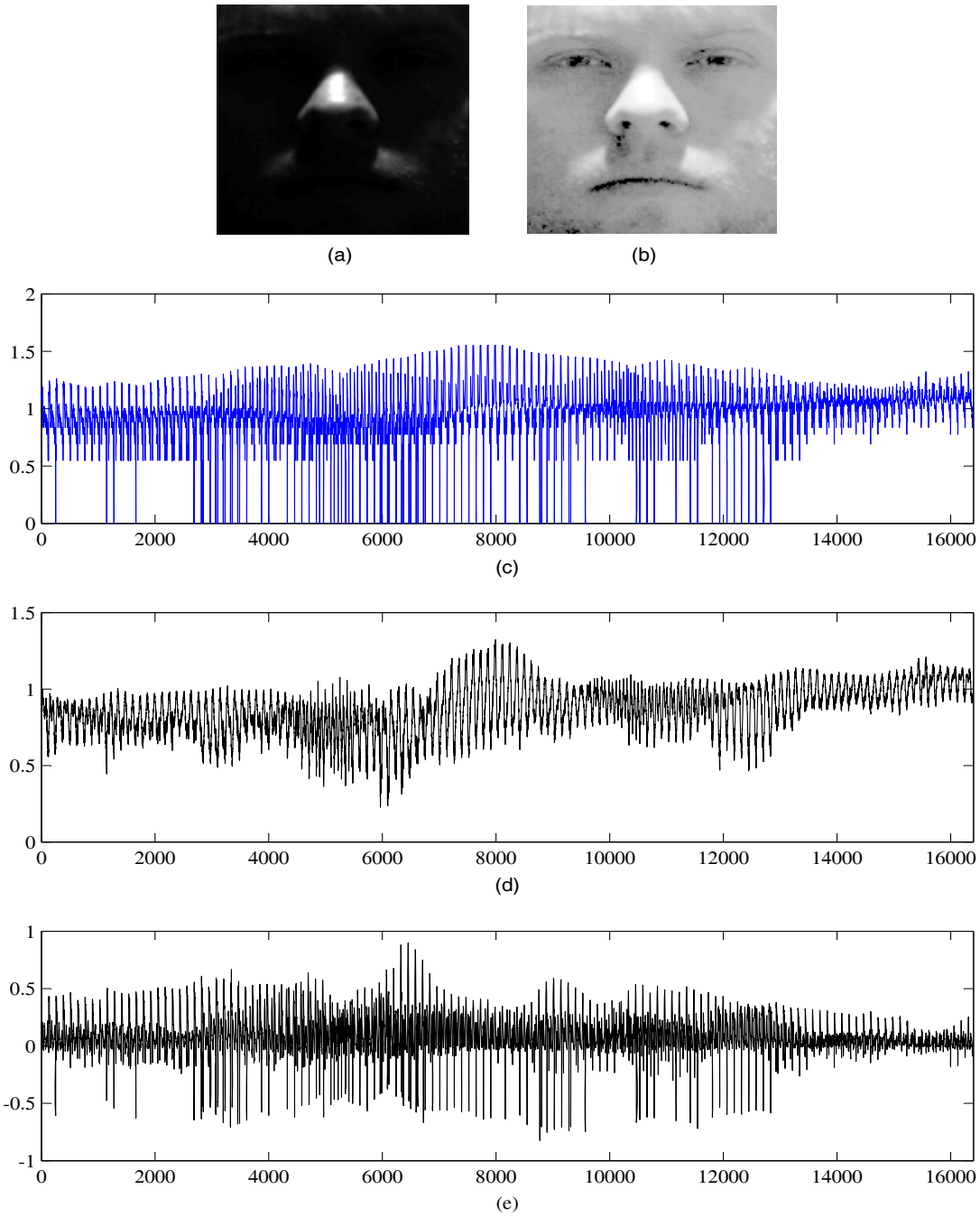


Figure 4.6 – Resonance based image decomposition to separate the high- and low-resonance components simultaneously. (a) A sample image from the Extended-Yale B database; (b) Equality in (4.19) is applied to the image in (a); (c) Signal representation for the image in (b); (d) High Q-factor channel (high-resonance components) via $Q_1 = 6$, $r_1 = 3$, $J_1 = 31$; (e) Low Q-factor channel (low-resonance components) via $Q_2 = 1$, $r_2 = 6$, $J_2 = 10$.

which both high and low Q-factors are employed jointly to represent the signal $\mathbf{x}$. In [88] the authors present a new method, known as morphological component analysis (MCA), to separate images into texture and piecewise smooth parts in which both the variational and sparsity mechanisms have been taken into consideration. Using MCA, the components $\mathbf{x}_1$ and $\mathbf{x}_2$ can be extracted individually and separately.

As it has been discussed in [15] traditional signal representation intends to decompose signals via superposition of sinusoids, the Fourier based representation, which possesses several drawbacks although the transformation is the most famous technique for signal representation. Therefore, there has been a remarkable attempt by authors to propose signal representation methods that can sparsely perform the decomposition. The development and progress on overcomplete multiresolution transformations, e.g., wavelets, curvelets, wavelet packets, higher-density dual-tree real and complex wavelets, chirplets, framelets, and the fact that decomposition into overcomplete systems is not necessarily unique, demands efficient methods to possibly represent the signals with less number of significant coefficients. In fact, traditional signal representation techniques mostly focus on gaining sufficient accuracy under reasonable computational complexity. However, sparsity of such representations remains undeveloped. Basis pursuit [15] is a recent and notable progress on signal representation that takes the overcomplete transformations into account in which the decomposition is achieved via the ℓ_1 norm minimization of the coefficients dealing with the representation. Being more specific, basis pursuit assumes that there are in general many representations due to number of overcomplete waveform bases as mentioned above.

Let us assume $\mathbf{x}_1$ and $\mathbf{x}_2$ are represented as weighted sum of matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. The basis pursuit analysis suggests the following convex problem where the approach indeed offers the sparsest solution amongst all representations [26][27][31], that is

$$\underset{\mathbf{w}_1,\mathbf{w}_2}{\operatorname{argmin}} \ \lambda_1\|\mathbf{w}_1\|_1 + \lambda_2\|\mathbf{w}_2\|_1 \ \text{ subject to } \ \mathbf{x} = \mathbf{S}_1\mathbf{w}_1 + \mathbf{S}_2\mathbf{w}_2. \tag{4.7}$$

Considering possible noise, the problem above can be written as

$$\underset{\mathbf{w}_1,\mathbf{w}_2}{\operatorname{argmin}} \lambda_1\|\mathbf{w}_1\|_1 + \lambda_2\|\mathbf{w}_2\|_1 \;\text{ subject to }\; \|\mathbf{x} - \mathbf{S}_1\mathbf{w}_1 - \mathbf{S}_2\mathbf{w}_2\|_2 \leq \varepsilon \tag{4.8}$$

where the unconstrained form of the problem

$$\mathbf{O}(\mathbf{w}_1, \mathbf{w}_2) := \|\mathbf{x} - \mathbf{S}_1\mathbf{w}_1 - \mathbf{S}_2\mathbf{w}_2\|_2^2 + \lambda_1\|\mathbf{w}_1\|_1 + \lambda_2\|\mathbf{w}_2\|_1 \tag{4.9}$$

must be minimized with respect to $\mathbf{w}_1$ and $\mathbf{w}_2$ where $\mathbf{S}_1\mathbf{w}_1$ and $\mathbf{S}_2\mathbf{w}_2$ are the estimated components to represent $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively.

Before investigating possible solutions for this problem, it is noted that the use of ℓ_2 norm cannot lead to prominent approximation. If the ℓ_2 norm is employed in (4.9) to impose a penalty in an unconstrained form of the problem, the minimizing $\mathbf{w}_1$ and $\mathbf{w}_2$ may be written in closed form [79], and consequently the estimated components $\mathbf{x}_1$ and $\mathbf{x}_2$ can be obtained in terms of λ_1, λ_2 and $\mathbf{x}$. It means that the estimated representatives, $\mathbf{x}_1$ and $\mathbf{x}_2$, are just a ratio of $\mathbf{x}$. It is also worth pointing out that the basis pursuit approach is not a simple replacement between ℓ_2 and ℓ_1 norms as it creates major consequences. Technically speaking, the traditional ℓ_2 norm based approximation is a quadratic problem with linear equality constraints, however, the basis pursuit is actually a nonquadratic problem while it still remains convex; solution to the basis pursuit requires more effort and thus it has been investigated in depth due to its importance [15].

The minimization of the objective function in (4.9) can be derived based on the morphological component analysis. That is, the equality in (4.9) can be written as

$$\mathbf{O}(\mathbf{w}_1, \mathbf{w}_2) = \|\mathbf{x} - \mathcal{S}\mathbf{w}\|_2^2 + \lambda_1\|\mathbf{w}_1\|_1 + \lambda_2\|\mathbf{w}_2\|_1 \tag{4.10}$$

where $\mathcal{S} = [\mathbf{S}_1 \quad \mathbf{S}_1]$, $\mathbf{w} = [\mathbf{w}_1 \quad \mathbf{w}_2]^t$, and t stands for transpose.

In [32], Figueiredo et al proposed a method known as the majorization-minimization

that is an iterative algorithm to find the minimizer argument

$$\begin{aligned} \boldsymbol{\upsilon}^i &= \mathbf{w}^i + \frac{1}{\varsigma}\mathcal{S}^t(\mathbf{x} - \mathcal{S}\mathbf{w}^i) && (4.11) \\ \mathbf{w}^{i+1} &= \underset{\mathbf{w}}{\operatorname{argmin}}\ [\varsigma\|\boldsymbol{\upsilon}^i - \mathbf{w}\|_2^2 + \lambda_1\|\mathbf{w}_1\|_1 + \lambda_2\|\mathbf{w}_2\|_1] && (4.12) \end{aligned}$$

where the value of ς is greater than the maximum eigenvalue of $\mathcal{S}^t\mathcal{S}$ to ensure stability and convergence [81]. It should be noted that because

$$\boldsymbol{\upsilon}^i = \begin{bmatrix} \mathbf{w}_1^i \\ \mathbf{w}_2^i \end{bmatrix} + \frac{1}{\varsigma}\begin{bmatrix} \mathcal{S}_1^t \\ \mathcal{S}_2^t \end{bmatrix}\left(\mathbf{x} - [\mathcal{S}_1 \quad \mathcal{S}_2]\begin{bmatrix} \mathbf{w}_1^i \\ \mathbf{w}_2^i \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{\upsilon}_1^i \\ \boldsymbol{\upsilon}_2^i \end{bmatrix} \tag{4.13}$$

it is sufficient to apply iterated soft-thresholding algorithm [24] on $\mathbf{w}$ to minimize $\mathbf{O}(\mathbf{w}_1, \mathbf{w}_2)$, that is

$$\begin{aligned} \mathbf{w}_1^{i+1} &= \mathbf{sft}(\boldsymbol{\upsilon}_1^i, \frac{\lambda_1}{2\varsigma}) && (4.14) \\ \mathbf{w}_2^{i+1} &= \mathbf{sft}(\boldsymbol{\upsilon}_2^i, \frac{\lambda_2}{2\varsigma}) && (4.15) \end{aligned}$$

where **sft** denotes the iterated soft-thresholding operator and it is defined as follows for a given input x and the threshold value of $\mathfrak{T}$

$$\mathbf{sft}(x, \mathfrak{T}) := \mathbf{sgn}\max(0, |x| - \mathfrak{T}) \tag{4.16}$$

where **sgn** is the signum function [81].

