



HAL
open science

Contributions to facial feature extraction for Face recognition

Huu Tuan Nguyen

► **To cite this version:**

Huu Tuan Nguyen. Contributions to facial feature extraction for Face recognition. Signal and Image processing. Université de Grenoble, 2014. English. NNT: . tel-01138363v1

HAL Id: tel-01138363

<https://hal.science/tel-01138363v1>

Submitted on 1 Apr 2015 (v1), last revised 20 Jul 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Signal, Image, Parole, Telecoms**

Arrêté ministériel : 7 août 2006

Présentée par **Huu-Tuan NGUYEN**

Thèse dirigée par **Alice CAPLIER**

préparée au sein du **GIPSA Laboratoire**

**Contributions to facial feature extraction for
Face recognition**

Thèse soutenue publiquement le **19 Septembre 2014**,

devant le jury composé de :

M. Pierre-Yves COULON

Professeur à Grenoble INP, Président

Mme. Sylvie LELANDAIS

Professeur à l'Université d'Évry-Val-d'Essonne, Rapporteur

M. Christophe GARCIA

Professeur à INSA Lyon, Rapporteur

M. Ngoc-Son VU

Maître de conférences à l'Université Cergy-Pontoise, Examineur

Mme. Alice CAPLIER

Professeur à l'INP Grenoble, Directrice de thèse



Dedicated to

Phuong Dung TRUONG, my beloved wife,
Bi Bon (Huu Thanh NGUYEN), my naughty boy,
Ủn Ủn (Minh Chau NGUYEN), my little girl,
Father, mother and my family.

Acknowledgements

Contents

Glossary	19
Nomenclature	23
1 Introduction	25
1.1 Face recognition problem	25
1.2 Stages in a feature based FR system	28
1.3 Why feature extraction?	30
1.4 Contributions of the present thesis	30
1.5 Thesis outline	32
2 State of the art of facial feature extraction	35
2.1 Global approaches	35
2.2 Local feature based approaches	37
2.2.1 Gabor wavelets based methods	37
2.2.2 LBP based methods	40
2.2.3 LPQ based methods	44
2.2.4 Multi-resolution/multi-scale methods	46
2.2.4.1 Simple multi-scale approaches	46
2.2.4.2 Gabor wavelets components based methods	46
2.2.4.3 Monogenic filter based methods	50
2.2.5 Sparse representation based methods	51
2.2.6 Other methods	52
2.3 Conclusions	53
3 Face recognition background	55
3.1 General FR framework	55
3.2 Face databases	56
3.2.1 AR (Aleix and Robert) database	57
3.2.2 FERET (Face Recognition Technology) database	58
3.2.3 SCface (Surveillance Camera face) database	59
3.3 Face cropping	60

3.4	Preprocessing techniques	62
3.4.1	Retinal filter	62
3.4.2	Histogram equalization	64
3.5	Template matching framework	66
3.6	Whitened PCA based framework	67
3.6.1	EVD (Eigenvalue decomposition) based WPCA	69
3.6.2	SVD (Singular value decomposition) based WPCA	70
3.7	Conclusions	71
4	Intensity-based feature extraction methods	73
4.1	ELBP, a novel variant of LBP	74
4.1.1	Motivations	74
4.1.2	ELBP in detail	75
4.1.3	Face recognition with ELBP	78
4.1.4	Experimental results	79
4.1.4.1	Results on AR database	79
4.1.4.2	Results on FERET database	83
4.1.4.3	Results on SCface database	84
4.1.4.4	ELBP parameters	86
4.1.4.5	Computational cost	87
4.1.5	Conclusions	88
4.2	LPQ as a facial feature extraction	89
4.2.1	Blur invariance of Fourier Phase spectrum	89
4.2.2	LPQ in detail	90
4.2.3	Face recognition with LPQ	92
4.2.4	Experimental results	93
4.2.4.1	Results on AR database	93
4.2.4.2	Results on FERET database	94
4.2.4.3	Results on SCface database	98
4.2.4.4	LPQ parameters	99
4.2.4.5	Computational cost	100
4.2.5	Conclusions	100
4.3	Conclusions	101
5	Patch based Local Phase Quantization of Monogenic components for Face recognition	103
5.1	Introduction	104
5.2	Monogenic filter	105
5.2.1	Log-Gabor filter	105
5.2.2	Image representation by Monogenic filter	106
5.3	Patched based LPQ of Monogenic bandpass components for FR	108

5.3.1	Patched based LPQ of Monogenic bandpass components feature extraction method	108
5.3.2	PLPQMC WPCA FR framework	110
5.4	Experimental results	110
5.4.1	Results on AR database	111
5.4.2	Results on FERET database	114
5.4.3	Results on SCface database	117
5.4.4	Computational performance	119
5.5	Conclusions	120
6	Gradient images based facial features	121
6.1	Introduction	121
6.2	Elliptical Patterns of Oriented Edge Magnitudes for Face recognition .	123
6.2.1	Elliptical Patterns of Oriented Edge Magnitudes feature extraction method	123
6.2.2	Using EPOEM for face recognition	126
6.2.3	Experimental results	126
6.2.3.1	Results on AR database	127
6.2.3.2	Results on FERET database	129
6.2.3.3	Results on SCface database	131
6.2.4	Conclusions	133
6.3	Local Patterns of Gradients for Face recognition	134
6.3.1	Local Patterns of Gradients feature extraction method	135
6.3.1.1	Block-wised ELBP: a novel variant of ELBP	135
6.3.1.2	LPOG in details	136
6.3.2	Using LPOG for face recognition	139
6.3.3	Experimental results	140
6.3.3.1	Results on AR database	140
6.3.3.2	Results on FERET database	145
6.3.3.3	Results on SCface database	150
6.3.3.4	Parameters setting	152
6.3.3.5	Computational cost	153
6.3.4	Conclusions	154
6.4	Conclusions	155
7	Conclusions and Future work	157
7.1	Conclusions	157
7.2	Future work	160
	Bibliography	162

List of Figures

1.1	What's in a Face?	27
1.2	General local feature-based FR framework	29
1.3	Main contributions of the thesis.	31
2.1	Eigenfaces scheme	36
2.2	An image and its Gabor wavelets components (5 scales and 8 orientations)	38
2.3	Bunch Graph structure proposed in [119].	39
2.4	LBP encoding scheme	40
2.5	LBP patterns	40
2.6	A face image and its LBPs	41
2.7	LBP description calculation	42
2.8	LPQ encoding scheme	44
2.9	A face image and its LPQs	45
2.10	General multi-resolution/multi-scale feature extraction scheme	47
2.11	Quadrant bit coding scheme	48
2.12	LXP encoding scheme	48
2.13	An image and its sparse representation	51
3.1	Stages in a local feature-based face recognition system	55
3.2	Sample cropped images from AR database.	57
3.3	Sample cropped images from FERET database.	59
3.4	Sample images from SCface database.	60

3.5	Face cropping based on eyes coordinates scheme.	61
3.6	Retinal filter scheme for illumination normalization	62
3.7	Retinal filter's illustration on illumination samples of FERET database.	64
3.8	Some face images from SCface database and their histogram equalization versions.	64
3.9	Histogram equalization mechanism.	65
3.10	General template matching framework	66
3.11	General WPCA based framework	67
4.1	Contributions presented in this chapter: ELBP and LPQ methods and their associated FR frameworks.	73
4.2	LBP and ELBP patterns	75
4.3	ELBP encoding scheme	75
4.4	An image (a) and its LBP^{8,1} (b), ELBP^{8,3,1} (c), ELBP^{8,4,3} (d)	76
4.5	ELBP bilinear interpolation scheme	77
4.6	ELBP feature vector computation	78
4.7	Accuracy performance of LBP and ELBP based systems on AR database.	80
4.8	The neighborhood and Fourier frequencies in LPQ.	90
4.9	LPQ feature vector computation	92
4.10	Accuracy performance of LPQ and ELBP(h+v) based systems on AR database.	94
5.1	Contributions presented in this chapter: Patch based LPQ of Monogenic components WPCA system.	103
5.2	A face image and its Monogenic components at 3 scales.	107
5.3	Steps in PLPQ of Monogenic components feature extraction method.	108
5.4	Comparison between bandpass components of illumination images from FERET database.	109
5.5	Accuracy performance of LPQ, LPQMC and PLPQMC WPCA systems on AR database.	112

List of Figures

6.1	Proposed methods in this chapter: EPOEM and LPOG.	121
6.2	An image and its gradient based components. The orientation component is visualized from its radiance values.	122
6.3	Steps in EPOEM encoding scheme for one pixel.	124
6.4	Facial representation computation by EPOEM method.	125
6.5	Some challenging SCface images captured at distance 1 where EPOEM fails to recognize.	132
6.6	BELBP encoding operators	135
6.7	BELBP operators	136
6.8	Steps in LPOG scheme	137
6.9	Comparison between gradient images of illumination images from AR database. The first image on the left is the gallery one while the rest are probe ones.	138
6.10	Comparison between histograms of images in Figs. 6.9.	138
6.11	Comparisons of recognition performance between LPOG and other gradient images based methods on AR database.	142
6.12	Comparisons of recognition performance between LPOG and PLPQMC on AR database.	143
6.13	Two Fb probe images (a), their wrongly assigned ones (b) and their correct gallery ones (c).	147
6.14	Original versions of images in Figs 6.13.	149

List of Tables

4.1	Rank-1 RRs (%) comparison between LBP and ELBP based methods on AR database	79
4.2	Rank-1 RRs (%) comparison with other contemporary systems on AR database using the same evaluation method	82
4.3	Rank-1 RRs (%) comparison with other state-of-the-art results on FERET database using the standard evaluation protocol [96]	83
4.4	Rank-1 RRs comparison between LBP and ELBP based methods on b-series of FERET database	84
4.5	Rank-1 RRs comparison with other leading methods on FERET b-series.	85
4.6	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]	86
4.7	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]	86
4.8	Computation time of ELBP in comparison with other feature extraction methods	87
4.9	Rank-1 RRs (%) comparison between ELBP(h+v) and LPQ based methods on AR database	93
4.10	Rank-1 RRs (%) comparison with other contemporary systems on AR database using the same evaluation method	95
4.11	Rank-1 RRs (%) comparison of LPQ based systems with other state-of-the-art results on the FERET database [96]	96
4.12	Rank-1 RRs comparison between ELBP(h+v) WPCA and LPQ based methods on b-series of FERET database	97
4.13	Rank-1 RRs comparison of LPQ WPCA with other leading systems on FERET b-series.	97

4.14	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]	98
4.15	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]	99
4.16	Computation time of LPQ in comparison with other feature extraction methods	100
5.1	Rank-1 RRs (%) comparison between LPQ WPCA and PLPQMC based WPCA methods on AR database	111
5.2	Rank-1 RRs (%) of PLPQMC WPCA in comparison with other contemporary systems on AR database using the same evaluation method . . .	113
5.3	Rank-1 RRs (%) comparison of PLPQMC based systems with other state-of-the-art results on the FERET database [96]	115
5.4	Rank-1 RRs comparison of PLPQMC WPCA with other leading systems on FERET b-series.	116
5.5	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]	118
5.6	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]	118
5.7	Computation time of PLPQMC in comparison with other feature extraction methods	119
6.1	Rank-1 RRs (%) comparison between POEM and EPOEM WPCA based methods on AR database	127
6.2	Rank-1 RRs (%) of EPOEM WPCA in comparison with other contemporary systems on AR database using the same evaluation method . . .	128
6.3	Rank-1 RRs (%) comparison of EPOEM WPCA with other state-of-the-art results on the FERET database [96]	129
6.4	Rank-1 RRs comparison of EPOEM WPCA with other systems on FERET b-series.	130
6.5	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]	131
6.6	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]	132

6.7	Rank-1 RRs (%) comparison between ELBP, BELBP, LPQ and LPOG (WPCA) on AR database	141
6.8	Rank-1 RRs (%) of LPOG WPCA in comparison with other contemporary systems on AR database using the same evaluation method	144
6.9	Rank-1 RRs comparison between ELBP, BELBP, LPQ and LPOG (WPCA) on FERET database using standard protocol [96]	145
6.10	Rank-1 RRs (%) comparison of LPOG based systems with other state-of-the-art results on the FERET database [96]	146
6.11	Rank-1 RRs comparison between ELBP, EBLBP, LPQ and LPOG (WPCA) on b-series of FERET database	148
6.12	Rank-1 RRs comparison of LPOG WPCA with other leading systems on FERET b-series.	150
6.13	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]	151
6.14	Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]	151
6.15	Details of divided sub-regions and window size used with LPOG	152
6.16	Computation time of LPOG in comparison with other feature extraction methods	153

Glossary

AR Aleix and Robert.

ATM Automated teller machine.

BELBP Block-wised Elliptical Local Binary Patterns.

CCS-POP Circular Center Symmetric-Pairs of Pixels.

CS-LBP Center Symmetric Local Binary Patterns.

DBC Directional Bandpass Component.

DLBP Discriminative Local Binary Patterns.

DoG Difference of Gaussians.

EBGM Elastic bunch graph matching.

ELBP Elliptical Local Binary Patterns.

EPFDA Ensemble of piecewise FDA.

EPOEM Elliptical Patterns of oriented edges magnitudes.

Eq Equation.

ESRC Extended sparse representation-based classification.

EVD Eigenvalue decomposition.

FDA Fisher Discriminant Analysis.

FERET Face Recognition Technology.

Fig Figure.

FLD Fisher's Linear Discriminant.

- FR** Face recognition.
- GOM** Gabor Ordinal Measures.
- GSF** Gabor surface feature.
- HGPP** Histogram of Gabor phase patterns.
- HOG** Histograms of Oriented Gradients.
- HR** High resolution.
- ICA** Independent Component Analysis.
- ILBP** Improved Local Binary Patterns.
- IP** Image Processing.
- IR** Infrared.
- k-NN** k-Nearest Neighbor.
- KDA** Kernel Discriminant Analysis.
- KPCA** Kernel Principal Component Analysis.
- LBP** Local Binary Patterns.
- LDA** Linear Discriminant Analysis.
- LFD** Local frequency descriptor.
- LGBPHS** Local Gabor binary pattern histogram sequence.
- LGOBP** Local gradient orientation binary pattern.
- LPOG** Local patterns of gradients.
- LPQ** Local Phase Quantization.
- LQP** Local Quantized Patterns.
- LR** Low resolution.
- LTP** Local Ternary Patterns.
- LXP** Local XOR pattern.
- MB-LBP** Multi-scale block LBP.

MBC Monogenic Binary Coding.

MBP Monogenic binary pattern.

MLPQ Multi-scale Local Phase Quantization.

OPL Outer plexiform layer.

PCA Principal Component Analysis.

PDO Patterns of dominant orientations.

PLPQ Patch-based Local Phase Quantization.

PLPQMC Patch-based Local Phase Quantization of Monogenic components.

POEM Patterns of oriented edges magnitudes.

PR Pattern recognition.

RLTP Relaxed Local Ternary Patterns.

RR Recognition rates.

SCface Surveillance Camera face.

SIFT Scale Invariant Feature Transform.

SLF Statistical local features.

SRC Sparse representation-based classification.

SSEC Structured sparse error coding.

SSPP Single sample (image) per person.

STFT Short-term Fourier transform.

SVD Singular value decomposition.

SVM Support Vector Machine.

WPCA Whitened Principal Component Analysis.

Nomenclature

- * convolution operation
- $\|X\|$ The Euclidean norm of vector X
- \odot Element-wise multiplication
- \bar{X} Mean value of X
- A^T The transpose of matrix A

Chapter 1

Introduction

This chapter presents an introduction about the Face recognition (FR) problem, its perspectives and challenges, and the most important parts are for the explanation about the purpose and the approaches of the thesis. An outline of the thesis is also included after the list of main contributions is given.

1.1 Face recognition problem

Face recognition, an attractive and full of challenges research field of Computer Vision and Biometrics concerning theoretical methods and software systems for machines to recognize people based on their digital face images, has been fuelled by many academic scientists and industrial developers for over twenty years. This is rooted in its various potential applications and the availability of human face in computer's images and videos, which can be encountered in many corners of life. Face recognition can be used for security applications (access control to authorized areas, computer, airports, etc), surveillance devices in public spaces (such as football stadiums, train stations, big trade centers), in forensic applications (identity verification/management for criminal justice system, disaster victim identification), querying person's identity in image/video databases, human machine interaction applications, smart card solutions (enhanced ATM's security, biometrics passport-also known as ePassport) [65], and targeted advertising. Face recognition, that exploits knowledge from many research disciplines such as Image processing (IP), Pattern recognition (PR), Machine learning (ML), Visual perception, psychophysics and neuroscience, is one of the most successful studies of biometrics, the two others being fingerprint and iris recognition. While fingerprint and iris recognition are mature technologies that can be deployed in real life applications, face recognition still has many challenges that need to be solved with more powerful methods even though numerous systems have been proposed for the

last two decades [137, 65]. To the best of our knowledge, what follows are the biggest challenges [137, 53] for face recognition:

- Pose variations. Face images are greatly varied by pose angles of people. When the pose angles are large (not in the range of $\pm 45^\circ$), the system accuracy is drastically degraded.
- Illumination variations. The acquired images are strongly affected by environment illumination and finally result in severe impairments on the overall accuracy performance of the systems since their intra-class dissimilarities are greater than the extra-class margins.
- Facial expressions. They cause deformations of crucial facial features such as eyes, eyebrows, mouth, nose and therefore affect the recognition results.
- Aging condition. As human faces are vastly changing over time, the identification process of face images under long term aging effects is a real challenge, even by human beings.
- Near infrared illumination (NIR). While reducing the bad effect of illumination variations, NIR images are strongly affected by the environment temperature, the pose variation, facial expressions and the health condition of the subject. This becomes more compounded when the system has to match images captured under NIR condition with ones acquired in natural illumination.
- Very large scale systems. Face database of a country's population can reach up to hundreds of millions or over one billion images. How to make a FR system working efficiently with such huge databases in real time is a hard question to answer.
- Low resolution images. Equipped with limited memory, usually average quality lens, and having to operate in real time, surveillance cameras consequently produce blur, small, and low quality images that are too challenging to obtain high recognition rates.

Unfortunately, these issues rarely come singly in realistic scenarios, however, grouped together they mark their impairment, and result in extremely varied face images of the same person, which are difficult to be recognized correctly even by humans. The source of such challenges stems from the uncontrolled conditions of input face images for a face recognition system. Unlike in fingerprint and iris recognition systems, which require strictly cooperations (via step-by-step interaction, physical contact or attention) of users to collect their biometrics features, the input face images of a face recognition system can be easily gathered without any real interaction with users and does not

1.1. Face recognition problem

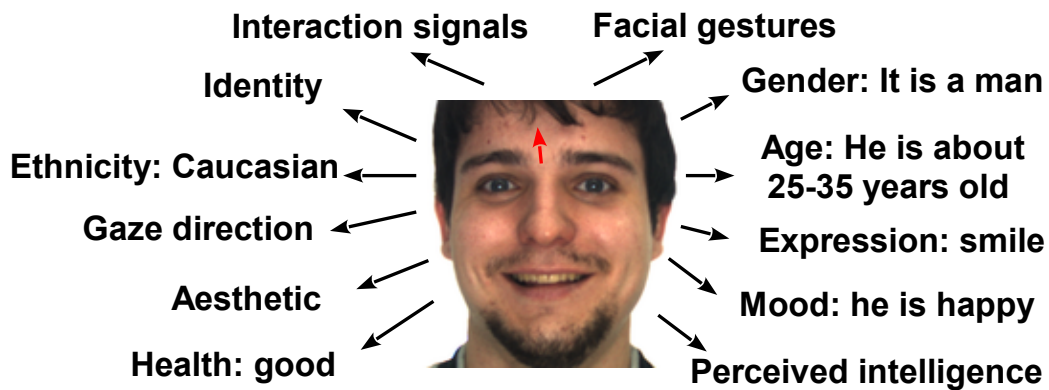


Figure 1.1: What's in a Face?

necessarily need to proceed under controlled environments. Another reason making face recognition more interesting is that face images contain a lot of useful information, including human gender, facial expressions, human age, human ethnic and gaze direction to name a few (see Fig. 1.1 for more details). Additionally, the rapid evolution of digital camera/camcoder devices, together with the appearance and blooming developments of image/video sharing services and social networks on the Internet, they have promoted and advanced the researches on face recognition. Because the most pervasive objects that can be found in image/video data are human faces.

The input data to a **FR** system can be 2D images, 3D images, video or image sequences but we constrain our focus on 2D images due to the following reasons:

- Eventually, the most common unit data processed at a time by a **FR** system is a 2D image. This is for achieving fast speed, efficient processing by many 2D algorithms proposed through years of **IP**.
- Despite their advantages, such as more useful information, less affected by pose and illumination variations, the widely usage of 3D images as input for a real **FR** system is prohibited since this entails 3D acquisition devices, which leads to more expensive and slower solutions.
- The obstacle caused by 3D acquisition process is just the tip of the iceberg. As 3D images mean more complicated image structures, thus many more efforts must be done to push on the development of 3D methods to exploit these structures as well as 3D information effectively.

Face recognition systems can fall into two kinds of tasks: face verification and face identification [137]. Face verification is a validating system that accepts or rejects a claimed identity based on a face image while a face identification system will identify

an individual from unknown input face images. The works presented in this thesis is confined to face identification only.

In a FR system, we have some still face images of known people (or labeled images) meaning that we know exactly which images belong to which person, these images are called reference or gallery images. The problem of FR is that when we have some new images, we must identify which person those images belong to. In this regard, those images are called probe images (also referred as test images).

The solution coming from PR is straightforward and spontaneous: we extract images' intrinsic features of gallery images, store them into a database and when a probe image arrives, we compare its intrinsic features with all the images in the database. The most similar gallery image's label is the probe image's label. This principle is the same as people's face recognizer mechanism: when we see a face image, we will try to capture the most intrinsic features of that face, such as characteristics of eyes, mouth, nose, ears, overall face shape and skin color, and search in our memory the name of the person that is most matched with the image. The unit used for facial features description in FR is called feature vector.

For over a quarter of a century, plethora of systems with various approaches have been developed for face recognition [137, 65]. These systems, based on the types of facial features they use, can be categorized into three types of approaches: local (feature-based), global (holistic) and hybrid [137]. In global methods, a single vector which extracts holistic information from the entire face image is used. Eigenfaces [114] and Fisherfaces [12] are the most representative systems that belong to holistic approaches. Unlike holistic methods, local feature based ones rely on the segmentation of a face image into different local facial features or components such as eyebrows, eyes, mouth, nose, etc. Then each image is represented as a feature vector obtained by applying a feature extraction algorithm to extract the most discriminant characteristics from those local facial features. Hybrid approaches combine both global and local methods to achieve better performance. Compared to global approaches, feature-based approaches have a significant advantage: they can perform much better under various uncontrolled conditions. More specifically, according to [47], when dealing with pose variation of pose angles within the range $\pm 40^\circ$ by Support Vector Machine (SVM) classification, a component-based feature extraction system could yield RRs about 60% higher than its global counterpart.

1.2 Stages in a feature based FR system

Like human visual perception process, which is a multilayered model, there are multiple major stages in a local feature-based FR system (see Fig. 1.2): face detection, face

1.2. Stages in a feature based FR system

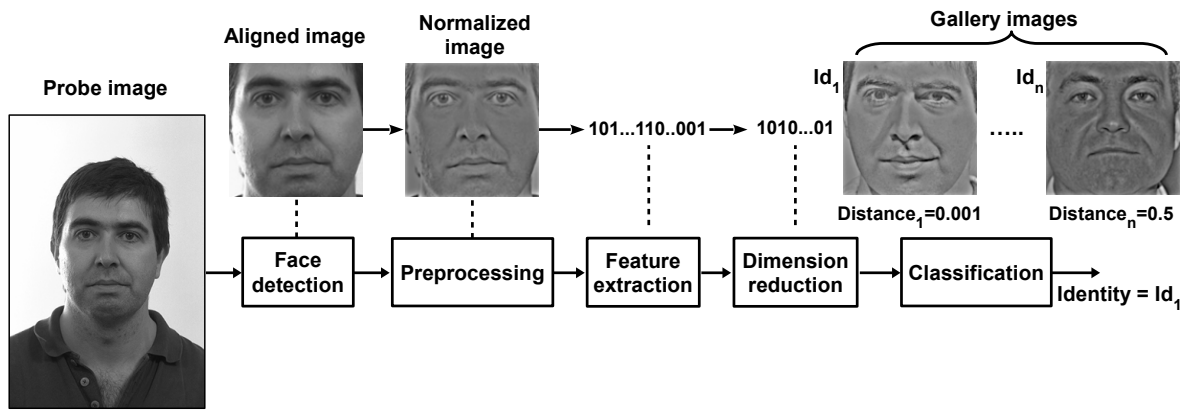


Figure 1.2: General local feature-based FR framework

preprocessing, feature extraction, dimension reduction and classification, and the results of each stage act as inputs to the following stages.

In the face detection stage, the system has to detect whether or not there are human faces in the input image. If the answer is yes, the face image's location must be exactly located. Then only the image region containing the face is cropped and aligned to have good frontal face image for the next stage. On account of the fact that we work on public face databases accompanied with annotation data about eyes coordinates and since our objective is not face detection, we utilize a simple cropping algorithm based on eyes' locations to crop and align face images. After that, cropped face images are processed by a preprocessing technique to eliminate the effect of illumination variations. A feature extraction method is applied on those preprocessed images to capture the most distinguishing features from the given images for classifying. Amongst all the current methods, Gabor wavelets [27] and Local Binary Patterns (LBP) [3] are the most widely used for facial feature extraction due to their efficiency. In order to make obtained feature vectors more discriminant and compact, a learning technique from Machine learning field is employed for dimensionality reduction task by projecting these vectors into a trained subspace built from so-called training images. Principal Component Analysis (PCA) [114] and Fisher's Linear Discriminant (FLD) [12] are the most well-known methods for dimension reduction purpose.

The identity (label) of the input face image is identified in the classification step based on projected vectors resulted from previous stage. For doing the classification, SVM [24] and k-nearest neighbor (k-NN) are the two most popular choices. Since FR's classification stage is a multi-label problem, SVM actually has to find multiple optimal hyper-planes (which is a costly operation), each for one of its native binary classification solutions to classify one test image. In the meantime, k-NN simply assigns each test image to the label of the gallery image that has at least $\frac{k}{2} + 1$ closest distances, which are computed by using a distance metric, to it. In this thesis, we use k-NN with $k = 1$ (as the gallery

set has one image for one person) due to its simplicity, fast implementation and good results.

Among all these stages, the feature extraction is the paramount step for building robust, reliable and viable face recognition systems [137, 100], because it is the only means to extract the most distinguishing characteristic features of face images to form feature vectors, which are then compared with each other to proceed the classification. This is reflected in the history of FR study when the development of FR systems can be viewed as the development of facial feature extraction methods.

1.3 Why feature extraction?

The performance of a FR system, with respect to both accuracy and computational speed, is not solely based on feature extraction algorithms, but as earlier pointed out, the most important stage is feature extraction. More importantly, as verified in [41], some local features are not varied with pose, facial expression variations and lighting direction. Additionally, evidences from [139] proved that local features are more appropriate for machines to recognize human faces than holistic ones. Also, pioneer studies in face perception [132, 36, 99] figured out that even with one local facial feature, such as the eyes or eyebrows, famous faces can be recognized correctly. Therefore, the works in this thesis are confined to *local feature extraction* methods.

Our goal in this thesis is to develop unified feature extraction methods which are *robust* in dealing with many, if not all, of aforementioned challenges (as they come in battalions) and *fast* enough to be applied in real world applications. Towards this end, we attempt to devise novel methods based on analyzing the disadvantages and advantages of contemporary approaches as well as harnessing results gained over years from visual perception researches and image processing algorithms, which are applicable for improving facial feature extraction efficiency.

1.4 Contributions of the present thesis

The main contributions of this dissertation (illustrated in Fig. 1.3) are the four novel feature extraction methods for FR succinctly described below:

1. Exploiting the fact that essential facial features, such as eyes, mouth, are ellipses and that human face contains more horizontal direction information than that of vertical direction, we propose Elliptical Local Binary Patterns (**ELBP**), a novel variant of LBP, by using horizontal ellipse when thresholding each image pixel

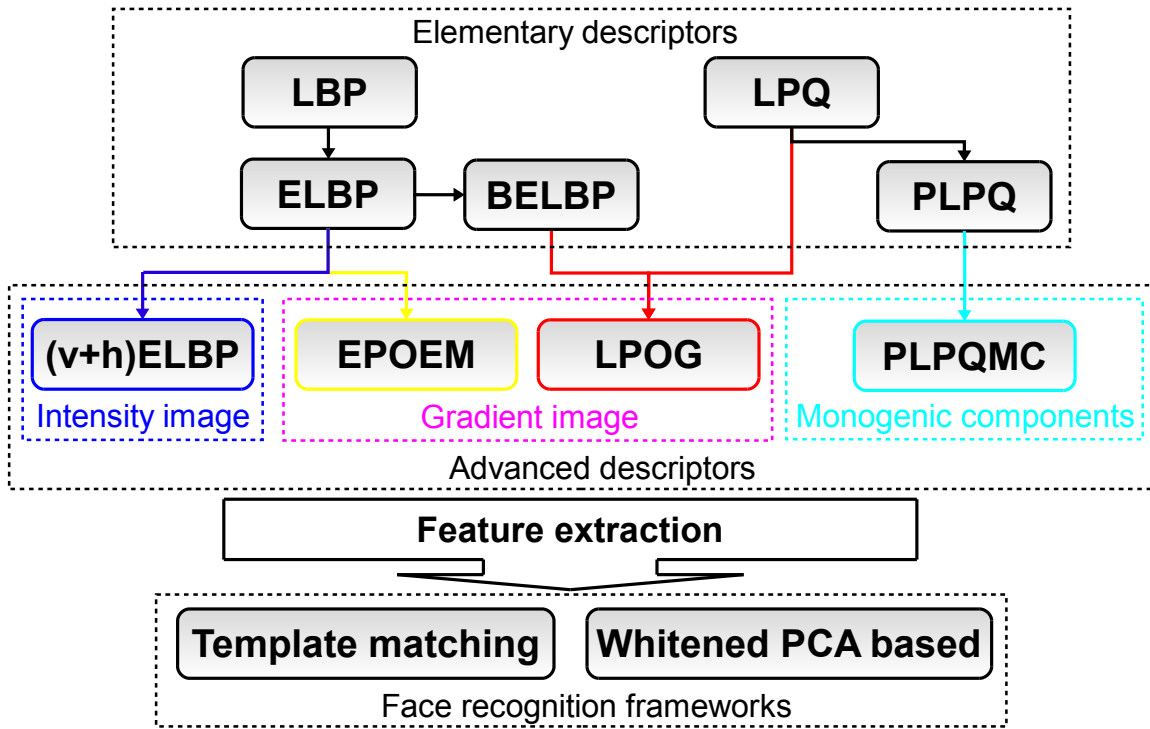


Figure 1.3: Main contributions of the thesis.

with its neighbors to encode micro textures from face images. Further, to capture both horizontal and vertical information, a symmetric pair of ELBPs are used and achieve excellent results while being fast in comparison with state of the art rivals. In addition to that, **ELBP**, acted as an elementary descriptor, can be further used to construct more robust feature extraction methods, such as **EPOEM** and **LPOG** in this thesis.

2. Based on Monogenic filter and benefited from Local Phase Quantization (**LPQ**), a novel multi-resolution feature extraction named as Patch based LPQ of Monogenic component (**PLPQMC**) is proposed. For feature extraction target, the Monogenic is employed to decompose the given input image into directional bandpass components upon which two Patch based LPQ (**PLPQ**) operators, a novel variant of **LPQ**, are applied to generate corresponding PLPQ images. The **PLPQMC** feature vector is then built by incorporating every PLPQ image's description into an augmented group. The method gains leading edge results when coping with various factors of **FR** and is competing with up-to-date methods. Meanwhile, it necessitates less computational cost than other Gabor wavelets based methods since only six Monogenic bandpass components are used compared to forty of Gabor wavelets ones.
3. Utilizing ELBP on edge magnitude images, a new variant so-called Elliptical Pat-

terns of Oriented Edge Magnitudes (EPOEM) of POEM is introduced. EPOEM is shown to attain higher accuracy performance than POEM while preserving its simplicity and without affecting the required computational time. In EPOEM, oriented edge magnitudes images are computed from magnitude image and different orientation qualification components of the input image. Then, descriptions of these images are aggregated to form a global feature vector.

4. A novel robust feature extraction method, namely Local Patterns of Gradients (LPOG), stemmed from applying Block-wised ELBP (BELBP), a refined variant of ELBP, and LPQ on gradient images, is proposed. In LPOG, we first present an enhanced variant of ELBP, entitled BELBP, by using horizontal and vertical blocks to build accumulated images from which their facial features are extracted. Then, inspired by their advantages over raw intensity images, we apply BELBP and LPQ directly upon gradient images to encode local patterns from them to constitute LPOG representation for face images. Extensive experiments on three public databases show that our method outperforms other state-of-the-art systems and is fast enough to be used in reality. Moreover, LPOG is verified to be robust against many challenging issues, including illumination, facial expressions, pose, time-lapse variations, occlusions and provide promising results when dealing with low resolution probe images captured by surveillance cameras under unconstrained conditions.

Alongside these feature extraction methods, their associated FR frameworks based on two general models "Template matching" and "Whitened PCA based" are also proposed and rigorously assessed upon three large public face databases following standard protocols by comparing with other state-of-the-art systems. Additionally, our results point out that FR is far from being a completed research, at least in video surveillance context. Besides, this suggests a pressing need of more attention from scientists on FR in video surveillance systems with more powerful methods.

1.5 Thesis outline

The rest of this thesis is organized into 6 chapters as follows.

Chapter 2 reviews state-of-the-art of facial feature extraction methods, including pure Gabor wavelets methods, LBP and its variants, LPQ based, multi-resolution/multi-scale based on Gabor wavelets and Monogenic filters, and Sparse representation based methods, each with key concepts upon which it is built, its drawbacks and advantages. By this, all the works related to our propositions (in chapters 4, 5, and 6) are grouped together in one place.

1.5. Thesis outline

Chapter 3 encompasses all the **FR** background materials being used throughout this thesis. The chapter contains the details of three public face databases and their standard protocols, the face cropping algorithm based on two eyes' coordinates, two preprocessing methods for illumination normalization, as well as two general **FR** frameworks namely "*Template matching*" and "*WPCA based*".

A novel **LBP** variant so-called Elliptical LBP (**ELBP**) and **LPQ**, another intensity-based descriptor, are presented and thoroughly assessed with both frameworks by means of a variety of experiments in chapter 4.

In chapter 5, a novel feature extraction method based on Monogenic filter and Patch based Local Phase Quantization (**PLPQ**), a variant of **LPQ**, is introduced.

Chapter 6 is dedicated to gradient images based approach with two proposed feature extraction methods named Elliptical Patterns of Oriented Edge Magnitudes (**EPOEM**) and Local Patterns of Gradients (**LPOG**). Moreover, the comparisons between each proposed method and other contemporary ones, its associated parameters and computational cost are included in the corresponding chapters.

Finally, the thesis ends with some conclusions and perspectives about future work.

Chapter 2

State of the art of facial feature extraction

This chapter is intended to give a state of the art review of facial feature extraction methods (global and local approaches). While this work concentrates on local feature based methods, a large portion of the chapter's content is hence mainly focused on that kind of approach.

2.1 Global approaches

The first and most well-known holistic method is Eigenfaces [114] (for more details, see Fig. 2.1), which is an expansion of Karhunen-Loève transform (also known as Principal Component Analysis-PCA). Kirby et al. [59] argued that each face image of a given set can be represented as a linear combination of basic orthogonal eigenvectors computed by PCA on the image set itself. Inspired from that, in Eigenfaces method [114], training images are first reshaped from their intensities values, in the form of 2-D integer matrices of the same size $M_{rows} \times N_{columns}$, to column vectors of length $M \times N$. These vectors are then normalized to have unit norm and mean-subtracted to have zero-mean. From normalized vectors, the PCA algorithm is employed to find the principal eigenvectors, corresponding to the largest eigenvalues, which are used as a seed set to represent for all other face images via a projection operation. This method is called Eigenfaces since these eigenvectors can be reconstructed and visualized as face images, as shown in Fig. 2.1. Eigenfaces [114] can work reasonably well with good quality images captured under strictly same conditions of light, pose, facial expressions and if there are neither aging variations nor occlusions. Conversely, its performance is dramatically degraded (e.g. it only offers 4.7% average RR on SCface database [42]) and thus obviously can not be applicable in reality.

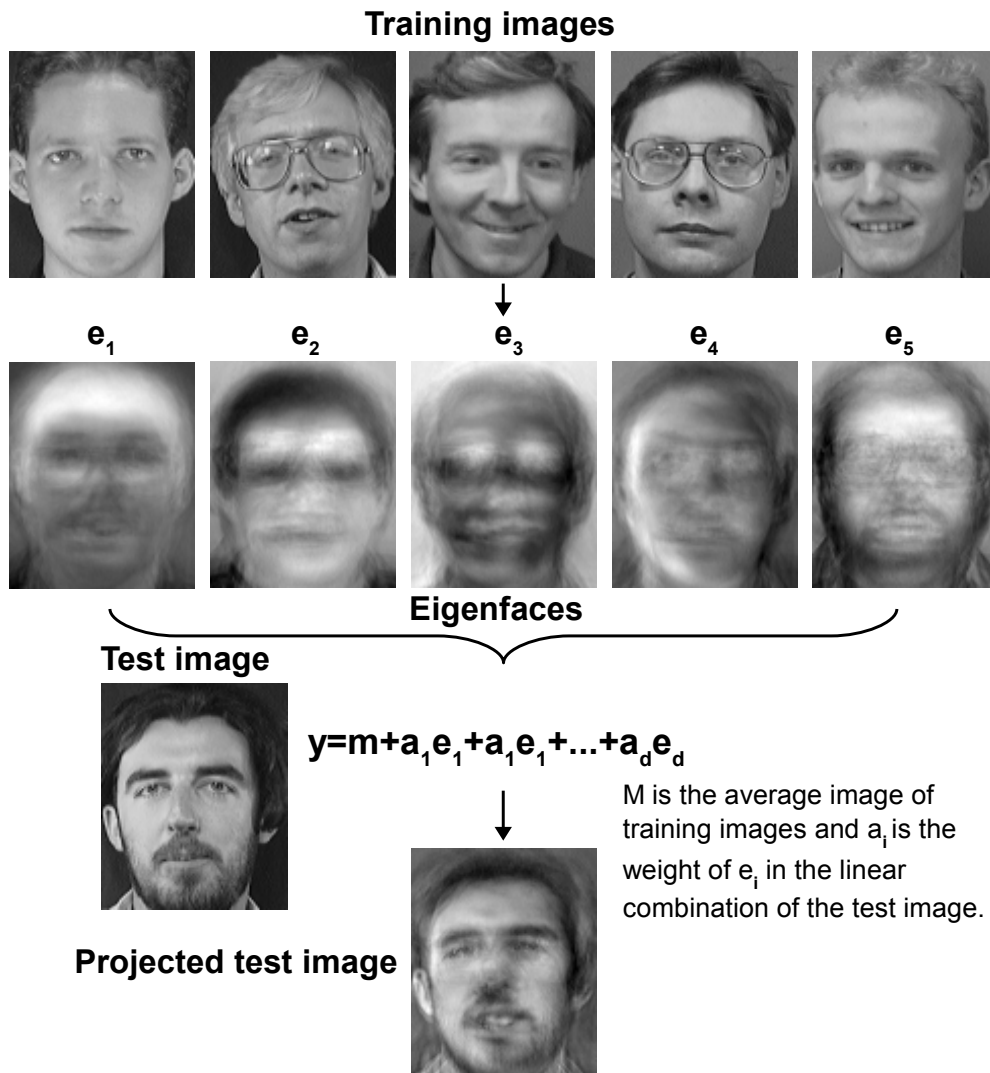


Figure 2.1: Eigenfaces scheme

When the training set has more than 1 image per subject, Eigenfaces [114] do not utilize such available information to improve system accuracy since PCA is an unsupervised learning technique. Motivated from that, Fisherfaces [12] was proposed by using Fisher's Linear Discriminant (FLD) [35] learning algorithm to maximize extra-class variations between images belonging to different people while minimizing the intra-class variations between those of the same person. Due to the fact that intra-class variations induced by challenging factors such as illumination, head pose and expression changes are almost always greater than extra-class variations come from the differences of face identities [2], thus can make images of a same person extremely different, the usage of FLD to reduce that impairment is valuable and leads to higher accuracy than Eigenfaces [12, 43]. But on the contrary, Fisherfaces can be applied only when the training set has more than 1 image per person. While this prerequisite is not always satisfied, it can be viewed as a

weakness of the method.

There also are some other global approaches that extend Eigenfaces and Fisherfaces, such as Independent component analysis (ICA) [11], 2D PCA [125] and 2D LDA [126], but their performance is far below those of local feature based methods, which are forthcoming presented.

2.2 Local feature based approaches

While holistic approaches are based on global features, feature based ones are built upon local facial features extracted from local components such as eyes, nose, mouth, etc. and local segmented regions. In this section, pure Gabor wavelets methods (no fusion strategy is used) are first surveyed. These methods exploit the fact that Gabor wavelets coefficients encode both facial shape and local appearance features. In contrast to these, LBP and LPQ, two intensity based elementary descriptors coming afterwards, capture micro appearance features from face images via their own operators. The usage of multiple LBPs, or LPQs with different core parameters creates new multi-scale methods. More efficiently, fusion strategies are used in multi-resolution/multi-scale methods to combine elementary methods with multi-resolution/multi-scale analysis tool. Recently sparse-representation based and some other methods are also covered.

2.2.1 Gabor wavelets based methods

Gabor wavelets transformation is a powerful joint time-frequency tool for image analysis based on Gabor filters [38]. On account of their strong similarity to perception mechanism of the human visual system [27] and their capability of providing multi-resolution/multi-orientation representations [28, 62] encompassing a large amount of meaningful salient visual features for FR, Gabor wavelets have been used for years in numerous feature extraction algorithms [103]. Using 2-D Fourier transformations on a set of Gabor kernels (also known as Gabor filters) parameterized by different orientations (usually 8) and scales (usually 5) and an input image, the method generates complex coefficients called Gabor wavelets components, which can be expressed by real and imagine parts, or alternatively, by magnitude and phase parts (as can be seen in Fig. 2.2). These components are representations of the given image and can be used for facial feature extraction.

The first and very famous Gabor wavelets based system is elastic bunch graph matching (EBGM) proposed by Wiskott et al. in [119] where face images are represented as labeled graphs, which are generated by utilizing a special data structure called bunch graph (see Fig. 2.3 for more details) to collect information from Gabor jets (all Gabor

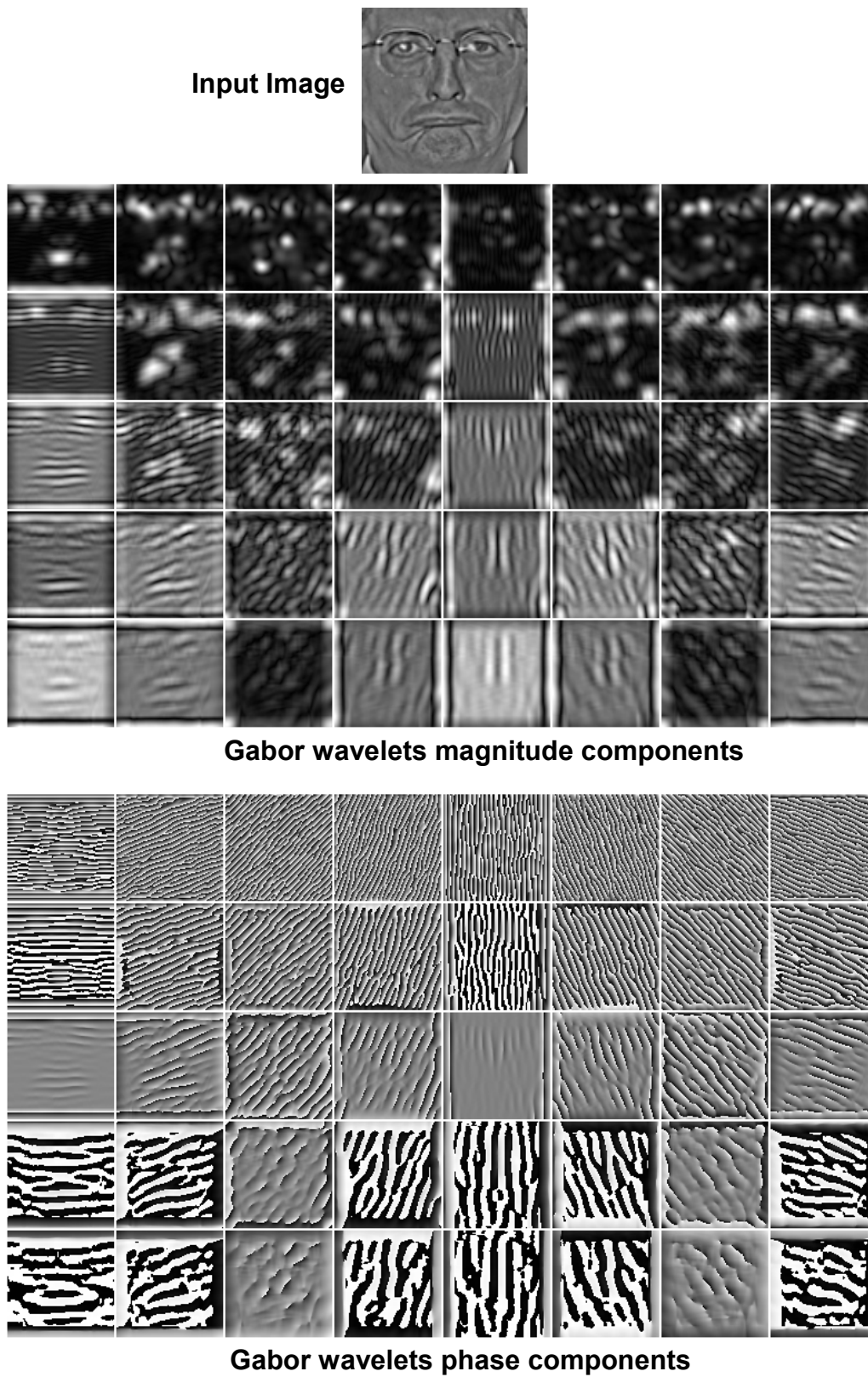


Figure 2.2: An image and its Gabor wavelets components (5 scales and 8 orientations)

2.2. Local feature based approaches

wavelets convolution values at a landmark location of local facial features). **EBGM** method yields a very encouraging recognition rate of 98% on the frontal subset (300 samples) of FERET database [96], the most popular face database whose details will be described in chapter 3, using a single sample (image) per person (SSPP) in the training stage.

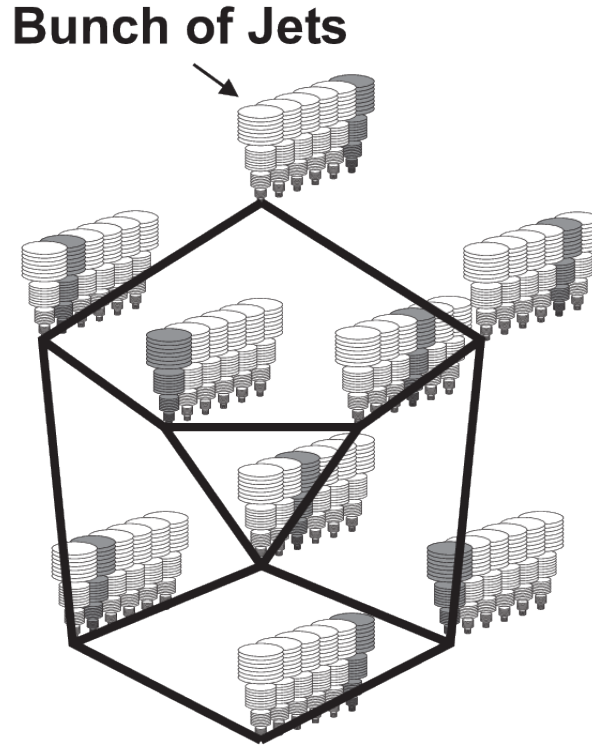


Figure 2.3: Bunch Graph structure proposed in [119].

Liu et al. [74] introduced augmented Gabor feature vector and used enhanced Fisher linear discriminant model (EFM) model [73] to form Gabor-Fisher-classifier (GFC) method for FR. The augmented feature vector is formed by first downsampling Gabor wavelets parts by a factor of 64, then normalizing and concatenating all of them as a whole representation. They showed that GFC could achieve very good results on a subset of 200 subjects (frontal images) of FERET database and is robust to illumination and facial expressions variations. The same technique was used in [72] to build Gabor feature representations, which were then projected into a sub space generated by Kernel Principal component analysis (KPCA) [101] with Fractional Power Polynomial Model. In [29], Deng et al. used Labeled-Graph for face representation by combining sampled Gabor magnitude values. The authors then applied whitened Principal component analysis (WPCA) for dimension reduction and cosine distance for classification to obtain encouraging performance on four standard frontal subsets of FERET database. Pang et al. [93] presented a two-fold method based on Gabor wavelets and Linear Discriminant Analysis (LDA-another name of FLD [35]). They used LDA directly on input image to

extract LDA features and then on Gabor features generated from selected discriminant pixels to produce Gabor-LDA features. Both types of features are finally fused to build a sum rule based classifier. The results on a FERET's subset of this method were higher than PCA and LDA.

Despite their promising results, above methods have an inherent drawback: they need huge calculation time for producing Gabor wavelets components to carry out feature extraction task.

2.2.2 LBP based methods

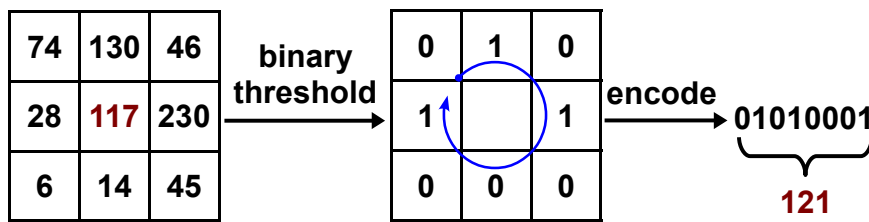


Figure 2.4: LBP encoding scheme

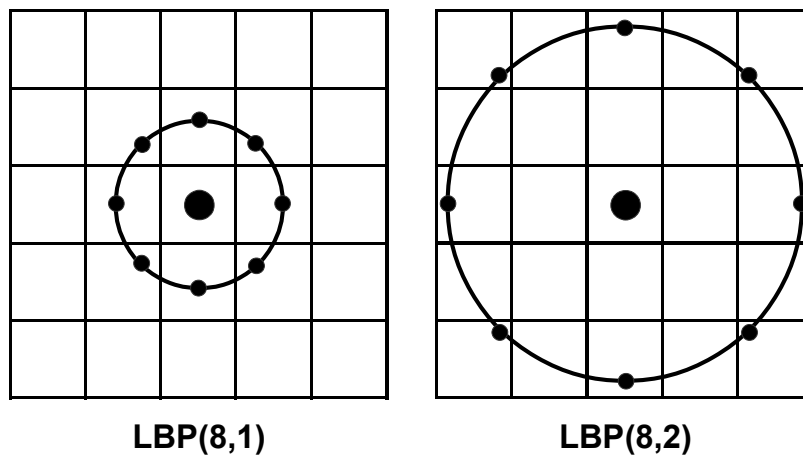


Figure 2.5: LBP patterns

Initially designed for texture classification problem as a texture descriptor, Local Binary Patterns (LBP) [91] has been quickly become one of the most popular features in FR literature. In original LBP [3], every pixel of an input image is assigned with a decimal number (called LBP label) which is computed by binary thresholding its gray level with its P neighbors sparsely located on a circle of radius r centered at the pixel itself. A bilinear interpolation is exploited to calculate the neighbor pixel values if they are not at the center of a pixel. This encoding scheme is called LBP operator and denoted as $LBP(P, r)$ (for more details, see Fig. 2.4 and Fig. 2.5). Applying LBP operator on

2.2. Local feature based approaches

every pixel of a face image produces a LBP image (Fig. 2.6 shows some samples) that contains very important information for FR: the local micro facial textures.

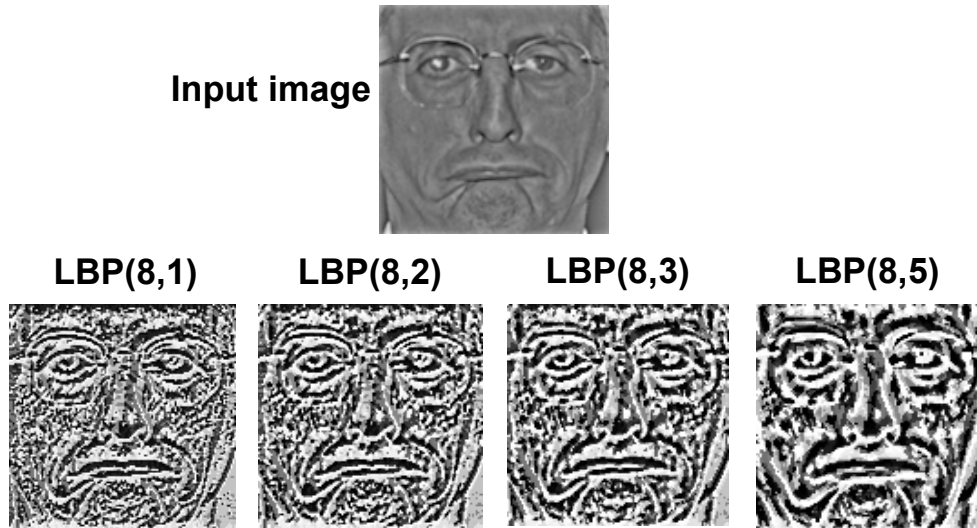


Figure 2.6: A face image and its LBPs

The LBP image obtained by LBP operator is then divided evenly into $W \times H$ ($3 \leq W, H \leq 9$) non-overlapped rectangular subregions to calculate their histograms. The LBP feature vector of the given face image is built by concatenating those histogram sequences. In doing so, LBP vector incorporates useful spatial information (spatial distribution of facial features in different subregions), that is widely known to have a key role in FR [137].

Basically, a LBP vector is a 2^P -bins representation, but statistical studies of LBP labels on different kinds of images revealed that some bins are more frequent than the others. Those principal bins, whose binary forms have no more than 2 bit transitions from 1 to 0 and vice versus, are called uniform patterns and are used to reduce the LBP feature vectors' size [3]. This compression operation results in shorter feature vectors and thus makes the classification faster, but on the contrary it causes a small decrement in accuracy performance. All of these steps can be seen in Fig. 2.7. It is worthwhile noting that LBP descriptions are usually formed with $P = 8$ and as a result, they are 59-bins feature vectors. Ahonen et al. [3] used LBP method to extract micro features of facial images and then used template matching for classification and got very promising results. Other applications of LBP (related to face recognition) include face detection [44], facial expression recognition [33], age estimation [57], gender classification [68], face spoofing detection [79], etc. But after all, LBP method was most successfully applied to face recognition. The merits of LBP are simple computation, small feature vector's size (in comparison with Gabor wavelets based methods) and robustness towards illumination variations.

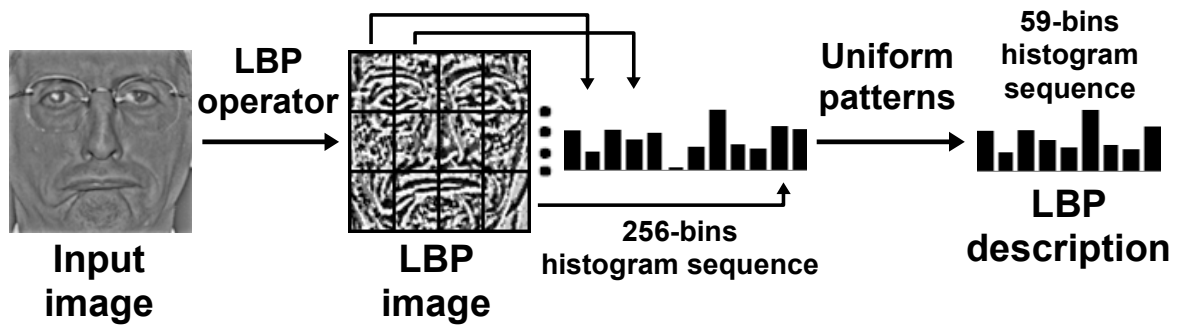


Figure 2.7: LBP description calculation

Following its first successful application [3], numerous variants of **LBP** have been proposed for face recognition in recent years. A boosting **LBP** was introduced by Zhang et al. in [134]. For one face image, over 7000 Chi Square distances of **LBP** patterns generated by shifting and scaling sub-window over the given image, are calculated. Adaboost [37] is next applied to select the most efficient **LBP** features. Boosting **LBP** gains higher overall **RR** than **LBP**. In [71], Multi-scale Block **LBP** (**MB-LBP**) is formed by using block regions instead of single pixel from input images. **LBP** can be considered as a special case of **MB-LBP** when block region is one pixel. **MB-LBP** encodes both micro-structures and macro-structures of face image and therefore provides a better representation for face images. Improved **LBP** (**ILBP**) is proposed in [55], the authors thresholds surrounding pixels of each pixel with theirs mean gray value. **ILBP** is proved more effective than **LBP** in face detection. In [46], Heikkilä et al. compared center symmetric pairs of pixels to form Center Symmetric **LBP** (**CS-LBP**). **CS-LBP** captures both micro features and gradient features of face images. **CS-LBP** feature vector' size is half of **LBP** feature vector' size using the same circular pattern. This technique was expanded by Choi et al. [22] when pairs of symmetric pixels in different orientations and various radii were compared to build up the Circular Center Symmetric-Pairs of Pixels (**CCS-POP**) representation. This way, **CCS-POP** captures pixel-wise local edge information and obtains higher accuracy than **LBP** when combining with Partial Least Squares (**PLS**) [102] for dimensionality reduction.

Based on a ternary threshold operator, Tan et al. [112] proposed a new **LBP** variant called Local Ternary Patterns (**LTP**) by using two **LBP** vectors for building one **LTP** description. **LTP** was verified to be more efficient than **LBP** against illumination and noise conditions. The downside of this approach are two times slower in speed and bigger in feature vector size. Further, the idea of the **LTP** was extended by Ren et al. [97] with a new variant called Relaxed **LTP** (**RLTP**). The authors used four **LBP** labels for encoding each image pixel and then they were accumulated in one **LBP** histogram of a **RLTP** vector. **RLTP** was claimed to bring improvement to **LTP** when dealing with image noise. The concept of applying multiple elliptical patterns in **LBP** on weighted facial regions was used by Liao et al. [70] in their Elongated **LBP** method. By using

2.2. Local feature based approaches

weighted factors for six regions of the face image and four different elliptical patterns (in four directions), Elongated LBP was argued to encode the anisotropic information of the image. While achieving better results than LBP, Elongated LBP had a shortcoming since its feature vector's size was four times longer than that of LBP.

Rather than digging into the fixed and predefined sets of neighbor pixels located on different patterns (circles, ellipses, symmetric pairs) or into threshold algorithms (binary, ternary or relaxed) as in above LBP variants, Maturana et al. [82] exploited a supervised learning technique to seek for most discriminative neighborhoods when computing the LBP label for one image pixel. This is done based on maximization of a Fisher-like class separability criterion. Although the method, named as discriminative Local Binary Patterns (DLBP), gained promising results, its computational cost for learning process is a real weakness. In [117] a descriptor called patterns of oriented edge magnitudes (POEM) was developed by applying multiple LBPs on accumulated magnitude images. The authors then combined POEM with patterns of dominant orientations (PDO) [118] and achieved better results.

Notwithstanding the fact that there is an abundance of its variants, LBP is still widely used in many multi-resolution/multi-scale feature extraction methods. This is rooted in its simplicity, computation effective and its compact representation for each input image it gives.

Originally, LBP is designed for *texture classification* problem and it is turned out that it has desirable properties for being an efficient facial representation in FR. While other variants of LBP tried to improve its powerfulness by using learning techniques (such as in [134, 82]) or different mechanisms in the thresholding step (for examples, MB-LBP, CS-LBP, and LTP), they tend to leave behind the fact that their main objective is for FR problem. For this goal, any kind of inspiration of a LBP variant must be based on aspects that evidently lead to higher accuracy results. Guided by this rule, Elliptical Local Binary Patterns (ELBP), one of our propositions in this work, is a LBP variant and is emanated from following observations of face images:

- Crucial facial components, eyes and mouth, are naturally elliptical. Plus, human faces contain more horizontal structures, which play important role in memorizing and recognizing faces [106, 40], than the local ones. Thus, the horizontal elliptical patterns are more efficient and more relevant than circular ones.
- When combining both horizontal and vertical information, the accuracy performance is improved [40]. So, instead of using just single horizontal ELBP description, we propose to fuse it with its vertical counterpart to enhance the discriminative of the resulted representation.

The details of ELBP as well as proofs to prove for its efficiency is described in chapter

4 of this document.

Further, employing [ELBP](#) as an primitive description, we constitute more advanced methods, Elliptical Patterns of Oriented Edge Magnitudes (EPOEM) and Local Patterns of Gradients (LPOG), by applying it on gradient based images. These two methods are presented in chapter 6 of this thesis.

2.2.3 LPQ based methods

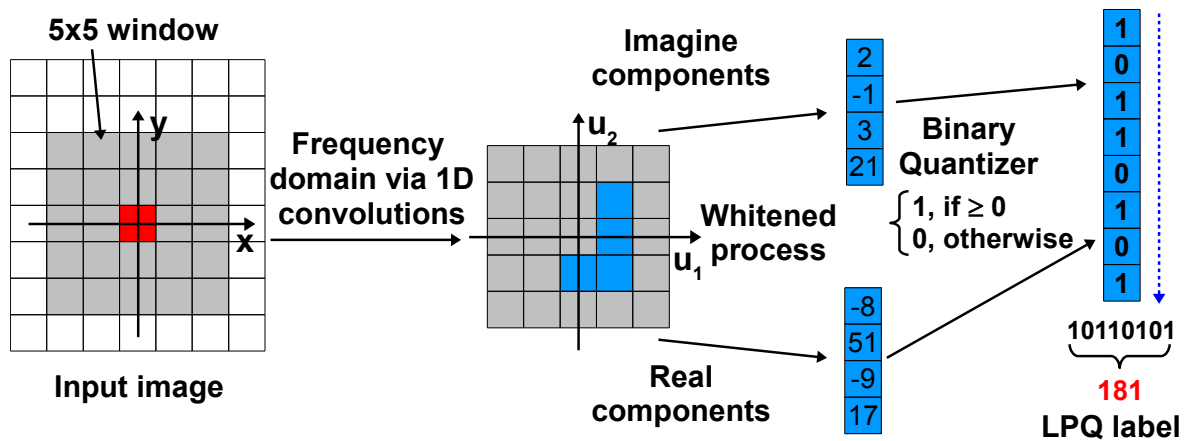


Figure 2.8: LPQ encoding scheme

Very recently, Local Phase Quantization ([LPQ](#)), a blur tolerant texture descriptor [92], has been further investigated for FR. In texture classification problem, LPQ gains better performance than Gabor wavelets based and LBP methods, particularly when working with blurred images. Having experiences with the case of LBP (also initially developed for texture classification but quickly gained its best notoriety as a feature extraction method in FR), many researchers have further investigated the use of LPQ for FR. While being proved to be blur insensitive [92], LPQ additionally reported promising results when dealing with blurred face images recognition [5].

Based on the blur invariance characteristic of the phase spectrum of image in the frequency domain, LPQ operator on an image pixel is done by using Short-term Fourier transform ([STFT](#)) over a window of size $M \times M$, whose centre is the image itself, with four scalar frequencies. Four imaginary components and four real components are then whitened based on a parameter ρ before being binary quantized to obtain the LPQ label for the given pixel. The process of applying a LPQ operator $LPQ(M, \rho)$ upon an image pixel is demonstrated in Fig. 2.8.

After employing a LPQ operator to produce a LPQ image (Fig. 2.9 shows examples) from input face image, Ahonen et al. [5] exploited the same technique as in [3] for building LPQ face representation by concatenating histogram sequences of LPQ image's

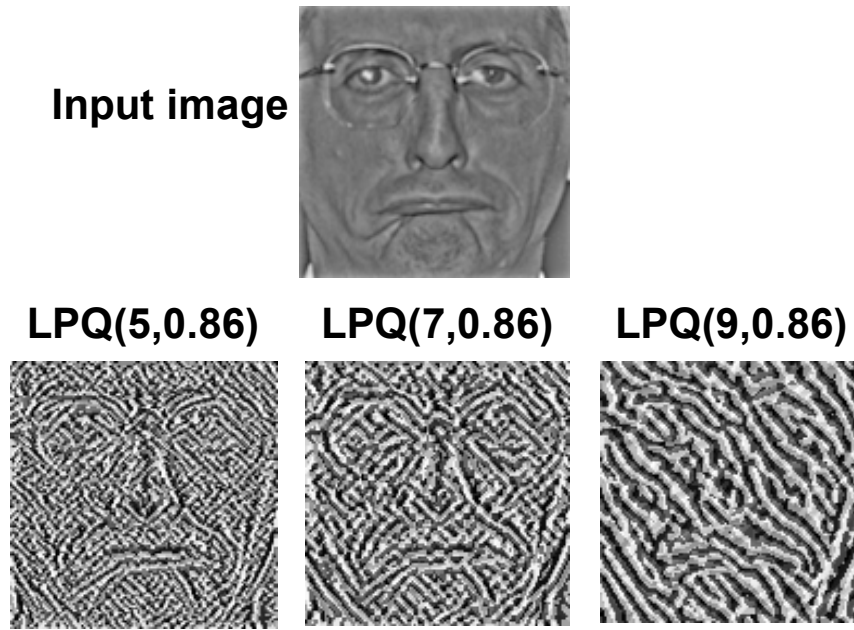


Figure 2.9: A face image and its LPQs

rectangular sub-regions. But unlike LBP [3] feature vector, whose size can be reduced efficiently by employing uniform patterns, each LPQ vector is a 256-bins description. Hence, each LPQ vector is about four times longer than an LBP vector with 8 neighbor pixels (a 59-bins representation), the most used LBP operator, when using the same divided sub-regions.

As far as we know in FR field, there has not been many variants of LPQ that have come up after its first appearance [5]. In LPQ, the magnitude information is not used. Taking into account both magnitude and phase features of images obtained from STFT, Lei et al. [63] presented a novel method called Local frequency descriptor (LFD), which could be considered as a LPQ variant, for low resolution FR. The same encoding technique as in LBP operator is applied on magnitude image whilst a binary qualification is used upon phase image for generating two encoded images. LFD feature vector is then built by concatenating sub-regions' histogram sequences of those images. This method brings higher performance than both LBP and LPQ under low resolution challenge.

As verified in [5], LPQ, when being used in a *Template matching* based FR scheme, is more robust than LBP for dealing with blur, illumination and facial expression variations images, but in FR literature, LPQ has not hitherto received the reputation it deserves and has been usually overshadowed by LBP and its variants. In this dissertation, we will show that LPQ is more efficient than LBP and its variants against all challenging issues and when engaging with WPCA, it outperforms many other state-of-the-art FR systems. Moreover, motivated from its efficiency against FR challenges and by delving further into Monogenic filter's components and gradient images, we exploit LPQ to constitute

two novel facial feature extraction methods: Patch based LPQ of Monogenic components (PLPQMC) and Local Patterns of Gradients (LPOG), which will be described in details in chapter 5 and 6 of the present thesis, respectively.

2.2.4 Multi-resolution/multi-scale methods

2.2.4.1 Simple multi-scale approaches

An intuitive and straightforward approach to enhance feature extraction performance is to use multiple elementary descriptors, such as LBP, LPQ and their variants, by varying associated parameters on the same input image. More specifically, circular patterns of various radii are used with LBPs and its variants while different windows sizes are employed for LPQs. The methods in this kind of approach are simple and easy to implement, but on the contrary, their performance improvements may not worth the computational cost and the memory required when pursuing them. Multi-scale is the general name for these methods.

Following this direction, in [17], multi-scale local phase quantization (MLPQ) was proposed by applying multiple LPQ operators of different filter's size and aggregating corresponding LPQ vectors into a final multi-resolution description. MLPQ feature vectors are then projected into a LDA subspace for FR. MLPQ has recently been fused with multi-scale LBP (MLPB) in [18]. Chan et al. [18] used kernel discriminant analysis (KDA) to improve recognition performance. MLPQ LDA [17] and MLPQ+MLBP KDA [18] gained impressive results but they have an obvious disadvantage: they need high computation time as using multiple LPQ operators (the authors used 7 ones in [17]) and multiple LBP operators (in [18]). Another MLPQ based method was introduced in [110] with linear regression classifier (LRC) for classification process. Some noticeable remarks about these multi-scale LPQ based systems are: their performance (MLPQ LDA [17] and MLPQ+MLBP KDA [18]) is not better than leading LBP, Gabor wavelets based methods (such as [104, 107, 52, 127, 16], see comparison tables in chapter 6 for more details), or they lack comparisons with other state-of-the-art systems on large public databases (LPQ [5]) so there is not enough evidence that these systems are really efficient and reliable at coping with challenging conditions of FR.

2.2.4.2 Gabor wavelets components based methods

A more efficient multi-resolution/multi-scale scheme for feature extraction than the simple general multi-scale approach described in previous section is illustrated in Fig. 2.10. In the methods complying to this model, an image multi-resolution/multi-scale analysis technique, such as Gabor wavelets (the most popular) or Monogenic (recently used),

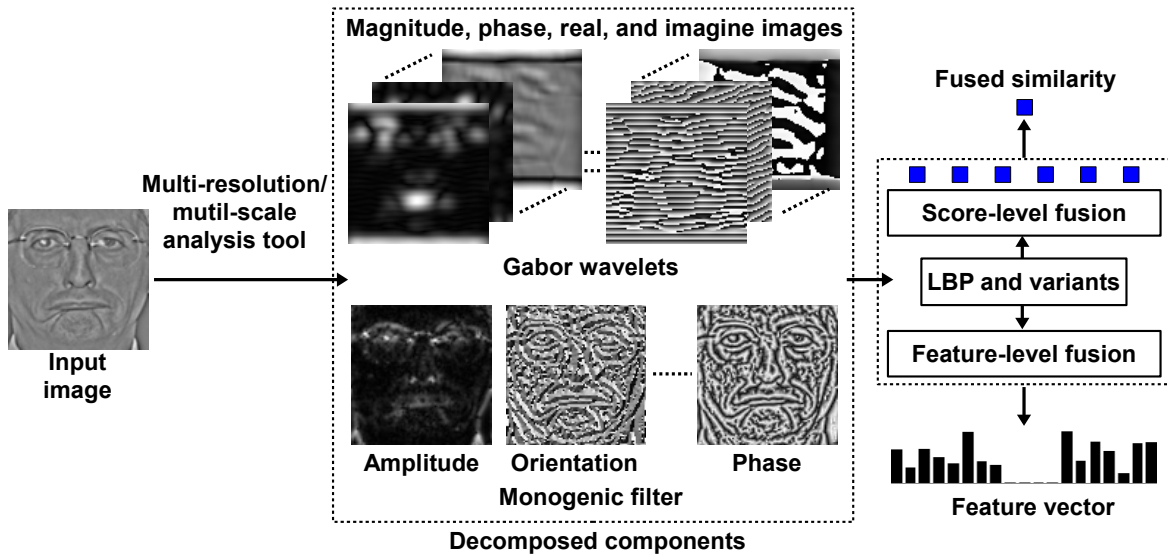


Figure 2.10: General multi-resolution/multi-scale feature extraction scheme

is first employed to decompose a face image into multiple components in the form of independent images. The number of images depends on the methodology in which they are used, for example Gabor wavelets based methods usually generate 40 complex parts at 5 scales and 8 orientations to encompass enough information from an input image, but we can generate 160 Gabor wavelets images (40 real images, 40 imagine images, 40 magnitude images, 40 phase images) in total from these 40 complex parts. Next, **LBP** and its variants are applied upon those component images to extract useful facial features for **FR** and result in different descriptions, each for one image. To combine all these separate descriptions, there are two fused strategies: score-level fusion and feature-level fusion. By score-level fusion methods, fused similarities between test images and gallery images are computed based on different scores and are used to determine the identities of the test images. In feature-level fusion algorithms, global feature vectors are obtained by incorporating all the descriptions from previous step. Consequently, these vectors are high dimensional with much redundant information and need to be projected into a subspace before proceeding the classification stage in an efficient manner.

Since its debut in FR literature [3], many researchers have attempted to combine LBP with Gabor wavelets by employing a multi-resolution/multi-scale model as mentioned above for improving recognition performance. Local Gabor binary pattern histogram sequence (**LGBPHS**) [135], ensemble of piecewise FDA (**EPFDA**) based on spatial histograms of local Gabor binary patterns [104], histogram of Gabor phase patterns (**HGPP**) [133], system in [113], fusing local patterns of Gabor magnitude and phase (**FLPGMP**) [107] and Gabor surface feature (**GSF**) [123] are the most representative methods. Following the feature-level fusion strategy, **LGBPHS** [135] vector is formed by using LBP operator on 40 Gabor magnitude pictures. The authors [135] then used

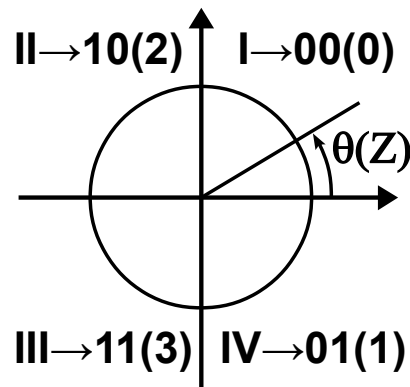


Figure 2.11: Quadrant bit coding scheme

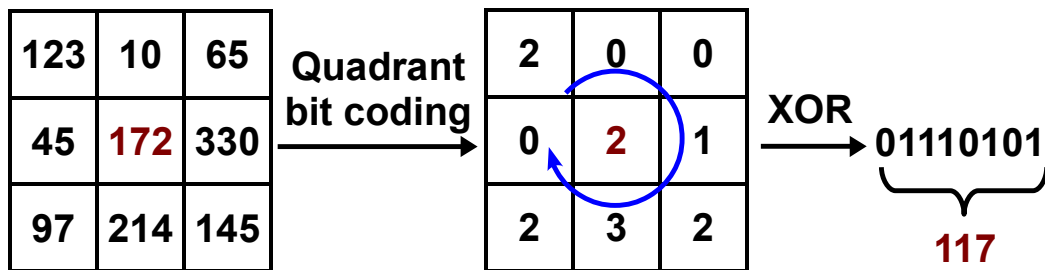


Figure 2.12: LXP encoding scheme

template matching with Chi Square distance for classification and attained good results on FERET [96] and AR [80] databases. To avoid direct FDA on very large size feature vectors, EPFDA [104] partitions each image into small blocks which are further divided into sub-blocks where Gabor filters are applied to produce feature segment of each block. Ensemble FDA training processes are done on these feature segments to build up FDA subspaces in which each face image is constituted as a sequence of projected feature segments. A sum rule for combining individual classifiers on projected feature segments is utilized in the classification stage. In HGPP [133], a quadrant bit coding scheme was first proposed to assign each phase value by a 2-bit number from 0 to 3 (for more details, see Fig. 2.11). Next, for extracting features from Gabor phase images, a LBP alike descriptor named local XOR pattern (LXP) (see Fig. 2.12 for more details) was introduced by applying XOR operator on quadrant bit codes. Then LXP was used on real and imaginary parts of Gabor phase images (the authors used totally 90 images) to encode both global and local Gabor phase patterns. High results of HGPP show that Gabor phase information also plays an important role in FR. Tan et al. [113] proposed a feature-level fusion method to fuse LBP with Gabor wavelets features. They first used PCA for reducing LBP and Gabor wavelets representations and then applied Kernel Discriminative Common Vectors [15] to project fused feature vectors into discriminant subspace for proceeding the classification task. In FLPGMP [107],

2.2. Local feature based approaches

the definition of **LXP** and **LBP** were used to exploit both Gabor wavelets magnitude and phase information, respectively. The resulted magnitude and phase feature vectors were then fed into a block-based **FDA** (**BFDA**) procedure to shorten their lengths and remove unnecessary information they carried. Both score-level (by a sum rule formula) and feature-level fusion strategies were assessed on projected vectors [107], and at the end, the former outperformed the latter by providing excellent results on **FERET** and other face databases. **GSF** [123] method uses **LBP** on combined maps of Gabor magnitude images and their 1st, 2nd derivatives. The **EPFDA** [104] was used to reduce **GSF** feature vectors' lengths and weighted scores based on cosine distances were computed for classification stage. **GSF** [123] achieved state-of-the-art results on **FERET** database.

Most recently, some novel Gabor wavelets based methods have been proposed and achieved very significant results. In [52], Hussain et al. propose a new feature extraction method so-called Local Quantized Patterns (**LQP**) by using vector quantization and lookup table to build facial description upon Gabor wavelets images. Having advantages over many existing **LBP** and Gabor wavelets based methods, **LQP** yields excellent performance for both face identification and verification, when incorporating with PCA and cosine metric. Statistical local features (**SLF**), a novel facial representation, has been proposed by Yang et al. [127]. The authors use a multi-partition max pooling technique for enhancing the invariance of **SLF** to image registration error first. After that, a kernel based representation model is adopted to thoroughly exploit discriminant features embedded in **SLF**. A **FR** framework named SLF-based robust kernel representation (**SLF-RKR**) is also proposed then. Extensive experiments show that **SLF-RKR** (using Gabor magnitude based **SLF**) is robust to occlusions and gains superior results in comparison with state-of-the-art systems, except when it has to face with pose variations challenge, even on small pose angles images (within the range $\pm 25^\circ$). Another novel Gabor wavelets based feature extraction method called Gabor ordinal measures (**GOM**) is introduced in [16]. Ordinal measures, which reflect the ordering relationship information between multiple variables (intensities or feature values), are used to encode facial feature from magnitude, orientation, real and imagine images of 40 Gabor wavelets components (90 images in total). Each **GOM** feature vector is refined by a block based partition strategy to have 5760 dimensions. **LDA** algorithm is used to further reduce **GOM** vectors' sizes and a sum rule score-level fusion of cosine distances is exploited in the classification. The results provided by **GOM** are very amazing but not higher than those of **SLF-RKR** [127]. Additionally, a downside of **GOM** is that it is relatively slow, when spending about 700ms for processing *one* face image.

The applications of **LBP**, **LXP** and other proposed techniques on Gabor wavelets images (magnitude, phase, etc.) in above multi-resolution/multi-scale feature extraction methods provide considerable performance as well as effectively reducing the size of resulted feature vectors but even with these latest attempts, the heavy computational

cost of this approach remains an unsolved problem. For real time systems, such as in video surveillance context, where computational speed is a primary objective, this is a *must be solved* issue.

Another important observation from the best results of the above cited works is that: a plain feature extraction method that works solely on intensity images (such as LBP, LPQ and their variants), even with the tuned parameters and at its best, is not sufficient to meet the requirement of a high accuracy and reliable system. By some kind of way, a robust facial representation must contain useful features that are subtly extracted from multi-resolution/multi-scale components. In the mean time, it should not be suffered from expensive computational cost.

2.2.4.3 Monogenic filter based methods

Recently, Monogenic filter [32], a multi-scale image analysis tool based on log-Gabor wavelets, has been used in FR (Monogenic Binary Pattern (MBP) [130] and Monogenic Binary Coding(MBC) [128] are examples) since it does not need huge calculations as Gabor wavelets while having good performance. Given an input image, a Monogenic filter will generate multiple component images of different types, including amplitude (also called as magnitude), orientation, phase and bandpass (see more in Fig. 2.10). The number of such images is established based on the number of scales used, and this one is usually set as 3 or 4, which leads to at most $4 * 6 = 24$ images. As a consequence, the Monogenic filter based feature extraction methods are more cost effective than the ones based on Gabor wavelets, with regard to both memory and computational aspects. Another advantage of Monogenic filter over Gabor wavelets is that its components preserve more image information than those of Gabor wavelets, which can be clearly seen in Fig. 2.10.

In [130], MBP representations are built by using LBP on Monogenic magnitude images, quadrant bit coding scheme on orientation images to generate MBP maps at 3 scales. A weighted intersection metric is used on MBP vectors to calculate the similarities between test and gallery images for classification. This way, MBP [130] exhibits higher RRs than LGBPHS [135], HGPP [133] but requires less computational cost and memory space. Further, in [128], Monogenic Binary Coding(MBC), a combination of applying LBP on Monogenic amplitude, LXP on Monogenic phase and the quadrant bit coding scheme on bandpass components, is proposed. BFDA [107] is again used to make MBC vectors more compact and a sum rule fusion tactic is applied to constitute a FR system named MBC-F. As shown in [128], MBC-F offers competing results with other leading edge Gabor wavelets based method while attaining a more cost effective property. These encouraging results, in our point of view, could open the door to many other FR researches based on Monogenic filter.

2.2. Local feature based approaches

Based on advantages of Monogenic filter in building robust facial representation, as proved in prior cited works, and based on benefits from LPQ's useful properties for FR, a novel multi-scale feature extraction method called Patch based LPQ of Monogenic components is presented in chapter 5.

2.2.5 Sparse representation based methods

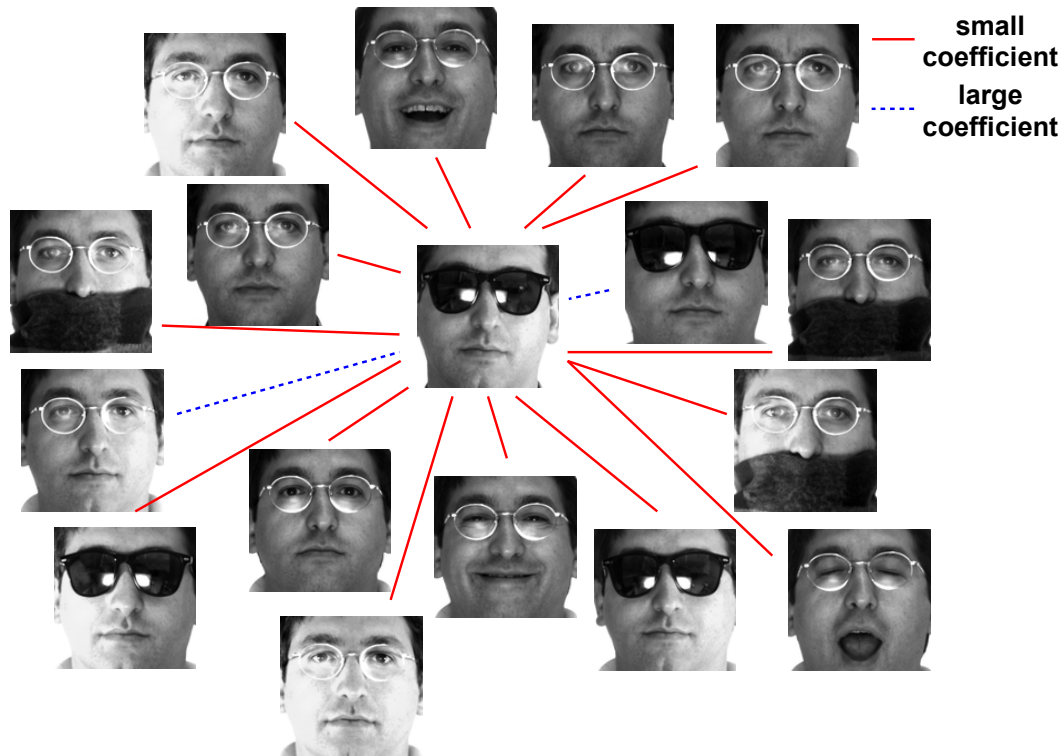


Figure 2.13: An image and its sparse representation

Recently, sparse representation has become a new kind of approach that attracts ever increasing attention from FR community. In a sparse representation-based classification (SRC) FR system, a test image is represented as a sparse linear combination of training images (Fig. 2.13 shows an illustration). This is done via an optimal problem whose sparsest solution can be found by solving an equivalent l^1 - Minimization problem [122]. In the classification stage, each test image is assigned to the label of training image whose has min l^2 - norm with it in the generated sparse feature space.