The derivations can be adopted in terms of the above mentioned resonance based discussion. In fact, the resonance based decomposition can be defined as solving the constraint optimization problem as

$$\underset{\mathbf{w}_1,\mathbf{w}_2}{\operatorname{argmin}}\ \lambda_1\|\mathbf{w}_1\|_1 + \lambda_2\|\mathbf{w}_2\|_1 \equiv \underset{\mathbf{w}_1,\mathbf{w}_2}{\operatorname{argmin}}\ \sum_{j=1}^{J_1+1}\lambda_{1,j}\|\mathbf{w}_{1,j}\|_1 + \sum_{j=1}^{J_2+1}\lambda_{2,j}\|\mathbf{w}_{2,j}\|_1 \tag{4.17}$$

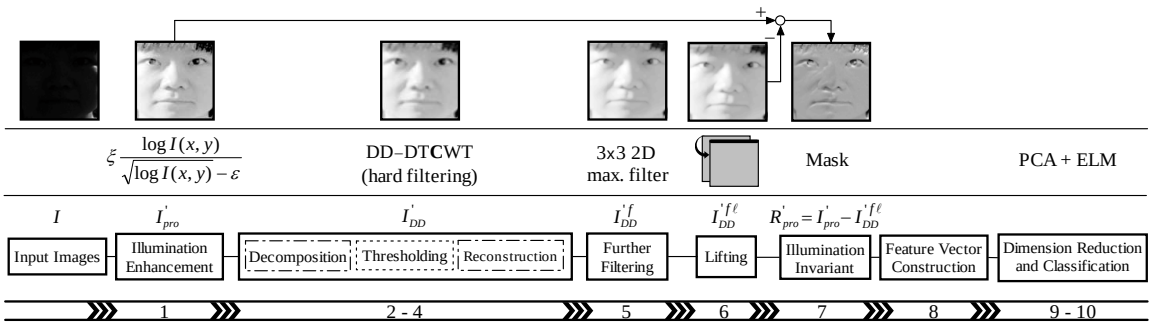
such that $\mathbf{x} = \mathrm{TQ}_1^{-1}(\mathbf{w}_1) + \mathrm{TQ}_2^{-1}(\mathbf{w}_2)$ where TQ stands for the tunable Q-factor wavelet transform, $\mathbf{w}_{i,j}$ for $i = 1, 2$ represent subband j in the transformation, $\mathbf{x}_1 =$

$\mathrm{TQ}_1^{-1}(\mathbf{w}_1)$ and $\mathbf{x}_2 = \mathrm{TQ}_2^{-1}(\mathbf{w}_2)$. It should be noted that, in practice and software coding, the rate of convergence of the algorithm used for optimization is a considerable factor. The discussion above regarding the use of soft-thresholding operator is to describe one possible simple solution. In [79] it has been shown that split augmented Lagrangian shrinkage algorithm converges faster and it can be more effective for resonance based decomposition in terms of software development. The following example visualizes the above formulations. For the given image in Fig. 4.6 (b), the image is first converted to a signal, $\mathbf{x}$, as is shown in Fig. 4.6 (d) and (e). Employing the solver function `dualQ` given in [82] one can solve the optimization problem in (4.17) to obtain the signal components $\mathbf{x}_1$ and $\mathbf{x}_2$, as well as the wavelet coefficients $\mathbf{w}_1$ and $\mathbf{w}_2$, respectively. As it can be seen, the signal is divided into two components. The high Q-factor channel contains mostly oscillatory parts of the signal where as the low Q-factor channel represents the signals that are not remarkably sustained in oscillations. Also, it is clear that both channels may consist of low- and high-frequency components. This is obviously different in comparison with a frequency based signal decomposition in which a given signal is divided based on frequency contents, that is, the low-frequency channel cannot support high-frequency information and vice versa. This is a kind of expertness to decide whether or not a frequency based or a resonance based representation might be suitable in an application.

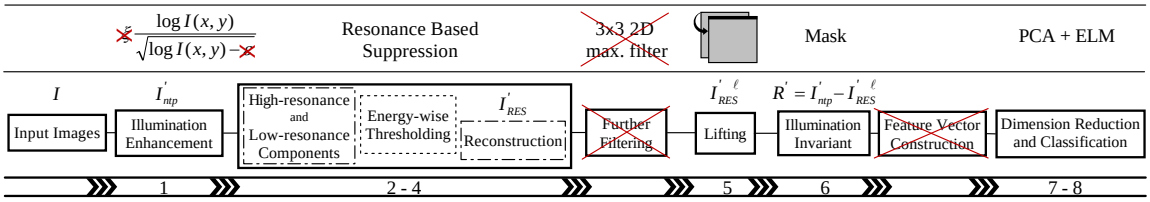
4.5 Illumination Invariant Representation and Concept of Resonance

4.5.1 Enhanced Black Box Approximation

In Section 2.3.1 of Chapter 2 we proposed the use of an enhanced version of log function to approximate the combination of reflectance and illumination. The common assumption in the literature is based on approximating $I(x, y) \propto R(x, y).L(x, y)$ with



(a)



(b)

Figure 4.7 – (a) and (b) are the block diagrams of the proposed methods in Chapter 2 and Chapter 4, respectively.

$I'_{\log}(x,y) = \log(I(x,y)) = \log(R) + \log(L) = R'_{\log}(x,y) + L'_{\log}(x,y)$ where $R(x,y)$ and $L(x,y)$ are the reflectance and illumination of a given image $I(x,y)$, and $I'_{\log}$, $R'_{\log}$ and $L'_{\log}$ stand for the logarithm of the image, reflectance and illumination, respectively. In order to enhance the recognition rates, we have used the following expression that can expand the range of dark pixel values and at the same time can be controlled by tuning parameters ξ and ϵ, that is,

$$I'_{pro}(x,y) = \xi \frac{\log(I(x,y))}{\sqrt{\log(I(x,y))} - \epsilon} \tag{4.18}$$

In [5] we have shown that satisfactory performance can be obtained with less number of parameters to reduce the computational time. Consequently, in this chapter we put steps forward and suggest to remove the tuning parameters if the resonance based representation is employed for illumination suppression and recognition purposes [6][7]. We show that using the proposed resonance based method, as described below, the

necessity of dealing with the time consuming process of finding tuning parameters is ignorable. Therefore, in this chapter we have used the following expression

$$I'_{ntp}(x,y) = \frac{\log(I(x,y))}{\sqrt{\log(I(x,y))}} \tag{4.19}$$

where the notation *ntp* stands for the no-tuning-parameters as the tuning parameters are removed and only the range of dark pixel values has been expanded.

The block diagram of the proposed frequency based method in Chapter 2 has been shown in Fig. 4.7 (a) followed by the resonance based illumination invariant framework in (b) proposed in this chapter. As it is shown in the next sections via results and discussions while the recognition accuracy remains satisfactory for the proposed system in Fig. 4.7 (b), the computational time is significantly reduced. At the same time, the most important note to be contributed is that the tuning parameters ξ and ϵ in (4.18) are entirely removed. It is a very important progress and finding in our study and investigation as the recognition process is now fully automatic.

4.5.2 Illumination Suppression via Energy of Subbands

Although it has been assumed by several authors that the illumination mostly resides in the low-frequency part of images, we would like to put a question mark as if these low-frequency signals need to be low-resonant or high-resonant components [6]. To analyze the effect of resonance on the performance of the face recognition system proposed in Chapter 2 we have studied the energy distribution of subbands of the high- and low-resonance components of the images in databases. Fig. 4.8 (a) shows the energy distribution of the resonance based components for an image from Subset 1 of the Extended-Yale B database. The images in Subset 1 contain low illumination variations and therefore as it can be seen in this figure, the energy distribution over the subbands remains focused on a single subband. In contrast, Fig. 4.8 (b) shows

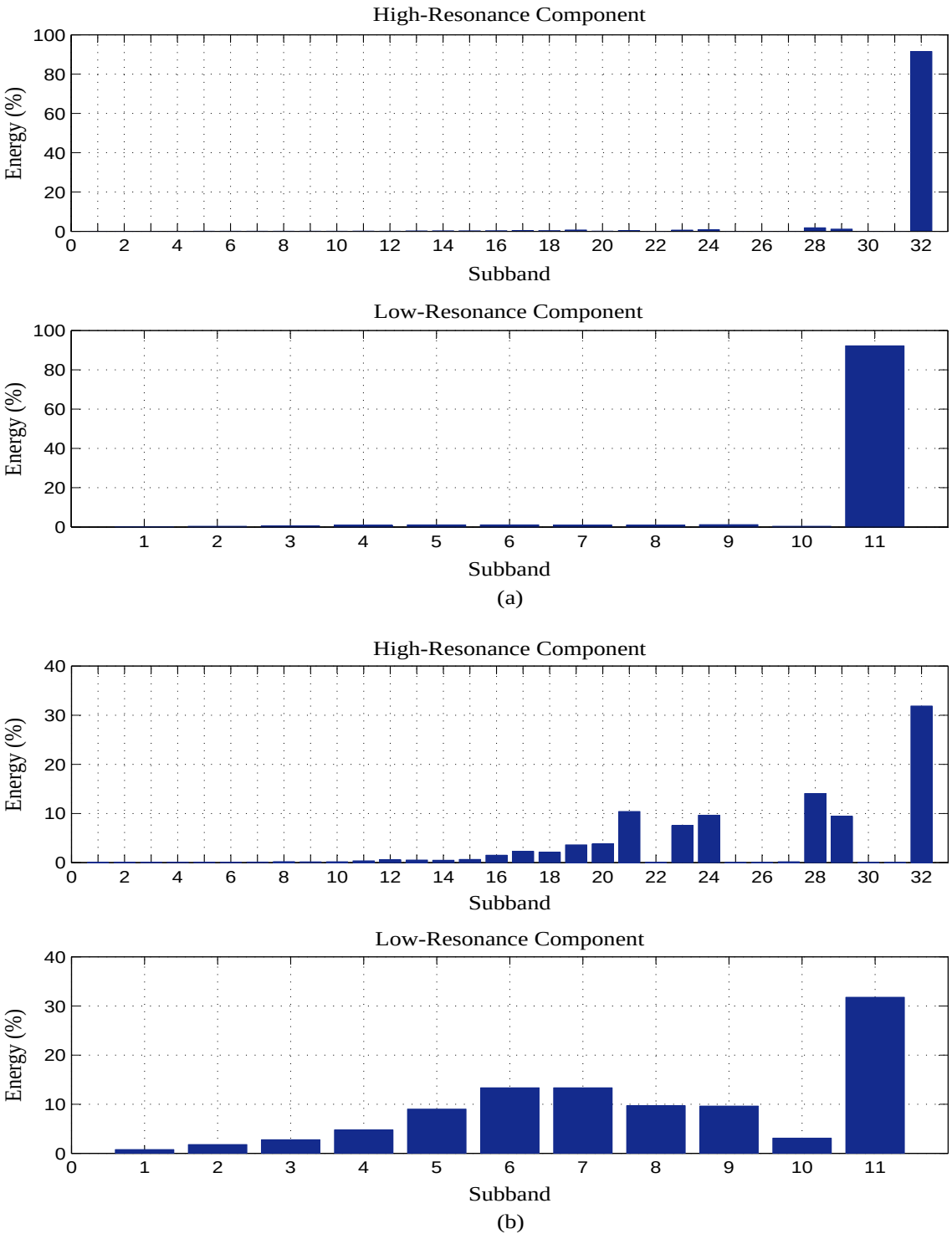


Figure 4.8 – Distribution of energy for sample images from the Extended-Yale B database. (a) An image from Subset 1 with low illumination variation; (b) An image from Subset 5 with high illumination effects (Fig.4.6 (a)). In this figure, $Q_1 = 4$, $r_1 = 3$, $J_1 = 31$ and $Q_2 = 1$, $r_2 = 3$, $J_2 = 10$.

an image from Subset 5 of the respective database. Subset 5 contains images with high variation in illumination and the images are partially or mostly covered by dark portions. To have a better visual perception, we have selected a Q-factor of degree four for the presentations in this figure. It is quite evident that the energy distribution varies considerably among the subbands. This unusual distribution is considered as the effect of unwanted illumination in our work.