Through a SRC FR system [122], Wright et al. pointed out that the problem of choosing the number of features for classification could be completely solved if the sparsity of the representation is properly computed. SRC [122] was shown to be able to yield noticeable results against occlusion and corruption. This conclusion is consistent with two other sparse coding based systems: extended SRC (ESRC) [30] and structured sparse error

coding (SSEC) [66]. ESRC [30] extends SRC by using an intraclass variant dictionary to depict the variation that may appear between training and probe images. In SSEC [66], a morphological graph model and an exponential probabilistic model are used for error support structure and error distribution structure, respectively. Applying sparse representation-based classification (SRC) with features extracted by using LPQ descriptor was the idea of LPQ+SRC facial expression recognition system in [138]. Beside achieving good results with occlusion and corrupted test images (when using multiple samples per person training sets), there is no evidence that a sparse representation-based FR system can outperform other leading edge ones based on Gabor wavelets and LBP in general. Additionally, a drawback of these sparse coding based methods is that they require multiple samples (at least 4 images) per person for the training stage. This prerequisite is actually not always fulfilled, even impossible, particularly in real-world situations.

2.2.6 Other methods

Some other local descriptors, such as Scale Invariant Feature Transform (SIFT) [77] and Histograms of Oriented Gradients (HOG) [26], have been commonly used in many real-world applications due to their efficient computations, resistance to partial occlusions, and being relatively insensitive to viewpoint changes. Even though SIFT and HOG have been evidently proved as two of the best methods for encoding edge or local shape information, there are not many contributions of them in building robust FR systems. Bicego et al. [13], Rosenberger and Brun [98] reported that the usage of SIFT for face authentication could yield promising results upon the BANCA [8] and AR [80] databases respectively, but no further results on bigger databases were published. According to [83], HOG features' performance on the FERET database is worse than that of LBP and Gabor wavelets. In summary, these evidences mean that SIFT and HOG can not pave the way to a robust facial feature extraction as our expectation with this thesis.

As face images captured under video surveillance context are low resolution (LR) while gallery images are often of high quality, it is preferable to have test images with better resolution. To do this, super-resolution techniques are employed to produce high resolution (HR) images before conducting the feature extraction for the hope of improving system accuracy. We have investigated three well-known super-resolution methods, which use different algorithms for building a HR image from one or multiple LR input images:

1. Bicubic Interpolation: the output image pixel is a weighted average of pixels in the nearest 4×4 neighborhood

2.3. Conclusions

2. Sparse representation based super-resolution [124]: the different patches of the HR image are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal atoms. The principle of compressed sensing is then applied to correctly recover the sparse representation from the down-sampled input image.
3. Regression based method [58]: Here, the basic idea is to learn a mapping from input LR images to target HR images based on pairs of example images using kernel ridge regression. To remove the blurring and ringing effects around strong edges introduced by the regression, a model that takes into account the discontinuity property of images is used for post-processing.

Once output images of these methods are generated, we use them with Eigenfaces [114] and LBP [3] upon SCface database to assess whether some accuracy improvements are achieved or not. With Eigenfaces, the RRs are not improved. The overall RR improvement from LBP is negligible, about 0.3%. As verified in [140], even when combining a super-resolution method with the relationship between a HR image and one of its LR version in training set for building a better mapping of LR testing images, the RRs on SCface are very low (average RR is only 20.2%). These results underline that super resolution techniques are not the right solution, at least at present, to handle LR test images acquired by surveillance cameras. Additionally, a drawback of super resolution algorithms is that they come with a significant computational cost for producing HR images. As an open question, we think that the right way to deal with low resolution FR may be to degrade the resolution of HR gallery images instead of trying to enhance that of LR probe images.

2.3 Conclusions

Through this chapter, an up-to-date survey of the state-of-the-art of major facial feature extraction approaches with their most representative methods is presented. Started with Eigenfaces and Fisherfaces, the two most popular global algorithms, but almost the content of the chapter is about local feature extraction methods. From pure Gabor wavelets methods, LBP and many of its variant, LPQ based methods, multi-scale/multi-resolution methods based on simple approach and some well-known multi-resolution/multi-scale transformations, such as Gabor wavelets and Monogenic filter, to Sparse representation based methods, all are described with emphasis on their key ideas, advantages and limitations. Some other local descriptors and super-resolution based approaches are also considered. While analysing the pros and cons of these methods, we also provide the main concepts behind our proposed approaches in this thesis, that are:

- Using elliptical samples to capture micro textures from face images to form Elliptical Local Binary Pattern (ELBP). Combining both horizontal and vertical ELBP in building a facial representation with a richer feature set to improve the recognition performance.
- Applying LPQ operator upon Monogenic components to build a multiresolution/multiscale face description.
- Employing ELBP upon oriented edge magnitude images to form Elliptical Pattern of Oriented Edge Magnitudes feature extraction method.
- Integrating two kinds of local patterns, ELBP and LPQ, directly on gradient images to take into account meaningful characteristics of local features and benefit the advantages of the gradient images over the raw intensity one.

Chapter 3

Face recognition background

In this chapter, all the details of fundamental materials for the thesis are provided. The chapter describes three public face databases and their experimental protocols used for evaluating the accuracy performance of the proposed feature extraction methods. The face cropping algorithm based on eyes' coordinates is given in detail. Two preprocessing techniques for illumination normalization and two general face recognition frameworks are also presented.

3.1 General FR framework

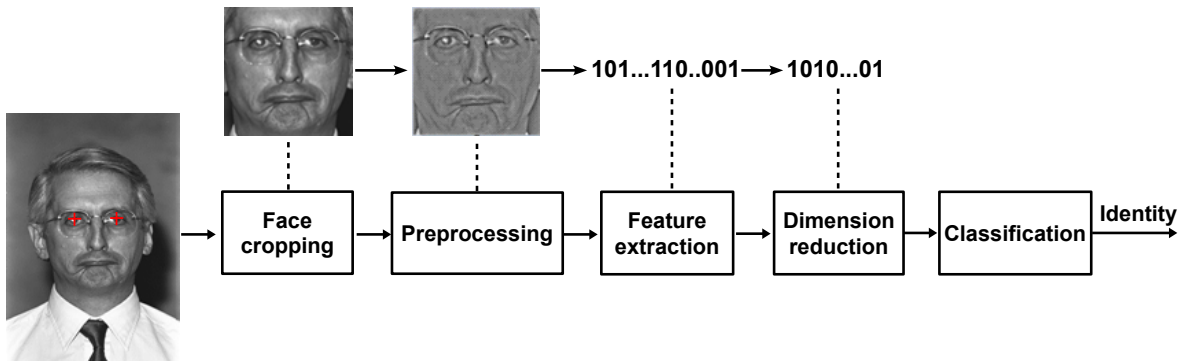


Figure 3.1: Stages in a local feature-based face recognition system

A general local feature-based FR system used in this thesis consists of several stages (see Fig. 3.1 for details). Face cropping based on eyes' locations from input image is the first stage, the image region that contains the face only is the result of this process. Next, the preprocessing stage normalizes those cropped face images to reduce the effects of illumination variations. Then a feature extraction method is applied to extract the most discriminative facial features from normalized face images. Once the feature extraction

stage is done, each face image is represented and stored as a high dimensional feature vector. Since a feature vector can be of very high dimension (ranging from thousands to hundreds of thousands dimensions) and conveys a lot of redundant information, a dimensionality reduction method is therefore needed to reduce the feature vector's length to a reasonable value and to eliminate unnecessary features. The dimension reduction stage also enhances the discrimination between face images of different people while descending those of images belonging to the same person. All the training images are used to generate the projection sub space in which the gallery and probe images are projected for recognition task. In the classification stage, a *k*-NN classifier and a distance metric are used to identify the identity of the probe image by assigning it to the label of the nearest gallery image.

3.2 Face databases

In order to rigorously evaluate the performance of a face recognition system, one has to use several public face databases with standard protocols. Results on one database is not adequate to conclude the reliability and the steady of the system while following standard experiments allows the comparisons between one FR system with other state-of-the-art ones. A good database should have a big enough number of subjects whose images are collected and used. More importantly, those images must be captured under many, if not all, variant conditions, such as illumination, facial expressions, occlusion, time-lapse, pose variations and video surveillance context.

With respect to above remarks, in this thesis, we consider three large public face databases namely *AR* [80], *FERET* [96] and *SCface* [42] to assess the accuracy and stability performance of all presented *FR* methods. Experiments on *AR* database are used to verify for the robustness of a *FR* system against illumination, occlusion, facial expression and time-lapse variations. Tests upon *FERET* database are for validating the system performance when coping with large scale (*FERET* has images from 1199 people) dataset under facial expression, illumination, time-lapse, and pose challenges. Different from these two ones, experiments performed upon *SCface* database is for investigating the effectiveness and efficiency of one *FR* system with low resolution images under unconstrained conditions. The database is more challenging as the probe images are of poor quality, small in size, have pose variation and are strong impaired by real lighting changes whilst the gallery images are high quality frontal ones acquired under controlled lighting conditions. Additionally, using single sample (image) per person (*SSPP*) training set is another purpose of all the tests we proceeded.

Besides images, each database has annotation data about the eyes coordinates of every face image. Based on that information, a simple face cropping algorithm is used to crop only the face region for carrying out experiments. Since images of *AR* and

3.2. Face databases

SCface databases are color, they are converted to gray scale format before being pre-processed by an illumination normalization algorithm. We report the results of all experiments in terms of rank-1 recognition rates (RR) and compare them with those of other contemporary systems.

3.2.1 AR (Aleix and Robert) database

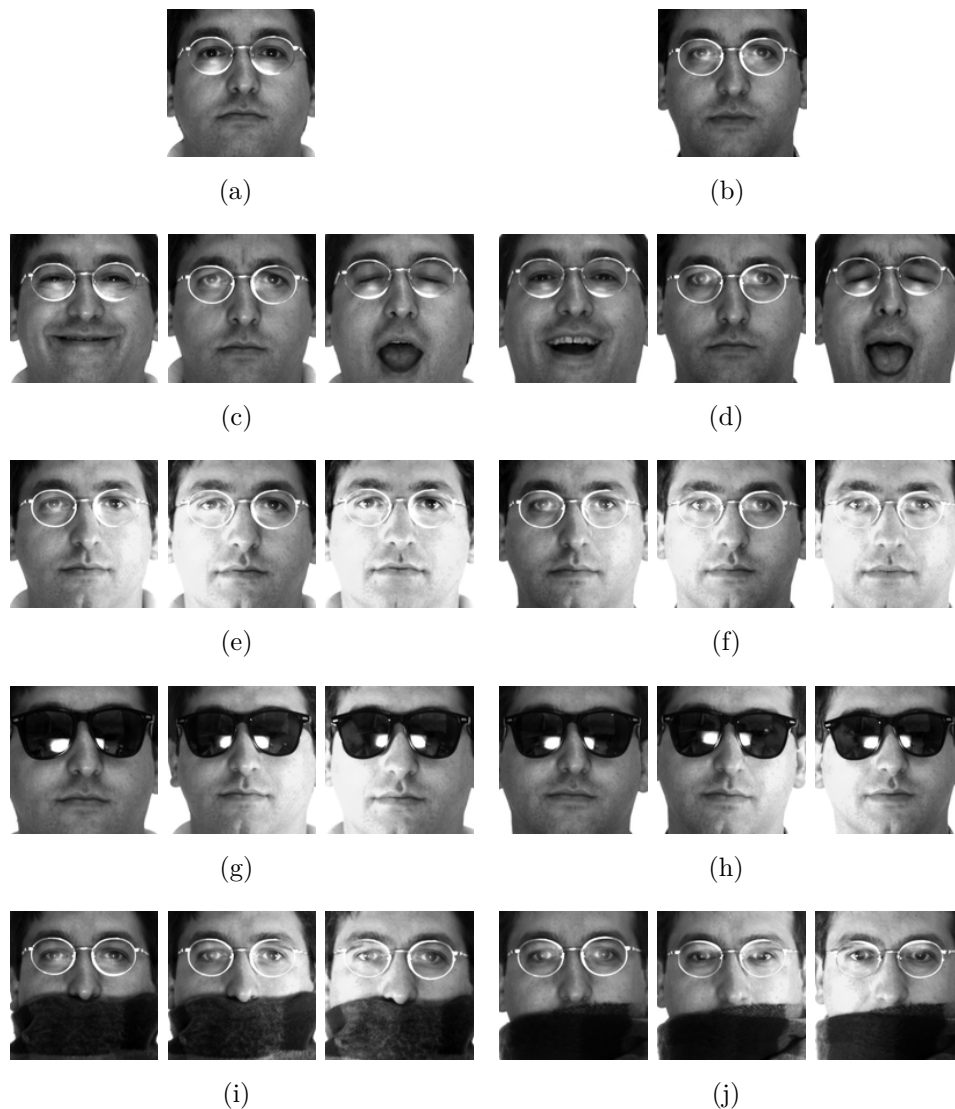


Figure 3.2: Sample cropped images from AR database.

The AR face database [80], created by Aleix Martinez and Robert Benavente, has over 4000 color face images from 134 people (75 men and 59 women). These images were collected under similar controlled conditions during two sessions, separated by 14 days (2 weeks), and were divided into 26 subsets with different facial expressions (smile, anger

and scream, see Figs. 3.2 (c-d)), illumination changes (right light on, left light on, and both sides light on, see Figs. 3.2 (e-f)) and occlusions (sun glasses and scarf, see Figs. 3.2 (g-j)). Because images in some subsets were missing or corrupted, we eventually have 1742 images in 13 subsets (each one has 134 images) of sessions 1 and 1534 images in 13 subsets (each one has 118 images) of session 2 for our experiments.

From these 3276 images, we set up four single sample per person face recognition experiments: experiment 1 (Exp 1) (134 images in reference and each probe set) uses all images from session 1, experiment 2 (Exp 2) (118 images in reference and each probe set) uses all images from session 2, experiment 3 (Exp 3) and experiment 4 (Exp 4) use images from both sessions. In Exp 1 and Exp 2 tests, the first neutral images (see Figs. 3.2 (a-b)) from each session are used for gallery set and all others corresponding images of the same session (see Figs. 3.2 (c-j)) are chosen for probe sets. Exp 3 and Exp 4 tests use first images of one session for the gallery and images of the other session for probe sets. Exp 1 and Exp 2 are conducted to investigate system performance under variations of expression, illumination and disguise whilst Exp 3 and Exp 4 are performed to validate each method against all that challenges plus time-lapse variation. Each experiment has 12 probe sets correspondingly named after their conditions as Smile, Anger, Scream, Neutral+Left light, Neutral+Right light, Neutral+both sides light, Sun glasses, Sun glasses+Left light, Sun glasses+Right light, Scarf, Scard+Left light, Scard+Right light. We label these probe sets by numbers from 1 to 12 for short notations. Each experiment uses one sample image (the first neutral image from each session) per subject for the training stage. All the images are cropped to 128x128 resolution and then preprocessed by retinal model [116] in order to remove the effects of illumination.

3.2.2 FERET (Face Recognition Technology) database

Frontal FERET image sets. FERET [96] is one of the most widely used face databases to evaluate the performance of a SSPP FR system since it has images captured under various conditions and from a large number of subjects (1199 people). The database has five frontal image sets namely Fa, Fb, Fc, Duplicate I (Dup I) and Duplicate II (Dup II) (see more at Figs. 3.3 (a-e)). Fa set, which is used for gallery set, has 1196 images of 1196 subjects. Fc, Fb, Dup I and Dup II sets, which consist of 1195, 194, 722 and 234 images respectively, are used for probe sets in the tests with the same names. The images of the Fb set are facial expression variations while Fc set contains images under different lighting conditions. Images in the Dup I and Dup II sets were taken about 1 year and two years respectively after the ones in the Fa set. Among others, Dup I and Dup II tests are more challenging as time-lapse is among the most difficult factors in the FR literature. The image size and preprocessing technique are identical to those applied upon images of AR database in previous subsection.

3.2. Face databases

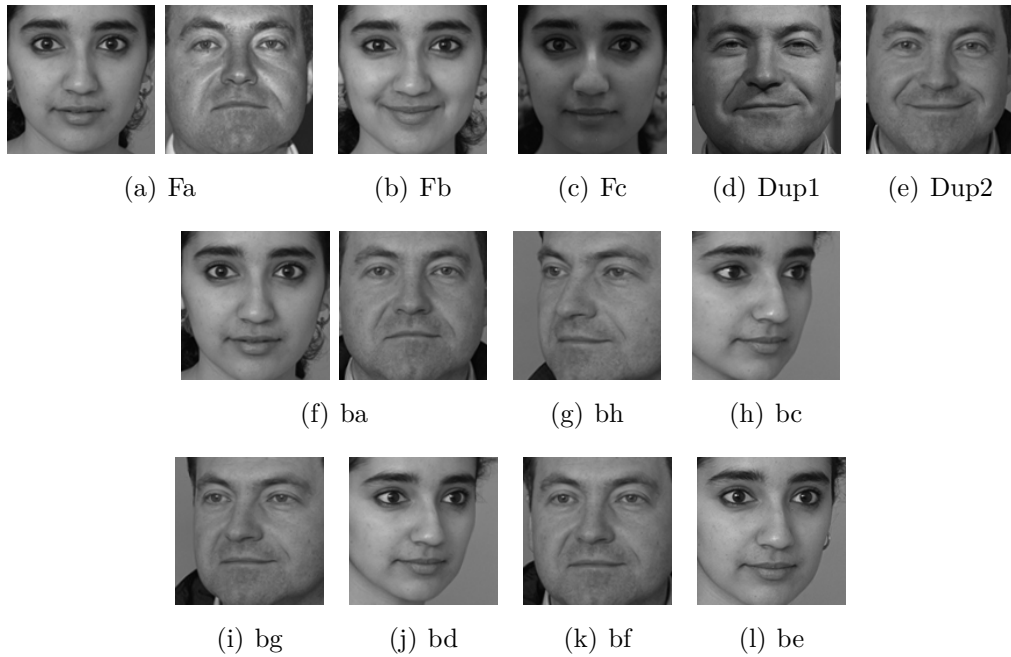


Figure 3.3: Sample cropped images from FERET database.

Non-frontal FERET image sets. Aside from frontal images, FERET has pose view images of 200 people (Figs. 3.3 (f-l) show examples). In this dissertation, we choose six subsets (each one contains 200 images) that have images with pose angles ranging from -40° to $+40^\circ$: **bh**, **bg**, **bf**, **be**, **bd**, **bc**, for probe sets. Frontal image set **ba** (200 images) is used for gallery while Fa set (1196 images) is used for training by WPCA. We also use the same image size and illumination normalization method as with frontal images. This experiment is used for verifying the performance of a system against pose variation, a major challenge of FR.

3.2.3 SCface (Surveillance Camera face) database

While previous experiments upon AR and FERET databases show adequate evidences to verify the predominance of one FR system based on a proposed feature extraction method over its related counterparts as well as other state-of-the-art systems, we give in this section the details of two experiments on SCface [42] database to validate the capability of our proposed frameworks against low resolution probe images.

The database contains color probe images of 130 people (Figs. 3.4 (a-h) show examples), which were captured by 7 different surveillance cameras named as cam1, cam2, cam3, cam4, cam5, cam6 and cam7, in which cam6 and cam7 are cam1 and cam5, respectively, but worked in infrared (IR) night vision mode (the rest worked in daylight condition). These cameras worked under *uncontrolled indoor conditions* at three distances of 4.2m,

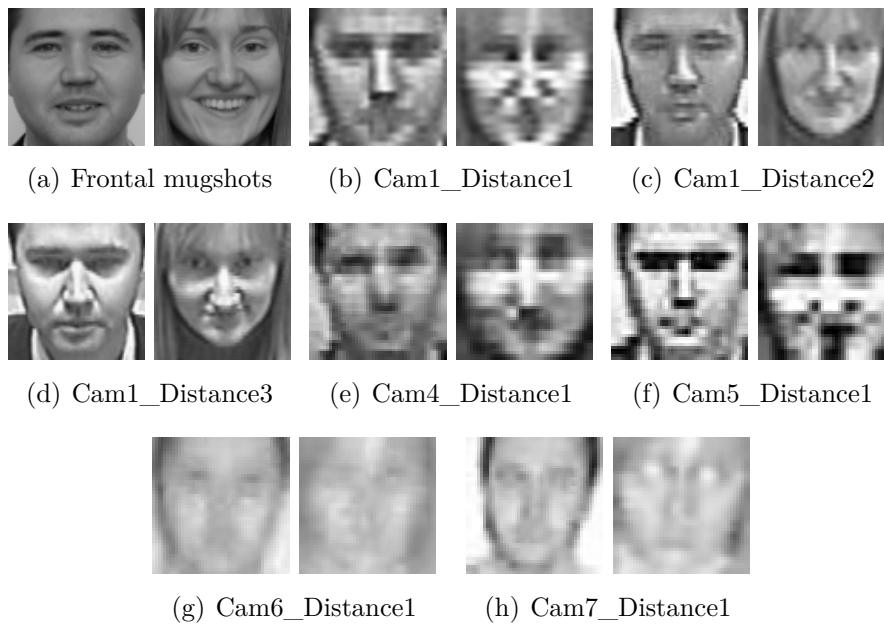


Figure 3.4: Sample images from SCface database.

2.6m and 1.0m, and each one has three image sets (of 130 images) corresponding to three distances. The challenge coming from its very low resolution probe images adversely affected by various unconstrained conditions is the cause of very few reported results on this database.

From 21 image sets, the authors [42] defined two experiments: DayTime and NightTime. DayTime experiment has 15 probe sets (see Figs. 3.4 (b-f)) while NightTime has 6 ones (see Figs. 3.4 (g-h)). The gallery set contains 130 high quality mug-shots (see Figs. 3.4a) that were taken under standard lighting indoor condition. In this work, we perform both experiments and report the results with two distinct training sets: the frontal Fa set of FERET database [96] like in [42] and frontal mug-shot images like in [140] and [89]. All the images are converted into grayscale format and then cropped (using the eyes' coordinates accompanied with the database) to have 48x48 pixels resolution. Next, standard histogram equalization algorithm is applied for illumination normalization.

3.3 Face cropping

The face cropping technique based on eyes' locations is depicted in Fig. 3.5. Considering (x_1, y_1) and (x_2, y_2) are left eye and right eye coordinates, the rotation angle for face alignment is computed as:

$$\phi = \text{atan}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) * \frac{180}{\pi}. \quad (3.3.1)$$

3.3. Face cropping

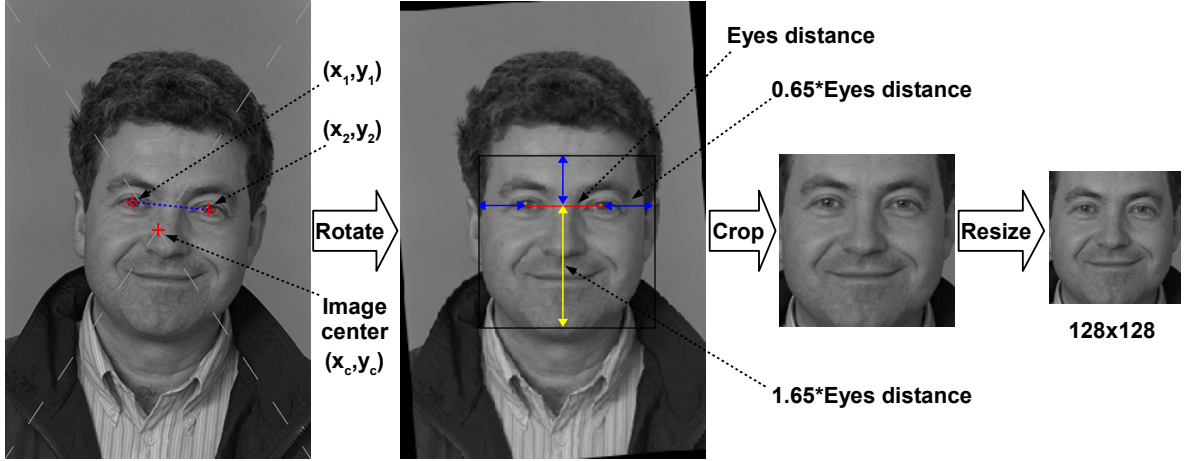


Figure 3.5: Face cropping based on eyes coordinates scheme.

The whole input image I of size $W \times H$ is then rotated by ϕ angle with the original coordinate, which is the center of the image itself, being calculated as:

$$\begin{cases} x_c = \frac{W}{2} \\ y_c = \frac{H}{2} \end{cases}. \quad (3.3.2)$$

The rotation operation at every pixel $I(x, y)$ at location (x, y) can be expressed by matrix multiplication as:

$$I_r(x, y) - [x_c \ y_c] = [(x - x_c) \ (y - y_c)] \cdot M_{rotate}. \quad (3.3.3)$$

Where $I_r(x, y)$ is the coordinates of $I(x, y)$ in the rotated image, M_{rotate} is the rotation matrix obtained from equation:

$$M_{rotate} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \quad (3.3.4)$$

Thus, new coordinates of two eyes are computed as:

$$\begin{cases} (x_{1r}, y_{1r}) = [x_c \ y_c] + [(x_1 - x_c) \ (y_1 - y_c)] \cdot M_{rotate} \\ (x_{2r}, y_{2r}) = [x_c \ y_c] + [(x_2 - x_c) \ (y_2 - y_c)] \cdot M_{rotate} \end{cases}. \quad (3.3.5)$$

Next, the eyes distance is calculated:

$$dist_{eyes} = \sqrt{(x_{1r} - x_{2r})^2 + (y_{1r} - y_{2r})^2}. \quad (3.3.6)$$

The image region contains the needed face is bounded by the rectangle that has upper-left and lower-right corners located at:

$$\begin{cases} (x_{ul}, y_{ul}) = (x_{1r} - rate * dist_{eyes}, y_{1r} - rate * dist_{eyes}) \\ (x_{lr}, y_{lr}) = (x_{2r} + rate * dist_{eyes}, y_{2r} + (1 + rate) * dist_{eyes}) \end{cases}. \quad (3.3.7)$$

In which *rate* is a constant to control the distance between eyes and boundaries of the face region. By empirical experiments, we fix $rate = 0.65$. The cropped face image is finally resized to 128×128 resolution (with [AR](#) and [FERET](#) databases) or 48×48 (for [SCface](#) database) before being fed into the subsequent stage.

Beside face cropping objective, this method also aligns the face images based on the rotation angle ϕ . Let us note that it is not a perfect approach and may not competing with other state-of-the-art alignment methods, but the resulting images are nevertheless sufficient for us to concentrate on the main goal of the thesis.

3.4 Preprocessing techniques

The purpose of the preprocessing algorithm is to remove the effects of illumination, the factor that can impair the performance of a [FR](#) system by producing extremely different images of one person. To select a preferable technique regarding system accuracy, many state-of-the-art ones [136] have been carefully examined by applying them upon images from the three face databases, which have been adopted earlier, before carrying out the experiments. The retinal filter [116] and standard histogram equalization were chosen in the end as they offer the best recognition rates.

3.4.1 Retinal filter

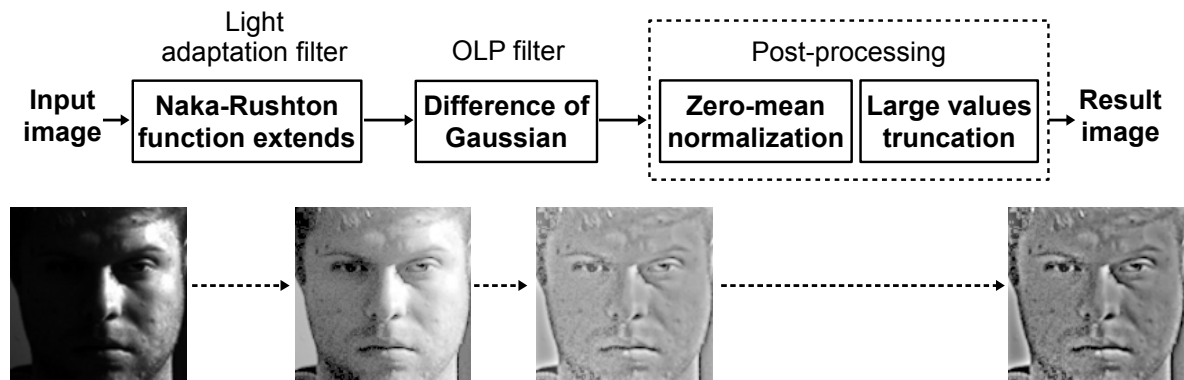


Figure 3.6: Retinal filter scheme for illumination normalization

Based on retinal modeling, Vu et al. [116] proposed retinal filter for illumination normalization. The method mimics the operation of two layers of the human retina, the photoreceptors and the outer plexiform layer (OPL), by applying two successive adaptive nonlinear functions, Difference of Gaussian (DoG) filter and a post-processing truncation (see Fig. 3.6 for more details).

3.4. Preprocessing techniques

The Naka-Rushton function [87], which is used to enhance image's local contrast, is expressed as:

$$Y = \frac{X}{X + X_0} \quad (3.4.1)$$

in which X is the input light intensity, X_0 is the adaptation factor (its value varies for each pixel), and Y is the result.

From the input image I , the light adaptation filter is proceeded on every pixel p by two lowpass filters stemmed from Naka-Rushton function (Eq. 3.4.1) as follow:

$$I_{la1} = (\max(I) + F_1(p)) \frac{I(p)}{I(p) + F_1(p)}, \quad (3.4.2)$$

$$I_{la2} = (\max(I_{la1}) + F_2(p)) \frac{I_{la1}(p)}{I_{la1}(p) + F_2(p)}. \quad (3.4.3)$$

The expressions $(\max(I) + F_1(p))$ and $(\max(I_{la1}) + F_2(p))$ act as normalization factors while \max is the function returning the maximal image intensity. Two adaptation factors $F_1(p)$ and $F_2(p)$ are computed by:

$$F_1(p) = I(p) * G_1 + \frac{\bar{I}}{2}, \quad (3.4.4)$$

$$F_2(p) = I_{la1}(p) * G_2 + \frac{\overline{I_{la1}}}{2}, \quad (3.4.5)$$

where $*$ is the notation of convolution operation, $\bar{\cdot}$ denotes the mean function, G_1, G_2 are Gaussian lowpass filters corresponding to two standard deviations $\sigma_1 = 1$ and $\sigma_2 = 3$:

$$G_1(x, y) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}}, \quad (3.4.6)$$

$$G_2(x, y) = \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}}. \quad (3.4.7)$$

The image I_{la2} is then processed by a DoG filter to enhance its edge information:

$$I_{dog} = DoG * I_{la2} \quad (3.4.8)$$

where DoG is obtained by:

$$DoG = \frac{1}{2\pi\sigma_{Ph}^2} e^{-\frac{x^2+y^2}{2\sigma_{Ph}^2}} - \frac{1}{2\pi\sigma_H^2} e^{-\frac{x^2+y^2}{2\sigma_H^2}} \quad (3.4.9)$$

with $\sigma_{Ph} = 0.5$ and $\sigma_H = 4$. As DoG filter inherent drawback is the reduction in global image contrast, a truncation of large values (with threshold $TH = 5$) followed by a zero-mean normalization is applied as:

$$I_{norm}(p) = \frac{I_{dog} - \overline{I_{dog}}}{std(I_{dog})} = \frac{I_{dog}}{std(I_{dog})} \quad (3.4.10)$$

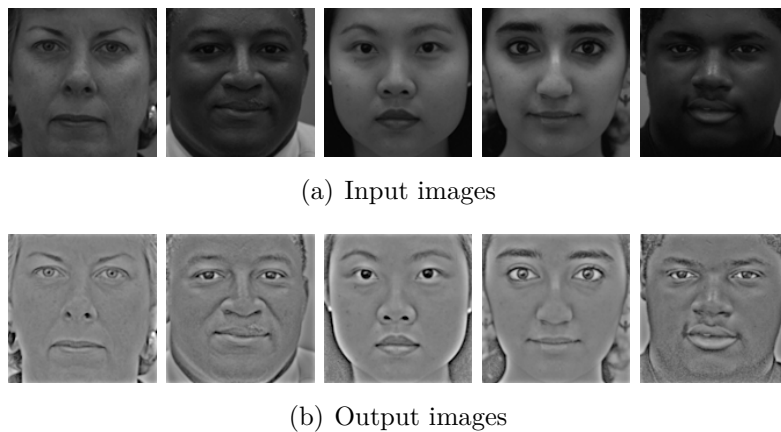


Figure 3.7: Retinal filter's illustration on illumination samples of FERET database.

$$I_{result}(p) = \begin{cases} \max(Th, |I_{norm}(p)|) & \text{if } I_{norm}(p) \geq 0 \\ -\max(Th, |I_{norm}(p)|) & \text{otherwise} \end{cases} \quad (3.4.11)$$

In Eq. 3.4.10, std is the standard deviation function and $\overline{I_{dog}}$ is very close to zero.

In Figs. 3.7(a-b), one can observe the results of applying the retinal filter on face images affected by illumination variations from FERET database: they (Figs. 3.7(a)) are normalized into the same lighting condition in the output images (Figs. 3.7(b)).

3.4.2 Histogram equalization

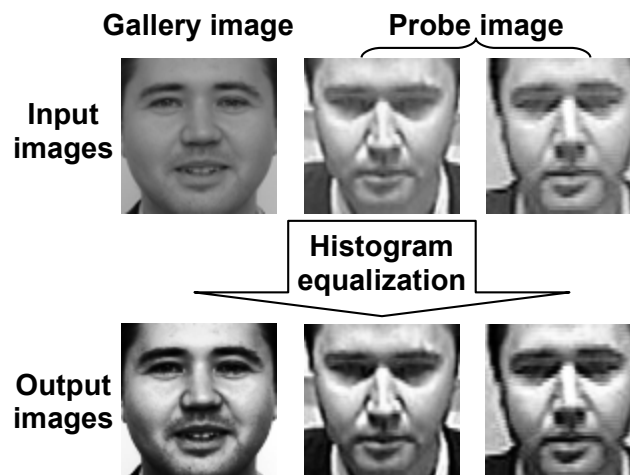


Figure 3.8: Some face images from SCface database and their histogram equalization versions.

This technique is dedicated for SCface database [42]. As aforementioned, many other

3.4. Preprocessing techniques

algorithms have been tested but they did not give better results than this simple histogram equalization technique. This occurs due to the fact that images from SCface database are low quality, blurred and contain very little fine image details (see Fig. 3.4 for more information). Plus, since the images are small in size, any noise filter of illumination methods (such as DoG in retinal filter [116] or proposed method in [112]) will inevitably blur the edges in them. On large images of higher quality (like the ones from FERET and AR databases), these degradations can be tolerated since the kernel size of the used filter is very small compared with the image size and the edge information is strong to be preserved after filtering. Therefore, applying other methods will discard such poor but crucial fine details from face images and thus no performance enhancement is achieved. In the meantime histogram equalization, while improving face images' global contrast for illumination normalization, does not affect images' visual features too much. This can be evidently seen in Fig. 3.8.

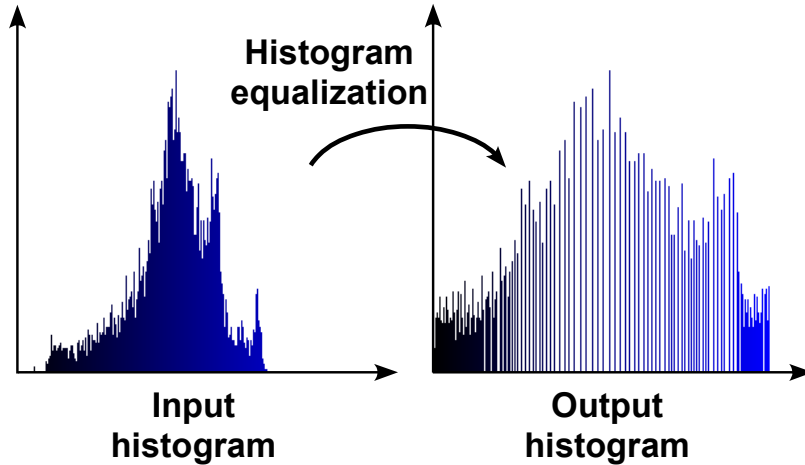


Figure 3.9: Histogram equalization mechanism.

Theoretically, histogram equalization algorithm redistributes the occurrences of intensities values in the input image to make them appear more equally in the output image (see Fig. 3.9 for more details). To do that, two steps, namely histogram normalization and intensity mapping, are performed as follows. Let $L - 1$ is the maximum intensity value, N is the number of intensity values, n_k is the number of pixels having intensity k , of the input image I . Then the probability of a pixel to have intensity k in the image I is estimated as:

$$p_x(k) = p(x = k) = \frac{n_k}{N}, k = [0, L) \quad (3.4.12)$$

Based on those probabilities, the histogram equalized image G is defined as:

$$G(x, y) = \text{floor}((L - 1) \sum_{i=0}^{I(x,y)} p_i), \quad (3.4.13)$$

where $\text{floor}()$ function rounds a real number down to the nearest integer, $G(x, y)$ and $I(x, y)$ are intensity values at pixel (x, y) of image I and G , respectively. This is equivalent to the intensity mapping T that transforms each intensity value k of image I to a new one s_k in image G as:

$$s_k = T(k) = \text{floor}\left((L - 1) \sum_{i=0}^k p_i\right) = \text{floor}\left((L - 1) \sum_{i=0}^k \frac{n_k}{N}\right) \quad (3.4.14)$$

Equalized image G is obtained by applying the above intensity transformation upon all the pixels of image I .

3.5 Template matching framework

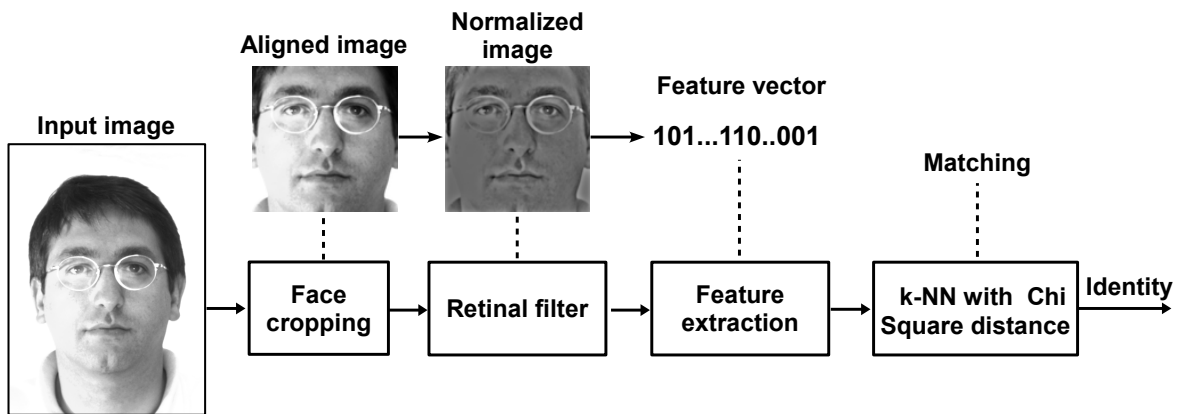


Figure 3.10: General template matching framework

Ahonen et al. [3] used template matching method with k-Nearest Neighbor (k-NN) and Chi Square distance functions (non-weighted and weighted) for classification. Inspired by that technique, a general template matching framework, whose steps are illustrated in Fig. 3.10, is used to investigate the discriminant capacity of a feature extraction method as follows. Firstly, the face images are cropped and aligned using their eyes coordinates. Then they are preprocessed by retinal filter [116] in order to diminish the bad effect of illumination variations. Next, a feature extraction method is used to extract the most distinguishing features from every normalized image. After the feature extraction stage, each face image is represented in terms of a feature vector and the Chi Square (non-weighted) distance is exploited to calculate the similarities between one test image and all the gallery ones. The Chi Square distance between two vectors $X = [x_1 x_2 \dots x_M]$ and $Y = [y_1 y_2 \dots y_M]$ is:

$$\text{dist}_\chi(X, Y) = \sum_{i=0}^M \frac{(x_i - y_i)^2}{x_i + y_i} \quad (3.5.1)$$

3.6. Whitened PCA based framework

The identity of one test image is assigned to the label which has highest similarity, also means the smallest distance, to it. There is no training stage to decrease the feature vector's length, hence this classification process will cost a long time for computing all needed distances. In every experiment, we will show that this unsupervised framework does not yield better accuracies than the [WPCA](#) based framework (which is presented in the next section) but it helps to verify the efficiency of one proposed feature extraction method. The weighted Chi Square function is not utilized since the performance improvement is negligible. Additionally, considering that experimental results upon [AR](#) and [FERET](#) are sufficient to conclude if one feature extraction algorithm is better than others or not, we do not evaluate this [FR](#) framework on [SCface](#) database.

3.6 Whitened PCA based framework

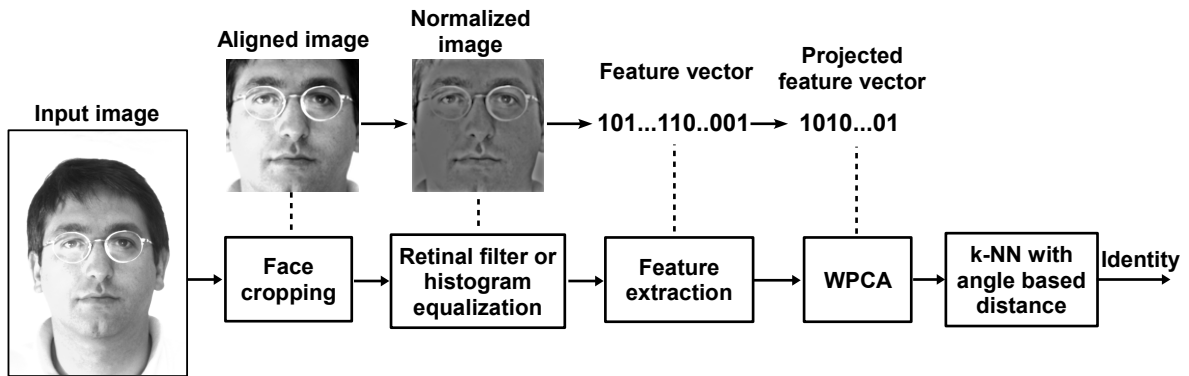


Figure 3.11: General [WPCA](#) based framework

Stages of the general Whitened PCA ([WPCA](#)) based framework used in this manuscript are showed in Fig. 3.11. All the stages before dimensionality reduction process are almost the same as in template matching framework, with a small addition that the preprocessing stage uses both retinal filter [116] (on [AR](#) and [FERET](#) databases) and histogram equalization technique (on [SCface](#) database). The main difference here are the use of [WPCA](#) and angle based distance functions (for classification).