The distributions such as the ones in Figs. 4.8 (b) and 4.11 (b) contain remarkable subband energies if compared to the same energy distribution representation for images with almost normal illumination effects as in Figs. 4.8 (a) and 4.11 (a), respectively. It is worth pointing out that as stated earlier, increasing the amount of Q-factor to some extent leads to more narrow frequency responses bands and therefore as it has been illustrated in Fig. 4.11 the energy distribution can be controlled via the Q-factor value. However, this improvement may not be satisfactory to significantly suppress the illumination effects. Figs. 4.9 and 4.10 show signal decompositions with a higher oscillation factor of degree six if compared to the decompositions already presented in Figs. 4.4 and 4.5 with $Q = 4$. To come up with this problem we propose a thresholding strategy to reduce or eliminate the role of unwanted illumination effects based on thresholding the energy of subbands that present a notable change in comparison to the last subband with the highest energy. It should be noted that a hard-thresholding approach would remove all subbands and keeps only the subband with the maximum energy. However, the subbands consisting of small amount of energies may need to remain unchanged as they are most probably the fine features rather than being redundant illumination. Therefore, we define the value of threshold such that if the value of the energy of a subband is less than two percent of the energy of the last subband then it is kept for the reconstruction phase, otherwise the wavelet coefficients associated with the corresponding subband are set to zero before reconstruction. Fig. 4.12 shows the thresholded subbands for the distributions already shown in Fig. 4.11 (b). This strategy has been applied to all images in the training

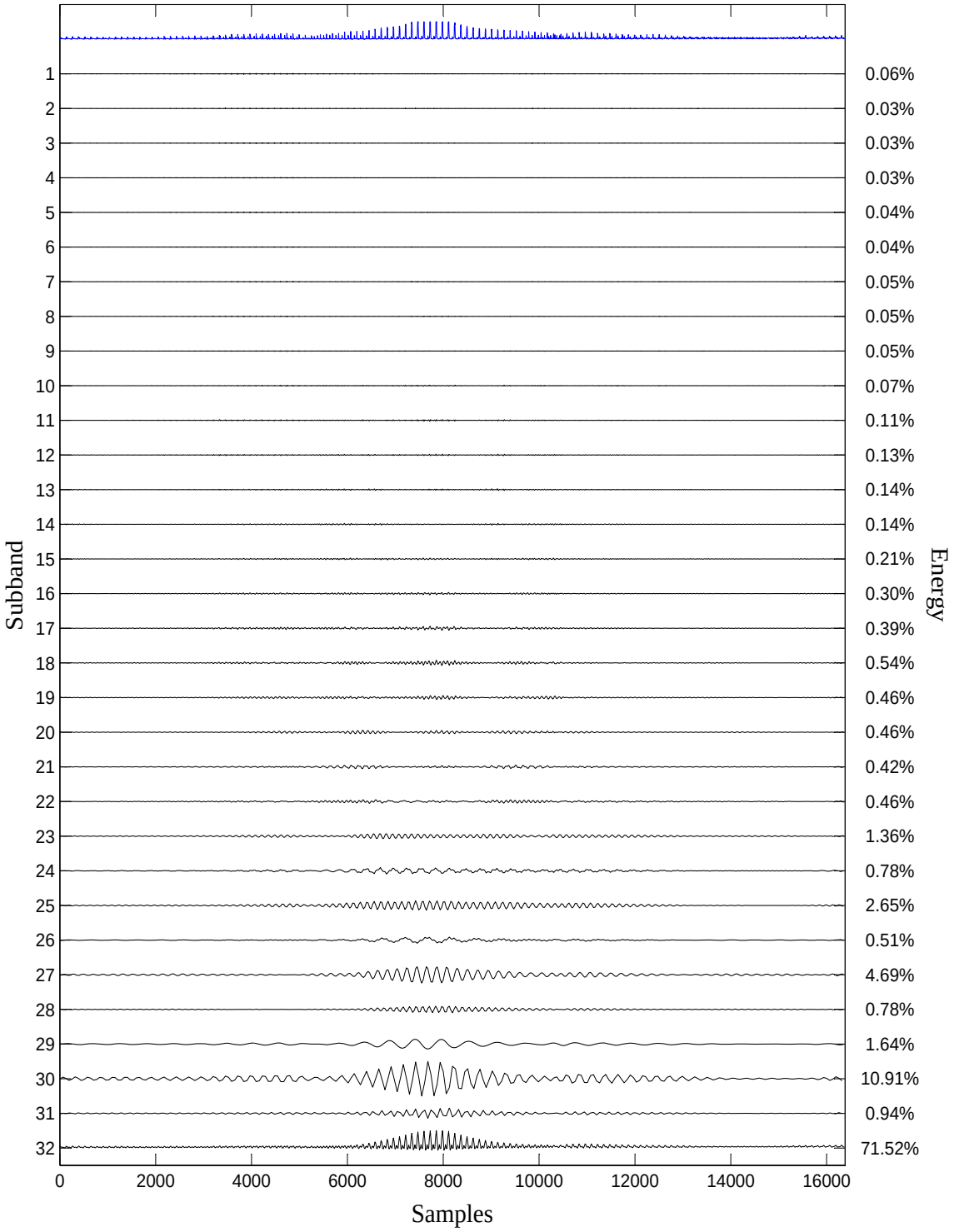


Figure 4.9 – Resonance based signal decomposition to separate the high-resonance components. A sample image is shown as a signal at top. High-resonance components have been presented via $Q = 6$, $r = 3$, $J = 31$.

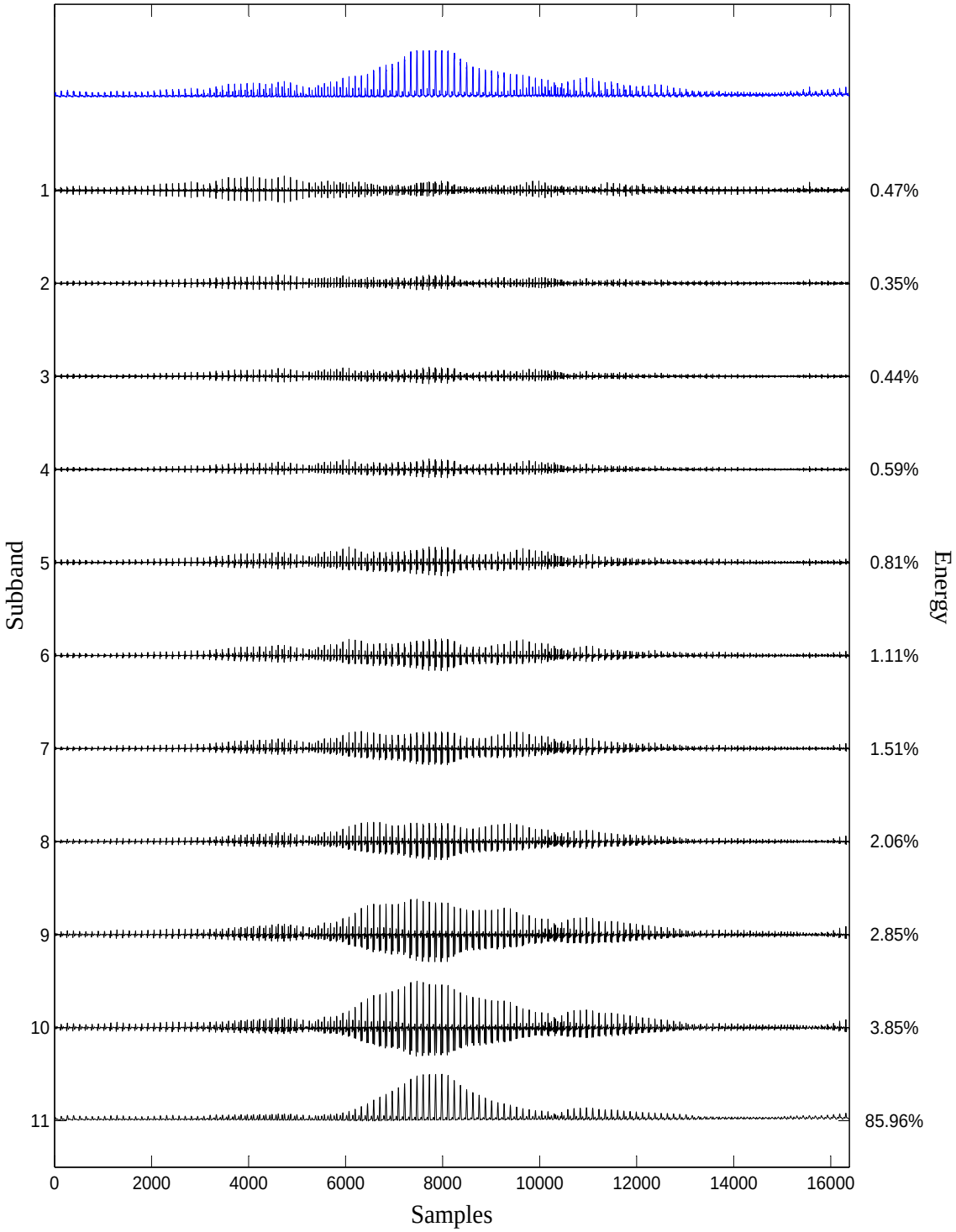


Figure 4.10 – Resonance based signal decomposition to separate the low-resonance components. A sample image is shown as a signal at top. Low-resonance components have been presented via $Q = 1$, $r = 6$, $J = 10$.

and testing galleries of all databases to construct a mask required for recognition. To sum up, the idea is proposed as preserving the illumination invariant information and at the same time suppressing the redundant data which mostly lie in the subbands of the resonance based decomposition.