For addressing the need of high performance (speed and accuracy) systems in reality, the template matching framework described in previous section is inadequate due to some reasons. Firstly, in their original forms, feature vectors obtained from the feature extraction stage are usually very big data thus making the classification task slow as it has to calculate the distances between them. Secondly, feature vectors contain redundant information that should be eliminated since it is meaningless to system performance. Hence, it is desirable to reduce their size and improve their distinctiveness by a learning algorithm. Towards this end, we adopt [WPCA](#) for dimensionality reduction stage. This

is a two-fold operation: the feature vector size is reduced to make the classification faster since it only works on smaller data and the recognition rates are higher as the projected vectors are more discriminative. WPCA is PCA [114] followed by a postprocessing operation called whitening step to weight eigenvectors by eigenvalues. Conventional PCA has two main disadvantages: the poor discriminative power in its sub space [109] if the feature vector length is larger than the samples number for training and the performance degradation when using its three leading eigenvectors [1]. The simple but effective whitening step rectifies these two main disadvantages of PCA whilst helping to enrich the discriminatory power in its projection space [29].

In order to determine the identity of the probe images in the classification stage, a distance function is needed for estimating the distance between projected feature vectors. Since the recognition performance of PCA based FR systems can varied significantly by using different functions, as reported by Moon et al. in [85], the best fit distance measures for our WPCA based systems have been selected via experiments by trying all available ones [95]. Among all distance measures involved [95], the negative angle distance and the weighted angle-based distance are found to bring the highest performance, depending on the used feature extraction method. Considering two vectors $X = (x_1, x_2, \dots, x_N)^T$, $Y = (y_1, y_2, \dots, y_N)^T$, their negative angle and weighted angle-based distances are computed as follow:

$$dist_{nang}(X, Y) = -\frac{X^T Y}{\|X\| \|Y\|} = -\frac{\sum_{i=1}^N x_i y_i}{\sqrt{\sum_{i=1}^N x_i^2} \sqrt{\sum_{i=1}^N y_i^2}} \quad (3.6.1)$$

$$dist_{wang}(X, Y) = -\frac{\sum_{i=1}^N z_i x_i y_i}{\sqrt{\sum_{i=1}^N x_i^2} \sqrt{\sum_{i=1}^N y_i^2}}, z_i = \sqrt{1/\lambda_i}, \quad (3.6.2)$$

here λ_i , ($i = 1..N$) are the eigenvalues of WPCA.

Also, it is worth indicating that in our WPCA based framework, each feature vector is standardized by firstly taking its square root and then the z-score normalization is applied on the obtained values. According to [120], this usage of the square root offers better results. Concretely, once taken its square root, a feature vector $X = (x_1, x_2, \dots, x_N)^T$ is normalized as:

$$norm(X) = \left(\frac{x_1 - \bar{x}}{std(X)}, \frac{x_2 - \bar{x}}{std(X)}, \dots, \frac{x_N - \bar{x}}{std(X)} \right)^T, \quad (3.6.3)$$

in which:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}, \quad (3.6.4)$$

and

$$std(X) = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (3.6.5)$$

3.6. Whitened PCA based framework

There are two steps in the usage of WPCA for FR: the training and the projecting ones. In the training step, a projection matrix is built from eigenvectors and eigenvalues, which are computed from feature vectors of training images. The projection matrix is then employed in the projecting step to project each feature vector of gallery or probe image into the WPCA subspace. In order to generate the projection matrix, there are two different ways that can be utilized: the one uses Eigenvalue decomposition algorithm and the one is based on Singular value decomposition method. Both of them are next described in detail and we will explain which one is better and therefore be used throughout this thesis.

3.6.1 EVD (Eigenvalue decomposition) based WPCA

From \mathbf{K} training samples, each one is represented as a column vector \mathbf{X}_i ($i = 1..\mathbf{K}$) of size N , the matrix \mathbf{L} is generated as:

$$\mathbf{L} = \frac{1}{K}A^T A, A = [\Phi_1 \Phi_2 \dots \Phi_K] \quad (3.6.6)$$

where

$$\Phi_i = \text{norm}(\sqrt{\mathbf{X}_i}). \quad (3.6.7)$$

In original PCA, we have to find eigenvectors and eigenvalues of covariance matrix $C = \frac{1}{K}AA^T$, a symmetric matrix of size $N \times N$, as:

$$C = W\Sigma W^T. \quad (3.6.8)$$

This is not feasible since N is large. Turk and Pentland [114] overcame this obstacle by a subtle trick (called Turk and Pentland's trick) that computes needed values from $L = \frac{1}{K}A^T A$, a smaller matrix of size $K \times K$, based on an observation when considering an eigenvector v_i of L as:

$$AA^T Av_i = A\lambda_i v_i = \lambda_i Av_i. \quad (3.6.9)$$

From above equation, we can see that Av_i is an eigenvector, corresponding to eigenvalue λ_i , of C . Following the Turk and Pentland trick, \mathbf{K} eigenvectors \mathbf{v}_i , ($i = 1..\mathbf{K}$) derived by eigenvalue decomposition of matrix \mathbf{L} are used to generate eigenvectors, which are then sorted decreasingly based on their associated eigenvalues λ_i , ($i = 1..\mathbf{K}$), of covariance matrix \mathbf{C} as:

$$\mathbf{u}_i = A\mathbf{v}_i. \quad (3.6.10)$$

\mathbf{K}' ($\mathbf{K}' < \mathbf{K}-1$) leading eigenvectors \mathbf{u}_i , ($i = 1..\mathbf{K}'$), also known as principal components, are retained to build the projection matrix \mathbf{U}_{proj} of PCA as:

$$\mathbf{U}_{proj} = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_{\mathbf{K}'}]. \quad (3.6.11)$$

The whitening step of WPCA normalizes eigenvectors of PCA by eigenvalues to construct its projection matrix \mathbf{W}_{proj} as:

$$\mathbf{W}_{proj} = \Lambda^{-1/2} \mathbf{U}_{proj}, \Lambda = [\lambda_1 \lambda_2 \dots \lambda_{K'}]. \quad (3.6.12)$$

This step is called whitening in the sense that it makes the projected data has an identity covariance matrix. Each feature vector \mathbf{y} of gallery or probe images is projected into WPCA subspace by the formula:

$$\tilde{\mathbf{y}} = \mathbf{W}_{proj}^T (\mathbf{y} - \bar{\mathbf{x}}). \quad (3.6.13)$$

3.6.2 SVD (Singular value decomposition) based WPCA

In **EVD WPCA** (also known as standard **WPCA**), we have to calculate the transpose matrix L of covariance matrix C and to use matrix A in conducting the projection matrix U_{proj} , these steps are time consuming if the features number of each vector \mathbf{x}_i and N are large. Besides, the multiplication between matrix A and its transpose A^T causes loss of precision with a lot of multiply and addition operations. So we propose to use Singular value decomposition (**SVD**) based **WPCA** (in the rest of this thesis when we refer to **WPCA**, it means **SVD** based **WPCA**) for boosting the computational performance. For this objective, the eigenvectors $u_i, i = 1..K$ and eigenvalues $\lambda_i, i = 1..K$ are calculated directly from the matrix A as:

$$[\mathbf{U}, \Sigma, \mathbf{V}] = svd(A), \Sigma^2 = \Lambda. \quad (3.6.14)$$

Then the whitening step is performed as:

$$W_{proj} = \Lambda^{-1/2} \mathbf{U} = \Sigma^{-1} \mathbf{U}. \quad (3.6.15)$$

This can be done due to the property of **SVD** that decomposes A as:

$$A = U \Sigma V^T, \quad (3.6.16)$$

in which V is an orthogonal matrix which entails $V^T V = I$. Covariance matrix C then can be expressed as:

$$C = AA^T = (U \Sigma V^T)(U \Sigma V^T)^T = (U \Sigma V^T)(V \Sigma U^T) = U \Sigma^2 U^T. \quad (3.6.17)$$

That means $\Sigma = \sqrt{\Lambda}$ are square roots of eigenvalues, which are associated with eigenvectors stored in matrix U , of the covariance matrix C .

3.7 Conclusions

In this chapter, all the background materials employed with the propositions for feature extraction stage are presented. We first discussed the general **FR** scheme in which a **k-NN** classifier is adopted to determine the identities of test images. Afterwards, with respect to criteria of a good database for using to assess the performance of a **FR** system in comparison with other state-of-the-art rivals, three large public face databases, namely **AR**, **FERET** and **SCface**, and their standard protocols are selected and described. It is worth highlighting that a broad range of challenges, including illumination, facial expressions, time-lapse, and pose variations, face occlusions, and low resolution probe images, is the purpose of experiments which will be conducted to verify our proposed feature extraction methods in the present dissertation. For face cropping, a simple yet efficient algorithm based on two eyes' coordinates, which are accompanied in the annotation data of the face databases, is used to crop and to align input face images. In the preprocessing stage, retinal filter [116], a powerful technique, is applied on **AR** and **FERET** databases while standard histogram equalization is wielded particularly upon images from **SCface** database as other methods do not bring accuracy improvement.

Based on the general **FR** scheme, we define two frameworks to be used with the feature extraction methods proposed in this thesis, which are called *Template matching* and *WPCA based*. In the "Template matching" framework, an unsupervised **FR** system is formed by using Chi Square distance and the **k-NN** classifier to recognize a face image. Meanwhile, in the "WPCA based" framework, we use **SVD** based **WPCA** method for dimensionality reduction and angle based distances with **k-NN** for classification.

Up to now, towards the purpose of building an efficient **FR** system, we have solutions for almost all of its stages, including face cropping, preprocessing, dimensionality reduction and classification, via adopting rational techniques for each stage, except the feature extraction one. By these, we next concentrate on the goal of devising robust facial feature extraction methods, as they play the most crucial part of a **FR** system. The content of the following three chapters will address this aim, step by step, by our propositions of novel facial descriptions.

Chapter 4

Intensity-based feature extraction methods

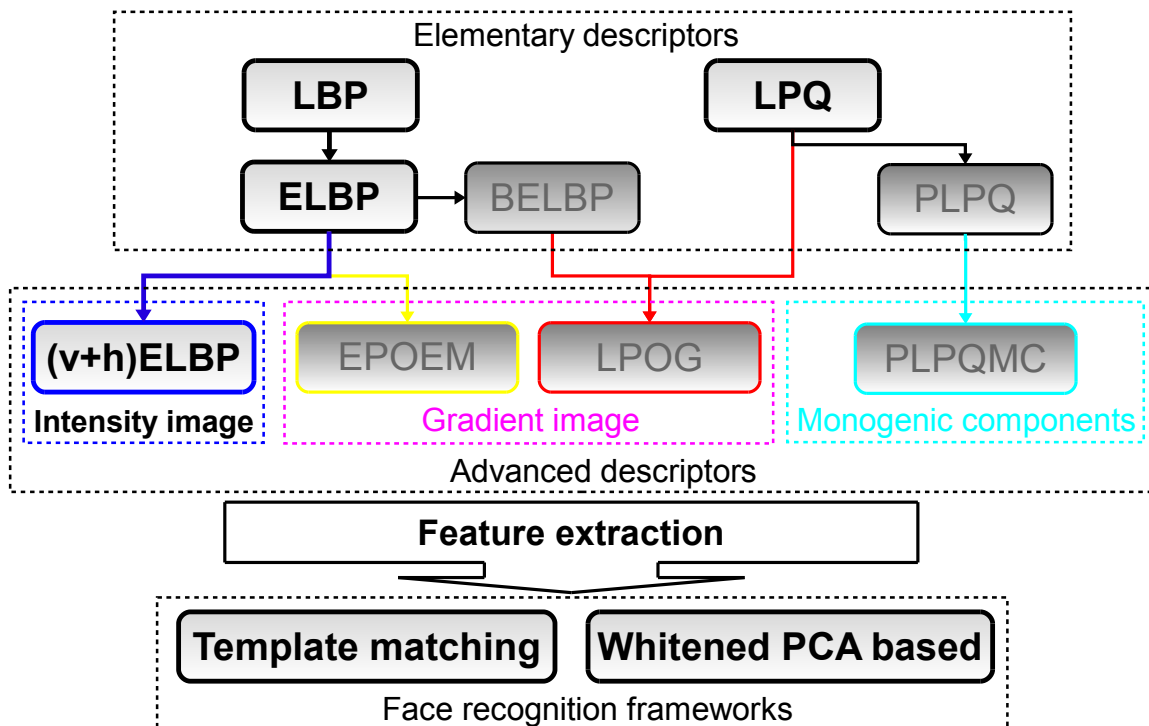


Figure 4.1: Contributions presented in this chapter: ELBP and LPQ methods and their associated FR frameworks.

With solutions for their stages of the two FR frameworks defined in chapter 3, the content of this thesis now moves to methods for facial feature extraction. In this chapter, two intensity-based local feature descriptors, Elliptical Local Binary Patterns (ELBP) and Local Phase Quantization (LPQ), are presented in detail (see Fig 4.1).

Both kinds of features are thoroughly explored via a long list of experiments upon preselected face databases (AR, FERET and SCface) and comparisons with other state-of-the-art systems. Additionally, the computational costs of the two methods are also investigated via a benchmark test upon images from Fa set of the FERET database. Since we primarily focus on local features to develop robust facial representations under challenging circumstances, the elementary descriptors scrutinized in this chapter play a key role for designing more sophisticated methods. Also importantly, via analyzing the results gained by ELBP and LPQ when they are plugged into *Template matching* and *WPCA based* frameworks, multiple useful conclusions, which drive the way we proceed in the subsequent two chapters, are made.

In the rest of this chapter, the ELBP feature extraction method is described in Section 4.1. Section 4.2 gives details of LPQ facial representation for face recognition. Experimental results of each method are provided in corresponding subsections and the conclusions are expressed in Section 4.3.

4.1 ELBP, a novel variant of LBP

In this section, we propose a novel variant of Local Binary Patterns (LBP) so-called Elliptical Local Binary Patterns (ELBP) which is dedicated to face analysis. In ELBP, we use horizontal and vertical ellipse patterns to capture micro facial feature of face images in both horizontal and vertical directions. ELBP is applied in face recognition with the *Template matching* framework and the *WPCA based* one in which dimension reduction step is done by Whitened Principal Component Analysis (WPCA). Our experiment results upon AR, FERET and SCface databases prove the advantages of ELBP over LBP for face recognition under different conditions and with ELBP WPCA we can get very remarkable results.

Before going further into details of the proposed method, it is worth recalling that an up-to-date survey on LBP and related works is presented in Section 2.2.2 of chapter 2.

4.1.1 Motivations

The purpose of a feature extraction method in a FR system is to capture the most intrinsic and discriminative facial features of face images in an efficient manner to form powerful face representations. As prior mentioned in chapter 1, these representations should be robust to as many as possible FR challenges. To meet this requirement, they must contains facial characteristics that maximize extra-class variations between face images of different identities while minimizing the intra-class variations between those of an individual.

4.1. ELBP, a novel variant of LBP

As far as we know, the most important facial parts of the human face are the eyes and the mouth [106, 45, 14, 108]. The natural shapes of human eyes and mouth are ellipses. Plus, there is more horizontal information in a face image than vertical one. Furthermore, horizontal information plays a very significant role in face recognition and the recognition performance is improved when we combine horizontal with vertical information [40]. As a consequence, in this chapter, we propose a novel variant of LBP so-called Elliptical LBP (ELBP) which uses horizontal and vertical ellipse patterns to form the ELBP feature representation for face recognition. The concept of applying elliptical patterns in LBP was also used by S. Liao and A.C.S Chung [70] to build the Elongated LBP. The authors used weighted factors for six regions of the face image and four different elliptical patterns (in four directions) to encode the anisotropic information of the image. Differently, in our ELBP scheme, we use only one horizontal ellipse and one vertical ellipse for capturing the micro facial features of face image and weighted factors are not used in producing the histogram sequence of ELBP images.

4.1.2 ELBP in detail

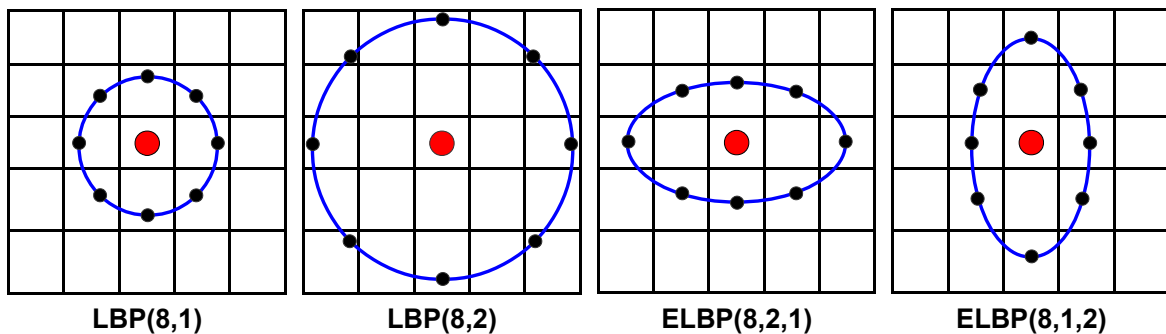


Figure 4.2: LBP and ELBP patterns

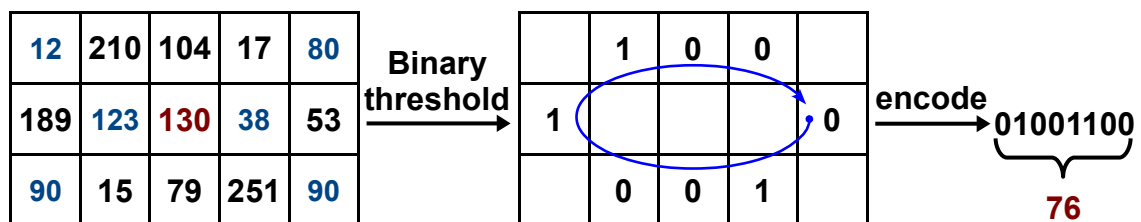


Figure 4.3: ELBP encoding scheme

In ELBP, at each pixel (x_c, y_c) of the input image, we consider its neighboring pixels that lie on an ellipse (see Fig. 4.2 for more details) with (x_c, y_c) is the center itself. The ELBP code of image pixel (x_c, y_c) with P surrounding pixels at $(R1, R2)$ distances is

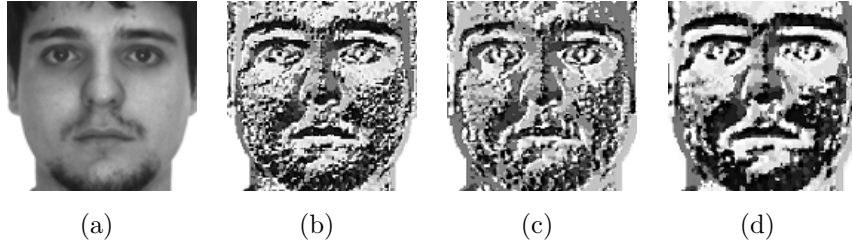


Figure 4.4: An image (a) and its $LBP^{8,1}$ (b), $ELBP^{8,3,1}$ (c), $ELBP^{8,4,3}$ (d)

computed as:

$$ELBP^{P,R1,R2}(x_c, y_c) = \sum_{i=1}^P s(g_i^{P,R1,R2} - g_c) 2^{i-1} \quad (4.1.1)$$

where $s(x)$ is a binary encoding function and is defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases} \quad (4.1.2)$$

In details, the coordinates of the i^{th} neighboring pixel of (x_c, y_c) are calculated using the formulas:

$$\begin{cases} angle_step = 2 * \pi / P \\ x_i = x_c + R1 * \cos((i - 1) * angle_step) \\ y_i = y_c - R2 * \sin((i - 1) * angle_step) \end{cases}, \quad (4.1.3)$$

Illustration of ELBP calculation for one pixel can be seen in Fig. 4.3. In Fig. 4.4 one can see a face image and its LBP and $ELBP$ versions.¹

When the coordinate of a pixel $\mathbf{p}_i = (x_i, y_i)$ on an ellipse computed from Eqs. 6.3.3 is not in the center of an image pixel, a bilinear interpolation of four the nearest pixels is applied to obtain its gray value \mathbf{g}_i . This process is illustrated in Fig. 4.5. Let the four nearest pixels of \mathbf{p}_i be a, b, c, d , whose centers are the black dots in Fig. 4.5, and g_a, g_b, g_c, g_d be their gray values, respectively. The distances between these pixels, more precisely their centers, is one pixel. Let dx denote the horizontal distance between \mathbf{p}_i (the red dot) and the centers of a and c and let dy denote the vertical distance between \mathbf{p}_i and the centers of b and d . Then the gray value of \mathbf{p}_i is calculated as:

$$g_i = g_a(1 - dx)(1 - dy) + g_b(dx)(1 - dy) + g_c(1 - dx)(dy) + g_d(dx)(dy), \quad (4.1.4)$$

¹As image pixels at the border do not have enough neighbors, the ELBP feature extraction procedure (the same as in original LBP method [3]) is not carried out on them.

4.1. ELBP, a novel variant of LBP

with d_x and d_y are computed as:

$$\begin{cases} d_x = 1 - (x_i - \text{floor}(x_i)) \\ d_y = 1 - (y_i - \text{floor}(y_i)) \end{cases}, \quad (4.1.5)$$

in which $\text{floor}()$ is the function that maps a real number to the largest previous integer.

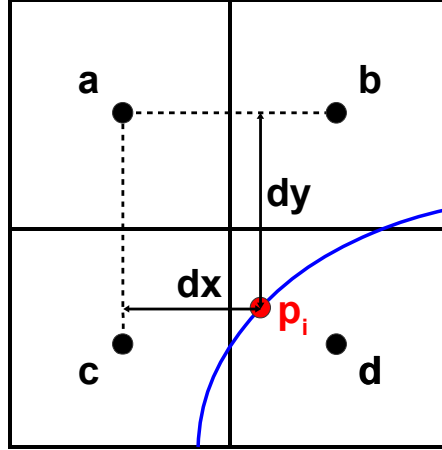


Figure 4.5: ELBP bilinear interpolation scheme

In [106, 14], authors indicated that eyes and mouth are the most important facial features in face recognition. The natural shapes of human eyes and mouth are ellipses. So horizontal ELBP is more suitable and more efficient than LBP in features extraction for face recognition. When $R1 = R2$, ELBP is LBP, when $R1 < R2$ we have a vertical ellipse and if $R1 > R2$ we have an horizontal ellipse, which matches most for human eyes and mouth. Additionally, studies in face perception [25, 40] verify that horizontal information drives the face identification of humans. Thus, by using horizontal ellipse to do the binary thresholding process in ELBP, the ELBP micro textures convey also the horizontal visual structures extracted from face images. Further, in this work, we use both horizontal and vertical ELBP to encode the micro facial feature in both directions because the combination of horizontal and vertical information of the face image gives the best recognition performance [40].

Building ELBP feature vector: For building the ELBP feature vector of input face images, we use ELBP operator to generate ELBP image (in Fig. 4.4 one can see an image and its ELBP images) and apply the similar method as Ahonen et al. [4]. When only horizontal ELBP is used, we firstly generate the ELBP image for the input image, then the ELBP image is divided into sub non-overlapped rectangular regions. In the next step, histogram sequences of sub regions are calculated and then concatenated to form the ELBP feature vector, uniform patterns [4] are employed in this step to reduce the vector's length. In the case of using both horizontal and vertical ELBP, we

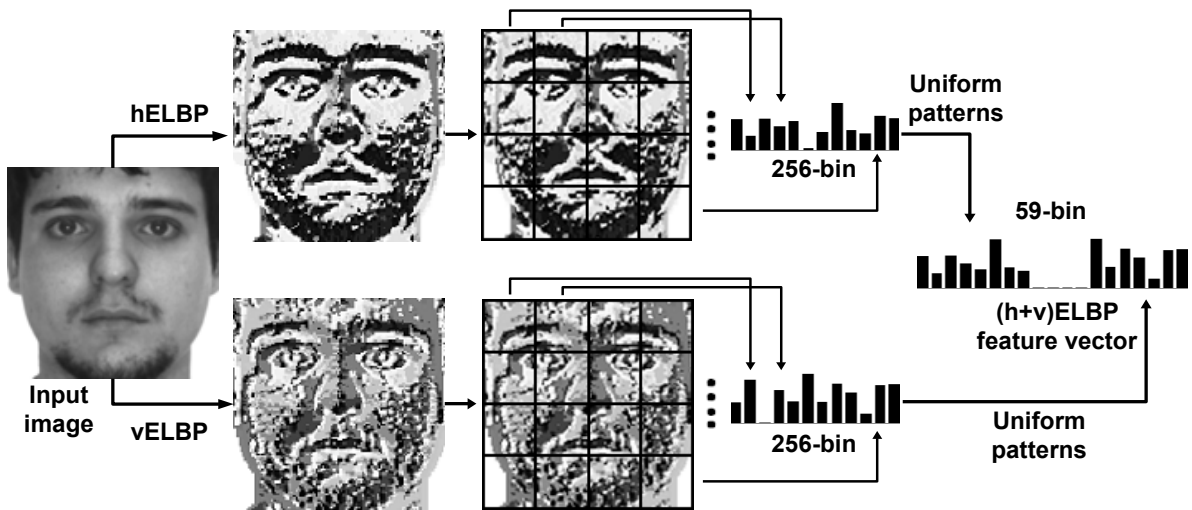


Figure 4.6: ELBP feature vector computation

apply two symmetric ELBP operators $ELBP^{P,R1,R2}$ and $ELBP^{P,R2,R1}$ to produce two ELBP images simultaneously. Then each ELBP feature vector corresponding to ELBP image is computed. After that, the two vectors are concatenated to form the complete horizontal and vertical ELBP feature vector for the given face image. All these steps are illustrated in Fig. 4.6.

The ELBP image is divided into $W \times H$ sub regions to build feature vector. So normally, with $(8, R1, R2)$ neighborhood patterns the horizontal ELBP feature vector length is $W * H * 256$ and the complete (both horizontal and vertical) ELBP feature vector length is $2 * W * H * 256$. Similar as in LBP [4], an ELBP value is called uniform pattern if its binary representation has no more than two bitwise transitions from 0 to 1 and vice versa. For example, the patterns 00000001 (1 transition) and 01111100 (2 transitions) are uniform but 01101001 (4 transitions) and 01010101 (7 transitions) are not. When they are used, each uniform ELBP value is assigned by a bin whereas only one bin is shared by all the rest ones (not uniform). In this document, ELBP feature vectors are only generated with $P = 8$ neighbor pixels, they are 59-bin histogram sequences as there are 58 uniform patterns. Thus, the ELBP feature vector length is reduced about 4 times (from $W * H * 256$ down to $W * H * 59$ and from $2 * W * H * 256$ down to $2 * W * H * 59$). For this reason, we use uniform patterns to speed up the ELBP calculations and to save required memory for storing ELBP feature vectors.

4.1.3 Face recognition with ELBP

In applying ELBP for face recognition, we use the template matching method (see Section 3.5 in chapter 3 for more details) and an advanced method that uses *negative cosine distance* function for classification and WPCA for dimension reduction, which is

4.1. *ELBP*, a novel variant of *LBP*

described in detail in Section 3.6 of chapter 3. With the template matching framework, we use *LBP*, *ELBP(h)* (means only horizontal *ELBP* is used), *ELBP(h+v)* (means a symmetric pair of both horizontal and vertical *ELBPs* is used) notations to indicate the corresponding feature extraction method while their equivalent *WPCA based* methods are named by adding the word "*WPCA*" as a suffix. The obtained recognition rates on *AR*, *FERET* and *SCface* databases are compared with other state-of-the arts *FR* systems.

4.1.4 Experimental results

4.1.4.1 Results on *AR* database

Table 4.1: Rank-1 RRs (%) comparison between *LBP* and *ELBP* based methods on *AR* database

Test/Method		Probe set												Avg
		1	2	3	4	5	6	7	8	9	10	11	12	
Exp1	LBP	100	100	74.4	100	100	97.8	76.9	55.2	45.4	98.5	76.9	71.6	83.1
	LBP WPCA	100	100	76.7	100	100	100	77.6	61.2	51.5	99.3	98.5	92.5	87.9
	ELBP(h)	100	100	79.7	100	100	98.5	85.1	55.2	46.3	98.5	86.6	76.9	85.6
	ELBP(h) WPCA	100	100	79.9	100	100	100	85.1	65.7	56.7	100	96.3	97.0	90.0
	ELBP(h+v)	100	100	81.2	100	100	98.5	91.0	55.2	47.0	99.3	86.6	79.1	86.5
	ELBP(h+v) WPCA	100	100	81.2	100	100	100	91.0	67.2	61.9	100	97.8	97.0	91.3
Exp2	LBP	100	100	74.8	100	100	100	81.5	55.1	46.6	97.5	76.3	68.6	83.4
	LBP WPCA	100	100	76.3	100	100	100	81.5	67.0	56.0	98.3	93.2	90.7	88.6
	ELBP(h)	100	100	75.6	100	100	100	84.9	57.6	46.6	98.3	85.6	81.4	85.8
	ELBP(h) WPCA	100	100	77.1	100	100	100	85.6	67.8	56.0	100	96.6	96.6	90.0
	ELBP(h+v)	100	100	79.0	100	100	100	87.3	58.5	49.2	98.3	85.6	81.4	86.6
	ELBP(h+v) WPCA	100	100	80.5	100	100	100	87.3	70.3	61.8	100	97.5	96.6	91.2
Exp3	LBP	95.0	96.6	56.3	94.9	94.9	82.2	57.2	50.0	43.2	95.0	46.6	46.6	71.5
	LBP WPCA	97.5	97.5	57.6	100	100	93.2	74.6	56.9	47.5	95.0	78.0	67.0	80.4
	ELBP(h)	95.0	98.3	57.2	97.5	94.9	83.9	57.2	52.5	44.9	95.0	55.9	49.2	73.5
	ELBP(h) WPCA	97.5	98.3	68.6	100	100	94.9	75.4	61.0	51.7	97.5	88.1	89.0	85.2
	ELBP(h+v)	96.6	98.3	62.2	97.5	94.9	83.9	65.6	52.5	47.5	95.8	58.5	52.5	75.5
	ELBP(h+v) WPCA	97.5	99.2	73.7	100	100	95.8	80.5	66.1	63.6	97.5	90.7	92.4	88.1
Exp4	LBP	92.3	95.7	45.7	97.4	97.4	90.6	66.7	55.6	49.6	72.7	58.1	41.9	72.0
	LBP WPCA	94.9	98.3	55.2	100	98.3	94.9	57.3	60.7	58.1	88.9	84.6	71.8	80.3
	ELBP(h)	94.9	95.7	45.7	98.3	98.3	90.6	66.7	59.0	50.4	78.6	61.5	56.4	74.7
	ELBP(h) WPCA	95.7	98.3	60.3	100	100	95.7	66.7	67.5	59.0	94.9	89.7	86.3	84.5
	ELBP(h+v)	95.7	95.7	48.3	98.3	98.3	90.6	68.4	63.3	55.6	82.1	62.4	58.1	76.4
	ELBP(h+v) WPCA	98.3	98.3	63.8	100	100	97.4	79.5	67.5	59.8	95.7	90.6	91.5	86.9

It is worth reminding that the probe sets of *AR* database, which are numbered from 1 to 12 in table 4.1, consist of images captured under various conditions of facial expressions (Smile, Anger, Scream), lighting changes (Neutral+Left light, Neutral+Right light, Neutral+both sides light), and occlusions (Sun glasses, Sun glasses+Left light, Sun glasses+Right light, Scarf, Scard+Left light, Scard+Right light). Results of *LBP* and

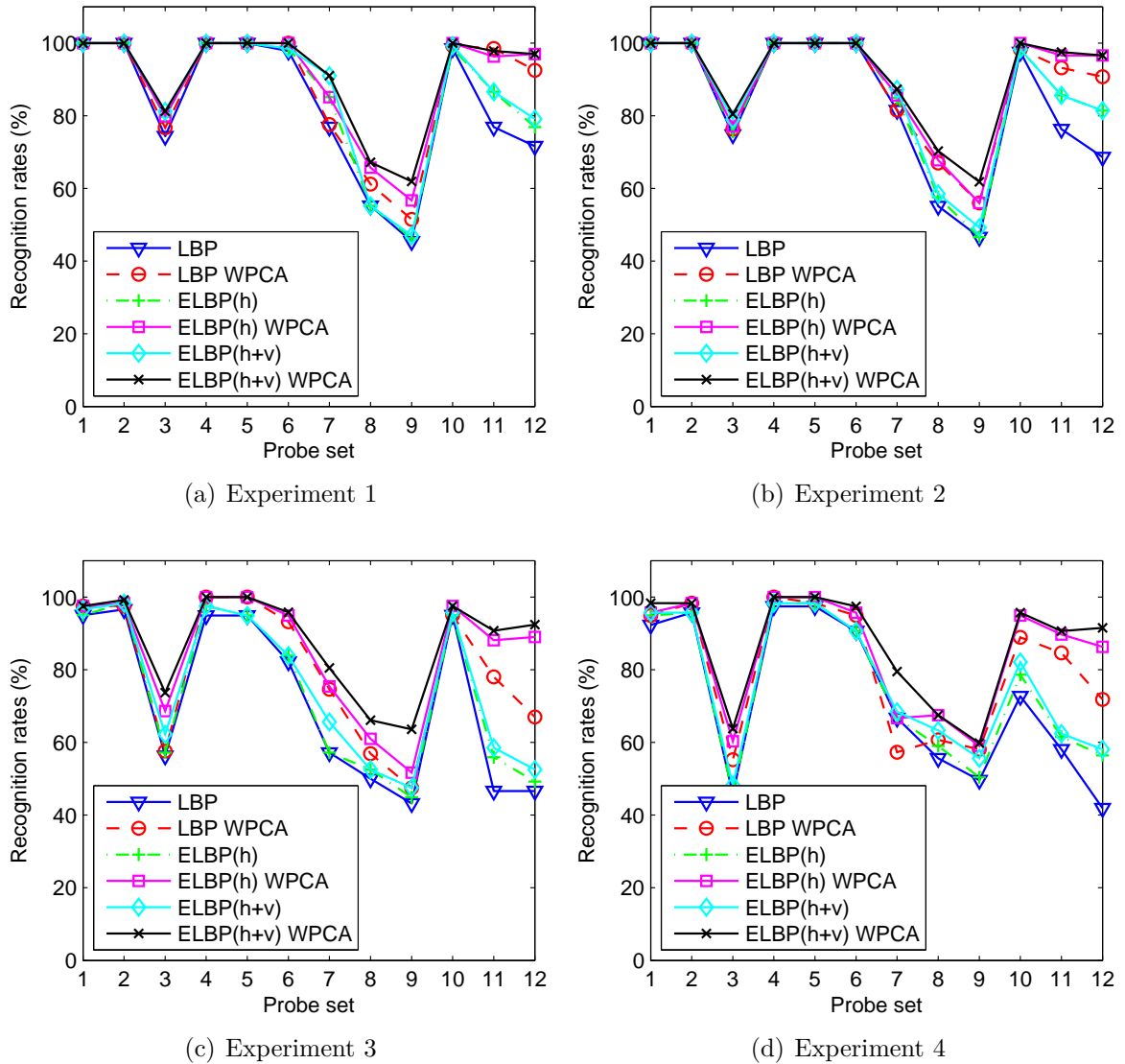


Figure 4.7: Accuracy performance of LBP and ELBP based systems on AR database.

ELBP based methods is provided in table 4.1 whilst the comparison between ELBP(h+v) WPCA and other well-known systems is shown in table 4.2. For clarifying purpose, we visualize the RRs of LBP and ELBP based systems as scatter plots in Figs. 4.7.

From results in table 4.1 and Figs 4.7, we can conclude that:

1. The horizontal ELBP ($ELBP(h)$) is more efficient than LBP in encoding micro facial features for FR. This superiority is consistent under all conditions (facial expressions, illumination variations and occlusions), in all experiments and with both template matching and WPCA based frameworks. The dominance of ELBP(h) over LBP is more significant on probe sets 3 (Scream), 8 and 9 (Sun glasses with illumination variations), 11 and 12 (Scarf with illumination variations), especially

4.1. *ELBP, a novel variant of LBP*

in Exp 3 and Exp 4, where the time-lapse variation is presented. In summary, these results confirm the efficiency of our proposition of using horizontal elliptical pattern to encode micro texture features in $ELBP(h)$.

2. The fusion at feature level of both horizontal and vertical ELBPs in feature extraction ($ELBP(h+v)$) gives better performance than using single horizontal ELBP ($ELBP(h)$). Again, this improvement is steady since it is achieved in all experiments with both template matching and WPCA based FR methods. We believe that this comes from useful horizontal and vertical texture features which are extracted by the combination of the symmetric pair of ELBPs. These results also point out that $ELBP(h+v)$ method is strong against facial expressions, illumination variations and occlusions.
3. Between the *Template matching* and *WPCA based* frameworks, the latter significantly outperforms the former. In all probe sets of four experiments and with all feature extraction methods, the usage of WPCA brings higher recognition rates than the direct matching method using the Chi Square distance. As mentioned in chapter 3, this improvement is gained since WPCA produces more discriminative feature vectors by eliminating redundant information from their original forms. Importantly, these improvements validate that if one feature extraction method is more powerful than another one in *Template matching* framework, the corresponding results it achieves with *WPCA based* paradigm will be higher than that of the compared method as a consequence.
4. Among all the facial expressions, the recognition rates with scream image sets (number 3) are lowest because the shapes of human eyes and mouth are changed most when screaming. Under occlusions conditions, the recognition rates upon scarf probe sets are higher than that on sun-glasses probe sets due to the fact that with glasses, the most important facial feature for face recognition, the eyes, are hidden. The combination of high and stable results on scarf sets and corresponding rates upon glasses probe sets point out that the upper part (above the mouth) of the human face is much more important than the lower part in face recognition. This is consistent with conclusion from face perception researches in [106, 45] that the upper portion of the face is more useful for FR than the lower one. The results of experiment 3 and 4 show that time-lapse conditions, even in a short period (about 2 weeks), but when appearing simultaneously with other variations (i.e. illumination, expression and occlusion), can degrade face recognition performance dramatically.

Although experiments on AR database of other methods in the literature usually used some probe sets only, their results were commonly compatible with those from our Exp 1, Exp 2, Exp 3 because the same SSPP protocol has been used. Hence, we report in

table 4.2 the most representative comparison results between ELBP(h+v) WPCA and other systems.

Table 4.2: Rank-1 RRs (%) comparison with other contemporary systems on AR database using the same evaluation method

	1	2	3	7	10	
Method	Smile	Anger	Scream	Glasses	Scarf	Classes ¹
Exp 1						
S-LNMF [90]	96.0	N/A ²	49.0	84.0	87.0	100
LGBP[135]	N/A ²			80.0	98.0	50
IRF [141]	N/A ²			87.5	91.7	120
String face [21]	87.5	87.5	25.9	88.0	96.0	100
Sparse coding [129]	N/A ²			94.7	91.0	100
DMMA[78]	99.0	93.0	69.0	N/A ²		100
Our	100		81.2	91	100	134
Exp 2						
DMMA[78]	85.0	79.0	45.0	N/A ²		100
Method in [90]	96.0	N/A ²	54.0	66.0	89.0	100
LGBP[135]	N/A ²			62.0	96.0	50
Sparse coding [129]	N/A ²			80.3	72.7	100
String face [21]	N/A ²			76.0	88.0	100
IRF [141]	N/A ²			82.5	84.0	120
SIS [75]	86.0			96.0	90.0	100
Sparse LF [76]	N/A ²			96.6	96.6	119
Our	100		80.5	87.3	100	118
Exp 3						
S-LNMF [90]	62.0	N/A ²	27.0	49.0	55.0	100
Method in [84]	N/A ²		52.3	54.2	81.3	80
PLD [54]	86.0			90.0	89.0	100
Our	97.5	99.2	73.7	80.5	97.5	118

¹ : The classes column is the number of persons whose images are used in experiments.

² N/A: Not available result.

It can be seen from table 4.2 that our system is comparable with other state-of-the-art methods. *ELBP(h+v) WPCA* is the only system that has perfect recognition rates on *Smile*, *Anger* and *Scarf* probe sets in Exp 1 and 2. While upon *Scream* and *Sunglasses* probe sets of Exp 1 and Exp 2 and all probe sets of Exp 3, our method also surpasses other rivals with higher recognition rates. These results are more interesting when considering that they are obtained with larger number of gallery/probe images

4.1. ELBP, a novel variant of LBP

(see the *Classes* column of table 4.2) in comparison with most of other methods as the recognition task becomes harder when there are more subjects involved.

4.1.4.2 Results on FERET database

This Section gives the rank-1 RRs of ELBP based systems in comparison with LBP and other Gabor wavelets based ones upon FERET database (for more details, see Section 3.2 in chapter 3). We report the comparative results of the standard protocol in table 4.3 whilst in table 4.4 and table 4.5, one can see those upon pose variation probe images.

Frontal FERET image sets

Table 4.3: Rank-1 RRs (%) comparison with other state-of-the-art results on FERET database using the standard evaluation protocol [96]

Method	Fb	Fc	Dup 1	Dup 2	Average
LBP	96.2	92.3	70.4	68.4	85.2
ELBP(h)	96.7	94.9	71.3	70.1	86.1
ELBP(h+v)	97.0	95.4	72.0	71.0	86.6
LGBPHS [135]	98.0	97.0	74.0	71.0	87.8
HGPP [133]	97.6	89.9	77.3	76.1	88.7
LGBP [88]	98.1	98.9	83.8	81.6	92.1
LBP WPCA	98.7	99.0	83.9	78.2	92.1
ELBP(h) WPCA	99.3	99.0	87.7	83.8	94.2
FGLBP [113]	98.0	98.0	90.0	85.0	94.2
ELBP(h+v) WPCA	99.4	100	89.1	86.8	95.0

The result of [70] is not included in this table because: the authors only provided the average RR (93.2%) and they did not follow the standard protocol [96] (They used a small subset of FERET database).

The comparison results in table 4.3 confirm that horizontal ELBP is more robust than LBP in micro facial features extraction (in both template matching and WPCA methods), especially in Dup 2 probe set. It is obvious that the usage of horizontal and vertical ELBPs brings very impressive improvement of recognition rates in comparison with original LBP and single horizontal ELBP (the most significant improvement cases are in the aging condition: Dup 1 and Dup 2 experiments). Once, with each method used for feature extraction (LBP, ELBP(h) and ELBP(h+v)), WPCA based framework yields higher accuracies than the *Template matching* one. The perfect recognition rate

(100%) upon Fc probe set of ELBP(h+v) WPCA illustrates the effectiveness of ELBP under illumination variations.

Non-frontal FERET image sets

Table 4.4: Rank-1 RRs comparison between LBP and ELBP based methods on b-series of FERET database

	LBP	ELBP	ELBP(h+v)	LBP WPCA	ELBP WPCA	ELBP(h+v) WPCA
-40°-bh	56.5	61.0	64.5	75.0	77.0	80.5
-25°-bg	89.0	91.0	91.5	97.0	98.0	98.5
-15°-bf	97.5	97.5	98.5	99.5	99.5	99.5
+15°-be	98.5	98.5	98.5	99.5	99.5	99.5
+25°-bd	90.0	91.5	92.0	98.0	98.5	99.0
+40°-bc	54.0	59.5	59.5	74.0	74.5	79.5
Avg	80.9	83.2	84.1	90.5	91.2	92.8

One can observe that the results upon probe sets under pose variations of FERET database in table 4.4 are agreed with those in table 4.1 and 4.3 when the recognition rates are improved gradually from LBP to ELBP(h), and from ELBP(h) to ELBP(h+v) in both *Template matching* and *WPCA based* frameworks. These improvements are more apparent when the head poses are larger (from $\pm 15^\circ$ to $\pm 40^\circ$). This is the confirmation for our approach to feature extraction with ELBP: firstly, the horizontal elliptical pattern in horizontal ELBP is more efficient than circular pattern in LBP, and secondly, the combination of both horizontal and vertical ELBPs yields a substantial improvement in accuracy performance. Besides, another conclusion is drawn from those results is that the learnt FR framework (*WPCA based*) is superior to the *Template matching* one.

Table 4.5 contains comparison results between ELBP(h+v) WPCA and other systems upon b-series images of FERET database. It can be noticed that our system achieves very promising accuracies as it outperforms many state-of-the-art counterparts. Being a general FR framework, ELBP(h+v) WPCA’s results are lower than leading-edge systems, which are dedicated for pose variations challenge, but they are evidences that an efficient feature extraction method can probably deal with head pose changes, at least when the pose angles are small (within $\pm 25^\circ$, our results are competing with the best ones in the FR literature).

4.1.4.3 Results on SCface database

This section gives results of *WPCA based* methods using LBP, ELBP(h) and ELBP(h+v) for feature extraction. Since experiments are evaluated with *WPCA based* methods, we name each method according to the feature extraction algorithm it employs. The high

4.1. ELBP, a novel variant of LBP

Table 4.5: Rank-1 RRs comparison with other leading methods on FERET b-series.

Method	-40° bh	-25° bg	-15° bf	$+15^\circ$ be	$+25^\circ$ bd	$+40^\circ$ bc	Avg
SLF-RKR [127]	N/A ¹	55.0	100	96.0	57.0	N/A ¹	N/A ¹
LSED [121] * ²	78.0	84.0	88.0	89.0	88.0	83.0	85.0
CCA [64] * ²	81.0	91.0	92.0	94.0	89.0	80.0	87.8
PAN [39]	81.5	93.0	97.0	98.5	91.5	78.5	90.0
RFC [20]	84.2	90.2	94.0	93.2	92.5	89.5	90.6
ADMCLS [105]	85.0	94.0	96.0	95.0	94.0	82.0	91.0
LMG [94]	N/A ¹	91.5	98.0	98.5	93.5	N/A ¹	N/A ¹
MRH [7] * ²	87.0	94.0	98.0	99.0	96.0	74.0	91.3
GLOH [100] ** ³	81.1	94.5	100	100	94.5	81.1	91.9
ELBP(h+v) WPCA	80.5	98.5	99.5	99.5	99.0	79.5	92.8
DWFF [86]	87.5	98.0	100	99.0	98.5	82.4	94.2
MRF [48]	91.0	97.3	98.0	98.5	96.5	91.5	95.5
3D Pose Norm [6]	90.5	98.0	98.5	97.5	97.0	91.9	95.6
CPN [31]	94.5	98.0	98.5	99.0	98.5	97.0	97.6

¹ N/A: Not available result.

² *: The RRs of the method are estimated from plotted figures.

³ **: The RRs on $\pm 25^\circ$ and $\pm 40^\circ$ subsets are average results.

quality mug-shots are used for training with WPCA in both DayTime and NightTime experiments (for more details, see Section 3.2 in chapter 3).

The results from table 4.6 and table 4.7 show that the ELBP(h+v) WPCA framework outperforms other state of the art systems, especially when compared to the baseline PCA [42] (our average result in DayTime experiment is about nine times higher than in [42]). These results (table 4.6 and table 4.7) also prove that horizontal ELBP descriptor is more robust than LBP in micro facial features extraction (under both day time and night time conditions at three distances) and again (as evaluations on AR and FERET databases), the combination of horizontal and vertical ELBP brings the best performance. To the best of our knowledge, our results on SCface database are the **first complete and highest** results reported in the literature so far.

It is clear that the results on SCface database are much lower than the recognition rates on AR database (table 4.1) and on FERET database (table 4.3). The very low resolution (small in size and very poor quality) of probe images in SCface database is the cause of those results.

Table 4.6: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]

Camera/Distance	PCA[42]	DSR[140]	LBP	ELBP(h)	ELBP(h+v)
cam1_1	2.3	N/A ¹	43.1	43.1	43.1
cam1_2	7.7		50.0	51.5	56.2
cam1_3	5.4		41.5	41.5	45.4
cam2_1	3.1		31.5	36.2	36.9
cam2_2	7.7		44.6	48.5	50.8
cam2_3	3.9		34.6	35.4	42.3
cam3_1	1.5		20.8	25.4	34.6
cam3_2	3.9		38.5	37.7	46.9
cam3_3	7.7		49.2	49.2	51.5
cam4_1	0.7		30.0	32.3	32.3
cam4_2	3.9		50.0	50.0	50.0
cam4_3	8.5		44.6	46.2	50.8
cam5_1	1.5		28.5	31.5	36.2
cam5_2	7.7		26.9	30.8	32.3
cam5_3	5.4		23.9	29.2	31.5
Average	4.7		20.2	37.2	39.2

¹ N/A: Not available result

Table 4.7: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]

Camera/Distance	PCA[42]	LBP	ELBP(h)	ELBP(h+v)
cam6_1	1.5	6.9	9.2	9.2
cam6_2	3.1	13.9	14.6	15.4
cam6_3	3.9	19.2	19.2	25.4
cam7_1	0.7	10.0	10.8	13.1
cam7_2	5.4	11.5	10.8	13.1
cam7_3	4.6	9.2	13.9	13.9
Average	3.2	11.8	13.1	15.0

4.1.4.4 ELBP parameters

The original LBP [3] for face recognition uses $LBP^{8,1}$ and $LBP^{8,2}$ operators on 7×7 sub regions of input images (128x128 resolution) to get the best performance. Our best results on AR database use $LBP^{8,5}$ (9x9 sub regions), $ELBP^{8,5,3}$ and $ELBP^{8,3,5}$ (9x9 sub regions). The $LBP^{8,5}$ (9x9 sub regions), $ELBP^{8,5,3}$ and $ELBP^{8,3,5}$ (9x9 sub regions) are used with FERET database. On SCface database, the $LBP^{8,3}$ (6x6 sub regions), $ELBP^{8,3,5}$ and $ELBP^{8,5,3}$ (6x6 sub regions) give the highest recognition rates.

4.1. ELBP, a novel variant of LBP

All this information about ELBP’s parameters indicates that the best ratio between horizontal radius and vertical radius of ELBP is 1.67(5/3). However, we think that the radii of ellipse samples used by ELBP probably depend on what kind of image it deals with (in this chapter, ELBP only works on intensity image) and a proof for this remark will be shown in chapter 6 when it is used upon accumulated oriented edge magnitudes images. Besides, the number of sub regions is definitely counted on the image size and it decides the length of the resulting feature vector.

With every feature extraction method presented in this dissertation, we first try to find and assign some core parameters to fixed values, e.g, the radii of ELBP in this part, via empirical experiments. Then, those values are used uniformly across 3 databases (AR, FERET and SCface) to report their corresponding results. Other depending parameters (on image size or kind of image used) may vary with appropriate explanations. By doing this way, we wish to show and highlight the true reliability of each method, which comes consistently from its inherent efficiency, rather than exhibiting the best RRs obtained after performing an exhausted search for the best fit parameters’ values of a specific database or experiment.

4.1.4.5 Computational cost

Table 4.8: Computation time of ELBP in comparison with other feature extraction methods

Method	Image size	Time (seconds)	Extraction time (milliseconds)	Images/second
ELBP(h)	128 × 128	4.44	3.71	269
ELBP(h+v)	128 × 128	7.28	6.09	164
Monogenic [128] ^{*1}	150 × 130	19.81	16.56	60
MBC-A [128] ^{*1}	150 × 130	30.54	25.54	39
MBC-O [128] ^{*1}	150 × 130	87.0	72.74	14
Gabor wavelets	88 × 80	96.23	80.46	12

¹ *: We used the Matlab code provided by the author.

When considering to deploy a FR system in real-life situations, its efficiency (high accuracy) is not the only prerequisite. In many scenarios, especially video surveillance, processing speed plays a *vital role*. For examining computational cost of feature extraction methods presented in this thesis, we proceeded some benchmark tests by running its Matlab implementation upon the whole Fa set (1196 images) of the FERET database and compared resulting metrics (total required time, extraction time for one image and its speed, which is measured as a number of images per second) with those of initial step of Gabor wavelets (just generating Gabor wavelets components at 5 scales and 8

orientations) based approaches, initial step of Monogenic filter (only producing its components at 3 scales) based methods, and some advanced feature extraction algorithms such as MBC-A, MBC-O [128].

To be fair, we used 88×80 resolution images for generating Gabor wavelets images like in [135] and 150×130 pixels images with monogenic based algorithms [128] while all other methods performed on 128×128 ones. More specifically, the same parameters as upon frontal FERET images were applied with ELBP(h), ELBP(h+v) while Monogenic based methods [128] use their default parameters.

All above experiments were performed on a Dell OptiPlex 790 desktop machine (CPU Core i7-2600 @3.4 GHz, 4Gb RAM) which was installed with Windows 7 64 bit SP1 and Matlab 2011b 32 bits programming environment. Although the machine has a multi-core CPU (4 cores), all tested implementations are not parallel. We ran each benchmark 100 times and reported the average results in table 4.8.

It can be seen from table 4.8 that ELBP(h) is very rapid when it can finish up to 269 images per second (just about 3.7 millisecond for one image). In addition, ELBP(h+v) is fast as it requires only 6.1 millisecond to process one image. The initial step of Gabor wavelets based method is about 13 times slower than ELBP(h+v) (although the image size in Gabor wavelets calculation is much smaller) while that of Monogenic filter based one is about 2.7 times slower than ELBP(h+v). Two Monogenic filter based feature extraction methods, MBC-A and MBC-O [128], are also slower than ELBP(h+v). With its fast speed, ELBP can be used as an elementary descriptor to form advanced multi-resolution/multi-scale facial representations.

4.1.5 Conclusions

This part of the dissertation introduces a novel variant of LBP operator so-called ELBP. We use a horizontal and a vertical ellipse patterns to form the ELBP face descriptor for feature extraction. Then ELBP images are divided into sub rectangular regions to build their ELBP histogram sequences. The ELBP feature vector is generated by concatenating sub regions' histogram sequences. In dimension reduction stage, we use WPCA for better recognition performance.

The experimental evaluations upon AR, and FERET databases show that, ELBP is more efficient than LBP in encoding micro facial features and ELBP can perform well under various conditions such as partial occlusion, facial expressions, time-lapse and pose variations. Additionally, the recognition performance on SCface database proves the effectiveness of ELBP for the problem of face recognition in video surveillance context. The original LBP is popular for its robustness to rotation because it uses circular patterns. While our results in this Section demonstrate advantages of ELBP

4.2. LPQ as a facial feature extraction

over LBP for face recognition, we do not suggest that ELBP is robust against rotation. Plus, ELBP code is an oriented feature that contains facial characteristics in horizontal orientation, the main direction of the face information, and also the information of vertical direction. This makes ELBP description has stronger discriminative power and hence gains higher performance than LBP. Without doubt, we strongly believe that ELBP can achieve better results in the research fields related to face recognition, where LBP was applied.

4.2 LPQ as a facial feature extraction

In this part, we present an intensive investigation of LPQ [5] as a facial feature extraction method by evaluating it in many experiments upon AR, FERET and SCface databases. These experiments are carried out with both *Template matching* and *WPCA based* frameworks. By doing so, we will indicate that LPQ is more powerful than LBP [3] and ELBP in extracting facial features from face images. More interestingly, the results show that the incorporation of LPQ with WPCA can be comparable with other state-of-the-art systems.

4.2.1 Blur invariance of Fourier Phase spectrum

According to [9], the model for expressing blur effects in digital image processing can be defined in the Fourier domain by using a convolution between an image and the point spread function (PSF) of the image acquisition system as:

$$G(\mathbf{u}) = F(\mathbf{u}) \cdot P(\mathbf{u}), \quad (4.2.1)$$

where $G(\mathbf{u})$, $F(\mathbf{u})$, $P(\mathbf{u})$ correspond to the discrete Fourier transforms of the blur image, the original image, the PSF of the blur, and \mathbf{u} is a 2D frequency $[u, v]^T$. When considering only the phase spectrum part of (4.2.1), we have:

$$\angle G(\mathbf{u}) = \angle F(\mathbf{u}) + \angle P(\mathbf{u}). \quad (4.2.2)$$

With the assumption that the PSF $P(\mathbf{u})$ is a positive and even function, its phase is consequently a binary values function, given by:

$$\angle P(\mathbf{u}) = \begin{cases} 0 & \text{if } P(\mathbf{u}) \geq 0 \\ \pi & \text{if } P(\mathbf{u}) < 0 \end{cases}. \quad (4.2.3)$$

This means:

$$P(\mathbf{u}) = \sum_{\mathbf{x} \in \mathcal{N}_0} \mathbf{p}(\mathbf{x}) \cos(2\pi \mathbf{u}^T \mathbf{x}), \quad (4.2.4)$$

where $\mathbf{p}(\mathbf{x})$ is the PSF representation in the spatial domain, \mathbf{x} is a vector of coordinates $[x, y]^T$, \mathcal{N}_0 is a window region of $M \times M$ pixels.

In other words, since the value of $P(\mathbf{u})$ is positive when $\angle P(\mathbf{u}) = 0$, the phase parts of $G(\mathbf{u})$ and $F(\mathbf{u})$ are equal, hence a blur invariant representation can be obtained by using phase information.

4.2.2 LPQ in detail

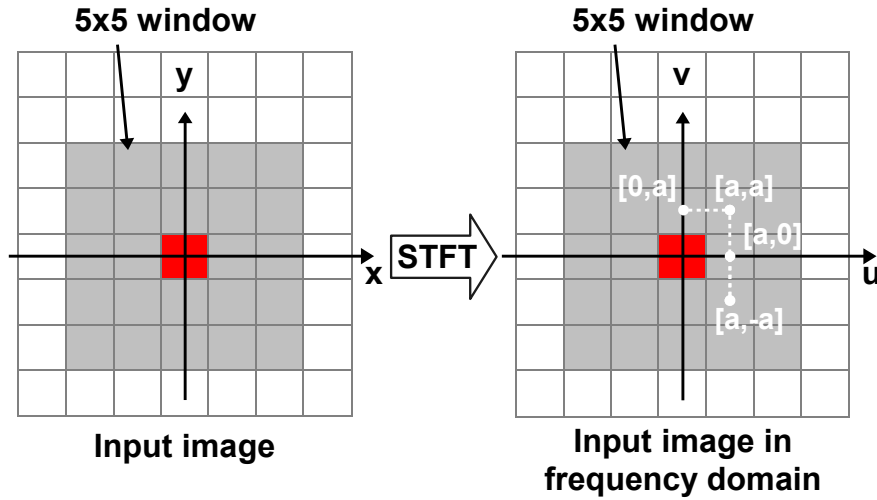


Figure 4.8: The neighborhood and Fourier frequencies in LPQ.

Based on the blur invariance of phase response from Fourier transformation, Ojansivu et al. [92] developed local phase quantization (LPQ) method to extract local phase information. At each pixel \mathbf{x} of the input image, a short-term Fourier transform (STFT, a special case of 2-D Discrete Fourier Transform) is applied over its rectangular $M \times M$ neighborhood by the formula:

$$F(\mathbf{u}, \mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{N}_{\mathbf{x}}} f(\mathbf{x} - \mathbf{y}) e^{-j2\pi \mathbf{u}^T \mathbf{y}} = \mathbf{w}_{\mathbf{u}}^T \mathbf{f}_{\mathbf{x}}, \quad (4.2.5)$$

where $\mathbf{w}_{\mathbf{u}}$ is the DFT's window function at 2-D frequency \mathbf{u} , $\mathbf{f}_{\mathbf{x}}$ is the vector that contains $M \times M$ pixels (image samples) at \mathbf{x} position, and $\mathcal{N}_{\mathbf{x}}$ is the window region associated with \mathbf{x} position. One can notice from Eq. 4.2.5 that STFT is separable, therefore it can be implemented by applying 1-D convolutions on rows and columns successively. Then the local coefficients $\mathbf{F}_{\mathbf{x}}^c$ are calculated for every pixel at four low frequencies (as shown in Fig. 4.8), corresponding to 2-D frequencies $\mathbf{u}_1 = [a, 0]^T$, $\mathbf{u}_2 = [0, a]^T$, $\mathbf{u}_3 = [a, a]^T$, and $\mathbf{u}_4 = [a, -a]^T$, in which a is the highest scalar frequency satisfying the condition $P(\mathbf{u}_i) \geq 0$. There are some ways to compute a from M but in this work, by empirical

4.2. LPQ as a facial feature extraction

experiments, we fix:

$$a = \frac{8}{M-1}. \quad (4.2.6)$$

With

$$\mathbf{F}_x^c = [\mathbf{F}(\mathbf{u}_1, \mathbf{x}), \mathbf{F}(\mathbf{u}_2, \mathbf{x}), \mathbf{F}(\mathbf{u}_3, \mathbf{x}), \mathbf{F}(\mathbf{u}_4, \mathbf{x})], \text{ and} \quad (4.2.7)$$

$$\mathbf{F}_x = [\text{Re}\{\mathbf{F}_x^c\}, \text{Im}\{\mathbf{F}_x^c\}]^T, \quad (4.2.8)$$

where $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ return the real and imaginary responses of \mathbf{F}_x^c , respectively. The STFT transform can be rewritten by vector notation as:

$$\mathbf{F}_x = \mathbf{W}\mathbf{f}_x, \quad (4.2.9)$$

where \mathbf{W} is a $8 \times M^2$ transformation matrix computed from four frequencies $\mathbf{u}_i, i = 1..4$ by the equation

$$\mathbf{W} = [\text{Re}\{\mathbf{w}_{\mathbf{u}_1}, \mathbf{w}_{\mathbf{u}_2}, \mathbf{w}_{\mathbf{u}_3}, \mathbf{w}_{\mathbf{u}_4}\}, \text{Im}\{\mathbf{w}_{\mathbf{u}_1}, \mathbf{w}_{\mathbf{u}_2}, \mathbf{w}_{\mathbf{u}_3}, \mathbf{w}_{\mathbf{u}_4}\}]^T. \quad (4.2.10)$$

To improve the performance of LPQ by maximally preserving scalar quantization information, Ojansivu et al. [92] applied a whitening transform to decorrelate \mathbf{F}_x

$$\mathbf{G}_x = \mathbf{V}^T \mathbf{F}_x, \quad (4.2.11)$$

where \mathbf{V} is an orthogonal matrix derived by using singular value decomposition of the matrix \mathbf{D} which is

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (4.2.12)$$

\mathbf{D} is the covariance matrix of the Fourier coefficients \mathbf{F}_x and can be expressed by

$$\mathbf{D} = \mathbf{W}\mathbf{C}\mathbf{W}^T, \quad (4.2.13)$$

where \mathbf{C} is the covariance matrix of $M \times M$ samples in \mathcal{N}_x and can be computed as:

$$\mathbf{C} = \begin{pmatrix} 1 & \sigma_{1,2} & \cdots & \sigma_{1,M^2} \\ \sigma_{2,1} & 1 & \cdots & \sigma_{2,M^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M^2,1} & \sigma_{M^2,2} & \cdots & 1 \end{pmatrix}. \quad (4.2.14)$$

In (4.2.14), $\sigma_{i,j} = \rho^{\|\mathbf{x}_i - \mathbf{x}_j\|}$ ($\|\cdot\|$ is the L_2 norm, ρ is the correlation coefficient between adjacent pixel values with assumption that the image function $f(\mathbf{x})$ is a result of a first-order Markov process, and the variance of each sample is 1) is the covariance between \mathbf{x}_i and \mathbf{x}_j . After decorrelating operation, the j th decorrelated coefficient g_j of \mathbf{G}_x is quantized by a binary quantizer

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (4.2.15)$$

Then the quantized values are represented as 8-bits decimal numbers (in the 0-255 range) by a simple binary coding

$$LPQ_{desc} = \sum_{j=1}^8 q_j 2^{j-1}. \quad (4.2.16)$$

As LPQ codewords are in the range 0-255, an LPQ image² containing local phase quantized information of the input image is obtained as a result. The LPQ image is divided into $R \times C$ non-overlapped rectangular sub-regions to calculate their histograms. These histogram sequences are then concatenated to build the Local Phase Quantization (LPQ) feature vector for FR. The whole process of building a facial representation from a given face image by using LPQ operator is illustrated in Fig. 4.9.

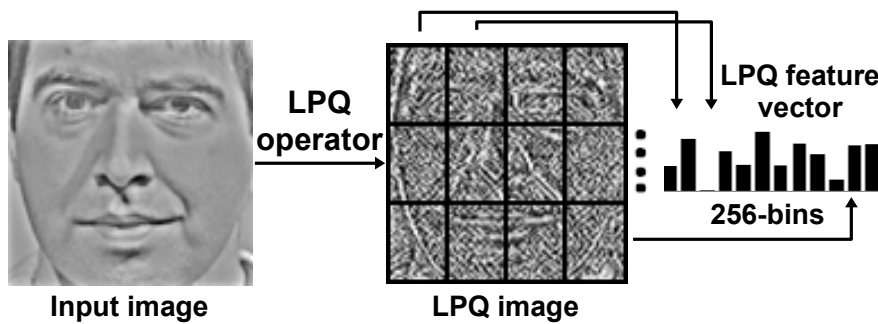


Figure 4.9: LPQ feature vector computation

It is worth indicating that there is no special technique to reduce the length of a LPQ feature vector as the uniform patterns effectively do with ELBP and LBP ones. Hence, when using the same number of sub-regions, a LPQ vector is a 256-bins representation and is about four times and two times longer than a ELBP(h) vector and a ELBP(h+v) one, respectively.

4.2.3 Face recognition with LPQ

Using LPQ for feature extraction, two FR frameworks, the *Template matching* (with details are presented in Section 3.5 of chapter 3) and the *WPCA based* that employs *negative cosine distance* function for classification and WPCA for dimension reduction (for more details, see Section 3.6 in chapter 3) are formed and assessed on AR, FERET and SCface databases. These systems are named as *LPQ* for the *Template matching* case and as *LPQ WPCA* when referring to the *WPCA based* framework in the comparison tables in the next Section.

²The same as in ELBP method in previous Section, the LPQ label of image pixels at the image border are not calculated.

4.2. LPQ as a facial feature extraction

4.2.4 Experimental results

4.2.4.1 Results on AR database

Table 4.9: Rank-1 RRs (%) comparison between ELBP(h+v) and LPQ based methods on AR database

Test/Method		Probe set												Avg
		1	2	3	4	5	6	7	8	9	10	11	12	
Exp1	ELBP(h+v)	100	100	81.2	100	100	98.5	91.0	55.2	47.0	99.3	86.6	79.1	86.5
	LPQ	100	100	82.7	100	100	100	85.1	69.4	56.7	99.3	97.0	94.8	90.4
	ELBP(h+v) WPCA	100	100	81.2	100	100	100	91.0	67.2	61.9	100	97.8	97.0	91.3
	LPQ WPCA	100	100	83.5	100	100	100	88.8	79.6	67.2	100	97.0	95.5	92.6
Exp2	ELBP(h+v)	100	100	79.0	100	100	100	87.3	58.5	49.2	98.3	85.6	81.4	86.6
	LPQ	100	100	77.1	100	100	100	89.0	72.0	65.3	99.2	98.3	99.2	91.7
	ELBP(h+v) WPCA	100	100	80.5	100	100	100	87.3	70.3	61.8	100	97.5	96.6	91.2
	LPQ WPCA	100	100	83.9	100	100	100	92.4	79.7	70.3	100	98.3	99.2	93.7
Exp3	ELBP(h+v)	96.6	98.3	62.2	97.5	94.9	83.9	65.6	52.5	47.5	95.8	58.5	52.5	75.5
	LPQ	95.8	97.5	67.0	99.2	97.5	93.2	78.0	64.4	57.6	94.1	89.0	83.9	84.8
	ELBP(h+v) WPCA	97.5	99.2	73.7	100	100	95.8	80.5	66.1	63.6	97.5	90.7	92.4	88.1
	LPQ WPCA	98.3	100	74.6	100	100	96.6	78.0	74.6	65.3	97.5	89.0	86.4	88.4
Exp4	ELBP(h+v)	95.7	95.7	48.3	98.3	98.3	90.6	68.4	63.3	55.6	82.1	62.4	58.1	76.4
	LPQ	94.0	99.2	69.0	100	100	98.3	69.2	71.8	63.3	94.0	89.0	89.7	86.5
	ELBP(h+v) WPCA	98.3	98.3	63.8	100	100	97.4	79.5	67.5	59.8	95.7	90.6	91.5	86.9
	LPQ WPCA	96.6	99.2	77.6	100	100	100	79.5	77.8	71.8	94.0	89.0	89.7	89.6

The full results of LPQ based FR frameworks on AR database are reported and compared with those of ELBP(h+v) based ones in table 4.9 while the comparisons between LPQ WPCA and other state-of-the-art methods is shown in table 4.10. Plus, the RRs of LPQ and ELBP(h+v) based systems are plotted in Figs. 4.10 for more visual comparisons.

It is apparent from table 4.9 and Figs. 4.10 that in the *Template matching* framework, LPQ significantly outperforms ELBP(h+v), particularly upon challenging probe sets, such as 8, 9, 11, 12 (Scarf and Sun glasses with illumination variations) of four experiments and 3 in experiments 3 and 4. On average recognition rate, in experiments 1 and 2, LPQ is about 4% and 5% higher than ELBP(h+v), respectively, while those numbers in experiments 3 and 4 are even more impressive: about 9% and 10%. The LPQ WPCA based framework also has higher overall results than ELBP(h+v) WPCA in all experiments. For more detailed comparison, in experiments 3 and 4, where there is the presence of time-lapse variations, on *Sun glasses* probe sets (number 10, 11 and 12), LPQ WPCA has lower, but not much, recognition rates than ELBP(h+v) WPCA. In summary, we conclude that LPQ is more robust than ELBP(h+v), ELBP(h) and LBP [3] feature extraction methods when coping with facial expressions, illumination, time-lapse variations and occlusions. Another conclusion from very high results of LPQ WPCA (all 4 experiments) on probe sets number 4, 5, and 6 (Neutral expression with lighting changes) is that LPQ is robust to illumination variations.

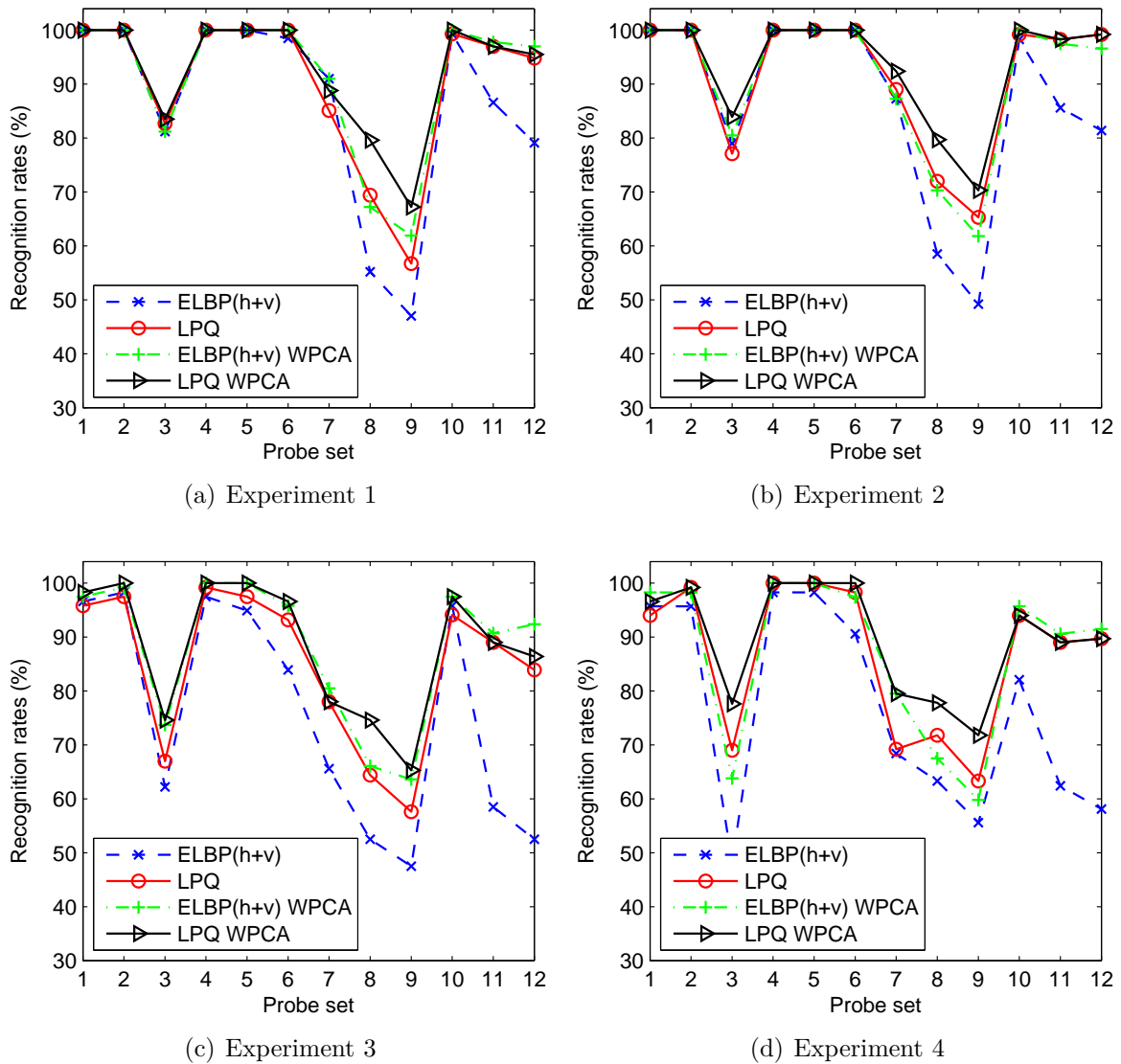


Figure 4.10: Accuracy performance of LPQ and ELBP(h+v) based systems on AR database.

Based on the comparison results in table 4.10, it is clear that LPQ WPCA framework is efficient against facial expressions, time-lapse variations and occlusions as its recognition rates are higher than many other contemporary systems. This superiority is more significant particularly upon *Scream*, *Scarf* and *Sun glasses* image sets in all experiments. Our system is the only one that gains 100% recognition rates on *Smile*, *Anger* and *Scarf* probe sets in experiment 1 and 2.

4.2.4.2 Results on FERET database

Frontal FERET image sets

4.2. LPQ as a facial feature extraction

Table 4.10: Rank-1 RRs (%) comparison with other contemporary systems on AR database using the same evaluation method

	1	2	3	7	10	
Method	Smile	Anger	Scream	Glasses	Scarf	Classes ¹
Exp 1						
S-LNMF [90]	96.0	N/A ²	49.0	84.0	87.0	100
LGBP[135]	N/A ²			80.0	98.0	50
IRF [141]	N/A ²			87.5	91.7	120
String face [21]	87.5	87.5	25.9	88.0	96.0	100
Sparse coding [129]	N/A ²			94.7	91.0	100
DMMA[78]	99.0	93.0	69.0	N/A ²		100
SIS [75]	99.0			99.0	98.0	100
PLD [54]	99.0			100	97.0	100
Our	100		83.5	88.8	100	134
Exp 2						
DMMA[78]	85.0	79.0	45.0	N/A ²		100
Method in [90]	96.0	N/A ²	54.0	66.0	89.0	100
LGBP[135]	N/A ²			62.0	96.0	50
Sparse coding [129]	N/A ²			80.3	72.7	100
String face [21]	N/A ²			76.0	88.0	100
IRF [141]	N/A ²			82.5	84.0	120
SIS [75]	86.0			96.0	90.0	100
Our	100		83.9	92.4	100	118
Exp 3						
S-LNMF [90]	62.0	N/A ²	27.0	49.0	55.0	100
Method in [84]	N/A ²		52.3	54.2	81.3	80
PLD [54]	86.0			90.0	89.0	100
Our	98.3	100	74.6	78.0	97.5	118

¹ : The classes column is the number of persons whose images are used in experiments.

² N/A: Not available result.

The results of LPQ based FR systems on FERET database (using the standard protocol) in comparison with ELBP based and other state-of-the-art rivals are presented in table 4.11. These results confirm that LPQ is more powerful than LBP, ELBP(h) and ELBP(h+v) against facial expressions, illumination and time-lapse variations in both *Template matching* and *WPCA based* frameworks. The dominance of LPQ over LBP (LGBPHS [135], HMBP [130], DLBP [82], Tan et al. [113] and POEM PDO [118]) and ELBP based methods is more convincing on Dup 1 and Dup 2 probe sets, whose

Table 4.11: Rank-1 RRs (%) comparison of LPQ based systems with other state-of-the-art results on the FERET database [96]

Method	Fb	Fc	Dup 1	Dup 2	Average
ELBP(h+v)	97.0	95.4	72.0	71.0	86.6
LGBPHS [135]	98.0	97.0	74.0	71.0	87.8
HMBP [130]	98.1	98.5	75.8	75.2	89.0
GEWC [29]	96.3	99.5	78.8	77.8	89.3
LPQ	97.7	97.9	79.5	77.8	90.1
HGPP [133]	97.5	99.5	79.5	77.8	90.2
DMMA [78]	98.1	98.5	81.6	83.2	91.6
LGBPWP [88]	98.1	98.9	83.8	81.6	92.1
CHG [22]	97.5	98.5	85.6	84.6	92.6
DLBP [82]	99.0	99.0	86.0	85.5	93.6
Tan et al. [113]	98.0	98.0	90.0	85.0	94.2
ELBP(h+v) WPCA [89]	99.4	100	89.1	86.8	95.0
LMG [94]	99.8	100	89.2	86.8	95.3
ESRC [30]	97.3	99.0	93.8	92.3	95.9
MS-LPQ [17]	99.2	100	92.0	88.0	95.9
EPFDA [104]	99.6	99.0	92.0	88.9	96.1
POEM PDO [118]	99.7	100	91.7	90.6	96.4
LPQ WPCA	99.5	100	92.9	91.0	96.7
FLPGMP [107]	99.0	99.0	94.0	93.0	96.9
G-LQP [52]	99.9	100	93.2	91.0	97.0
MBC-F [128]	99.7	99.5	93.6	91.5	97.0
GSF [123]	99.6	99.5	94.0	91.5	97.1

images are affected by time-lapse variation, one of the most challenging factors of FR. Also, the high recognition rate of LPQ based frameworks lead us to new findings that LPQ is an efficient feature extraction method under facial expression, illumination and time-lapse variations and the combination of LPQ with WPCA can constitute an excellent FR system.

Apparently, the results of LPQ WPCA are lower than leading systems (such as FLPGMP [107], G-LQP [52], MBC-F [128], GSF [123]) as they all employ advanced multi-resolution/multi-scale feature extraction methods based on Gabor wavelets or Monogenic filter.

Non-frontal FERET image sets

In this Section, recognition rates of LPQ based frameworks are exhibited and compared with ELBP (h+v) WPCA in table 4.12 while the comparison between them and other

4.2. LPQ as a facial feature extraction

Table 4.12: Rank-1 RRs comparison between ELBP(h+v) WPCA and LPQ based methods on b-series of FERET database

	ELBP(h+v) WPCA	LPQ	LPQ WPCA
-40°-bh	80.5	83.5	89.0
-25°-bg	98.5	97.0	99.0
-15°-bf	99.5	99.0	99.5
+15°-be	99.5	99.5	100
+25°-bd	99.0	98.0	99.0
+40°-bc	79.5	81.5	86.5
Avg	92.8	93.1	95.5

Table 4.13: Rank-1 RRs comparison of LPQ WPCA with other leading systems on FERET b-series.

Method	-40° bh	-25° bg	-15° bf	+15° be	+25° bd	+40° bc	Avg
SLF-RKR [127]	N/A ¹	55.0	100	96.0	57.0	N/A ¹	N/A ¹
LSED [121] * ²	78.0	84.0	88.0	89.0	88.0	83.0	85.0
CCA [64] * ²	81.0	91.0	92.0	94.0	89.0	80.0	87.8
PAN [39]	81.5	93.0	97.0	98.5	91.5	78.5	90.0
RFC [20]	84.2	90.2	94.0	93.2	92.5	89.5	90.6
ADMCLS [105]	85.0	94.0	96.0	95.0	94.0	82.0	91.0
LMG [94]	N/A ¹	91.5	98.0	98.5	93.5	N/A ¹	N/A ¹
MRH [7] * ²	87.0	94.0	98.0	99.0	96.0	74.0	91.3
GLOH [100] ** ³	81.1	94.5	100	100	94.5	81.1	91.9
DWFF [86]	87.5	98.0	100	99.0	98.5	82.4	94.2
LPQ WPCA	89.0	99.0	99.5	100	99.0	86.5	95.5
MRF [48]	91.0	97.3	98.0	98.5	96.5	91.5	95.5
3D Pose Norm [6]	90.5	98.0	98.5	97.5	97.0	91.9	95.6
CPN [31]	94.5	98.0	98.5	99.0	98.5	97.0	97.6

¹ N/A: Not available result.

² *: The RRs of the method are estimated from plotted figures.

³ **: The RRs on $\pm 25^\circ$ and $\pm 40^\circ$ subsets are average results.

systems can be seen in table 4.13.

As shown in table 4.12, LPQ convincingly outperforms ELBP(h+v) method when its *Template matching* framework can even achieve higher results than the ELBP(h+v) WPCA framework. Moreover, the LPQ WPCA framework attains very encouraging recognition rates. Based on these numbers, we conclude that LPQ is better than ELBP(h+v) in coping with pose variations and the LPQ WPCA framework is more

efficient than the LPQ based *Template matching* one.

Although LPQ WPCA is not devoted to pose variation, one can observe from table 4.13 that its results are very interesting when they are higher than many other state-of-the-art systems. While having lower recognition rates than some leading systems (such as 3D Pose Norm [6] or CPN [31]), which are dedicated to handle head pose challenge by employing special tactics, it is not overrated to conclude that LPQ is robust to this difficulty. Upon probe sets of relatively small head pose (in the range of $\pm 25^\circ$), LPQ WPCA even outperforms the best one (CPN [31]). This characteristic of LPQ should be highlighted as pose variation is widely regarded as one of the greatest challenges of the Face recognition problem.

4.2.4.3 Results on SCface database

Table 4.14: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]

Camera/Distance	PCA[42]	DSR[140]	ELBP(h+v) [89]	LPQ
cam1_1	2.3		43.1	54.6
cam1_2	7.7		56.2	58.5
cam1_3	5.4		45.4	41.5
cam2_1	3.1		36.9	43.9
cam2_2	7.7		50.8	54.6
cam2_3	3.9		42.3	40.8
cam3_1	1.5		34.6	39.2
cam3_2	3.9	N/A ¹	46.9	54.6
cam3_3	7.7		51.5	47.7
cam4_1	0.7		32.3	37.7
cam4_2	3.9		50.0	56.9
cam4_3	8.5		50.8	48.5
cam5_1	1.5		36.2	46.2
cam5_2	7.7		32.3	40.8
cam5_3	5.4		31.5	33.1
Average	4.7	20.2	42.7	46.6

¹ N/A: Not available result

This Section presents the results of LPQ WPCA on the SCface [42] database following its DayTime and NightTime protocols. The gallery images are also exploited for training stage as in ELBP based systems (for more details, see Section 3.2 of chapter 3 and Section 4.1.4.3 in this chapter). As our results are all yielded by *WPCA based* frameworks, we label each system by the name of the facial description it used.

4.2. LPQ as a facial feature extraction

Table 4.15: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]

Camera/Distance	PCA[42]	ELBP(h+v)	LPQ
cam6_1	1.5	9.2	11.5
cam6_2	3.1	15.4	16.2
cam6_3	3.9	25.4	25.4
cam7_1	0.7	13.1	13.9
cam7_2	5.4	13.1	15.4
cam7_3	4.6	13.9	14.6
Average	3.2	15.0	16.2

It is clear from table 4.14 and table 4.15 that LPQ WPCA method outperforms other state of the art systems in both experiments. Our average result is about 10 times higher than PCA baseline [42] in DayTime tests. In NightTime tests, the average recognition rate of PCA baseline [42] is 5 times lower than ours. LPQ WPCA also significantly outperforms ELBP(h+v) WPCA [89] on most of the probe sets and obtains better overall recognition rates by 3.9% (46.6% vs. 42.7%) in DayTime experiment and 1.2% (16.2% vs. 15.0%) in NightTime one. According to the best of our knowledge, LPQ WPCA’s results upon SCface database are the best reported results to date in face recognition literature. These results are consistent with those upon AR and FERET face databases in previous Sections.

We argue that the LPQ WPCA’s good results upon SCface databases as well as its superiority over ELBP based methods are mainly stemmed from meaningful phase based patterns in LPQ feature vector and the blur tolerant characteristic of LPQ description.

4.2.4.4 LPQ parameters

There are several parameters in building a LPQ vector: the window size of the STFT transform- M , the whitening argument- ρ and the $R \times C$ sub-regions for capturing the final spatial histogram sequence. Empirically, ρ is fixed at 0.91 for all experiments. The divided sub-regions are assigned as 8×9 and 10×10 for AR and FERET databases, respectively. On SCface database, since the images are of small sizes, they are divided into 4×4 non-overlapped rectangular sub-regions for doing the feature extraction with LPQ. Also, the M parameter depends on the image’s size when it is set at 7 for SCface database and at 9 for AR and FERET ones.