The block diagram of the proposed method has been shown in Fig. 4.7. Given an image I, illumination is initially enhanced using (4.19) to obtain I'_{ntp}. It is worth pointing out that the proposition is free of tuning parameters and the computational complexity of searching for the optimal range for parameters has been removed from the whole procedure. I'_{ntp} is then decomposed into the high- and low-resonance components as depicted in Fig. 4.6 (d) and (e) for instance. The two components for each image are analyzed in terms of energy of subbands. Noting that the tunable Q-factor wavelet transform satisfies Parseval's theorem, the total energy of the wavelet coefficients equals the energy of the signal. The obtained energy distributions are thresholded based on the fact that for an image with less illumination variation the significant portion of energy resides in the last subband of the transformation. In other words, and in contrast, for an image with higher changes in illumination the energy of the last subband is decreased and instead, visible amount energy is seen for some of the subbands. Therefore, thresholding is performed based on this finding, that is, the value of threshold is set to the two percent of the energy of the last subband to keep or kill a subband. Note that the last subband anyways is kept for reconstruction. This procedure is applied to both the high- and low-resonance components. The thresholded subbands of the two components for each image are reconstructed via an inverse tunable Q-factor transformation to obtain I'_{RES}. Unlike the proposed method in Chapter 2, the maximum filter is removed from the procedure which in turn reduces the computational complexity. Similarly, the lifting stage is carried out on I'_{RES} to amplify the role of edges and fine features to obtain ${I'_{RES}}^{\ell}$ where the notation ℓ denotes the lifting operation. As discussed before, the value of shift depends on the size of images, e.g., for images of size 32×32 to 128×128, the

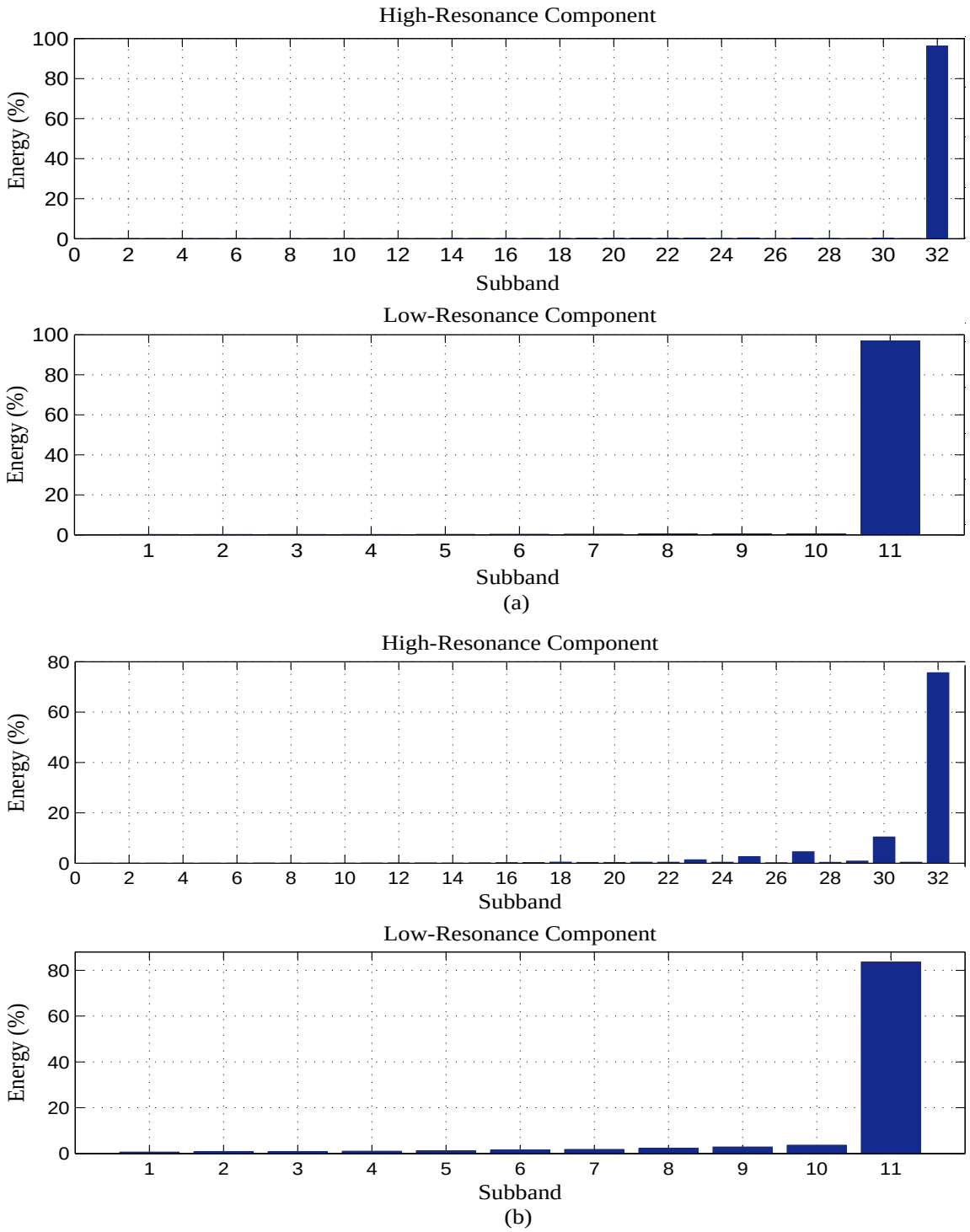


Figure 4.11 – Distribution of energy for via dual Q-factor decomposition with $Q_1 = 6$, $r_1 = 3$, $J_1 = 31$ and $Q_2 = 1$, $r_2 = 6$, $J_2 = 10$. (a) an image from Subset 1 of the Extended-Yale B database with low illumination variation; (b) an image from Subset 5 of the Extended-Yale B database with high illumination effects.

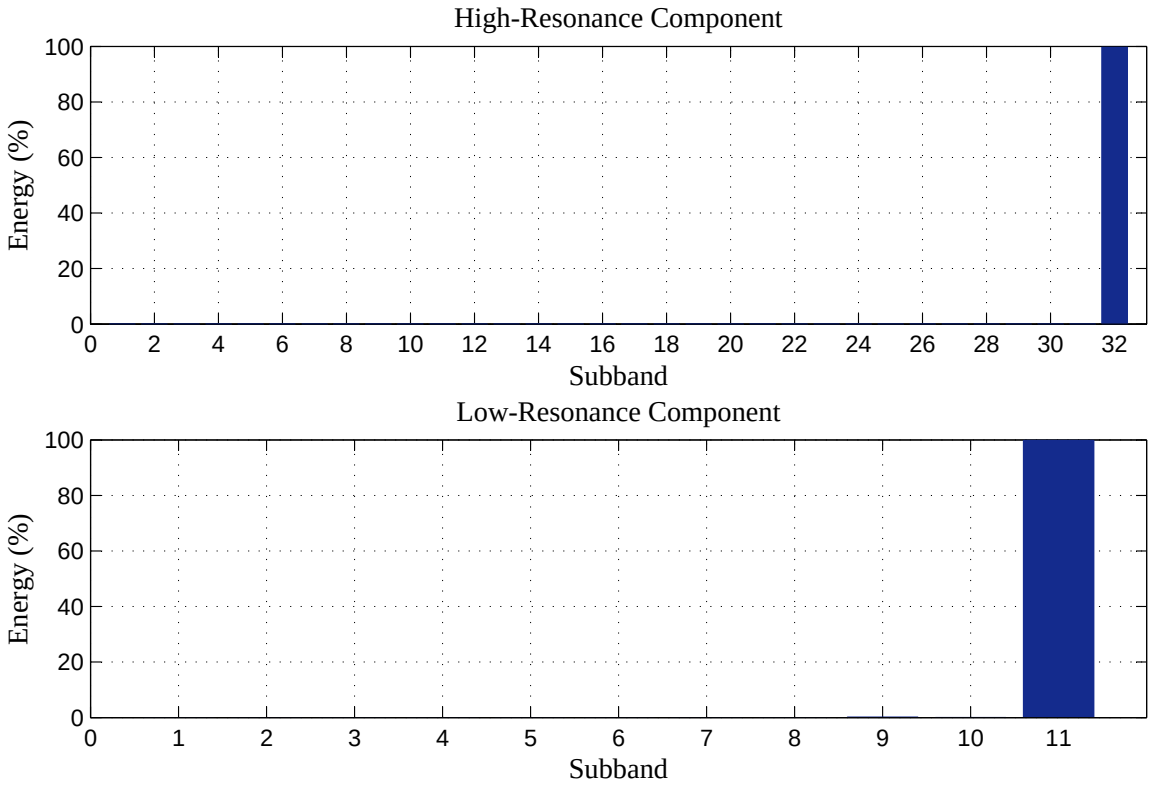


Figure 4.12 – Thresholded distribution of energy for the image shown in Fig. 4.6 (b) (See the distribution of energy in Fig. 4.11 (b) before thresholding) with $Q_1 = 6$, $r_1 = 3$, $J_1 = 31$ and $Q_2 = 1$, $r_2 = 6$, $J_2 = 10$.

effective value of shift is only one to four rows. In order to extract a mask to obtain the illumination invariant image ${I'_{RES}}^{\ell}$ is subtracted from I'_{ntp} to complete the mask extraction procedure. Unlike our approach in Chapter 2 in which R'_{pro} is decomposed into frequency subbands via DD-DT$\mathbb{C}$WT to form a feature vector, we have totally removed this phase as well. The features are only selected based on the use of PCA; the computational complexity is further reduced. In fact, we apply PCA to reduce the dimension of feature space and to make a feature vector. Finally, the extreme learning machine (ELM) is used for classification to evaluate the effectiveness and performance of the proposed resonance based recognition system. ELM has been already discussed in details in Section 2.4.

In the next section we employ the proposed resonance based method in this chap-

ter using the well known databases in the literature used for the problem of the illumination invariant human face recognition. It is shown that the technique can significantly facilitate the redundant illumination suppression task.

4.6 Experiments and Results

Several face image databases are available to study and develop experiments to evaluate the performance of a face recognition method. The databases in the literature are basically divided into different groups in terms of the corresponding application each database can have a remarkable impact. For instance pose, illumination, age, occlusion, blur and several other factors can significantly change the accuracy and performance of a face recognition system. For the problem of the illumination invariant face recognition, which is the main interest in this dissertation, the most famous and well known databases such as the Yale B, Extended-Yale B and the CMU-PIE databases have been used by several pioneer researchers in recent years. Similarly and to keep consistency with previous works we have performed the experiments and simulations using the same databases and settings. During the experiments, the lighting conditions, camera and flash angles, the number and type of images in each subset of each database, and all other experimental setups are exactly the same as the ones in the literature unless otherwise stated. In the following section the performance of the proposed resonance based illumination invariant face recognition system is analyzed employing the Yale B, Extended-Yale B and the CMU-PIE databases.

4.6.1 The Yale B and Extended-Yale B Databases

The two databases, already explained in Section 2.5, are extensively discussed in [105] and [106], respectively. The Yale B database contains 5760 images with nine poses and 64 illumination conditions for each of the ten subjects in the database. The difference between the Yale B and Extended-Yale B is that the former contains

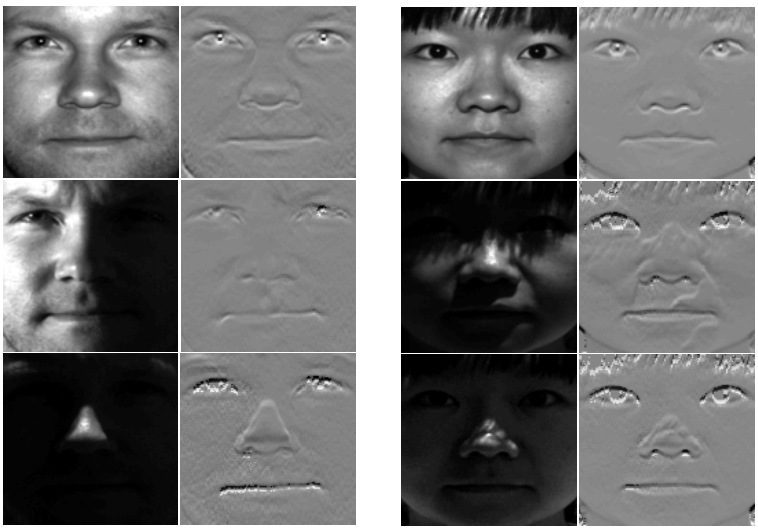


Figure 4.13 – Sample images and the corresponding masks for the Yale B (first column) and the Extended-Yale B (third column) databases respectively using the proposed resonance based method.

ten individuals while the latter includes 38 subjects with some images under high illumination changes. The Extended-Yale B database therefore consists of 21,888. Similar to previous works the same setup and number of images have been allocated to the training and testing galleries in our experiments. Each of the two databases contain five subsets representing the different light source directions and flash; the angle τ between the light source and camera axis. With this configuration for $\{\tau < 12^\circ, 20^\circ < \tau < 25^\circ, 35^\circ < \tau < 50^\circ, 60^\circ < \tau < 77^\circ, \tau > 78^\circ\}$ there are {70, 120, 120, 140, 190} and {266, 456, 456, 532, 722} images in each of the subsets, Subsets 1–5, for the Yale B and Extended-Yale B databases respectively. It yields 64 images per individual in the databases.