Table 4.16: Computation time of LPQ in comparison with other feature extraction methods

Method	Image size	Time (seconds)	Extraction time (milliseconds)	Images/second
ELBP(h)	128×128	4.44	3.71	269
LPQ	128×128	5.45	4.56	219
ELBP(h+v)	128×128	7.28	6.09	164
Monogenic [128]* ¹	150×130	19.81	16.56	60
MBC-A [128]* ¹	150×130	30.54	25.54	39
MBC-O [128]* ¹	150×130	87.0	72.74	14
Gabor wavelets	88×80	96.23	80.46	12

¹ *: We used the Matlab code provided by the author.

4.2.4.5 Computational cost

To investigate the computational performance of LPQ, the same benchmark tests as described in Section 4.1.4.5 of this chapter are carried out with it and the obtained metrics are compared with those of ELBP based and some other feature extraction methods.

From the benchmark results (in table 4.16), it is apparent that LPQ is very fast, despite the fact that its speed is a little slower than ELBP(h). It is because of the eight 1D convolution operations that are required to build the LPQ feature vector while in ELBP(h) (and also LBP [3]), only thresholding between each image pixel and its neighbors are done. LPQ is faster than ELBP(v+h) [89] since ELBP(v+h) needs a pair of ELBP operators to form a facial representation. Gabor wavelets images generation are fairly slow, although it is only an initial step and it works on much smaller images (80x88 resolution). Based on the comparison between LPQ and Gabor wavelets initial step (in table 4.16), we conclude that LPQ is about $96.23/5.45 = 17.7$ times faster than every Gabor wavelets based feature extraction method (such as LGBP [135], HGPP [133], FLPMP [107], GFS [123], G-LQP [52], MFR [48], DWFF [86], 3D Pose Norm [6], CPN [31], etc.). With its 219 images per second processing speed, LPQ could certainly be applied in more efficient multi-resolution/multi-scale feature extraction methods.

4.2.5 Conclusions

We summarize here the intriguing properties of the LPQ feature extraction method. For each pixel, LPQ captures local phase based patterns based on STFT transformation over its square neighborhood whose center is the pixel itself. The LPQ representation has the following properties:

4.3. Conclusions

1. LPQ patterns are phase based features and they are blur invariant. As the LPQ codeword of an image pixel is calculated based on all the pixels in a square window centered at the pixel, the dependence of that pixel and its neighbors is stronger than in LBP and ELBP methods. Thus, LPQ is more robust than LBP and ELBP methods when applying them on intensity image. This superiority of LPQ is proved to be stable and consistent under a wide range of FR issues such as facial expression, illumination, pose, time-lapse variations, partial occlusions and with low resolution images.
2. LPQ is an efficient facial feature extraction method when dealing with various FR challenges which are mentioned in the above remark. The extensive experimental results upon AR, FERET and SCface databases are apparent evidences for this conclusion.
3. LPQ is fast to compute and consequently can be further harnessed to develop more advanced feature extraction methods by using a multi-resolution/multi-scale strategy.
4. When joining LPQ with WPCA, the LPQ WPCA framework achieves excellent results and outperforms many state-of-the-art LBP and Gabor wavelets based systems in the FR literature.

Besides those precious characteristics, LPQ method has one drawback: its feature vector is a dense description of 256-bins histogram sequence, and thus it requires more memory space to be stored than ELBP based methods but does not affect its processing speed. How to resolve this issue to reduce LPQ vector's length before feeding it into the training and projecting steps of WPCA is still an open question.

Obtained results presented through out this Section once again confirm the advantages of the WPCA based framework over the *Template matching* one when all the comparisons between the two systems in terms of accuracy performance are led to that unassailable conclusion.

4.3 Conclusions

On the way to devise an elite facial representation, this chapter presents two intensity based feature extraction methods: the Elliptical Local Binary Patterns (ELBP) and the Local Phase Quantization (LPQ). Both of them are local feature based approaches that extract local micro patterns from the face images to build their feature vectors. In ELBP, a novel variant of LBP, an ELBP label is computed from its scattered neighbor pixels lied on an horizontal ellipse instead of a circle as in LBP. Its inspiration comes

naturally from the shape and structures of the human faces: the eyes, mouth are ellipses and there is more horizontal information, which drives the face perception, than vertical one. By this, ELBP is more domain-specific than LBP, but, on the contrary it may neither be rotation invariant since the sampling patterns are not circular, nor be good for texture classification task as original LBP.

Based on extensive experiments upon three large public face databases, [AR](#), [FERET](#) and [SCface](#), we have pointed out that horizontal ELBP (using an horizontal ellipse sample) is more efficient than LBP. Further, we have shown that the fusion of horizontal ELBP with its vertical one, under the form of a symmetric pair of descriptions, achieves higher accuracy than using it solely. This results from a richer set of valuable features contained in the fused vector in comparison with that of a horizontal ELBP one.

From the comparisons of recognition rates obtained on the same databases as ELBP, LPQ is proved to be an efficient facial feature extraction method under a wide spectrum of [FR](#) challenges as its accuracy performance surpasses many contemporary LBP and its variants, including the ELBP, and Gabor wavelets based systems. Hence, we suppose that LPQ is worthy of receiving more attentions from [FR](#) researchers.

The computational cost of ELBP and LPQ is also practically assessed via benchmark experiments upon the Fa set of the [FERET](#) database. The comparative results show that both of them are computation efficient even though their tested implementations are neither optimized nor parallelized.

The excellent performance, with respect to both high accuracy and fast processing speed requirements, of ELBP and LPQ makes them good candidates for being primitive descriptors that can be further applied in building more robust multi-resolution/multi-scale feature extraction methods.

Importantly, the experiment results in this chapter evidently validate that, when using the same facial representation, the supervised [FR](#) framework, namely *WPCA based*, significantly outperforms its *Template matching* counterpart and if one feature extraction method is superior over another one, this superiority will be reflected in both frameworks' outcomes. Thus, from now on, in this thesis, experiments for evaluating the efficiency of a feature extraction method are conducted only by the *WPCA based* framework.

The comparison results in this chapter also lead us to another critical conclusion that the preeminence of one feature extraction method over other ones, as well as that of a [FR](#) system over other competitors, is illustrated more apparently and in a more credible way under the toughest challenges such as *Scream* expression (among other ones), time-lapse (Exp 3 and 4 on [AR](#) database, Dup 1 and Dup 2 tests on [FERET](#) database) and pose variations (on b-series images of [FERET](#) database), and especially low resolution probe images (from [SCface](#) database).

Chapter 5

Patch based Local Phase Quantization of Monogenic components for Face recognition

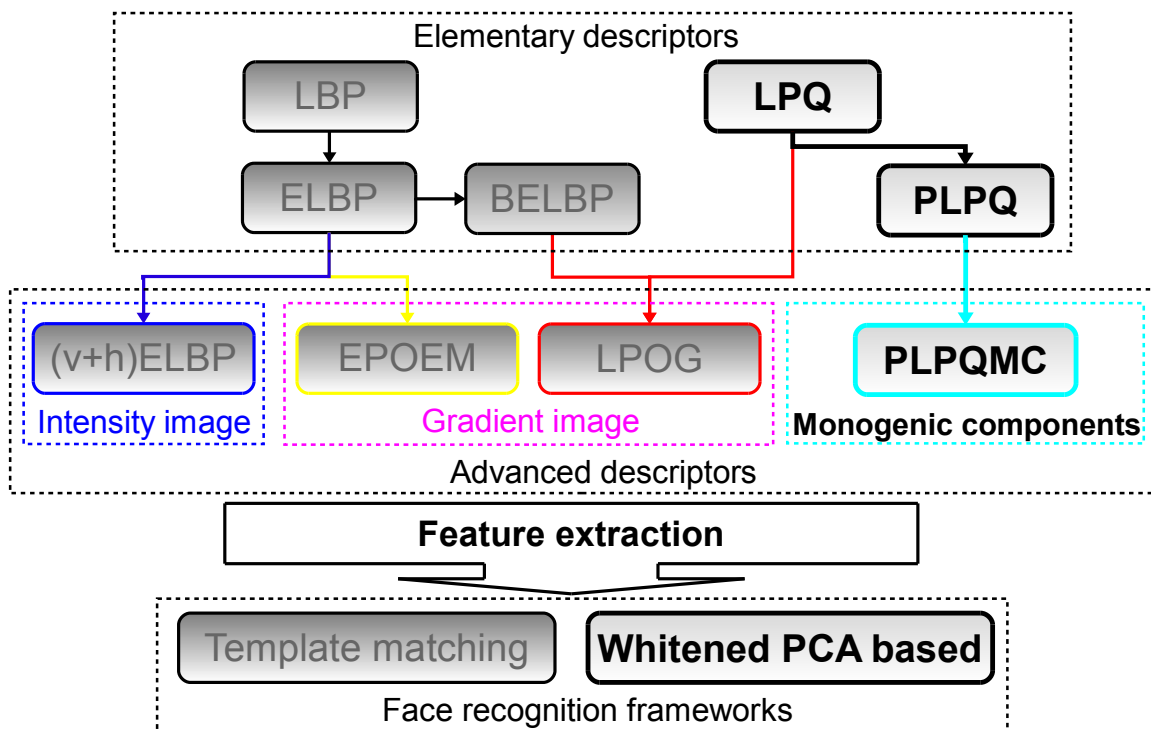


Figure 5.1: Contributions presented in this chapter: Patch based LPQ of Monogenic components WPCA system.

5.1 Introduction

The in-depth survey in chapter 2 provides a lot of clues that an elementary feature extraction method does not fulfil the need of a high accuracy FR system. This is reinforced in the previous chapter by numerous comparisons between results of ELBP and LPQ based frameworks¹, in which the facial features are extracted directly from intensity images, and those of other state-of-the-art systems. Besides, the best results upon the FERET database (both on frontal images and pose varied ones), the most widely adopted benchmark for evaluating the recognition performance of a FR system, are all delivered by multi-resolution/multi-scale feature extraction methods (for more details, refer to Section 4.2.4.2 in chapter 4). While almost all of these methods are Gabor wavelets based, Monogenic filter is also used by a fusion strategy in MBC-F system [128] and in MBP [130]. Targeting a fast feature extraction algorithm, we do not turn our attention to Gabor wavelets based direction due to its heavy computational burden. Compared with Gabor wavelets transformation, Monogenic filter requires substantially less computational cost (see table 4.16 in chapter 4 for more details) but it is proved to be capable to offer excellent accuracy. Additionally, owning desirable characteristics and fast speed as shown in chapter 4, LPQ is a standout elementary descriptor for building up a multi-resolution/multi-scale feature extraction method.

Inspired by the advantages of Monogenic filter and LPQ descriptor, in this part of the present dissertation, we propose a novel feature extraction method for Face recognition called patch based Local Phase Quantization of Monogenic components (PLPQMC) by applying patch based LPQ (PLPQ), a new variant of LPQ operator, upon Monogenic directional bandpass images (see Fig. 5.1). From the input image, the directional Monogenic bandpass components are generated. Then, each pixel of a bandpass image is substituted by the mean value of its rectangular neighborhood. Next, LPQ histogram sequences are computed upon those images. Finally, these histogram sequences are concatenated for constituting a global representation of the face image. Using the proposed method for feature extraction, a new WPCA based face recognition system is constructed with Whitened Principal Component Analysis (WPCA) for dimensionality reduction, k-nearest neighbor classifier and *weighted angle distance* for classification. Performance evaluations on the three public face databases, AR, FERET and SCface, show that the PLPQMC feature extraction method is efficient against a broad range of FR challenges, for instance, expressions, illumination, time-lapse, and pose variations, and partial occlusions, as well as low resolution probe images. In the mean time, by comparing the results from these experiments with those of other state-of-the-art counterparts, we evidently verify that the PLPQMC WPCA based framework is competing with the best systems in the FR literature so far. Especially, from these comparisons,

¹In chapter 4, we have shown that the ELBP and LPQ methods are efficient facial representations with remarkable RRs but there are still large gaps between their results and that of the leading systems.

5.2. Monogenic filter

one of the most outstanding characteristic of our method, which is its robustness to illumination changes² when it attains excellent RRs without employing any preprocessing algorithm to normalize the illumination conditions of the face images, is empirically shown. In addition to that, comparative results from timing experiments validate that the proposed method has inexpensive computational cost and is feasible for practical applications in real life.

As the related works of Monogenic filter and LPQ based methods are covered in chapter 2, the remainder of this chapter is organized as follows. We first present in Section 5.2 the steps of Monogenic filter as a multi-scale image analytic technique. Then all the details of the proposed method are given in Section 5.3. Experimental results and comparative studies between them and those of state-of-the-art systems are provided in Section 5.4. Conclusions are finally expressed in Section 5.5.

5.2 Monogenic filter

5.2.1 Log-Gabor filter

Proposed by Field [34], Log-Gabor filter is a bandpass filter and is an alternative of the Gabor function [38] based on the observation that the coding procedure of natural images is more efficient by filters whose Gaussian transformation functions work on logarithmic frequency scale instead of the scale value itself. Keeping the focus on the topic of Monogenic transformation, hereafter, we briefly describe the steps of constructing multi-scale Log-Gabor filters for producing Monogenic components. For more information about Log-Gabor filter, works in [34, 60], and references therein are recommended.

Let the size (rows and columns) of the input image be $rows \times cols$ pixels, then two matrices of the same size, u_1 and u_2 , which contain horizontal and vertical frequency in the range $[-0.5, 0.5]$, are first computed as:

$$\begin{cases} u_1(i, j) &= -0.5 + (j - 1) * colStep \\ u_2(j, i) &= u_1(i, j) \end{cases}, \quad (5.2.1)$$

in which (i, j) is the location of one image pixel, $colStep$ is a stepping value calculated from $cols$ by the following formula:

$$colStep = \begin{cases} \frac{1}{cols} & \text{if } cols \text{ is even} \\ \frac{1}{cols-1} & \text{otherwise} \end{cases}. \quad (5.2.2)$$

²Within the FR domain, illumination variations is widely accepted as one of the biggest challenges that a FR system must face with [137, 65].

Then u_1 and u_2 are quadrant shifted to move zero frequency to their corners:

$$\begin{cases} u_1 = \text{ifftshift}(u_1) \\ u_2 = \text{ifftshift}(u_2) \end{cases}, \quad (5.2.3)$$

where *ifftshift* is a function that swaps the first and the third quadrants, and the second and fourth quadrants of the input matrix.

Next, a matrix, namely *radius*, whose values are frequencies as radii from the center, is generated as:

$$\text{radius} = \sqrt{u_1^2 + u_2^2} \quad (5.2.4)$$

To avoid trouble caused by zero radius value when taking its logarithm or dividing by it, the top left corner of *radius* is assigned to 0:

$$\text{radius}(1, 1) = 0. \quad (5.2.5)$$

In the ensuing step, at each scale s (in the range $[1, \text{maxScale}]$), one Log-Gabor filter, named as logGabor_s , is computed as:

$$\begin{cases} \text{wavelength} = \text{minWaveLength} * \text{mult}^{(s-1)} \\ \text{fo} = \text{wavelength}^{-1} \\ \text{exp} = \frac{-\log(\frac{\text{radius}}{\text{fo}})^2}{2 * \log(\text{sigmaOnf})^2} \\ \text{logGabor}_s = e^{\text{exp}} \\ \text{logGabor}_s(1, 1) = 0 \end{cases}, \quad (5.2.6)$$

where *minWaveLength* is the wavelength of the smallest scale filter, *mult* is scaling factor between successive filters, while *sigmaOnf* is the parameter used to control the bandwidth of the filter, and the last Eq. is for setting the value at the 0 frequency point back to zero.

With these Log-Gabor filters, each input image is decomposed into multiple components by Monogenic transformation whose details are presented in the subsequent Section.

5.2.2 Image representation by Monogenic filter

Monogenic [32] is a signal analysis tool towards decomposing an image into multiple components of different types. These components are local amplitude A (also known as local magnitude), local orientation $\theta \in [0, \pi)$, local phase $\phi \in [0, 2\pi)$, and bandpass components h, h_x, h_y (for more details, see Fig. 5.2). By using the Riesz transform, for each scale value, Monogenic components of an input image I is computed as:

$$\begin{cases} A = \sqrt{h^2 + h_x^2 + h_y^2} \\ \theta = \text{atan}(h_y/h_x) \\ \phi = -\text{sign}(h_x)\text{atan2}(\sqrt{h_x^2 + h_y^2}, h) \end{cases}, \quad (5.2.7)$$

5.2. Monogenic filter

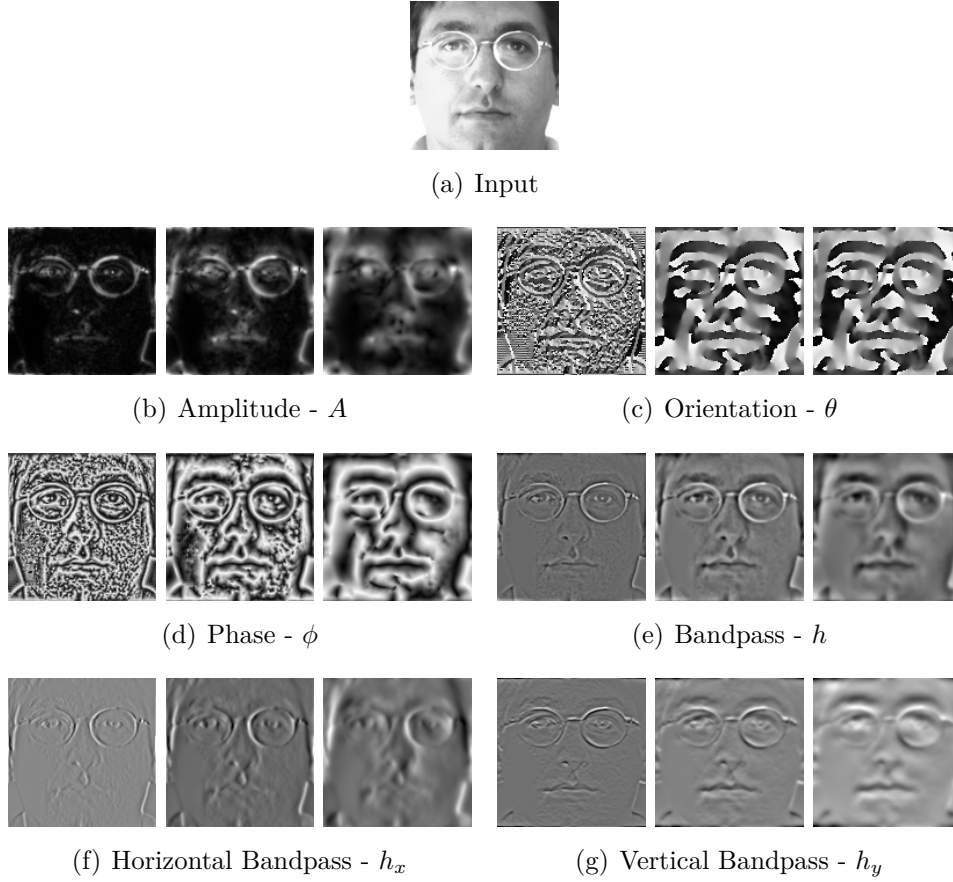


Figure 5.2: A face image and its Monogenic components at 3 scales.

where,

$$\begin{cases} h = I * F^{-1}(G(\omega)) \\ h_d = F^{-1}((i \cdot \omega_d / \sqrt{\omega_x^2 + \omega_y^2}) F(h)), d \in \{x, y\} \end{cases}, \quad (5.2.8)$$

in which “*” denotes the convolution operator, $i = \sqrt{-1}$, F is the 2D Fourier transform, $G(\omega)$ is the bandpass Log-Gabor filter response, ω_x and ω_y are the oriented frequencies (horizontal and vertical).

With one Log-Gabor filter $\log Gabor_s$ produced from previous Section, we have:

$$\begin{cases} \omega_x = u_1 \\ \omega_y = u_2 \end{cases}, \quad (5.2.9)$$

Then, by applying filtering operations in the frequency domain and retaining the spatial results, h , h_x , and h_y are generated as:

$$\begin{cases} h = \text{real}(F^{-1}(F(I) \odot \log Gabor_s)) \\ h_d = \text{real}(F^{-1}(F(I) \odot ((i \cdot \omega_d / \text{radius}) \odot \log Gabor_s))), d \in \{x, y\} \end{cases}, \quad (5.2.10)$$

where $radius$ is calculated from u_1 and u_2 by Eq. 5.2.4 and \odot is the notation of the element-wise multiplication between two matrices.

As Log-Gabors are bandpass filters, a multi-scale (from 3 to 5) Monogenic filter is mandatory to have a complete representation of an image. Our approach in this thesis uses three scales as they give the best results.

5.3 Patched based LPQ of Monogenic bandpass components for FR

5.3.1 Patched based LPQ of Monogenic bandpass components feature extraction method

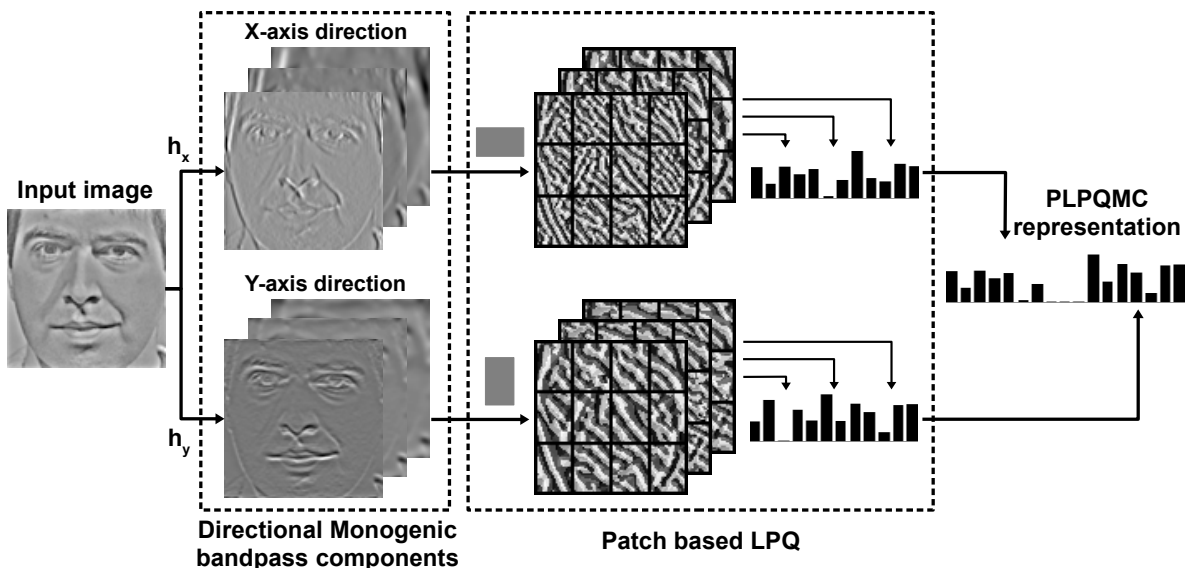
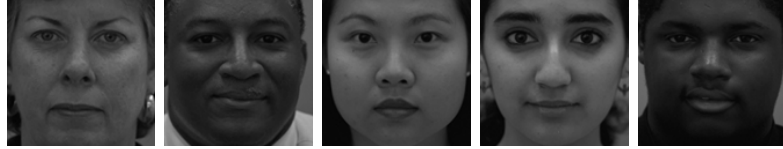


Figure 5.3: Steps in PLPQ of Monogenic components feature extraction method.

The procedure to compute the PLPQMC feature vector of a given face image is illustrated in Fig. 5.3. In the first step, Monogenic directional bandpass components (DMBC) at three scales are generated from the input image. Then, a new variant of LPQ called patch based LPQ (PLPQ) is applied on each of these DMBCs to produce their own descriptions. In PLPQ, we replace each image pixel with the average value of the pixels in a rectangular neighborhood whose center is the pixel itself before applying the LPQ operator. With this step, the relation between one pixel and its neighbors is taken into account, thus making the feature extraction more robust. Here, we use two kinds of PLPQ namely horizontal and vertical, corresponding to two patch pattern types (horizontal and vertical rectangles), for x-axis and y-axis DMBCs respectively. Specifically,

5.3. Patched based LPQ of Monogenic bandpass components for FR

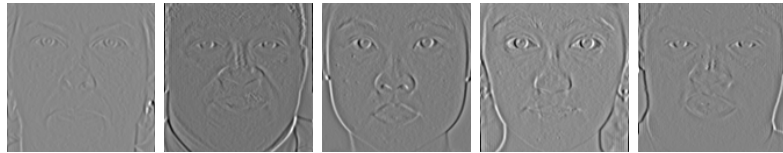
3×5 pattern is used for x-axis DBCs and 5×3 is used for y-axis DBCs. After applying PLPQ operators, each PLPQ image is split into non-overlapped rectangular sub-regions to calculate their histogram sequences. Then each PLPQ description of one PLPQ image is constituted by concatenating its corresponding histogram sequences. At the last step, the PLPQMC representation is built up by aggregating all the individual PLPQ descriptions as a whole.



(a) Input images



(b) Bandpass component



(c) Horizontal bandpass component



(d) Vertical bandpass component

Figure 5.4: Comparison between bandpass components of illumination images from FERET database.

Different from other Monogenic based methods ([130, 128]), which use LBP and its variants on amplitude, phase and orientation components, our method employs PLPQ upon DBCs for three reasons:

- Firstly, since LPQ was proved to be strong against many FR challenges in chapter 4, we suppose that PLPQ is more efficient for FR since each PLPQ pixel, being the mean value of an image patch, carries more discriminating information than the corresponding intensity value of the DBCs. Furthermore, by using correspondingly horizontal and vertical patches for horizontal and vertical DBCs, the distinctive property of the PLPQMC description is accelerated.

- Secondly, the bandpass components contain finer details from input image than the rest ones (for more details, refer to Fig. 5.2) so the feature extraction process performs on these components is better.
- The last reason is that the feature vector extracted from directional bandpass components (h_x and h_y) is more discriminative than the ones computed just from h or from all the bandpass components (h , h_x and h_y). Moreover, as intuitively illustrated in Fig. 5.4, while the bandpass components h (Fig. 5.4b) of images affected by lighting changes are in varied illumination conditions, their directional versions (Fig. 5.4 (c-d)), without losing much image information, are less influenced when having the similar illumination conditions. Thus, the facial representation extracted from DBCs is robust to illumination variations. This will be empirically justified by comparing the results of PLPQMC for FR when no illumination normalization technique is employed with recognition accuracies from other systems in Section 5.4.

Besides, since our method uses 3 scales for generating Monogenic directional bandpass components, it therefore requires much less computational cost in comparison with other Gabor wavelets based ones which usually work with 5 scales and 8 orientations. More precisely, the PLPQMC feature extraction procedure operates upon 6 DBCs of an input image to form its feature vector.

5.3.2 PLPQMC WPCA FR framework

Using PLPQMC method for feature extraction stage, a new FR system is built based on WPCA based framework described in Section 3.6 of chapter 3. For doing the classification task, the *weighted angle-based* function (Eq. 3.6.2) is used to calculate the similarities between probe images and gallery ones as it gives the highest recognition rates.

5.4 Experimental results

In this Section, comparative studies of the RRs of the proposed method and that of many existing systems are given to prove for its efficiency and the key concepts based on which it is built. Hence, beside reporting the results of PLPQMC WPCA framework, we provide those from LPQMC WPCA one, which uses LPQ instead of PLPQ on DBCs, and PLPQMC WPCA **NP**, which is PLPQMC WPCA but no preprocessing method is adopted for illumination normalization. To be fair, all these systems are evaluated on AR and FERET databases with the same parameters.

5.4. Experimental results

5.4.1 Results on AR database

Table 5.1: Rank-1 RRs (%) comparison between LPQ WPCA and PLPQMC based WPCA methods on AR database

Test/Method		Probe set												Avg
		1	2	3	4	5	6	7	8	9	10	11	12	
Exp1	LPQ	100	100	83.5	100	100	100	88.8	79.6	67.2	100	97.0	95.5	92.6
	LPQMC	100	100	92.5	100	100	100	97.0	96.3	87.3	98.5	98.5	97.8	97.3
	PLPQMC NP	100	100	93.2	100	100	100	97.8	96.3	86.6	99.3	99.3	97.0	97.5
	PLPQMC	100	100	94.7	100	100	100	97.8	96.3	90.3	100	100	97.8	98.1
Exp2	LPQ	100	100	83.9	100	100	100	92.4	79.7	70.3	100	98.3	99.2	93.7
	LPQMC	100	100	94.1	100	100	100	96.6	93.2	91.5	100	98.3	95.7	97.5
	PLPQMC NP	100	100	94.9	100	100	100	97.5	94.9	94.9	99.2	98.3	95.8	98.0
	PLPQMC	100	100	94.9	100	100	100	98.3	94.9	94.9	100	98.3	99.2	98.4
Exp3	LPQ	98.3	100	74.6	100	100	96.6	78.0	74.6	65.3	97.5	89.0	86.4	88.4
	LPQMC	99.2	99.2	77.1	100	100	98.3	88.1	78.8	77.1	97.5	89.8	86.4	91.0
	PLPQMC NP	99.2	100	82.2	100	100	96.6	90.7	81.4	73.7	96.6	91.5	87.3	91.6
	PLPQMC	100	100	82.2	100	100	100	90.7	81.4	78.0	97.5	95.8	91.5	93.1
Exp4	LPQ	96.6	99.2	77.6	100	100	100	79.5	77.8	71.8	94.0	89.0	89.7	89.6
	LPQMC	99.2	99.2	78.5	100	100	98.3	90.6	81.2	79.7	94.9	89.0	89.7	91.7
	PLPQMC NP	99.2	99.2	81.9	100	100	99.2	88.9	81.2	79.7	97.4	91.5	87.2	92.1
	PLPQMC	100	100	81.9	100	100	100	90.6	81.2	81.2	97.4	92.3	89.7	92.9

The full results of four *WPCA based* frameworks, which use LPQ, LPQMC and PLPQMC methods for feature extraction, upon *AR* face database are compared in table 5.1. The LPQ WPCA’s RRs are presented as they are the highest results before the appearance of PLPQMC WPCA in this part of the thesis. Also, the comparisons between these results are visualized in Figs 5.5.

Based on results in table 5.1 and Figs. 5.5, we have several remarks:

1. The multi-scale feature extraction methods based on Monogenic filter and LPQ (LPQMC and PLPQMC) significantly outperform intensity based methods whose best delegation is the LPQ method. This is demonstrated when the overall RRs of the formers are much higher than those of the latters in all experiments. More concretely, the advances of multi-scale approaches over LPQ method are more apparent in more challenging conditions, such as in probe set number 3 (Scream expression), and the ones from number 7 to 12 (sun glasses and scarf occlusions, occlusions with illumination effects), when the RRs of LPQ are far below that of LPQMC and PLPQMC.
2. PLPQ is more efficient than LPQ when they are exploited to extract facial features from DBCs of Monogenic filter. The proofs for this conclusion are more convincing in the results upon difficult probe sets mentioned in the previous comments, where

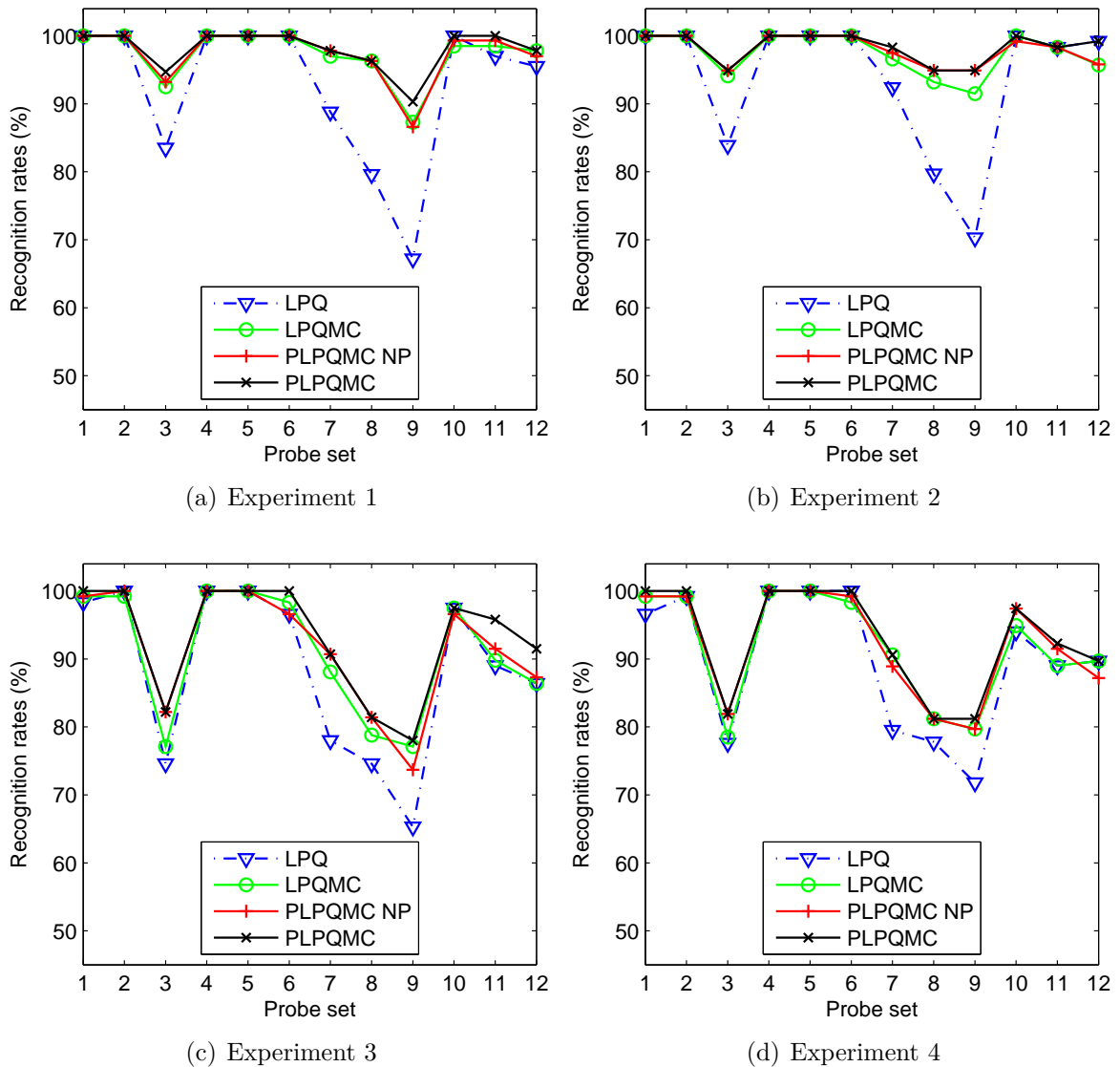


Figure 5.5: Accuracy performance of LPQ, LPQMC and PLPQMC WPCA systems on AR database.

PLPQMC achieves its most noticeable performance improvements in comparison with LPQ and LPQMC. These verify the conjecture we made in the paragraph about the PLPQMC methodology that a patch based pixel of a DBC conveys more image features than its original intensity value.

3. With negligible differences of overall recognition accuracies in four experiments between using or not the retinal filter for illumination normalization, as can be observed in the lines PLPQMC NP and PLPQMC, and the comparability in results of LPQMC, which uses retinal filter, and PLPQMC NP (in fact, PLPQMC NP's average RRs are even slightly higher than LPQMC), as well as the superiority of

5.4. Experimental results

PLPQMC NP over LPQ, where retinal filter is also employed, PLPQMC is shown to be robust towards illumination variations. Also, the perfect accuracy upon image sets 4, 5, and 6 in Exp 1 and Exp 2 without the usage of any illumination normalization technique is a strong evidence for this conclusion.

Table 5.2: Rank-1 RRs (%) of PLPQMC WPCA in comparison with other contemporary systems on AR database using the same evaluation method

	1	2	3	7	10	
Method	Smile	Anger	Scream	Glasses	Scarf	Classes ¹
Exp 1						
String face [21]	87.5	87.5	25.9	88.0	96.0	100
Sparse coding [129]	N/A ²			94.7	91.0	100
DMMA[78]	99.0	93.0	69.0	N/A ²		100
SIS [75]	99.0			99.0	98.0	100
PLD [54]	99.0			100	97.0	100
Sparse LF [76]	N/A ²			100	99.2	126
3D [10]	100	100	97.0	N/A ²		100
PLPQMC NP	100		93.2	97.8	99.3	134
PLPQMC	100		94.7	97.8	100	134
Exp 2						
LGBP[135]	N/A ²			62.0	96.0	50
Sparse coding [129]				80.3	72.7	100
String face [21]				76.0	88.0	100
IRF [141]				82.5	84.0	120
SIS [75]	86.0			96.0	90.0	100
Sparse LF [76]	N/A ²			96.6	96.6	119
PLPQMC NP	100		94.9	97.5	99.2	118
PLPQMC	100		94.9	98.3	100	118
Exp 3						
S-LNMF [90]	62.0	N/A ²	27.0	49.0	55.0	100
Method in [84]	N/A ²		52.3	54.2	81.3	80
PLD [54]	86.0			90.0	89.0	100
3D [10]	99.0	99.0	82.0	N/A ²		100
Bag of words [67]	97.5	97.5	77.3	77.3	89.9	119
PLPQMC NP	100		82.2	90.7	96.6	118
PLPQMC	100		82.2	90.7	97.5	118

¹ : The number of persons whose images are used in experiments.

² N/A: Not available result.

In table 5.2, the most compatible results upon probe sets number 1 (Smile), 2 (Anger),

3 (Scream), 7 (Sun glasses) and 10 (Scarf) in experiments 1, 2, and 3 of PLPQMC (with and without using retinal filter to reduce the bad effects of illumination conditions) are compared with other existing systems. From that table, one can see that even when the retinal filter is not used, PLPQMC outperforms the state-of-the-art systems as its accuracy performance is better than theirs. Some systems, such as the ones in [54, 76, 10], may have higher or comparable RRs, but since the numbers of probe images (in the column *Classes*) are much smaller (100 and 126 versus 134 in Exp 1, 100 versus 118 in Exp 3), so at the end, their recognition performance is less precise than our proposed method as the correct number of probe images recognized by them are lower than ours³, especially when retinal filter is applied in the preprocessing stage for illumination normalization to achieve the best accuracy. In addition to that, when there are more subjects in the gallery/probe set, there are more gallery images to confuse one probe with, thus the recognition task becomes harder. For this, we conclude that PLPQMC is an efficient facial representation against facial expression, and time-lapse variations, and occlusions.

5.4.2 Results on FERET database

Similar as with ELBP and LPQ based systems in chapter 4, the performance of PLPQMC WPCA framework is assessed upon frontal and pose-variant image sets of FERET database via results comparisons with state-of-the-art approaches. These comparisons are reported in tables 5.3 and 5.4 for frontal and non-frontal images, respectively.

Frontal FERET image sets

From the comparisons in table 5.3, one can observe that PLPQMC WPCA outperforms all other existing systems in the FR literature in overall RR and upon Fc, Dup 1, and Dup 2 probe sets. The most considerable results are on Dup 1 and Dup 2 subsets, whose probe images are more challenging than that of Fb and Fc, where our system has reached up to 96.8% and 95.7% accuracy while the highest published ones are 96.3% and 94.4%, correspondingly. This means our method is robust to time-lapse variations, which is one of the most challenging factors of FR study. Besides, the RRs of Fb and Fc tests prove the efficiency of the proposed system against expression and illumination variations.

Importantly, there are three other main conclusions inferred from table 5.3:

³For the clarity of the comment, let's consider, for instance, the system in [10] with *Scream* set, its RRs on 100 probe images of in Exp 1 and Exp 3 are 97.0% and 82.0%, respectively, when ours (no preprocessing method) are 93.2% (for 134 images) and 82.2% (for 118 images). Hence, our method recognizes correctly 125 and 97 probe images in Exp 1 and Exp 2, while its corresponding numbers are 97 and 82, which are obviously far below.

5.4. Experimental results

Table 5.3: Rank-1 RRs (%) comparison of PLPQMC based systems with other state-of-the-art results on the FERET database [96]

Method	Fb	Fc	Dup 1	Dup 2	Average
HMBP [130]	98.1	98.5	75.8	75.2	89.0
CHG [22]	97.5	98.5	85.6	84.6	92.6
Tan et al. [113]	98.0	98.0	90.0	85.0	94.2
LMG [94]	99.8	100	89.2	86.8	95.3
ESRC [30]	97.3	99.0	93.8	92.3	95.9
MS-LPQ [17]	99.2	100	92.0	88.0	95.9
EPFDA [104]	99.6	99.0	92.0	88.9	96.1
POEM PDO [118]	99.7	100	91.7	90.6	96.4
LPQ WPCA	99.5	100	92.9	91.0	96.7
FLPGMP [107]	99.0	99.0	94.0	93.0	96.9
G-LQP [52]	99.9	100	93.2	91.0	97.0
MBC-F [128]	99.7	99.5	93.6	91.5	97.0
GSF [123]	99.6	99.5	94.0	91.5	97.1
PLPQMC WPCA NP	99.6	100	95.4	94.0	97.8
GOM [16]	99.9	100	95.7	93.1	97.9
LPQMC WPCA	99.6	100	96.0	94.4	98.0
SLF-RKR [127]	99.7	99.5	96.3	94.4	98.1
PLPQMC WPCA	99.7	100	96.8	95.7	98.4

1. The superiority of PLPQ over LPQ when performing feature extraction on Monogenic DBCs is validated as results of PLPQMC are better than LPQMC, particularly when facing with time-lapse variation of images in Dup 1 and Dup 2 sets. This is consistent with the remark coming from results upon AR database in the previous Section.
2. As expected, PLPQMC is verified to be robust to illumination variations. There are two evidences for this conclusion. The first one is the perfect identification rate on illumination variant images (the Fc probe set) of PLQMC even when it does not require any preprocessing technique for the illumination normalization purpose. The second evidence is the high results it gains without using an illumination normalization algorithm. While the images in Fc set is strongly affected by illumination changes, this does not mean other images (from Fb, Dup 1 and Dup 2 sets) are not. In fact, illumination variations, being a major challenge, influence generally and thus sharply decrease the overall accuracy performance of any FR system. But, by performing feature extraction on Monogenic DBCs, PLPQMC method, without employing any technique to alleviate the effect of lighting changes, achieves comparable results with the best reported ones (GOM [16] and SLF-

RKR [127]).

3. The concept of applying PLPQ upon Monogenic DBCs shows its efficiency to prove for the correctness of our proposition. This is illustrated via excellent results of PLPQMC WPCA in general, and through the substantial outperformance of the proposed method over other Monogenic and LPQ based systems (such as HMBP [130], MBC-F [128], MS-LPQ [17]) in particular.