All the images of size 192×168 are first eye-aligned, cropped and then resized into 128×128 gray level pixels. In all experiments in this section with the Yale B and the Extended-Yale B databases, $\{Q_1 = 6, r_1 = 3, J_1 = 31\}$ and $\{Q_2 = 1, r_2 = 6, J_2 = 10\}$ are set to represent the signal in terms of the high- and low-resonance components. The proposed method is applied on images to obtain the masks required for the rest of procedure. Similar to previous approaches, Subset 1 is used for training and the

Table 4.1 – Yale B: Recognition rate for different techniques (%) employing the proposed resonance based method.

Method	Subset 2	Subset 3	Subset 4	Subset 5
PP+LTP [92]	100	100	99.20	97.20
MQI [112]	100	100	100	98.40
S–P [30]	100	100	87.70	45.68
S–P+PP [30]	100	100	100	100
II+PCA [66]	100	100	98.60	98.90
II+PCA+DVS [66]	100	100	99.29	99.47
II+PCA+WVS [66]	100	100	100	99.47
Wavelet+PCA [110]	100	100	100	100
Wavelet+PCA [14]	100	100	100	100
Proposed in Chapter 2	100	100	100	100
Proposed in this Chapter	100	100	100	100

Table 4.2 – Extended-Yale B: Recognition rate for different techniques (%) employing the proposed resonance based method.

Method	Subset 2	Subset 3	Subset 4	Subset 5
PCA[11]	90.16	41.23	6.37	3.24
LDA[10]	100	98.12	38.35	5.13
II+PCA+SVM [66]	100	99.78	95.44	94.68
II+PCA+1NN [66]	100	100	96.01	92.44
II+PCA+DVS [66]	100	100	96.60	95.40
II+PCA+WVS [66]	100	100	97.91	96.54
Proposed in Chapter 2	100	100	98.68	99.03
Proposed in this Chapter	100	100	99.21	99.17

remaining subsets are employed during the testing phase. Fig. 4.13 shows two individuals from the Yale B and the Extended-Yale B databases, respectively. In this figure, column one and three show the two subjects under different illumination conditions, respectively. The obtained masks for each individual are presented in columns two and four, respectively. The average of 50 separate execution of simulations are given in Tables 4.1 and 4.2 for the Yale B and the Extended Yale B databases, respectively. It is seen that the proposed method performs reasonably well compared to the approaches in the literature as well as the proposed method in Chapter 2. It

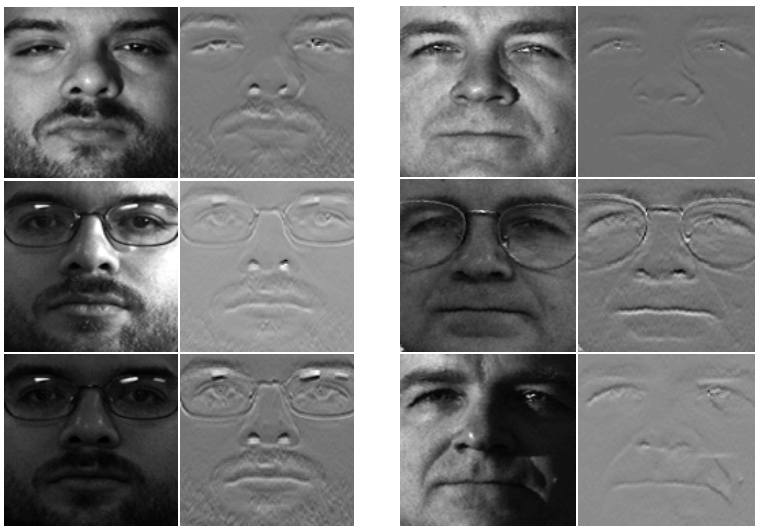


Figure 4.14 – Sample images and the corresponding masks for the CMU-PIE database using the proposed resonance based method. The images have been randomly selected from both the lights-off and lights-on settings.

is important to note that due to eliminating the tuning parameter selection phase, removing the maximum filtering stage as well as the feature vector construction step, the implementation and computational complexity of the proposed method has been greatly improved. It has been discussed numerically in Section 4.7 in terms of the required time to analyze an input image.

4.6.2 Results for the CMU-PIE Database

The Carnegie Mellon University Pose, Illumination and Expression (CMU-PIE) database which consists of 41,368 images from 68 individuals with 13 different poses and 43 illumination conditions already presented in Section 2.5 and has been discussed extensively in details in [2]. Similar to the previous methods in the literature such as [66] frontal images have been used in our experiments. All images in this database are eye-aligned and resized into 96×96 gray level pixels and the proposed algorithm is applied to the images to extract the corresponding mask. For the two lighting conditions, lights-off and lights-on [87], the images {i08, i11, i20} and {I06, I08, I11 I20} of each individual are used for training, respectively. The remaining images are

Table 4.3 – CMU-PIE: Recognition rate for different lighting conditions (%) employing the proposed resonance based method. The second row in each experiment shows the results obtained by using our proposed method in Chapter 2. The third row indicates the results reported in [66].

lighting condition	NTR	NTS	Total	%
off	3	18	21	100
				99.59
				94.85
on	4	20	24	100
				100
				100
on/off	7	38	45	100
				99.92
				99.74
	6	39	45	99.97
				99.77
				99.55
	5	40	45	99.94
				99.70
				98.99

considered for testing. Fig. 4.14 presents some randomly selected individuals from the lights-off and lights-on settings, respectively. Similar representation can be offered for the lights-off/on setting. The result are given in Table 4.3 and it is seen that the proposed method can significantly improve the recognition accuracy compared to one of the recent approaches [66] in the literature as well as our proposed method in [3].

4.7 Robustness and Training Sample Reduction

Similar to the experiments we carried out in Chapter 2 to evaluate the robustness of our proposed method against the reduction in number of training images we have investigated the performance of our proposed resonance based approach under the condition in which the number of training samples are decreased to one image only. In general, reduction in the number of training samples is directly related to the

reduction in the probability of obtaining desired features which can negatively affect the accuracy of a recognition system and consequently can be considered as a potential problem where there are only few training images. In order to analyze this issue employing the resonance based method and to compare the results with the ones obtained in Chapter 2, we have reduced the number of images in the training gallery for the Yale B, Extended-Yale B and the CMU-PIE databases to reach NTR=1 where NTR denotes the number of training samples and TIN indicates the training image name. Obviously, the remaining images in each experimental setting are transferred to the testing gallery. Table 4.4 shows the results for the Yale B database. As it can be seen, and if compared to the previous approaches already shown in Table 4.1, both the frequency based and resonance based methods proposed in this dissertation not only significantly improve the recognition results in terms of accuracy but also persist against any reduction in the number of training images. Table 4.5 illustrates the results obtained via the Extended-Yale B database. While the Extended-Yale B database consists of some images that almost fully dark, it is difficult to reach high recognition rates on this database. However, as it can be seen, the proposed resonance based method can further improve the results compared to our frequency based method in Chapter 2 as well as the results reported in previous papers given in Table 4.2.

The experiments with the CMU-PIE database have been given in Table 4.6. For the lights-on setting and as it was expected, the perfect accuracy is achieved even if a single image is used for training. The experiments under the lights-off and lights-on/off conditions show that the results obtained by using the proposed resonance based method compare favorably to the frequency based approach in Chapter 2. Overall, it is concluded that the recognition rates are improved by using the resonance based method. Furthermore, the method presents a robust behavior for any reduction in the number of training samples. A detailed discussion about particular examples to elaborate the robustness of the proposed frequency based method in Chapter 2 has

Table 4.4 – Recognition rate (%) with reduction in the number of training samples for the Yale B database. The second row in each experiment shows the results obtained by using our proposed method in Chapter 2.

NTR	Subset 2	Subset 3	Subset 4	Subset 5	TIN
7	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10, +0E+0, +5E+10, +5E-10
	100	100	100	100	
6	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10, +0E+0, +5E+10
	100	100	100	100	
5	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10, +0E+0
	100	100	100	100	
4	100	100	100	100	-10E+0, -5E-10, +10E+0, -5E+10
	100	100	100	100	
3	100	100	100	100	-10E+0, -5E-10, +10E+0
	100	100	100	100	
2	100	100	100	100	-10E+0, -5E-10
	100	100	100	100	
1	100	100	99.29	99.48	-10E+0
	100	100	99.29	100	

Table 4.5 – Recognition rate (%) with reduction in the number of training samples for the Extended-Yale B database. The second row in each experiment shows the results obtained by using our proposed method in Chapter 2.

NTR	Subset 2	Subset 3	Subset 4	Subset 5	TIN
7	100	100	99.21	99.17	-5E-10, +0E+0, -5E+10, -10E+0, +5E-10, +5E+10, +10E+0
	100	100	98.68	99.03	
6	100	100	99.28	99.00	-5E-10, +0E+0, -5E+10, -10E+0, +5E-10, +5E+10
	100	100	98.50	98.89	
5	100	100	99.24	98.45	-5E-10, +0E+0, -5E+10, -10E+0, +5E-10
	100	100	98.31	98.75	
4	100	100	99.32	98.47	-5E-10, +0E+0, -5E+10, -10E+0
	100	100	98.12	98.61	
3	100	99.78	98.72	98.12	-5E-10, +0E+0, -10E+0
	100	99.78	97.74	98.47	
2	97.50	99.82	95.48	97.76	-5E-10, -10E+0
	100	99.56	96.80	98.20	
1	97.63	98.16	93.97	97.48	-5E-10
	98.27	98.90	95.87	97.37	

Table 4.6 – Recognition rate (%) with reduction in the number of training samples for the CMU-PIE database. The second row in each experiment shows the results obtained by using our proposed method in Chapter 2.

lighting condition	NTR	NTS	Total	%	TIN
lights-off	4	17	21	100	i8, i9, i11, i20
				100	
	3	18	21	100	i8, i11, i20
				99.59	
	2	19	21	100	i11, i20
				99.42	
	1	20	21	99.60	i11
				98.97	
lights-on	4	20	24	100	I8, I9, I11, I20
				100	
	3	21	24	100	I8, I11, I20
				100	
	2	22	24	100	I8, I11
				100	
	1	23	24	100	I8
				100	
lights-on/off	7	38	45	100	i8, i11, i20, I8, I9, I11, I20
				99.92	
	6	39	45	99.97	i8, i11, i20, I8, I11, I20
				99.77	
	5	40	45	99.94	i8, i11, I8, I11, I20
				99.70	
	4	41	45	99.96	i8, i11, I8, I11
				99.64	
	3	42	45	99.74	i8, I8, I11
				99.02	
	2	43	45	99.25	i8, I11
				98.70	
	1	44	45	84.26	I20
				84.59	

been given in Section 2.6. As the resonance based method compares reasonably well with the frequency based technique in general, the same discussion in Section 2.6 is also valid here. Finally, it should be noted that to the best of our knowledge none of the techniques in the literature, including our proposed methods, are fully invariant

against any reduction in the number of training samples to reach the perfect accuracy. Taking the trade-off between complexity of design, recognition rate and number of training samples into account, one may investigate the role of various classifiers to further improve the robustness of a respective technique.

In Chapter 2 we indicated that the computational time required to process an incoming image of size 128×128 pixels to obtain a mask was 58 milliseconds (ms/image) using a Core 2 Duo 2.40 GHz CPU and 2GB RAM. Although single Q-factor based decomposition of a signal into low- or high-resonance components is so fast, in order to decompose an image via a dual Q-factor representation, one needs to solve an iterative optimization problem to obtain the low- and high-resonance components simultaneously. This procedure is not that fast and takes 0.87 seconds for an input image of size image 128×128. While it may seem to be slow at the first glance the key point here is that, unlike the proposition in Chapter 2, and as extensively discussed in in this chapter we have removed the two tuning parameters ξ and ϵ from the design procedure for the resonance based method (See (4.18) and (4.19) in Sections 4.5.1 and 2.3.1). It is very important that although it may seem that the technique in Chapter 2 is faster, finding the optimal values for the two parameters is not an automatic process. The resonance based method is in fact one step further to reach a fully automated high performance recognition system.

4.8 Conclusions and Summary

This chapter proposes a novel multiresolution based method for the problem of illumination invariant face recognition that relies on illumination suppression via the concept of resonance. This is the first time in the literature that the concept of resonance is introduced to the community of face recognition. It is assumed that an image can be represented as a combination of illumination and reflectance. Unlike our proposed frequency based model in Chapter 2, the resonance based technique does

not need tuning parameters to control the image representation and enhancement step. Furthermore, the approach is free of using extra filtering and fine-tuning phase, that is, the maximum filter has been removed from the procedure. Additionally, the feature vector construction introduced and used in Chapter 2 has been removed as well. Because this is the first attempt to employ the concept of resonance in this field, indeed there is room for further development and design in several aspects. Our analysis and investigation show that the energy distributions of subbands via resonance decomposition are different for an image with normal illumination effects and an image with high illumination variation. This key point has been the main philosophy behind our proposed method in this chapter. Being more specific, we have thresholded the energy of subbands of a dual Q-factor based decomposition to first deactivate the subbands with unwanted energy distribution created by illumination effects. The transformation facilitates redundant illumination effect removal task by thresholding unwanted components. For dimensionality reduction and classification, and to keep consistency with experiments in Chapters 2 and 3, the principal component analysis and the extreme learning machine have been used, respectively. Experiments and comparison with the recent best results in the literature as well as the findings in previous chapters suggest that the proposed resonance based method significantly outperforms the current available techniques for the problem of illumination invariant face recognition.

Chapter 5

Conclusions

5.1 Summary

In reality and especially in implementation of real world applications, the cost of design is the most important factor whereas the difference between 99% and 100% recognition rates is negligible in most commercial applications. In other words, computational complexity plays an important role and deserves more attention from the community. As already stated through chapters, we believe that face recognition systems in recent years are well developed in terms of required accuracy for most of the current databases available for research. In our opinion, further investigation would be more valuable if new and complex databases could be generated by the pioneers, and in the case of working on current available databases, the priority of a technique should reflect the simplicity and low-complexity of the design rather than a minor and negligible improvement in recognition accuracy only.

In Chapter 2, we proposed a novel method to deal with the problem of illumination invariant face recognition. The proposed method is based on enhancing the lighting effects of images first and then suppressing the illumination part via multiresolution decomposition and filtering of input images. It can enhance the face image masks and facilitates the illumination invariant face recognition task. It can be achieved in view of the unique properties of the double-density dual-tree complex wavelet transform

and especially due to increased density of the wavelets in this transformation. Several experiments have been done on available well documented databases used in illumination invariant face recognition, as well as the databases used in pose invariant and facial expression based recognition techniques. The results indicate that the proposed method outperforms the previously reported results both in terms of the recognition rate and computational complexity, and thus performs reasonably well compared to the current best results in the literature. Unlike most of the recent techniques, the proposed method is systematic for softcomputing simulations and implementations, significantly faster, more accurate even if trained with fewer images, and not dependent of any prior information about the face shape or illumination. Also, the method does not need a complicated feature extraction or dimensionality reduction process, and a PCA based feature dimensionality reduction phase can outperform the recent results in the literature.

The proposed method in Chapter 2 is further investigated in Chapter 3 to see whether or not frequency-selectivity of the subbands of a multiresolution transformation may affect the performance of a frequency domain based face recognition system. This has been the core topic in Chapter 3 where we found that internal characteristics of a multiresolution transformation such as the transition band width and magnitude of ripples of the frequency responses of the filters involved in the structure of the transformation can considerably change the recognition accuracy. To this end, first it is required to have reliable control over the frequency specifications of the filters. Several filters with various characteristics have been designed using the proposed formulation and design in Chapter 3. The filters are substituted in the structure of a two channel filter bank, and performance of each filter pair is evaluated under the illumination invariant recognition framework proposed in Chapter 2. We found that there is a compromise between the characteristics of filters of a multiresolution transformation and the recognition accuracy. These characteristics mainly include sharpness of the frequency response, amplitude of ripples, filter length and smoothness of the

frequency response at vanishing points. Generally, for a fixed filter length, frequency selectivity has an important impact on multiresolution based imaging. Also, shorter transition band with relatively small ripples can offer higher accuracy. The trade-off between the accuracy and frequency selectivity of the subbands of the transformation has been discussed extensively in 3.5.2.

Chapter 4 presents a challenging idea in which the concept of resonance has been admitted to the face recognition area. As it was stated earlier in the introduction, the common assumption regarding the low-frequency nature of illumination led us into doubt to think about if these low-frequency contents belong to high- or low-resonance components. While any image can be considered as a mixture of illumination and reflectance, we found that unlike our proposed frequency based model in Chapter 2 there is no need to extra tuning parameters to represent the combination. Additionally, feature selection phase of the system can be simplified and limited to the principal component analysis and the maximum filtering step can be entirely removed. Despite the frequency based subband thresholding approach used in Chapter 2 and Chapter 3, the resonant components have been thresholded in terms of the energy of high- and low-resonance components. This strategy is proposed in view of remarkable energy distribution differences between a good quality image and a too dark or too bright image under variety of illumination effects. The results show that the proposed resonance based approach performs better than the frequency based methods proposed and analyzed in Chapter 2 and Chapter 3. It should be noted that our frequency based method already outperformed the current best results in the literature [3].

5.2 Future Research

One of the main goals of this research is to answer a question we first proposed it in Chapter 2, that is, to initiate the concept of embedded illumination invariant face recognition in which it is expected that the whole procedure could be replaced with a system-level design approach. In other words, the idea is basically to design and develop a method in which the output of the system in Chapter 2 or Chapter 4 is the input to an unknown system defined by a z-domain multitask function. The aim here is to determine the polynomial coefficients for which the two systems, i.e., the separate algorithm and the embedded one, share almost the same output. It means that, in terms of recognition accuracy the two systems are almost identical, where the complexity of design and processing has been significantly reduced. We believe that it can be done if an illumination index or metric could be defined first. Then the metric may be characterized using a similar approach as in Chapter 3. It should be noted that such a metric can be formulated and controlled either by energy based thresholding of resonance components or frequency selectivity of subbands of a multiresolution transformation. If the embedded design could be implemented, attribute and nature of features could be tailored application-wise that can lead to the design of specific classifiers in complex applications such as space imaging and biomedical instrumentation.

Another step in direction of this research will be finding an analytical model to represent the reflectance and illumination part of an image separately. One should note that our propositions in Chapter 2 and Chapter 4 enhance the illumination effects and at the same time provide more flexibility to tune the approximation per image. However, it is worth pointing out that up to now and to the best of our knowledge there has been no direct and exact model to elaborate reflectance or illumination individually. Therefore, in view of the findings in this research it is hoped to go one step further and investigate such challenging issues in the future.

Further investigation may include robustness analysis using few numbers of training images. It would be an impressive case if one could reach the same accuracy with competitively low computation time using only one sample image for each subject within the training cycle. In addition to resistivity for lower training samples, sensitivity analysis and robustness in presence of noise and resolution changes of facial image is another interesting topic needs to be studied in details.

References

[1] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, IEEE Trans. Pattern Analys. Mach. Intell., 19 (7) (1997) 721–732.

[2] S. Baker, T. Sim, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. Pattern Analys. Mach. Intell., 25 (12) (2003) 1615–1618.

[3] A. Baradarani, Q. M. Jonathan Wu, M. Ahmadi, An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTℂWT filtering, Pattern Recognition, 46 (1) (2013) 57–72.

[4] A. Baradarani, Q. M. Jonathan Wu, M. Ahmadi, P. Mendapara, Tunable halfband-pair wavelet filter banks and application to multifocus image fusion, Pattern Recognition, 45 (2) (2012) 657–671.

[5] A. Baradarani, Q. M. Jonathan Wu, M. Ahmadi, Illumination suppression for illumination invariant face recognition, Accepted.

[6] A. Baradarani, Q. M. Jonathan Wu, Design of an efficient illumination invariant face recognition system via resonance based image representation, Submitted.

[7] A. Baradarani, Q. M. Jonathan Wu, Illumination invariant human face recognition: frequency or resonance?, Accepted.

[8] R. Basri, D. W. Jacobs, Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Analys. Mach. Intell., 25 (2) (2003) 218–233.

[9] I. Bayram, I. W. Selesnick, Frequency-domain design of overcomplete rational-dilation wavelet transforms, IEEE Trans. Signal Processing, 57 (8) (2009) 2957–2972.

[10] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces versus fisherfaces: Recognition using class specific linear projection, IEEE Trans. Pattern Analys. Mach. Intell., 19 (7) (1997) 711–720.

[11] P. N. Belhumeur and D. J. Kriegman, What is the set of images of an object under all possible lighting conditions? Proc. IEEE Intl. Conf. Comp. Vis. Pattern Rec. (CVPR96), San Francisco, CA, (1996) 270–277.

[12] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, 2008.

[13] M. Bresch, Optimizing filter banks for supervised texture recognition, Pattern Recognition, 35 (2002) 783–790.

[14] X. Cao, W. Shen, L. G. Yu, Y. L. Wang, J. Y. Yang, Z. W. Zhang, Illumination invariant extraction for face recognition using neighboring wavelet coefficients, Pattern Recognition, 45 (4) (2012) 1299–1305.

[15] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, Jour. SIAM Rev., 43 (1) (2001) 129–159.

[16] J. L. Chen, A. KunduRotation, Rotation and gray scale transform invariant texture identification using wavelet decomposition and hidden Markov model, IEEE Trans. Pattern Analys. Mach. Intell., 16 (2) (1994) 208–214.

[17] T. Chen, X. S. Zhou, D. Comaniciu, T. S. Huang, Total variation models for variable lighting face recognition, IEEE Trans. Pattern Analys. Mach. Intell., 28 (9) (2006) 1519–1524.

[18] W. Chen, M. J. Er, S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithmic domain, IEEE Trans. Syst., Man, and Cybernetics - Part B, 36 (2) (2006) 458–466.