Non-frontal FERET image sets

Table 5.4: Rank-1 RRs comparison of PLPQMC WPCA with other leading systems on FERET b-series.

Method	-40° bh	-25° bg	-15° bf	$+15^\circ$ be	$+25^\circ$ bd	$+40^\circ$ bc	Avg
SLF-RKR [127]	N/A ¹	55.0	100	96.0	57.0	N/A ¹	N/A ¹
LSED [121] * ²	78.0	84.0	88.0	89.0	88.0	83.0	85.0
CCA [64] * ²	81.0	91.0	92.0	94.0	89.0	80.0	87.8
PAN [39]	81.5	93.0	97.0	98.5	91.5	78.5	90.0
RFC [20]	84.2	90.2	94.0	93.2	92.5	89.5	90.6
ADMCLS [105]	85.0	94.0	96.0	95.0	94.0	82.0	91.0
LMG [94]	N/A ¹	91.5	98.0	98.5	93.5	N/A ¹	N/A ¹
MRH [7] * ²	87.0	94.0	98.0	99.0	96.0	74.0	91.3
GLOH [100] ** ³	81.1	94.5	100	100	94.5	81.1	91.9
DWFF [86]	87.5	98.0	100	99.0	98.5	82.4	94.2
MRF [48]	91.0	97.3	98.0	98.5	96.5	91.5	95.5
3D Pose Norm [6]	90.5	98.0	98.5	97.5	97.0	91.9	95.6
PLPQMC WPCA NP	92.0	99.5	100	99.5	99.5	91.5	97.0
LPQMC WPCA	93.0	99.5	100	99.5	99.0	92.0	97.2
CPN [31]	94.5	98.0	98.5	99.0	98.5	97.0	97.6
PLPQMC WPCA	95.0	100	100	100	99.5	95.0	98.3
PAF [131]	98.0	98.5	99.25		98.5	98.0	98.6

¹ N/A: Not available result.

² *: The RRs of the method are estimated from plotted figures.

³ **: The RRs on $\pm 25^\circ$ and $\pm 40^\circ$ subsets are average results.

Based on comparable results in table 5.4, we draw the following conclusions: 1) PLPQMC method is efficient when coping with pose variant probe images since its *WPCA* based framework outperforms almost other contemporary systems, except the one in [131] (PAF). This is more interesting when considering that PLPQMC WPCA is a general FR framework while all other systems are dedicated for pose variation

5.4. Experimental results

challenge by equipping special tactics to tackle this difficulty⁴. Further, with relatively small pose angles (within $\pm 25^\circ$), our system is the only one that has nearly perfect RR: it only misses one image in the **bd** set. It is also preferable to expect that when the head poses are not large (bounded by $\pm 40^\circ$), an elite feature extraction could effectively solve the problem. 2) Once again, these results verify that PLPQ is more efficient than LPQ (upon Monogenic DBCs). This is clearer when the head pose increases from $\pm 15^\circ$ to $\pm 40^\circ$. 3) PLPQMC is robust towards illumination variations. Without using any illumination normalization procedure, PLPQMC WPCA framework still obtains very high RRs in comparison with other competitors. In such a challenging situation, amongst all other listed systems in the literature, there are only two ones that surpass ours: CPN [31] and PAF [131].

5.4.3 Results on SCface database

As extensive experiments upon **AR** and **FERET** databases in prior Sections are adequate to justify the efficiency and the effectiveness of the PLPQMC method for facial feature extraction, in this Section, the performance of the PLPQMC WPCA framework against low resolution images is estimated via DayTime and NightTime protocols upon **SCface** database. The results are given with two different training sets: the Fa set of the **FERET** database (denoted as *Our-F*) and the high quality frontal images from **SCface** database (denoted as *Our-S*). They are compared with reported results of PCA [42], DSR [140], ELBP(h+v) WPCA [89] and LPQ WPCA in chapter 4.

It can be seen from tables 5.5 and 5.6 that our system outperforms all state-of-the-art systems when dealing with low resolution probe images. The reasons for this excellent results are probably that PLPQMC is robust to blur images, one property it has by using PLPQ, a variant of LPQ [92], and the illumination invariant strength rooted in the feature extraction process upon Monogenic DBCs. These two meaningful characteristics play a critical role when dealing with such low resolution images of **SCface** database as they were acquired under *unconstrained indoor illumination conditions* and are *real blurred* images. Our average RRs, with training images from **FERET** database, are about 10 and 5 times higher than those of the baseline PCA method in the DayTime and NightTime tests, respectively (50.5% versus 4.7% and 18.2% versus 3.2%). The proposed system, when using the training images from **SCface** database, surpasses the best known method (ELBP [89]) in the FR literature by 12.6% and 5.8% on these two experiments. It also outperforms LPQ WPCA system with large margins in overall RRs. This confirms that the multi-resolution approach based on LPQ in the PLPQMC method is more efficient than using LPQ on intensity appearance only. Besides, when training with mug-shots of the same individuals whose probe images are captured by

⁴It is widely known that along with illumination variations, face recognition across pose is one of the hardest factors of **FR**.

Table 5.5: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]

Camera/Distance	PCA	DSR	ELBP(h+v)	LPQ	Our-F	Our-S	
cam1_1	2.3	N/A ¹	43.1	54.6	64.6	70.0	
cam1_2	7.7		56.2	58.5	63.9	68.5	
cam1_3	5.4		45.4	41.5	42.3	43.9	
cam2_1	3.1		36.9	43.9	46.9	54.6	
cam2_2	7.7		50.8	54.6	57.7	65.4	
cam2_3	3.9		42.3	40.8	37.7	41.5	
cam3_1	1.5		34.6	39.2	42.3	49.2	
cam3_2	3.9		46.9	54.6	59.2	64.6	
cam3_3	7.7		51.5	47.7	43.1	49.2	
cam4_1	0.7		32.3	37.7	46.9	47.7	
cam4_2	3.9		50.0	56.9	68.5	72.3	
cam4_3	8.5		50.8	48.5	46.2	51.5	
cam5_1	1.5		36.2	46.2	54.6	55.4	
cam5_2	7.7		32.3	40.8	51.5	56.2	
cam5_3	5.4		31.5	33.1	31.5	40.0	
Average	4.7		20.2	42.7	46.6	50.5	55.3

¹ N/A: Not available result

Table 5.6: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]

Camera/Distance	PCA[42]	ELBP(h+v)	LPQ	Our-F	Our-S
cam6_1	1.5	9.2	11.5	10.0	14.6
cam6_2	3.1	15.4	16.2	22.3	24.6
cam6_3	3.9	25.4	25.4	25.4	25.4
cam7_1	0.7	13.1	13.9	15.4	15.4
cam7_2	5.4	13.1	15.4	15.4	22.3
cam7_3	4.6	13.9	14.6	20.8	22.3
Average	3.2	15.0	16.2	18.2	20.8

surveillance cameras, the achieved RRs are higher than using a set of images from totally different people. In summary, we conclude that the PLPQMC WPCA system is efficient to cope with low resolution images.

In spite of very encouraging results brought by PLPQMC WPCA framework, one can notice that the recognition performance upon SCface database is still poor and can not be comparable with that of AR and FERET datasets. This means that there is immense room for the research on the topic of FR by machine with low resolution images under

5.4. Experimental results

video surveillance context.

5.4.4 Computational performance

Table 5.7: Computation time of PLPQMC in comparison with other feature extraction methods

Method	Image size	Time (seconds)	Extraction time (milliseconds)	Images/second
ELBP(h)	128×128	4.44	3.71	269
LPQ	128×128	5.45	4.56	219
ELBP(h+v)	128×128	7.28	6.09	164
PLPQMC	128×128	27.91	23.34	43
MBC-A [128]* ¹	150×130	30.54	25.54	39
MBC-O [128]* ¹	150×130	87.00	72.74	14
Gabor wavelets	88×80	96.23	80.46	12

¹ *: We used the Matlab code provided by the author.

Keeping in mind that one of the key points in designing an elite feature extraction method is processing speed, the same timing benchmark described in Section 4.1.4.5 of chapter 4 is conducted with PLPQMC to measure its computational performance in practice. The obtained results are compared with some other methods and the initial step of a Gabor wavelets based approach in table 5.7.

It is transparent from table 5.7 that the proposed method has fast computational speed when it can process 43 images within a second. In PLPQMC, the feature extraction procedure performs on 6 DBCs, hence it is much slower than ELBP, LPQ and ELBP(h+v) methods, in which only one intensity image is processed. Even though MBC-A and MBC-O just work with 3 amplitude images and 3 orientation images, respectively, but they are slower than our method since they need to generate $3 \text{ scales} \times 5 \text{ components} = 15 \text{ images}$ while that of PLPQMC is 6. Additionally, MBC-A and MBC-O require images of larger resolution than PLPQMC (150×130 versus 128×128). Our method is about 3.5x faster than *the initial step* of Gabor wavelets based methods, which produces 40 Gabor wavelets components at 5 scales and 8 orientations. This means PLPQMC is much faster than any Gabor wavelets based feature extraction methods. Requiring only 23.34 milliseconds for processing one image by an unoptimized and un-parallelized implementation, we believe that PLPQMC is capable of being used in real world applications.

5.5 Conclusions

As one step further, in this chapter, a novel multi-resolution feature extraction method, named as Patch based Local Phase Quantization of Monogenic components (PLPQMC), is presented. The proposed method is based on Monogenic filter and Patch based LPQ, a new variant of LPQ. By exploring the insights of different Monogenic components, we have found that its directional bandpass components (DBC) are good candidates to perform the facial feature extraction task. On the other hand, for accelerating the discriminatory power of local patterns captured from those DBCs, we have proposed to apply PLPQ, where local phase patterns are computed from a patch based version of each DBC, instead of LPQ on them. Since there are two kinds of DBC, the horizontal one and the vertical one, their patch based images are computed with two oriented patches of sizes 3×5 and 5×3 , respectively. Once the PLPQ images of 6 DBCs are generated, they are divided into non-overlapped sub-regions whose histograms are computed and concatenated to form the final PLPQMC facial representation. Our method, while benefiting many useful properties from LPQ (by using PLPQ, such as the robustness to blur images), is inherited reasonable computation cost from Monogenic filter. More importantly, as its features are extracted from DBCs, which are strong to illumination variations, PLPQMC has been shown to be robust to that challenge. Also, since many image details are preserved in DBCs, the features extracted from them contribute much to make the proposed method an efficient facial representation for coping with challenging issues. Employing PLPQMC method for feature extraction, a new FR system has been also constituted with WPCA for dimension reduction, k-NN classifier and weighted angle based distance for classification.

Comparative experimental studies on three databases AR, FERET and SCface prove the efficiency of our method against various challenging factors, such as illumination, facial expression, time-lapse, and pose variations, and occlusion, and low resolution images also, since it yields excellent results and outperforms other state-of-the-art systems. Additionally, PLPQMC is fast to compute and can certainly be used in real life applications.

Compared with intensity-based methods shown in chapter 4, ELBP and LPQ, PLPQMC is a multi-scale description whose features are extracted by PLPQ from DBCs of Monogenic filter. By itself, PLPQMC representation is highly insensitive to both illumination variations and blur, two major difficulties a FR system must confront, especially under video surveillance context. As a result, it significantly outperforms ELBP and LPQ in terms of accuracy performance while sustaining a notably fast processing speed. However, its results upon *unconstrained* low resolution probe images in video surveillance context of the SCface database are still far from adequate for practical applications. For this, more robust approaches are needed and that is the goal of the next chapter's content.

Chapter 6

Gradient images based facial features

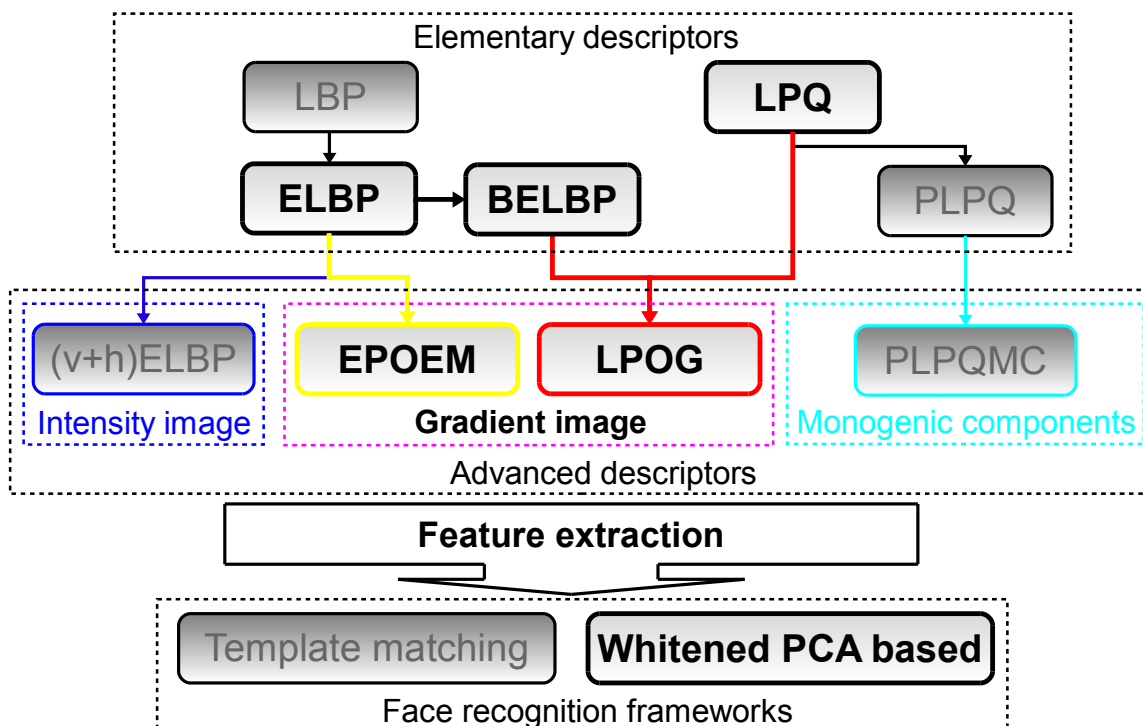


Figure 6.1: Proposed methods in this chapter: EPOEM and LPOG.

6.1 Introduction

The intensity appearance of a human face captured and recognized by machines contains raw image information together with a lot of unexpected variations produced by the effects of unpredictable environment conditions, e.g., light, pose, occlusions and

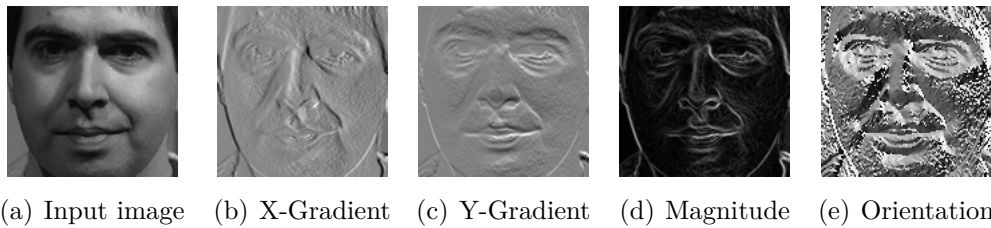


Figure 6.2: An image and its gradient based components. The orientation component is visualized from its radiance values.

background, to name a few. Since these variations are the source of the persistent challenges that hinder high and reliable recognition performance, it is hence desirable to have methods for eliminating them and retaining as many as possible discriminative features of the face images. To the best of our knowledge, illumination variations is one of the major factors that affect the accuracy performance of a FR system the most [137, 65]. This is justified by the fact that the within-class variations caused by lighting changes are almost always greater than the inter-class variations due to the difference of identity [2]. To counteract that, a quite large number of methods/strategies have been developed. Amongst all, using gradient images is one of the most effective solutions.

Gradient images, as showed in pioneer works of visual perception [61, 56, 49], play an important role for human visual perception system. This is on account of the fact that there is a wealth of helpful image features carried inside them, such as edge information, discontinues, orientations, local changes of intensity values. Additionally, one crucial advantage of gradient images for FR, as we will point out, is that they are strong to illumination variations. From an intensity image, its horizontal and vertical gradients are usually computed for building the magnitude and orientation components. At each edge pixel of the input image, the magnitude of gradient conveys information about the strength of the edge while its orientation counterpart contains information about the direction of the edge. In Fig. 6.2, one can see a face image and its x-axis gradient (b), y-axis gradient (c), magnitude (d) and orientation (e) images. In comparison with magnitude and orientation components, the gradient images along the x-axis and y-axis direction preserve more image details.

Within the computer vision domain, Scale Invariant Feature Transform (SIFT) [77] and Histograms of Oriented Gradients (HOG) [26] are the most well-known local descriptors based on gradient images. However, as figured out in chapter 2, they did not help much in forming an efficient facial representation, at least until now regarding the context of the present thesis. This is comprehensible since they are not designed specifically for FR task. Inspired by the insensitivity to the illumination variations of the orientation component [19], the Gradientfaces method, in which the L_1 distance is used to calculate the similarity between two orientation images, was proposed [111].

6.2. Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

Further, in [115], Tzimiropoulos et al. proposed to use PCA with orientation images to construct the Image Gradient Orientations PCA (IGO-PCA) system and gained higher RRs than Gradientfaces. Also, local gradient orientation binary pattern (LGOBP) [69], where gradient angles of salient regions are first assigned as one of the four labels according to the quadrant they belong to and then these assigned values are encoded by a similar way as LBP does with image intensities, is another feature extraction method rooted in gradient orientations. Patterns of Oriented Edge Magnitudes (POEM) [117] is an efficient feature extraction method that exploits the discriminative power from accumulated oriented magnitude images, which are computed from both magnitude and orientation components, by applying LBP descriptor on them.

With the arrival of the new ELBP method in chapter 4, it is preferable to believe that by replacing LBP, where it is used for feature extraction, with ELBP, the recognition performance will be improved. Hence, a new variant of POEM called Elliptical POEM (EPOEM) is proposed by taking into account useful characteristics from POEM and the goodness of ELBP (illustrated in Fig. 6.1). Furthermore, motivated from the advantages of gradient images over intensity appearance as well as magnitude and orientation components, and from the concept of combining two kinds of local descriptors: Block-wise ELBP (BELBP), a new variant of ELBP, and LPQ, we propose a novel feature extraction method named as Local Patterns of Gradients (LPOG) (as can be seen in Fig. 6.1). Comparative studies of obtained results and that of existing systems upon the AR, FERET and SCface databases prove the efficiency and effectiveness of the proposed methods under a diverse range of FR challenges.

The content of this chapter is structured as follows. The details of the EPOEM method, its results and the corresponding comparative studies are presented in Section 6.2. In Section 6.3, we describe the LPOG facial representation alongside with the experimental results upon public databases and show its superiority in recognition accuracy performance over other reported results in the literature. Finally, the conclusions constitute Section 6.4.

6.2 Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

6.2.1 Elliptical Patterns of Oriented Edge Magnitudes feature extraction method

The underpinning idea of EPOEM is shown in Fig. 6.3, where the procedure of computing the EPOEM code for an input pixel of the orientation image is illustrated. First,

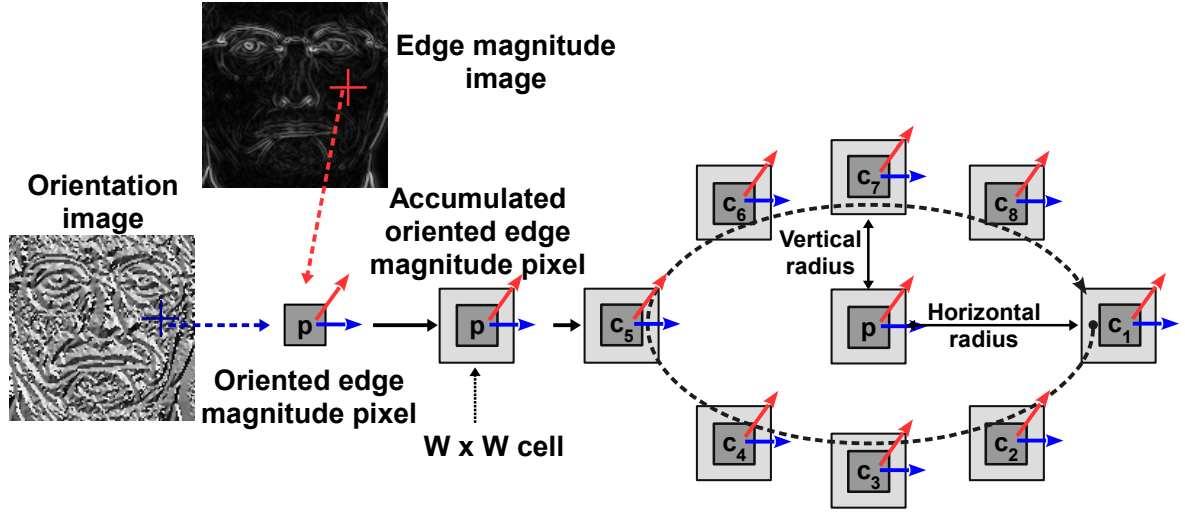


Figure 6.3: Steps in EPOEM encoding scheme for one pixel.

each pixel from orientation image is evenly discretized over the $[0, \pi]$ range. Hence, at each pixel, the magnitude and its discretized orientations are hold. Then, every discretized orientation is weighted by its corresponding magnitude to generate oriented edge magnitudes values. In Fig. 6.3, the current pixel is referred as \mathbf{p} , the local magnitude is represented by the red arrow emitted from \mathbf{p} whilst the blue arrow is the discretized orientation. Next, each oriented pixel is replaced by the mean value of its neighbors within a *cell*, which is a square neighborhood of the pixel. By this, the accumulated pixels are obtained and the EPOEM code of the given pixel is computed by a horizontal ELBP descriptor as:

$$EPOEM_m^{N,R1,R2}(p) = \sum_{i=1}^N f(c_i^{N,R1,R2} - I_p)2^{i-1} \quad (6.2.1)$$

where N is the number of neighbors of p , which is fixed as 8 in this work¹, m is the current orientation, $c_i, i \in [1, N]$ is the intensity value of the i^{th} neighboring pixel of p , I_p is the intensity at p location in the accumulated image, $R1, R2$ are the radii of the ellipse sample used with ELBP, and $f(x)$ is a binary encoding function and is defined as:

$$f(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases} \quad (6.2.2)$$

The whole process of building the EPOEM feature vector from a given face image is depicted in Fig. 6.4. First, the edge magnitude and orientation images are generated from the input image. Next, each orientation angle is evenly discretized into M partitions

¹In fact, POEM [117] results were reported with $N = 6$ but we have found that higher RRs are achieved when $N = 8$. Hence all related results presented in this thesis are yielded with that value.

6.2. Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

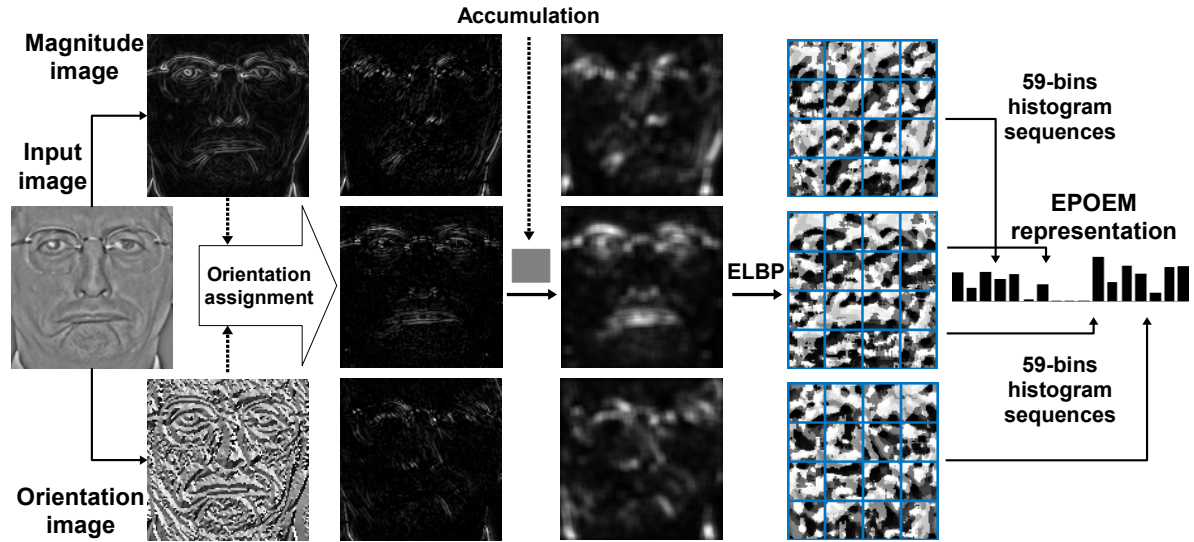


Figure 6.4: Facial representation computation by EPOEM method.

(the number of partitions used in this work is 3 or 4, depending on the tested database). Then, M oriented edge magnitudes images are computed from edge magnitude image and discretized orientation ones by a element-wise multiplication of them. These two steps are called as *Orientation assignment* process in Fig. 6.4. By doing this, the information of both magnitude and orientation images are incorporated and thus makes the feature extraction stage more efficient. Each pixel in an oriented edge magnitude image is then substituted by its neighbors' average value to obtain an accumulated image, this is also a boosting step for the discriminatory power of EPOEM features since each accumulated value not only carries its original intensity from oriented image, but also holds that of its neighbors. Once this accumulation step is finished, the ELBP descriptor is applied on all accumulated images to form their own histogram sequences. Finally, the EPOEM vector is constructed by concatenating all the histogram sequences of the accumulated images. To be a compact facial representation, uniform patterns are used to reduce the size of every EPOEM description. By using $N = 8$ neighbors when computing a EPOEM code for a given pixel, the EPOEM vector is a 59-bins histogram sequence.

Due to the advantages of magnitude and orientation components, the efficiency of ELBP, as well as the manner it is built, the EPOEM descriptor has the following interesting properties:

1. It has oriented features. By varying the number of partitions discretized by the orientation of each pixel, EPOEM is capable to capture features in different orientations.
2. As prior mentioned in the first part of this chapter, orientation image is insensitive

to lighting changes. Plus, one of the most significant of ELBP (as proved in Section 4.1 of chapter 4) is invariance to illumination variations. As a consequence, EPOEM is even more robust to illumination changes.

3. EPOEM features are encoded from accumulated oriented edge magnitudes images, they encompass many meaningful characteristics from both orientation and magnitude images, thus make them more efficient against FR challenges that are ubiquitous when dealing with face images.
4. By thresholding accumulated pixels when they are computed, EPOEM codes extend the locality property of ELBP features, which are calculated based on intensity pixels, to enhance their discriminatory power.

Regarding to the EPOEM's parameters, we use 7×7 and 5×5 cells for images of 128×128 and 48×48 resolutions when performing the accumulation step, respectively. Also, with 128×128 pixels images, the radii of horizontal ellipse used with ELBP are 7 (horizontal radius) and 5 (vertical radius) while those for 48×48 pixels images are 5 and 3. Each accumulated image of size 128×128 is divided into 8×8 non-overlapping sub-regions to compute their histograms. Since 48×48 pixel images (from SCface database) are small, they are split into 5×5 sub-images. The 128×128 orientation images are partitioned into $M = 3$ partitions whereas that number for 48×48 images is 4. Almost these values are used with POEM, except that the radii used with LBP are 5 and 3 for 128×128 and 48×48 images, correspondingly.

6.2.2 Using EPOEM for face recognition

In order to apply EPOEM representation for FR, the WPCA based framework, whose details are presented in Section 3.6 of chapter 3, is used. The *negative angle based* distance function (Eq. 3.6.1) is adopted in the classification stage for measuring the similarity scores between probe images and gallery ones as it offers the best accuracy results.

6.2.3 Experimental results

In this Section, the recognition performance of EPOEM WPCA framework is evaluated via experiments on AR, FERET and SCface databases. The obtained results are then compared with that of POEM WPCA and other state-of-the-art systems using the same protocols. With respect to the order of appearance in the FR literature, some leading systems published after the year 2012, when EPOEM was shown, are not listed in the comparisons.

6.2. Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

6.2.3.1 Results on AR database

Table 6.1: Rank-1 RRs (%) comparison between POEM and EPOEM WPCA based methods on AR database

Test/Method		Probe set												Avg
		1	2	3	4	5	6	7	8	9	10	11	12	
Exp1	POEM	100	100	83.6	100	100	100	91.0	70.9	64.9	98.5	96.3	97.0	91.9
	EPOEM	100	100	85.1	100	100	100	91.0	72.4	66.4	100	97.8	97.0	92.5
Exp2	POEM	100	100	83.9	100	100	100	89.8	72.9	63.6	99.2	96.6	96.6	91.9
	EPOEM	100	100	86.4	100	100	100	91.5	74.6	66.9	100	97.5	97.5	92.9
Exp3	POEM	97.5	100	74.6	100	100	97.5	80.5	65.3	63.6	96.6	90.7	90.7	88.1
	EPOEM	99.2	100	78.0	100	100	97.5	80.5	66.9	65.3	97.5	91.5	91.5	89.0
Exp4	POEM	97.4	100	69.8	100	100	99.2	78.6	66.7	60.7	95.7	90.6	90.6	87.4
	EPOEM	98.3	100	71.6	100	100	99.2	79.5	67.5	61.5	96.6	91.5	91.5	88.1

It can be seen in table 6.1, where the comparisons between full results of POEM and EPOEM on AR database are tabulated, that EPOEM gains higher RRs than POEM. This performance improvement is not so significant but it is indisputable and consistent, particularly in RR of every probe set and generally in overall average RR of each experiment. The most notable cases are shown in probe sets 3 (*Scream*), 8 (*Sun glasses+Left light*) and 9 (*Sun glasses+Right light*). Thus, we conclude that EPOEM is more efficient than POEM when handling face images under variations of illumination, facial expression, time-lapse and occlusions.

Another noteworthy observation about those results is the distinction in recognition accuracies between the images disguised by *Sun glasses* (in probe sets 7, 8, and 9) and the ones partially masked by *Scarves* (in probe sets 10, 11, and 12). The relatively high RRs offered by EPOEM on *Scarves* images and that of far lower on *Sun glasses* images mean that EPOEM is very sensitive when dealing with face images whose eyes and eyebrows, the two most crucial features, are largely obscured. This is even worse when there is the presence of other factors, such as illumination and time-lapse variations. The accumulated oriented images might be a reason for this weakness of EPOEM since they retain few image details to form a robust representation that can rectify the lack of such important facial features.

Similarly to other proposed methods in this thesis, the results of EPOEM in experiments 1, 2, and 3 on probe sets 1, 2, 3, 7, and 10 are compared with other existing systems in table 6.2. From these results, it becomes evident that EPOEM WPCA substantially outperforms other state-of-the-art systems on AR database. This performance superiority of EPOEM WPCA is more interesting when considering the fact that most of the listed works are dedicated to tackle with FR under occlusions by various approaches, such as Sparse representation based [129, 76], Bag of words [67] or 3D [10], whilst the proposed

Table 6.2: Rank-1 RRs (%) of EPOEM WPCA in comparison with other contemporary systems on AR database using the same evaluation method

	1	2	3	7	10	
Method	Smile	Anger	Scream	Glasses	Scarf	Classes ¹
Exp 1						
String face [21]	87.5	87.5	25.9	88.0	96.0	100
Sparse coding [129]	N/A ²			94.7	91.0	100
DMMA[78]	99.0	93.0	69.0	N/A ²		100
SIS [75]	99.0			99.0	98.0	100
PLD [54]	99.0			100	97.0	100
Sparse LF [76]	N/A ²			100	99.2	126
3D [10]	100	100	97.0	N/A ²		100
EPOEM	100		85.1	91.0	100	134
Exp 2						
LGBP[135]	N/A ²			62.0	96.0	50
Sparse coding [129]				80.3	72.7	100
String face [21]				76.0	88.0	100
IRF [141]				82.5	84.0	120
SIS [75]	86.0			96.0	90.0	100
Sparse LF [76]	N/A ²			96.6	96.6	119
EPOEM	100		86.4	91.5	100	118
Exp 3						
S-LNMF [90]	62.0	N/A ²	27.0	49.0	55.0	100
Method in [84]	N/A ²		52.3	54.2	81.3	80
PLD [54]	86.0			90.0	89.0	100
3D [10]	99.0	99.0	82.0	N/A ²		100
Bag of words [67]	97.5	97.5	77.3	77.3	89.9	119
EPOEM	99.2	100	78.0	80.5	97.5	118

¹ : The classes column is the number of persons whose images are used in experiments.

² N/A: Not available result.

framework is a general system. In addition to that, these RRs are attained under more diffusing conditions as there are more gallery/probe images used in our experiments (in the last column of table 6.2) than those of cited systems. In combination, a conclusion is drawn that EPOEM is efficient to facial expression and time-lapse variations, and partial occlusions.

6.2. Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

6.2.3.2 Results on FERET database

The comparison results between EPOEM WPCA framework and state-of-the-art systems upon FERET database can be observed in table 6.3 for standard protocol of frontal images, and in table 6.4 for pose variations images.

Frontal FERET image sets

Table 6.3: Rank-1 RRs (%) comparison of EPOEM WPCA with other state-of-the-art results on the FERET database [96]

Method	Fb	Fc	Dup 1	Dup 2	Average
LGBPHS [135]	98.0	97.0	74.0	71.0	87.8
HMBP [130]	98.1	98.5	75.8	75.2	89.0
GEWC [29]	96.3	99.5	78.8	77.8	89.3
HGPP [133]	97.5	99.5	79.5	77.8	90.2
DMMA [78]	98.1	98.5	81.6	83.2	91.6
LGBPWP [88]	98.1	98.9	83.8	81.6	92.1
CHG [22]	97.5	98.5	85.6	84.6	92.6
DLBP [82]	99.0	99.0	86.0	85.5	93.6
IGO-PCA [115]	N/A ¹		88.9	85.4	N/A ¹
Tan et al. [113]	98.0	98.0	90.0	85.0	94.2
ELBP(h+v) WPCA [89]	99.4	100	89.1	86.8	95.0
LMG [94]	99.8	100	89.2	86.8	95.3
POEM WPCA	99.3	100	90.4	90.2	95.3
ESRC [30]	97.3	99.0	93.8	92.3	95.9
MS-LPQ [17]	99.2	100	92.0	88.0	95.9
EPFDA [104]	99.6	99.0	92.0	88.9	96.1
POEM PDO [118]	99.7	100	91.7	90.6	96.4
EPOEM WPCA	99.6	100	92.4	92.3	96.7
FLPGMP [107]	99.0	99.0	94.0	93.0	96.9
G-LQP [52]	99.9	100	93.2	91.0	97.0
GSF [123]	99.6	99.5	94.0	91.5	97.1

¹ N/A: Not available result.

According to the results in table 6.3, some conclusions can be stated as follows:

1. EPOEM outperforms POEM on FERET database. This is more evident on RRs of Dup 1 and Dup 2 tests where EPOEM achieves a performance improvement of about 2.0% when dealing with images affected by time-lapse variations. Plus, EPOEM also surpasses another system based on POEM in [118].

2. Inherited advantages from both ELBP and POEM, EPOEM is strong to illumination variations as it gains perfect accuracy on Fc probe set.
3. The results provided by EPOEM is encouraging, despite the fact that they are lower than those of three leading Gabor wavelets based systems (FLPGMP [107], G-LQP [52] and GSF [123]). On average RR, there is only a slight difference between ours and theirs. More importantly, since ELBP is significantly faster than any Gabor wavelets based method (about 22×, refer to table 4.8 in chapter 4 for more details) and EPOEM uses an ELBP operator on 3 accumulated images, the proposed method therefore requires much less computational cost than the three mentioned systems.

Non-frontal FERET image sets

Table 6.4: Rank-1 RRs comparison of EPOEM WPCA with other systems on FERET b-series.

Method	-40° bh	-25° bg	-15° bf	+15° be	+25° bd	+40° bc	Avg
LSED [121] *2	78.0	84.0	88.0	89.0	88.0	83.0	85.0
CCA [64] *2	81.0	91.0	92.0	94.0	89.0	80.0	87.8
PAN [39]	81.5	93.0	97.0	98.5	91.5	78.5	90.0
RFC [20]	84.2	90.2	94.0	93.2	92.5	89.5	90.6
ADMCLS [105]	85.0	94.0	96.0	95.0	94.0	82.0	91.0
LMG [94]	N/A ¹	91.5	98.0	98.5	93.5	N/A ¹	N/A ¹
MRH [7] *2	87.0	94.0	98.0	99.0	96.0	74.0	91.3
GLOH [100] **3	81.1	94.5	100	100	94.5	81.1	91.9
ELBP(h+v) WPCA	80.5	98.5	99.5	99.5	99.0	79.5	92.8
POEM	84.0	99.0	99.5	99.5	99.0	83.0	94.0
DWFF [86]	87.5	98.0	100	99.0	98.5	82.4	94.2
EPOEM	87.5	99.5	100	99.5	99.0	86.0	95.3
MRF [48]	91.0	97.3	98.0	98.5	96.5	91.5	95.5
3D Pose Norm [6]	90.5	98.0	98.5	97.5	97.0	91.9	95.6
CPN [31]	94.5	98.0	98.5	99.0	98.5	97.0	97.6

¹ N/A: Not available result.

² *: The RRs of the method are estimated from plotted figures.

³ **: The RRs on $\pm 25^\circ$ and $\pm 40^\circ$ subsets are average results.

It can be evidently seen from table 6.4 that EPOEM, once again, yields higher accuracy performance than POEM upon probe images of varying head poses. And the larger the pose angles are, higher the improvements are. Compared with other methods, there are few ones that can surpasses EPOEM (MRF [48], 3D Pose Norm [6] and CPN [31]).

6.2. Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

However, when the head poses are in small scope (of $\pm 25^\circ$), EPOEM is more stable and attains higher RRs with at least 99.0% of accuracy, which is the best result of those 3 leading systems. Plus, considering that MRF [48], 3D Pose Norm [6] and CPN [31] are all Gabor wavelets based approaches, they are obviously slower than our method. As a consequence, we argue that EPOEM is a promising solution to the pose variations challenge.

6.2.3.3 Results on SCface database

From comparative studies between results of EPOEM and POEM, as well as other existing system in the literature via a large set of experiments upon AR and FERET database in previous sections, there are sufficient evidences to validate the improvement of EPOEM over POEM and the efficiency of EPOEM when coping with a wide range of challenges. Hence, in this Section, tests on SCface database are conducted to assess the performance of the EPOEM WPCA framework in confronting with low resolution image captured under video surveillance context. The results are obtained with high quality mug-shot images from the same database and then are compared with those of PCA [42], DSR [140] and ELBP(h+v) WPCA [89].

Table 6.5: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]

Camera/Distance	PCA	DSR	ELBP(h+v)	EPOEM
cam1_1	2.3	N/A ¹	43.1	40.0
cam1_2	7.7		56.2	62.3
cam1_3	5.4		45.4	47.7
cam2_1	3.1		36.9	33.9
cam2_2	7.7		50.8	46.2
cam2_3	3.9		42.3	46.9
cam3_1	1.5		34.6	28.5
cam3_2	3.9		46.9	48.5
cam3_3	7.7		51.5	55.4
cam4_1	0.7		32.3	26.2
cam4_2	3.9		50.0	59.2
cam4_3	8.5		50.8	53.1
cam5_1	1.5		36.2	30.8
cam5_2	7.7		32.3	41.5
cam5_3	5.4		31.5	37.7
Average	4.7		20.2	42.7

¹ N/A: Not available result

Table 6.6: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]

Camera/Distance	PCA[42]	ELBP(h+v)	EPOEM
cam6_1	1.5	9.2	8.5
cam6_2	3.1	15.4	16.2
cam6_3	3.9	25.4	24.6
cam7_1	0.7	13.1	10.8
cam7_2	5.4	13.1	14.6
cam7_3	4.6	13.9	17.7
Average	3.2	15.0	15.4

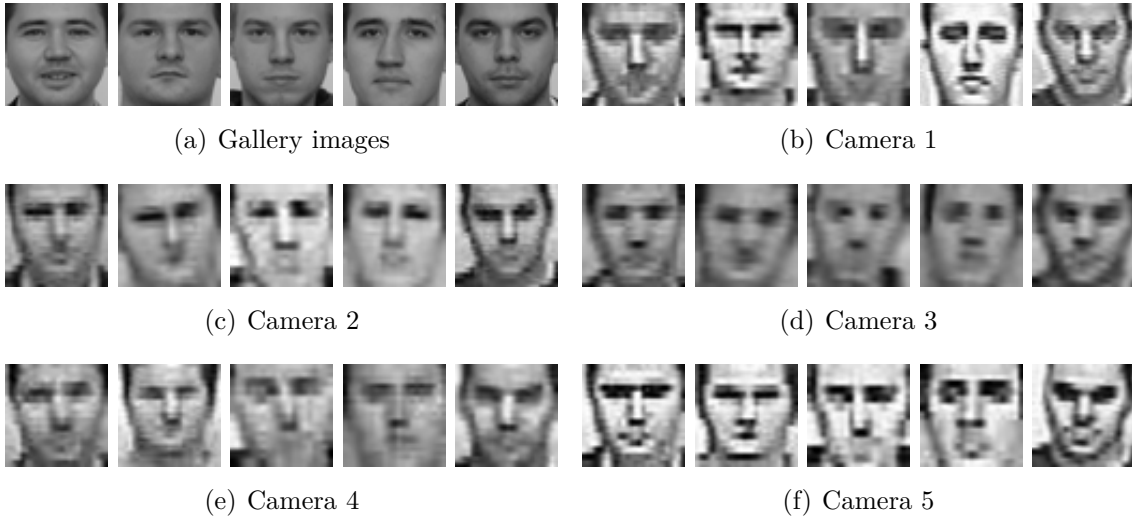


Figure 6.5: Some challenging SCface images captured at distance 1 where EPOEM fails to recognize.

It can be clearly observed in tables 6.5 and 6.6 that EPOEM outperforms other reported results on SCface database followed both DayTime and NightTime protocol. Generally, comprehensive performance of the proposed method, in terms of average RR, is higher than ELBP(h+v) [89] (43.9% versus 42.7% in DayTime test, 15.4% versus 15.0% in NightTime tests), the best results so far. It seems obvious since EPOEM significantly surpasses ELBP(h+v) on FERET database (for more details, refer to tables 6.3 and 6.4). However, surprisingly, this superiority is not appeared in probe sets of distance 1, where the RRs of ELBP(h+v) are all better than EPOEM's results. The anomaly occurs not only with DayTime test, but also with NightTime one.