[19] K.-O. Cheng., N.-F. Law, W.-C. Siu, Multiscale directional filter bank with applications to structured and random texture retrieval, Pattern Recognition, 40 (2007) 1182–1194.

[20] T. S. Choi, M. Asif, J. Yun, Three-dimensional shape recovery from focused image surfaces, IEEE Int. Conf. Accoust., Speech Signal Processing, 6 (1999) 3269–3272.

[21] R. R. Coifman, D. L. Donoho, Translation-invariant de-noising, in wavelets and statistics, New York: Springer-Verlag, 1995.

[22] H. Çağlar and A. N. Akansu, A generalized parametric PR-QMF design technique based on Bernstein polynomial approximation, IEEE Trans. on Signal Processing, 41 (7) (1993) 2314–2321.

[23] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

[24] I. Daubechies, M. Defriese, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun., Pure Appl. Math, (LVII) (2004) 1413–1457.

[25] T. N. Davidson, Z. Q. Luo, J. F. Sturm, Linear matrix inequality formulation of spectral mask constraints with applications to FIR filter design, IEEE Trans. on Signal Processing, 50 (11) (2002) 2702–2715.

[26] D. L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inform. Theory 47 (7) (2001) 2845–2862.

[27] D. L. Donoho, M. Elad, Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization, Proc. Natl. Acad. Sci. 100 (2003) 2197–2202.

[28] B. Dumitrescu, SDP approximation of a fractional delay and the design of dual-tree complex wavelet transform, IEEE Trans. on Signal Processing, 56 (9) (2008) 4255–4262.

[29] B. Dumitrescu, C. Popeea, Accurate computation of compaction filters with high regularity, IEEE Signal Processing Lett., 9 (9) (2002) 278–281.

[30] M. El Aroussi, M. El Hassouni, S. Ghouzali, M. Rziza, D. Aboutajdine, Local appearance based face recognition method using block based steerable pyramid transform, Signal Processing, 91 (1) (2011) 38–50.

[31] M. Elad, A. M. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. Inform. Theory, 48 (9) (2002) 2558–2567.

[32] M. Figueiredo, J. Bioucas-Dias, R. Nowak, Majorization-minimization algorithms for wavelet-based image restoration, IEEE Trans. Image Processing, 16 (12) (2007) 2980–2991.

[33] R. A. Fisher, The statistical utilization of multiple measurements, Annals of Eugenics, 8 (1938) 376–386.

[34] K. Fukunaga, *Introduction to statistical pattern recognition*, Second Edt., Academic Press, 1990.

[35] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, Trans. Pattern Analys. Mach. Intell., 23 (6) (2001) 643–660.

[36] G. A. Papakostas, Y. S. Boutalis, D. A. Karras, B. G. Mertzios, A new class of Zernike moments for computer vision applications, Information Sciences, 177 (13) (2007) 2802–2819.

[37] B. K. P. Horn, *Robot Vision*, MIT Press, MA, 1986.

[38] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing, 70 (1-3) (2006) 489–501.

[39] G.-B. Huang, X. Ding, H. Zhou, Optimization method based extreme learning machine for classification, Neurocomputing, 74 (1-3) (2010) 155–163.

[40] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst., Man, and Cybernetics - Part B, 45 (2) (2012) 513–529.

[41] G.-B. Huang, D. H. Wang, Y. Lan, Extreme learning machines: a survey, Int. Jour. Machine Learning and Cybernetics, 2 (2) (2011) 107–122.

[42] J.-C. Huang, T.-S. Su, L.-J. Wang, W.-S. Hsieh, Double-change-detection method for wavelet-based moving-object segmentation, Electronics Letters, 40 (13) (2004) 798–799.

[43] G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments, University of Massachusetts, Amherst, Technical Report, (2007) 07–49. Available online at `http://vis-www.cs.umass.edu/lfw/`.

[44] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1989.

[45] E. Jones, P. Runkle, N. Dasgupta, L. Couchman, L. Carin, Genetic algorithm wavelet design for signal classification, IEEE Trans. on Pattern Analys. Machine Intell., 23 (8) (2001) 890–895.

[46] V. Kecman, *Learning and soft computing*, MIT Press, MA, 2001.

[47] H. H. Kha, H. D. Tuan, T. Q. Nguyen, Optimal design of FIR triplet halfband filter bank and application in image coding, IEEE Trans. on Image Processing, 20 (2) (2011) 586–591.

[48] S. Lakshmanan, A. K. Jain, Y. Zhong, Multiresolution image representation using Markov random fields, IEEE Int. Conf. on Image Processing, ICIP, (1) (1994) 855–860.

[49] S. Lakshmanan, P. Watta, Yu Lin Hou, N. Gandhi, Comparison between Eigenfaces and Fisherfaces for estimating driver pose, IEEE Int. Conf. Intell. Transpo. Syst., Oakland, California, (2001) 889–894.

[50] E. H. Land, The Retinex, American Scientist, 52 (2) (1964) 247–264.

[51] E. H. Land, J. J. McCann, Lightness and retinex theory, Jour. Optical Society of America, 61 (1) (1971) 1–11.

[52] Khoa N. Le, K. P. Dabke and G. K. Egan, Hyperbolic wavelet family, Review of Scientific Instruments, 75 (11) (2004) 4678–4693.

[53] Khoa N. Le, A mathematical approach to edge detection in hyperbolic-distributed and Gaussian-distributed pixel-intensity images using hyperbolic and Gaussian masks, Digital Signal Processing, 21 (1) (2011) 162–181.

[54] Khoa N. Le, K. P. Dabke, G. K. Egan, On mathematical derivations of auto-term functions and signal-to-noise ratios of Choi-Williams, first- and nth-order hyperbolic kernels, Digital Signal Processing, 16 (1) (2006) 84–104.

[55] K.-C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Analys. Mach. Intell., 27 (5) (2005) 684–698.

[56] K.-C. Lee, J. Ho, D. Kriegman, Nine points of light: acquiring subspaces for face recognition under variable lighting, IEEE Intl. Conf. Comp. Vis. Pattern Rec. (CVPR01), Kauai, HI, 1 (2001) 519–526.

[57] H. Li, B. S. Manjunath, S. K. Mitra, Multi-sensor image fusion using the wavelet transform, Proc. IEEE Int. Conf. on Image Processing, ICIP94, 1 (1994) Austin, TX, 51–55.

[58] J. Löfbrerg, YALMIP: a toolbox for modeling and optimization in MATLAB, IEEE Int. Symp. Computer Aided Control System Design, Taipei, 284–289 (2004). Available online at
`http://control.ee.ethz.ch/~joloef/yalmip.php`.

[59] S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[60] S. G. Mallat, A theory of multiresolution signal decomposition: the wavelet representation, IEEE Trans. on Pattern Analys. Machine Intell., 11 (7) (1989) 674–693.

[61] T. Mandal, Q. M. Jonathan Wu, Y. Yuan, Curvelet based face recognition via dimension reduction, Signal Processing, 89 (12) (2009) 2345–2353.

[62] P. Mendapara, A. Baradarani, Q. M. Jonathan Wu, An efficient depth map estimation technique using complex wavelets, IEEE Int. Conf. on Multimedia and Expo, ICME10, (2010) 1409–1414.

[63] S. K. Nayyar, Y. Nakagawa, Shape from Focus, IEEE Trans. on Pattern Analys. Machine Intell., 16 (1994) 824–831.

[64] Y. Nesterov, *High Performance Optimization*, Chapter 17, Squared functional systems and optimization problems, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[65] N. G. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, Applied and Computational Harmonic Analys., 10 (3) (2001) 234–253.

[66] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi, Illumination invariant feature extraction and mutual information based local matching for face recognition under illumination variation and occlusion, Pattern Recognition, 44 (10-11) (2011) 2576–2587.

[67] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1) (1996) 51–59.

[68] G. Pajares, J. Cruz, A wavelet-based image fusion tutorial, Pattern Recognition, 37 (9) (2004) 1855–1872.

[69] P. A. Papachristodoulou, P. Seiler, P. A. Parrilo, SOSTOOLS: sum of squares optimization toolbox for MATLAB, (2006). Available online at `http://mit.edu/~parrilo/sostools/index.html`.

[70] P. A. Parrilo, Semidefinite programming relaxations for semialgebraic problems, Jour. Mathematical Prog. Ser. B, 96 (2) (2003) 293–320.

[71] B. D. Patil, P. G. Patwardhan, V. M. Gadre, On the design of FIR wavelet filter banks using factorization of a halfband polynomial, IEEE Signal Processing Lett., 15 (9) (2008) 485–488.

[72] P. J. Phillips, H. Moon, P. J. Rauss, S. Rizvi, The FERET evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Analys. Mach. Intell., 22 (10) (2000) 1090–1104 .

[73] S. M. Phoong, C. W. Kim, P. P. Vaidyanathan, R. Ansari, A new class of two-channel biorthogonal filter banks and wavelet bases, IEEE Trans. on Signal Processing, 43 (3) (1995) 649–655.

[74] V. Powers and B. Reznick, Polynomials that are positive on an interval, Trans. Amer. Math. Soc., 352 (10) (2000) 4677–4692.

[75] V. Powers and T. Wörmann. An algorithm for sums of squares of real polynomials, Journal of Pure and Applied Algebra, 127 (1998) 99–104.

[76] O. Rioul and P. Duhamel, A Remez exchange algorithm for orthonormal wavelets, IEEE Trans. Circuits Syst. II: Analog and Digital Signal Processing, 41 (8) (1994) 550–560.

[77] I. W. Selesnick, The double-density dual-tree DWT, IEEE Trans. Signal Processing, 52 (5) (2004) 1304–1314.

[78] I. W. Selesnick, R. G. Baraniuk, N. G. Kingsbury, The dual-tree complex wavelet transform – a coherent framework for multiscale signal and image processing, IEEE Signal Processing Magazine, 22 (6) (2005) 123–151.

[79] I. W. Selesnick, Resonance-based signal decomposition: A new sparsity-enabled signal analysis method, Signal Processing, 91 (12) (2011) 2793–2809.

[80] I. W. Selesnick, Wavelet transform with tunable Q-factor, IEEE Trans. Signal Processing, 59 (8) (2011) 3560–3575.

[81] I. W. Selesnick, I. Bayram, Oscillatory + transient signal decomposition using overcomplete rational-dilation wavelet transforms, Proc. SPIE, 7446 (Wavelets XIII) (2009).

[82] I. W. Selesnick, Software, Available online at `http://eeweb.poly.edu/iselesni/software.index.html`.

[83] I. W. Selesnick, A new sparsity-enabled signal separation method based on signal resonance, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), (2010) 4150–4153.

[84] G. Shakhnarovich, T. Darrell, P. Indyk, *Nearest-neighbor methods in learning and vision: Theory and practice*, MIT Press, MA, 2006.

[85] S. Shan, W. Gao, B. Cao, D. Zhao, Illumination normalization for robust face recognition against varying lighting conditions, IEEE Intl. Workshop Analys. Modeling Faces Gestures (AMFG03), (2003) 157–164.

[86] A. Shashua, T. R. Raviv, The quotient image: class-based re-rendering and recognition with varying illuminations, IEEE Trans. Pattern Analys. Mach. Intell., 23 (2) (2001) 129–139.

[87] J. R. D. Solar and J. Quinteros, Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different preprocessing approaches, Pattern Recognition Lett., 29 (14) (2008) 1966–1979.

[88] J.-L. Starck, M. Elad, D. L. Donoho, Image decomposition via the combination of sparse representation and a variational approach, IEEE Trans. Image Processing, 14 (10) (2005) 1570–1582.

[89] G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge, 1996.

[90] J. F. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optimization Methods Software, (11-12) (1999) 625–653. Available online at `http://sedumi.ie.lehigh.edu`.

[91] Y. Taigman, L. Wolf, T. Hassner, Multiple one-shots for utilizing class label information, The British Machine Vision Conference (BMVC), (2009) 1–12.

[92] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Intl. Workshop Analys. Modeling Faces Gestures (AMFG07), (2007) 168–182.

[93] D. Tao, X. Li, X. Wu, S. J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, IEEE Trans. Pattern Analys. Mach. Intell., 29 (10) (2007) 1700–1715.

[94] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, S.J. Maybank, Bayesian tensor approach for 3D face modeling, IEEE Trans. Circuits and Syst. Video Technol., 18 (10) (2008) 1397–1410.

[95] D. B. H. Tay, Two-stage least squares design of biorthogonal filter banks, IEE Vision Image Signal Processing, 149 (2002) 341–346.

[96] D. B. H. Tay, ETHFB: a new class of even-length biorthogonal wavelet filters for Hilbert pair design, IEEE Trans. on Circt. Syst. I, 55 (6) (2008) 1580–1588.

[97] M. Turk, A. P. Pentland, Eigenfaces for recognition, Journal Cognitive Neuroscience, 3 (1) (1991) 71–86.

[98] V. Vapnik, *The nature of statistical learning theory*, 2nd Edt., Springer, 2000.

[99] M. Vetterli and C. Herley, Wavelets and filter banks: theory and design, IEEE Trans. on Signal Processing, 40 (9) (1992) 2207–2232.

[100] H. Wang, S. Z. Li, Y. Wang, Face recognition under varying lighting conditions using self quotient image, IEEE Intl. Conf. Automatic Face Gesture Recognition (AFGR04), (2004) 819–824.

[101] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error measurement to structural similarity, IEEE Trans. on Image Processing, 13 (4) (2004) 600–612.

[102] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Analys. Mach. Intell., 31 (2) (2009) 210–227.

[103] X. Xie, K.-M. Lam, An efficient illumination normalization method for face recognition, Pattern Recognition Lett., 27 (6) (2006) 609–617.

[104] C. S. Xydeas and V. Petrovic, Objective image fusion performance measure, Electronics Letters, 36 (4) (2000) 308–309.

[105] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Analys. Mach. Intell., 23 (6) (2001) 643–660. The Yale B Face Database, Available online at `http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html`.

[106] K. C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Analys. Mach. Intell., 27 (5) (2005) 684–698. The Extended-Yale B Face Database, Available online at `http://vision.ucsd.edu/~leekc/ExtYaleDatabase`.

[107] J. Yang, D. Zhang, A. F. Frangi, J.-Y Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Analys. Mach. Intell., 26 (1) (2004) 131–137.

[108] R. Yu, A. Baradarani, Design of halfband filters for orthogonal wavelets via sum of squares decomposition, IEEE Signal Processing Lett., 15 (2008) 437–440.

[109] D. Zhang, G. H. Qu, P. Yan, Information measure for performance of image fusion, Electronics Letters, 38 (7) (2002), 313–315.

[110] T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li, F. Lang, Multiscale facial structure representation for face recognition under varying illumination, Pattern Recognition, 42 (2) (2009) 251–258.

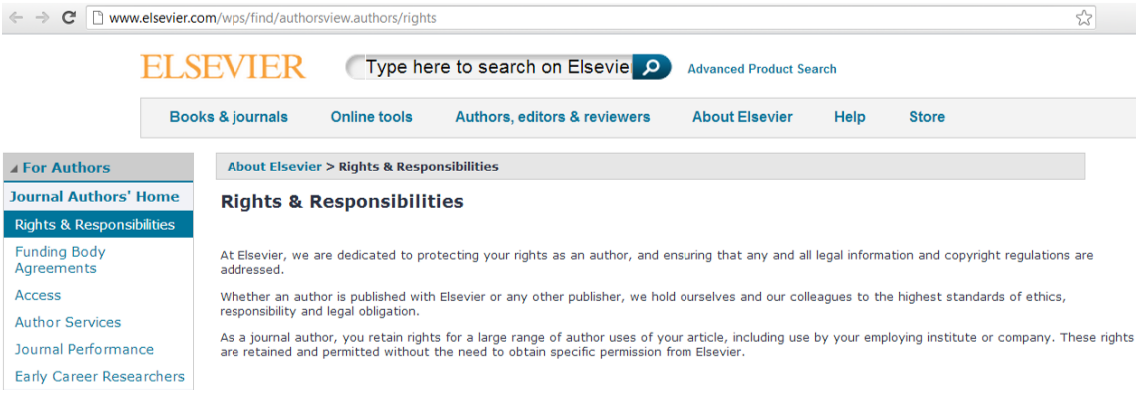
[111] X. Zhang, Design of FIR halfband filters for orthonormal wavelets using Remez exchange algorithm, IEEE Signal Processing Lett., 16 (9) (2009) 814–817.

[112] Y. Zhang, J. Tian, X. He, X. Yang, MQI based face recognition under uneven illumination, Springer Advances in Biometrics, 4642 (2007) 290–298.

[113] J. Zhao, Y. Su, D. Wang, S. Luo, Illumination ratio image: synthesizing and recognition with varying illuminations, Pattern Recognition Lett., 24 (15) (2003) 2703–2710.

Appendix A

Copyright Permissions



Intellectual property | Your role | Permissions | Publishing ethics | Other policies

Copyright
Intellectual property, in particular copyright (rights in editorial content), trademarks (rights in brands for services or journals), and database rights (rights in compilations of information), form the foundation of Elsevier's publishing services and communications businesses. We in Elsevier embrace the opportunities the digital environment offers for communication and access, while at the same time we recognize the new risks that this environment poses, that being the ease with which unauthorized copies can be made and distributed worldwide. Download your practical guide to Elsevier's copyright policy.

Our objective
We aim to manage digital rights and brands amidst the structural changes that the "information society" represents, while at the same time recognizing the shared goals we have with our customers and authors. These include providing the widest possible distribution of scientific and medical content and services in a financially sustainable business model.

Elsevier wants to ensure a proper balance between the scholarly rights which authors retain (or are granted/transferred back in some cases) and the rights granted to Elsevier that are necessary to support our mix of business models. We routinely analyse and modify our policies to ensure we are responding to authors' needs and concerns, and to the concerns in general of the research and scholarly communities.

What rights do I retain as a journal author*?

- the right to make copies (print or electronic) of the journal article for your own personal use, including for your own classroom teaching use;
- the right to make copies and distribute copies of the journal article (including via e-mail) to research colleagues, for personal use by such colleagues for scholarly purposes*;
- the right to post a pre-print version of the journal article on Internet websites including electronic pre-print servers, and to retain indefinitely such version on such servers or sites for scholarly purposes* (with some exceptions such as The Lancet and Cell Press. See also our information on electronic preprints for a more detailed discussion on these points)*;
- the right to post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on your personal or institutional website or server for scholarly purposes*, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article (but not in subject-oriented or centralized repositories or institutional repositories with mandates for systematic postings unless there is a specific agreement with the publisher. Click here for further information);
- the right to present the journal article at a meeting or conference and to distribute copies of such paper or article to the delegates attending the meeting;
- for your employer, if the journal article is a 'work for hire', made within the scope of the author's employment, the right to use all or part of the information in (any version of) the journal article for other intra-company use (e.g. training);
- patent and trademark rights and rights to any process or procedure described in the journal article;
- the right to include the journal article, in full or in part, in a thesis or dissertation;
- the right to use the journal article or any part thereof in a printed compilation of your works, such as collected writings or lecture notes (subsequent to publication of the article in the journal); and
- the right to prepare other derivative works, to extend the journal article into book-length form, or to otherwise re-use portions or excerpts in other works, with full acknowledgement of its original publication in the journal.

***Commercial purposes and systematic distribution**
Authors of Elsevier-published articles may use them only for scholarly purposes as set out above and may not use or post them for commercial purposes or under policies or other mechanisms designed to aggregate and openly disseminate manuscripts or articles or to substitute for journal-provided services. This includes the use or posting of articles for commercial gain or to substitute for the services provided directly by the journal including the posting by companies of their employee-authored works for use by customers of such companies (e.g. pharmaceutical companies and physician-prescribers); commercial exploitation such as directly associating advertising with such postings; the charging of fees for document delivery or access; the systematic distribution to others via e-mail lists or list servers (to parties other than known colleagues), whether for a fee or for free; the posting of links to sponsored articles by commercial third parties including pharmaceutical companies; institutional, funding body or government manuscript posting policies or mandates that aim to aggregate and openly distribute the accepted, peer reviewed manuscripts or published journal articles authored by its researchers or funded researchers; and subject repositories that aim to aggregate and openly distribute accepted peer reviewed manuscripts or published journal articles authored by researchers in specific subject areas.

For a more detailed discussion of our article posting policies and the different stages of a journal article development that are relevant from a policy perspective, please see the Article Posting Policies information page.

When Elsevier changes its journal usage policies, are those changes also retroactive?
Yes, when Elsevier changes its policies to enable greater academic use of journal materials (such as the changes several years ago in our web-posting policies) or to clarify the rights retained by journal authors, Elsevier is prepared to extend those rights retroactively with respect to articles published in journal issues produced prior to the policy change.

We are pleased to confirm that, unless explicitly noted to the contrary, all policies apply retrospectively to previously published journal content. If, after reviewing the material noted above, you have any questions about such rights, please contact Global Rights.

How do I obtain a Journal Publishing Agreement?
You will receive a form automatically by post or e-mail once your article is received by Elsevier's Editorial-Production Department. View a generic example of the agreement. Some journals will use another variation of this form.

Why does Elsevier request transfer of copyright?
The research community needs certainty with respect to the validity of scientific papers, which is normally obtained through the editing and peer review processes. The scientific record must be clear and unambiguous. Elsevier believes that, by obtaining copyright transfer, it will always be clear to researchers that when they access an Elsevier site to review a paper, they are reading a final version of the paper which has been edited, peer-reviewed and accepted for publication in an appropriate journal. This eliminates any ambiguity or uncertainty about Elsevier's ability to distribute, sub-license and protect the article from unauthorized copying, unauthorized distribution, and plagiarism.

Can you provide me with a PDF file of my article?
Many Elsevier journals are now offering authors e-offprints – free electronic versions of published articles. E-offprints are watermarked PDF versions, and are usually delivered within 24 hours, much quicker than print copies. These PDFs may not be posted to public websites. For more information, please see your journal's Guide to Authors or contact authorsupport@elsevier.com

Vita Auctoris

Aryaz Baradarani moved to Canada in 2008 and started his Ph.D. program in Biometrics and Pattern Recognition in the Department of Electrical Engineering at the University of Windsor, Ontario, Canada. During his graduate studies at Windsor he was the recipient of the University of Windsor IGE and IGS scholarships, 3M Company Bursary Award in 2009, and President's Excellence Scholarship. Aryaz completed his Ph.D. with a GPA of 13/13 (A^+). He was the winner of the prestigious Ontario Graduate Scholarship in Science and Technology, and the Endowment Graduate Award in 2012. His main interests are biometrics, medical imaging, signal processing, pattern recognition and the design of embedded classifiers.