As an effort to pinpoint and understand the source of that incident poor performance of the proposed method, we gather some common images which were acquired by first 5 cameras at distance 1 and were wrongly identified in Figs. 6.5. The probe images in SCface database were captured at three distances: distance 1 is 4.2m, distance 2 is

6.2. Elliptical Patterns of Oriented Edge Magnitudes for Face recognition

2.6m and distance 3 is 1.0m. Amongst all probe images, the ones of distance 1 are most blurred as they were taken at the farthest distance. The blurring level of those images can be intuitively seen in Figs 6.5(b-f). Also, one can notice the extreme contrast between them and the gallery images (Figs 6.5a) with respect to several important aspects, e.g., image quality, resolution, blurring level, misalignment level. But why under the same conditions, ELBP(h+v) can harvest better performance than EPOEM? The answer for this question might be rooted in the fact that EPOEM patterns are extracted by an horizontal ELBP descriptor from **accumulated oriented edge magnitudes images**. While computing accumulated oriented images from oriented and magnitude ones, many image details are unavoidably lost (refer to Fig. 6.4 for an intuitive illustration). As a result, the accumulated oriented images do not preserve much edge information. This image degradation is specially critical with probe images of distance 1 as they are so blurred. Plus, ELBP is not a blur tolerant method. Thus, the EPOEM representation, in this particular case, is not efficient as ELBP(h+v) one, which extracts features straightly from intensity images.

There are probably several solutions to the above problem:

1. Using a blur insensitive descriptor, such as LPQ or PLPQ, rather than ELBP, to strengthen the robustness of EPOEM against blurred images. Due to the lack of time, we have not tried this approach but we believe it is a worth pursuing direction.
2. In order to avoid the image degradation issue, the resulting features should be extracted from other types of image, which are capable of retaining more image information than accumulated images. In the next Section, a novel method followed this manner is presented by performing the feature extraction task efficiently upon horizontal and vertical gradient images.

6.2.4 Conclusions

As an attempt to develop an efficient facial representation, the horizontal ELBP descriptor is used to extract local micro patterns on oriented edge magnitudes components, which are generated from gradient images by varying orientations. In this way, the EPOEM feature extraction method has been built. Having many intriguing characteristics, EPOEM, when used in *WPCA based* framework, has been extensively assessed via experiments on three public database ([AR](#), [FERET](#) and [SCface](#)). The results show that the proposed method outperforms POEM and gives competitive accuracy performance in comparison with state-of-the-art algorithms.

Nevertheless, the drawback of EPOEM was exposed when it faced some extremely challenging conditions, such as important facial features (eyebrows and eyes) hidden

and under variations of illumination and time-lapse, or very blurred images captured by surveillance cameras.

6.3 Local Patterns of Gradients for Face recognition

In this Section, we propose a novel feature extraction method so-called Local Patterns of Gradients (LPOG) for FR. It provides an unified way to capture local patterns from gradient images.

We firstly introduce a novel variant of ELBP [89] named Block-wised ELBP (BELBP), which is more efficient than ELBP for encoding micro facial features. In BELBP, the average value of the rectangular block at every image pixel is calculated, then a thresholding over that values of each pixel and its neighbors, who lie on an elliptical pattern, is undertaken. For vertical ellipse, a vertical rectangular block is used while a horizontal rectangular block is applied for horizontal ellipse, that is the reason why we name this method Block-wised ELBP.

As the researches in visual perception [61, 56, 49] showed that human visual system is more sensitive to local changes of intensity (gradient images) than to the image intensity itself, in our LPOG method, feature extraction is done upon gradient images instead on intensity images as many other algorithms usually do. Towards this purpose, we use BELBP and LPQ operators on gradient images to encode local patterns in the form of histogram sequences. Finally, a global feature vector (LPOG) is built by concatenating BELBP and LPQ descriptions.

Using LPOG as facial features, a novel single sample per person FR framework called Local Patterns of Gradients Whitened Principal Component Analysis (LPOG WPCA) is proposed with WPCA for training stage and weighted angle-based distance and k-Nearest Neighbor (k-NN) classifier for classification.

Our LPOG WPCA system is evaluated and compared with other competitors through extensive experiments upon large public face databases including AR [80], FERET [96] and SCface [42]. Comparisons between our experimental results and other state-of-the-art systems confirm that our framework consistently outperforms other contemporary systems while illustrating its robustness by setting the best ever performance against major challenges like facial expression changes, occlusion, pose and time-lapse variations as well as low resolution images.

Moreover, practical timing tests show that LPOG is faster than many leading feature extraction algorithms and is feasible for constraints of real life applications.

6.3. Local Patterns of Gradients for Face recognition

6.3.1 Local Patterns of Gradients feature extraction method

6.3.1.1 Block-wised ELBP: a novel variant of ELBP

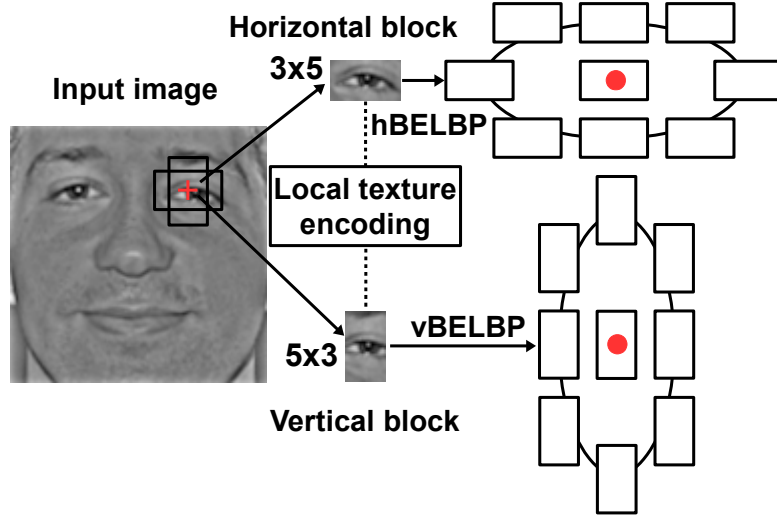


Figure 6.6: BELBP encoding operators

Inspired by ELBP [89], a so-called block-wised ELBP (BELBP) (see Fig. 6.6) operator of an image I is formed by first generating its two accumulated images (AI) corresponding to two block patterns (BP) as:

$$AI_{v,h}^{BP}(x,y) = \frac{1}{N} \sum_{(x,y) \in BP} I(x,y), \quad (6.3.1)$$

where v and h are notations for vertical and horizontal directions, N is the number of pixels in pattern BP and $I(x,y)$ is the intensity value at location (x,y) of I . The block patterns BPs are directional rectangulars: horizontal one for horizontal AI and vertical one for vertical AI. After that, the BELBP code of a given pixel (a decimal value) in each AI image is computed by comparing its value with surrounding pixels which are located on an ellipse, whose center is at the current pixel itself. In details, with K neighboring pixels, horizontal radius $R1$ and vertical radius $R2$, the formula for calculating BELBP label of one pixel $P=AI(x_c,y_c)$ is:

$$BELBP^{K,R1,R2}(x_c,y_c) = \sum_{i=1}^K s(g_i^{K,R1,R2} - g_c)2^{i-1}, \quad (6.3.2)$$

where g_c is the gray scale value of P and g_i is the gray scale value of its i^{th} neighbor whose coordinates are generated by using the formulas:

$$angle = 2 * \pi / K, \quad (6.3.3)$$

$$x_i = x_c + R1 * \cos((i - 1) * \text{angle}), \quad (6.3.4)$$

$$y_i = y_c - R2 * \sin((i - 1) * \text{angle}). \quad (6.3.5)$$

The binary encoding function $s(x)$ is defined as

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}. \quad (6.3.6)$$

If $R1 > R2$ and if considering a horizontal AI, one has horizontal BELBP (hBELBP) operator, vertical BELBP (vBELBP) operator is obtained on a vertical AI when $R2 > R1$ (see Fig. 6.7). A symmetric pair of BELBP (see Fig. 6.7) operators consists of two BELBPs where the first one's horizontal radius is the other's vertical radius and vice versa.

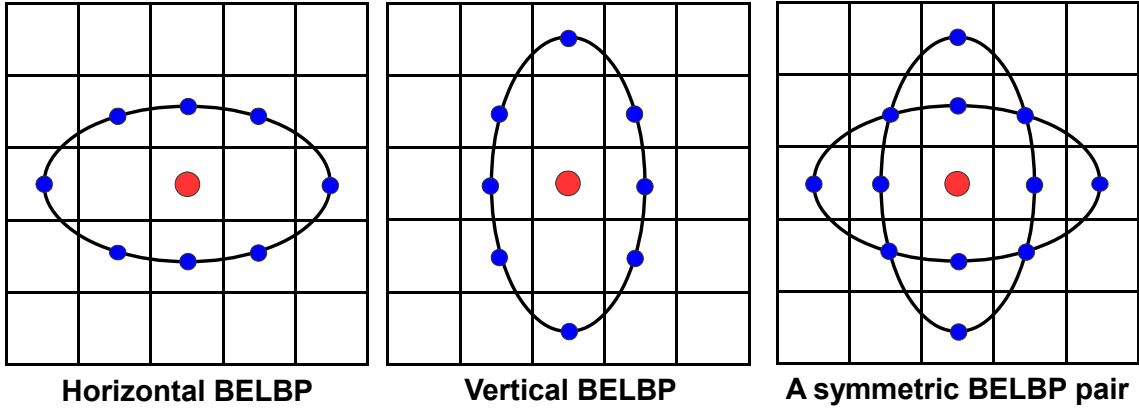


Figure 6.7: BELBP operators

In this work, we use a symmetric pair of BELBP with two block patterns size 3×5 and 5×3 . By doing this, the dependence between each pixel of the input image $I(x,y)$ and its neighbors is taken into account when extracting its local micro texture, thus making the feature extraction process more efficient. Besides, the usage of a symmetric pair of BELBP to encode both horizontal and vertical texture information gives a more discriminative representation of the face image than just using a single horizontal BELBP. Consequently, two BELBP images are generated from every input image (see Fig. 6.8 for more details).

6.3.1.2 LPOG in details

$$G_x = -\frac{1}{2} \cdot I(x - 1, y) + 0 \cdot I(x, y) + \frac{1}{2} \cdot I(x + 1, y) \quad (6.3.7)$$

$$G_y = -\frac{1}{2} \cdot I(x, y - 1) + 0 \cdot I(x, y) + \frac{1}{2} \cdot I(x, y + 1) \quad (6.3.8)$$

6.3. Local Patterns of Gradients for Face recognition

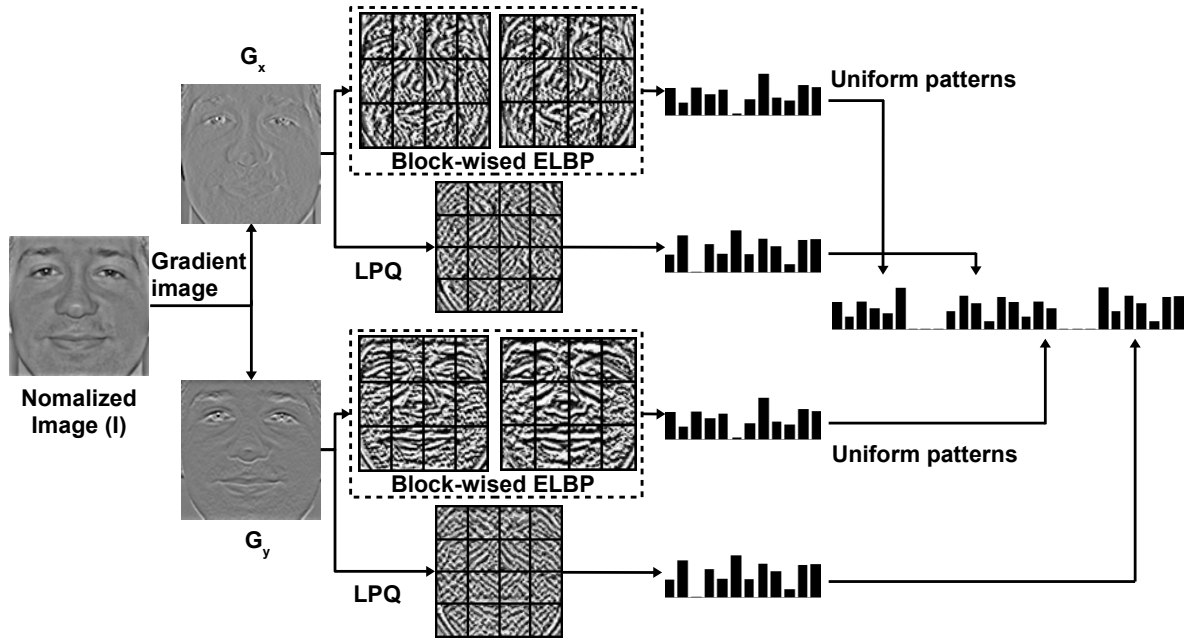


Figure 6.8: Steps in LPOG scheme

The proposed feature extraction method called local patterns of gradients (LPOG) in this work (Fig. 6.8 illustrates steps of LPOG) is a powerful face description for dealing with challenging issues of FR. From the input image I , two directional gradient images along x-axis and y-axis (G_x , G_y) are generated using the formulas (6.3.7)-(6.3.8). Then, a symmetric pair of BELBP and a LPQ operator, whose details are in Section 4.2 of chapter 4, are used (the details of used parameters will be given at the end of Section 6.3.3) to extract local patterns from each gradient image under the form of BELBP and LPQ images (we call them local patterns images). Next, each local pattern image is divided into disjoint rectangular sub-regions to calculate their histogram sequences, which are then concatenated to constitute a description of the image. Uniform patterns [3], which are binary strings with no more than two bitwise transitions from 0 to 1 or 1 to 0, are utilized to reduce the BELBP descriptions' length. As a last step, all the local patterns image's representations are aggregated to form a global LPOG feature vector of the given image.

The LPOG method of this work is primarily stemmed from two inspirations.

The first motivation comes from the observation of early perception researches results ([61, 56, 49]). While these results prove the more sensitivity of visual human system to local intensity differences, than to raw intensity values, they suggest that extracting texture information from gradient images is more efficient than from intensity image. Fig. 6.9 illustrates the advantages of this approach: the illumination condition effect on gradient images is less than on raw intensity ones, so the feature extraction stage on gradient images will be more illumination invariant. In addition to that, the gradient

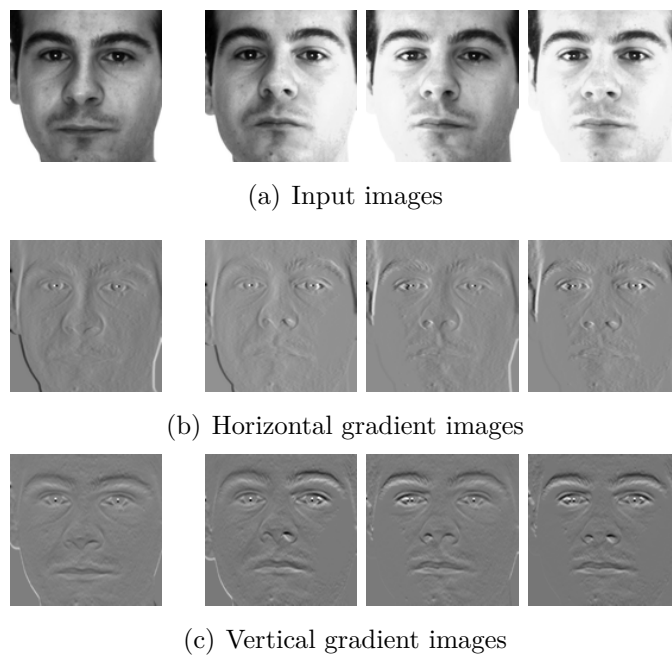


Figure 6.9: Comparison between gradient images of illumination images from AR database. The first image on the left is the gallery one while the rest are probe ones.

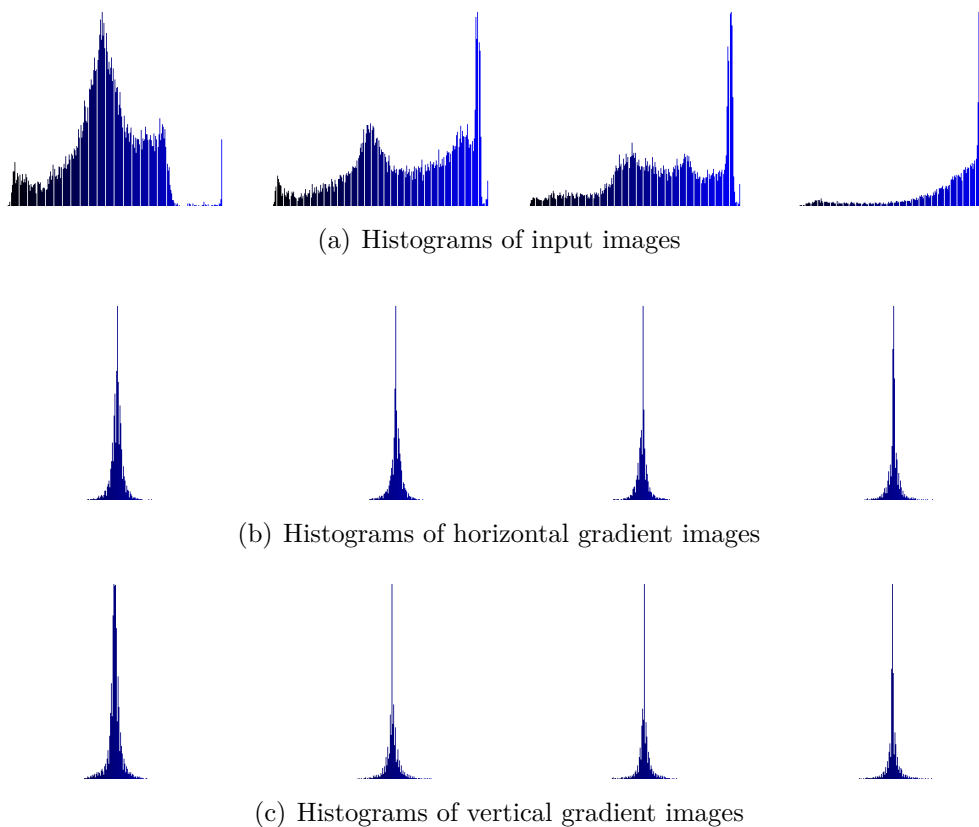


Figure 6.10: Comparison between histograms of images in Figs. 6.9.

6.3. Local Patterns of Gradients for Face recognition

images contain enhanced edge information, which is very important for building a strong facial representation. Indeed, one can observe in Figs. 6.9 (b-c) that the gradient images (of gallery and probe images) are in a much more similar illumination condition than their intensity appearances Figs. 6.9 (a), and the details of important facial features, such as eyebrows, eyes, nose and mouth, are also strengthened. Plus, it can be seen in Figs 6.10 that the similarities between the gallery image (the left one in Figs. 6.9 (a)) and its probe ones (the right ones in Figs. 6.9 (a)), when they are represented as histograms, are more marked within the gradient domain (Figs 6.10 (b-c)) than in the intensity domain (Figs 6.10 (a)). Therefore, these leads us to apply LPOG on gradient images for building LPOG representation. This process makes LPOG different from many other intensity-based methods as well as gradient magnitudes and gradient orientations based ones (e.g. [111, 117, 118, 115]) since it is gradient images based. It also enriches discrimination power of LPOG feature vector with a wealth of meaningful visual features from gradient images, including local contrast, edges, discontinuities properties, thus making LPOG more robust against FR's challenges such as illumination, pose and time-lapse variations.

Secondly, the discriminant power of LPOG is built upon the usage of BELBP and LPQ operators to extract local texture patterns from gradient images. The novel BELBP operator inherits the efficiency of ELBP, since it is a ELBP variant, and encodes macroscopic relation between one gradient pixel and its neighbors. The ideas of ELBP about the nature orientation (horizontal) of important facial features (such as eyes, eyebrows, mouth) and using a symmetric pair of ELBP for integrating both horizontal and vertical information from a face image are augmented in BELBP with the oriented block patterns when computing the accumulated images as they are not only present on the used ellipse patterns but also on the block-wised neighborhoods. All of those aspects result that LPOG contains rich micro patterns of gradient images as BELBP provides it with such useful properties in a more efficient manner than ELBP [89]. Beside the role of BELBP, LPQ operator has a significant impact on the robustness of LPOG method as well. While BLEBP is based on binary thresholding of each pixel from an accumulated gradient image with its neighbor, LPQ is based on quantization of STFT phase responses and it extracts the local phase patterns from gradient images. LPQ was proved to be strong against blurred faces [5], as a result its presence in LPOG equips our approach with *blur invariant* property. On the other hand, LPOG is also strong to uniform illumination, an attribute derived from LPQ [92].

6.3.2 Using LPOG for face recognition

Using LPOG method for feature extraction, a new FR system called LPOG whitened PCA (LPOG WPCA) is formed by employing the *WPCA based* framework, whose are detailed in Section 3.6 of chapter 3. In the classification stage, the *weighted angle-based*

function, which is used to estimate the distances between projected feature vectors of gallery and probe images, is applied due to the highest results it delivers.

6.3.3 Experimental results

For evaluating the performance of the proposed LPOG WPCA system, we proceed experiments (with standard protocols) upon three large public face databases: AR [80], FERET [96] and SCface [42]. Details of these experiments are presented in Section 3.2 of chapter 3.

To validate the illumination invariant characteristic of the LPOG method, the recognition performance of its *WPCA based* framework without the usage of retinal filter for illumination normalization, denoted as *LPOG NP* or *LPOG WPCA NP* in result tables, is evaluated and analyzed in comparison with other systems.

Alongside LPOG WPCA's results, the comparisons between them and those of other gradient based methods, including ELBP WPCA, BELBP WPCA, and LPQ WPCA (which are referred as ELBP, BELBP, LPQ in the comparison tables for short notations), are also presented. ELBP WPCA, BELBP WPCA, LPQ WPCA are formed by using a symmetric pair of ELBP, BELBP and LPQ **on gradient images** for feature extraction, respectively, while other details of dimension reduction and classification stages are the same as in LPOG WPCA. Lastly, tests with very low resolution probe face images from SCface database are done to validate LPOG WPCA's performance under video surveillance context. All LPOG WPCA's results are compared with other state-of-the-art systems using the same evaluation protocols.

Besides recognition performance evaluations and details about LPOG's parameters, the speed of LPOG is also assessed and compared with some other feature extraction methods via computational benchmarks on Fa image set of FERET database.

6.3.3.1 Results on AR database

It is apparent from table 6.7 and Figs. 6.11 that LPOG has the highest RRs (almost in all probe sets of all experiments) and the performance differences with other gradient based methods are even more obvious on difficult probe sets (such as 3, 7, 8, 9, 11, 12) as well as when there is the influence of time-lapse (Exp 3 and Exp 4). This approves for the approach we use in LPOG method when combining both BELBP and LPQ to extract local patterns from gradient images rather than using them separately. Along with the above conclusion, the increasing average accuracies from ELBP to BELBP (in all four experiments) make a confirmation that BELBP is better than ELBP.

Another important conclusion based on comparison results in table 6.7 is that the

6.3. Local Patterns of Gradients for Face recognition

Table 6.7: Rank-1 RRs (%) comparison between ELBP, BELBP, LPQ and LPOG (WPCA) on AR database

Test/Method		Probe set												Avg
		1	2	3	4	5	6	7	8	9	10	11	12	
Exp1	ELBP	100	100	91.7	100	100	100	94.8	92.5	94.8	97.8	97.8	94.8	97.0
	BELBP	100	100	94.0	100	100	100	100	97.8	97.0	100	98.5	97.8	98.8
	LPQ	100	100	91.7	100	100	100	97.8	97.0	97.0	99.3	99.3	99.3	98.4
	PLPQMC NP	100	100	93.2	100	100	100	97.8	96.3	86.6	99.3	99.3	97.0	97.5
	PLPQMC	100	100	94.7	100	100	100	97.8	96.3	90.3	100	100	97.8	98.1
	LPOG NP	100	100	94.0	100	100	100	97.0	95.5	86.6	100	99.3	99.3	97.6
	LPOG	100	100	94.0	100	100	100	100	100	98.5	97.0	100	100	99.3
Exp2	ELBP	100	100	93.2	100	100	100	97.5	94.9	97.5	97.5	94.9	92.4	97.3
	BELBP	100	100	95.8	100	100	100	100	96.6	97.5	100	100	96.6	98.9
	LPQ	100	100	94.9	100	100	100	97.5	94.9	96.6	100	100	100	98.7
	PLPQMC NP	100	100	94.9	100	100	100	97.5	94.9	94.9	99.2	98.3	95.8	98.0
	PLPQMC	100	100	94.9	100	100	100	98.3	94.9	94.9	100	98.3	99.2	98.4
	LPOG NP	100	100	94.1	100	100	100	96.6	94.9	91.5	100	99.2	100	98.0
	LPOG	100	100	96.6	100	100	100	100	100	97.5	100	100	100	99.5
Exp3	ELBP	98.3	100	73.7	98.3	100	94.9	82.2	71.2	73.7	94.1	88.1	83.1	88.1
	BELBP	100	100	81.4	100	100	95.8	92.4	80.5	78.8	99.2	93.2	89.8	92.6
	LPQ	99.2	100	78.8	100	100	97.5	85.6	78.8	74.6	98.3	95.8	91.5	91.7
	PLPQMC NP	99.2	100	82.2	100	100	96.6	90.7	81.4	73.7	96.6	91.5	87.3	91.6
	PLPQMC	100	100	82.2	100	100	100	90.7	81.4	78.0	97.5	95.8	91.5	93.1
	LPOG NP	100	100	80.5	100	100	97.5	89.0	82.2	72.0	99.2	95.8	92.4	92.4
	LPOG	100	100	81.4	100	100	97.5	92.4	82.2	83.9	99.2	95.8	92.4	93.7
Exp4	ELBP	97.4	97.4	74.1	100	100	97.4	87.2	74.4	65.8	96.6	92.3	88.0	89.2
	BELBP	100	100	81.4	100	100	98.3	93.2	84.6	78.8	97.5	92.3	88.1	92.9
	LPQ	98.3	99.2	77.6	100	100	99.2	87.2	82.9	75.2	96.6	94.9	89.7	91.7
	PLPQMC NP	99.2	99.2	81.9	100	100	99.2	88.9	81.2	79.7	97.4	91.5	87.2	92.1
	PLPQMC	100	100	81.9	100	100	100	90.6	81.2	81.2	97.4	92.3	89.7	92.9
	LPOG NP	100	100	83.1	100	100	99.2	89.7	84.6	71.8	97.4	94.9	91.5	92.7
	LPOG	100	100	83.1	100	100	99.2	93.2	84.6	83.9	97.4	94.9	91.5	94.0

proposed method is robust to lighting changes as it achieves very high RRs when no illumination normalization is employed. The performance improvement brought by retinal filter is clearly seen only on *Sun glasses* probe sets (numbers 7, 8 and 9), especially with the one of number 9 (Sun glasses with right light on), while it is negligible or not existed upon the rest ones, in all four experiments.

More concretely, it can be seen from table 6.7 that LPOG shows a performance distinction against other gradient based methods when reaching over 99.1% of accuracy in Exp 1, Exp 2, and over 93.7% in Exp 3, Exp 4. The average result of LPOG WPCA is 99.3% when Exp 3 and Exp 4 are excluded. According to the best of our knowledge, these are the best results on AR database with just SSPP for both training and gallery sets. It has perfect accuracy on 8 probe sets of Exp 1 and 10 probe sets of Exp 2. These excellent results confirm the efficiency of LPOG WPCA against variations of

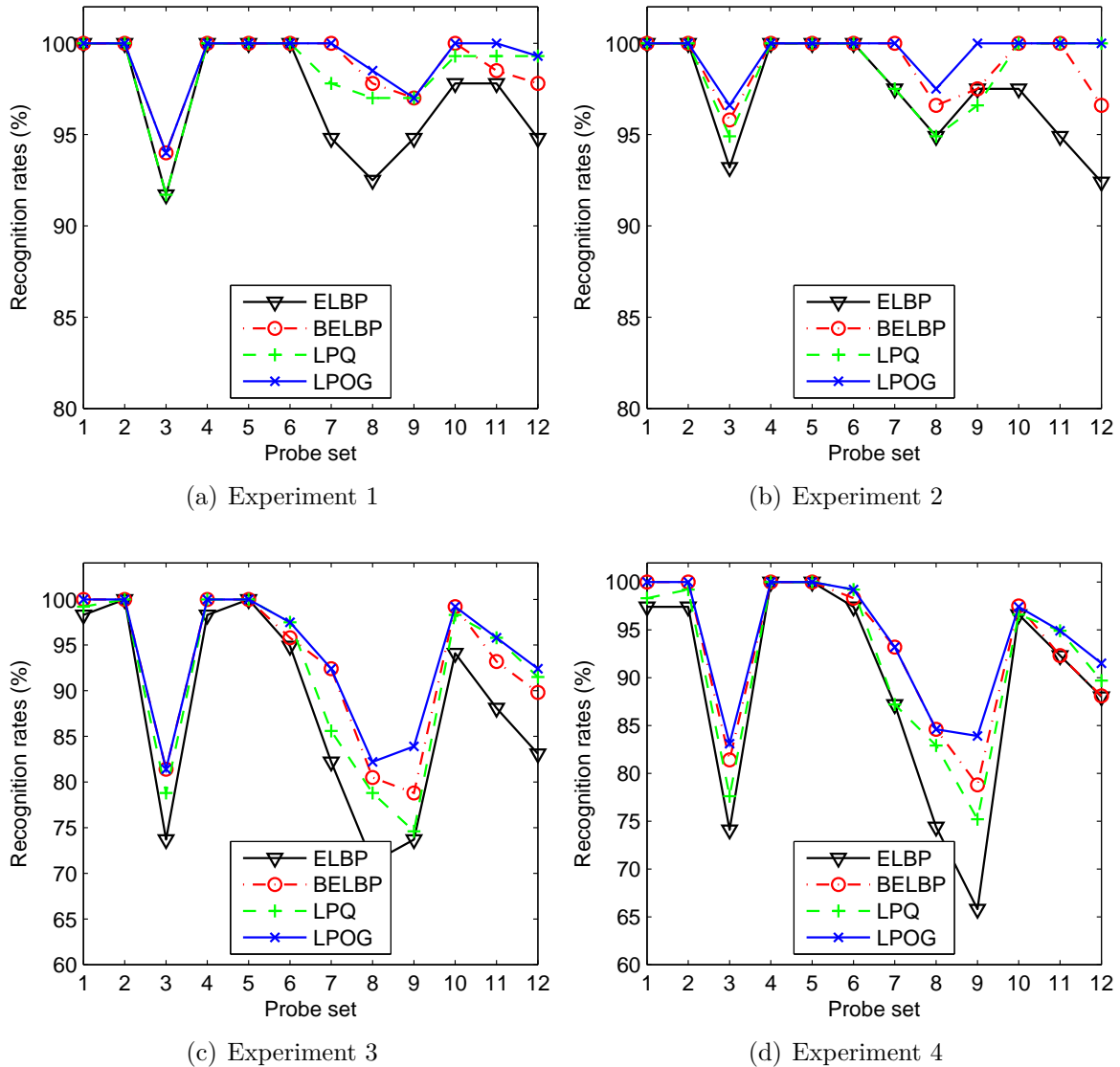


Figure 6.11: Comparisons of recognition performance between LPOG and other gradient images based methods on AR database.

illumination, facial expression, occlusion and time-lapse. More concretely, our system gains perfect accuracy with neutral images under the effect of illumination (probe sets 4-6) in Exp 1 and Exp 2 while missing only 4/236 images in Exp 3 and Exp 4. When occlusions are present (probe set 7-12), the performance reduction is small, from 1.5% to 3.0%, in spite of the fact that the test images are partially occluded about 25%, 40% by sun glasses and scarves, respectively, if time-lapse variation is excluded (Exp 1 and Exp 2).

Among all facial expressions (Smile, Anger, Scream), LPOG WPCA's results on Scream probe sets are the lowest (all experiments). They are even lower than those of all other

6.3. Local Patterns of Gradients for Face recognition

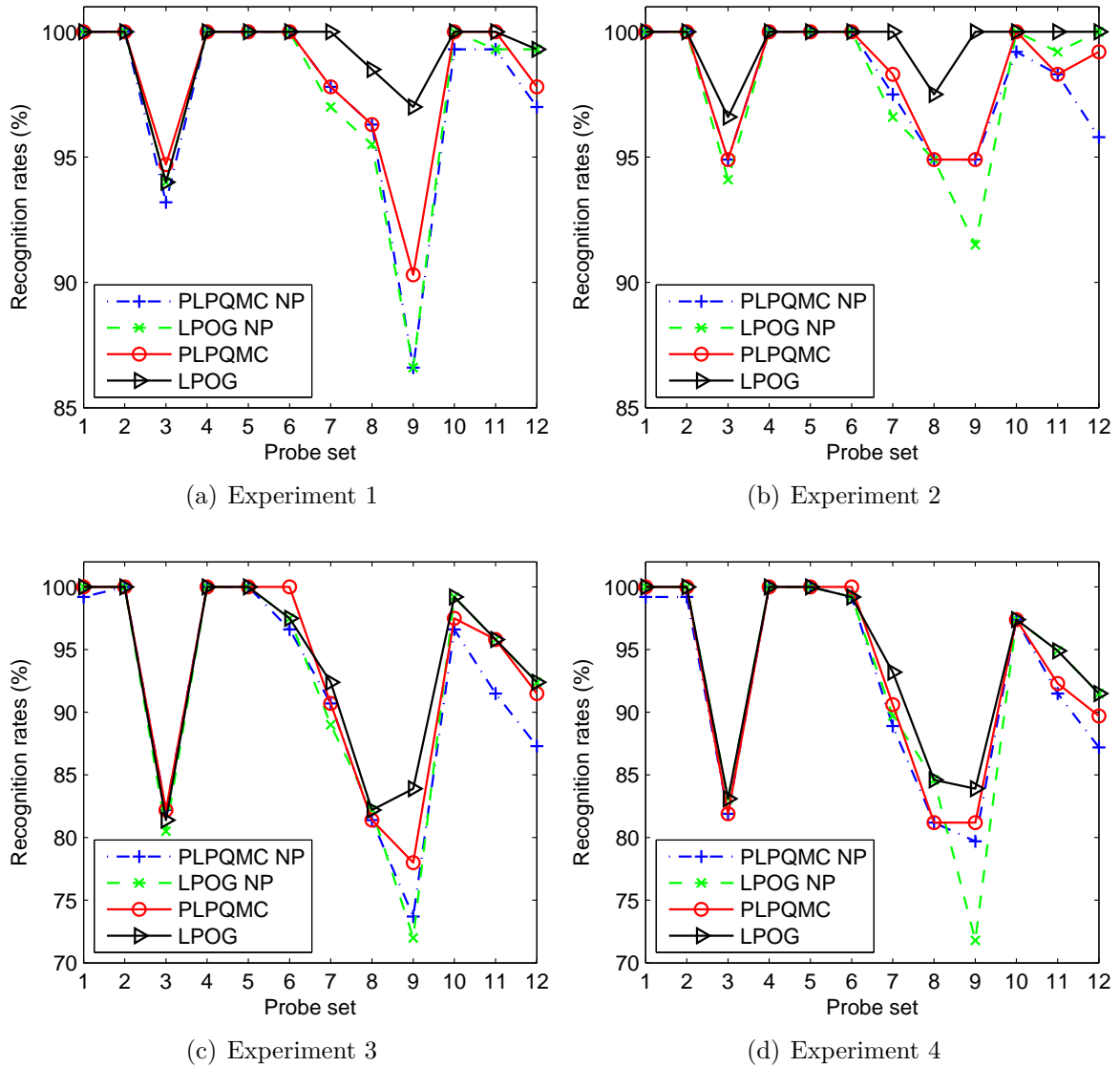


Figure 6.12: Comparisons of recognition performance between LPOG and PLPQMC on AR database.

probe sets when there is no presence of time-lapse variation (only in Exp 1 and Exp 2). This is due to the radical deformations of important facial features (eyes, eyebrows, mouth, etc.) and the overall shape of the face when people scream. Even though the time-lapse interval is only 14 days, it makes sharp performance degradations from Exp 1, Exp 2 to Exp 3, Exp 4, respectively (about 5.4%). The worst performance declinations are on the two illumination affected Sun glasses probe sets (8 and 9).

Additionally, one can observe in table 6.7 and Figs. 6.12 that LPOG surpasses PLPQMC, one of our propositions based on Monogenic components and Patch based LPQ (PLPQ) within chapter 5, in all four experiments. This performance superiority presents under

both cases when the retinal filter is used with the two methods (LPOG versus PLPQMC) or not (LPOG NP versus PLPQMC NP).

Table 6.8: Rank-1 RRs (%) of LPOG WPCA in comparison with other contemporary systems on AR database using the same evaluation method

	1	2	3	7	10	
Method	Smile	Anger	Scream	Glasses	Scarf	Classes ¹
Exp 1						
String face [21]	87.5	87.5	25.9	88.0	96.0	100
Sparse coding [129]	N/A ²			94.7	91.0	100
DMMA[78]	99.0	93.0	69.0	N/A ²		100
SIS [75]	99.0			99.0	98.0	100
PLD [54]	99.0			100	97.0	100
Sparse LF [76]	N/A ²			100	99.2	126
3D [10]	100	100	97.0	N/A ²		100
LPOG NP	100		94.0	97.8	100	134
LPOG	100		94.0	100		134
Exp 2						
LGBP[135]	N/A ²			62.0	96.0	50
Sparse coding [129]				80.3	72.7	100
String face [21]				76.0	88.0	100
IRF [141]				82.5	84.0	120
SIS [75]	86.0			96.0	90.0	100
Sparse LF [76]	N/A ²			96.6	96.6	119
LPOG NP	100		94.1	96.6	100	118
LPOG	100		96.6	100		118
Exp 3						
S-LNMF [90]	62.0	N/A ²	27.0	49.0	55.0	100
Method in [84]	N/A ²		52.3	54.2	81.3	80
PLD [54]	86.0			90.0	89.0	100
3D [10]	99.0	99.0	82.0	N/A ²		100
Bag of words [67]	97.5	97.5	77.3	77.3	89.9	119
LPOG NP	100		80.5	89.0	99.2	118
LPOG	100		81.4	92.4	99.2	118

¹ : The classes column is the number of persons whose images are used in experiments.

² N/A: Not available result.

Results in table 6.8 clearly shows that LPOG WPCA system substantially outperforms all other state-of-the-art competitors in all three experiments. The margins between

6.3. Local Patterns of Gradients for Face recognition

LPOG WPCA’s results and other methods increase from Exp 1 to Exp 3, particularly on Scream and Glasses probe sets, which are the most difficult cases. Only our system that yields perfect RRs on 4/5 probe sets of Exp 1 and Exp 2 as well as 2/5 probe sets of Exp 3. This is an important superiority of our method over the others, especially considering that it uses all available images from AR database while almost others did not (refer to the column *Classes* of table 6.8 for details). Besides, surprisingly, even when the retinal filter is not used, the LPOG WPCA framework (with results in the *LPOG NP* rows) consistently surpasses all cited systems. By this, the illumination invariant of the proposed method is convincingly verified.

Apart from above observations, our results are not compared with SRC based systems ([122], [30], [66]) and SLF-RKR [127], which were claimed to be robust to occlusion variations, because of some common shortcomings:

1. The authors did not follow a well-defined protocol when using random ([30], [66]) or unrevealed [122] selections of images for their experiments. Besides, only at most 120 subjects’ images were used, thus making it unfair to compare them with other methods, including ours.
2. As noted earlier (refer to Section 2.2.5 in chapter 2 for more details), methods in ([127], [122], and [66]) used multiple samples per person for training stage (at least 4 images) or the results were not provided as RRs [30].

6.3.3.2 Results on FERET database

Frontal FERET image sets

Table 6.9: Rank-1 RRs comparison between ELBP, BELBP, LPQ and LPOG (WPCA) on FERET database using standard protocol [96]

	ELBP	BELBP	LPQ	LPOG
Fb	99.4	99.7	99.6	99.8
Fc	100			
Dup I	92.8	96.5	95.3	97.4
Dup II	92.3	97.0	94.9	97.0
Avg	96.7	98.5	97.8	98.8

Results in table 6.9 evidently demonstrate the effectiveness of the LPOG method as it substantially surpasses ELBP, BELBP, LPQ on all four probe sets. The dominance of LPOG over those methods are clearer and more cogent on Dup I and Dup II, the most difficult probe sets of FERET database (see table 6.10 for more details), on which great performance improvements are shown. These findings are exactly in accordance with

those upon AR database (table 6.7) and our aforementioned expectations from LPOG. Again, one can observe without difficulty the increasing performance enhancements from ELBP to BELBP as well as the advantages of fusing both BELBP and LPQ (at feature level) in LPOG over just using them separately.

Table 6.10: Rank-1 RRs (%) comparison of LPOG based systems with other state-of-the-art results on the FERET database [96]

Method	Fb	Fc	Dup 1	Dup 2	Average
HMBP [130]	98.1	98.5	75.8	75.2	89.0
CHG [22]	97.5	98.5	85.6	84.6	92.6
Tan et al. [113]	98.0	98.0	90.0	85.0	94.2
LMG [94]	99.8	100	89.2	86.8	95.3
ESRC [30]	97.3	99.0	93.8	92.3	95.9
MS-LPQ [17]	99.2	100	92.0	88.0	95.9
EPFDA [104]	99.6	99.0	92.0	88.9	96.1
POEM PDO [118]	99.7	100	91.7	90.6	96.4
FLPGMP [107]	99.0	99.0	94.0	93.0	96.9
G-LQP [52]	99.9	100	93.2	91.0	97.0
MBC-F [128]	99.7	99.5	93.6	91.5	97.0
GSF [123]	99.6	99.5	94.0	91.5	97.1
PLPQMC NP	99.6	100	95.4	94.0	97.8
GOM [16]	99.9	100	95.7	93.1	97.9
LPOG WPCA NP	99.7	100	96.1	94.4	98.1
SLF-RKR [127]	99.7	99.5	96.3	94.4	98.1
PLPQMC	99.7	100	96.8	95.7	98.4
LPOG WPCA	99.8	100	97.4	97.0	98.8

According to results in table 6.10, it is indisputable that LPOG WPCA convincingly and significantly outperforms other state-of-the-art systems on FERET database using standard protocol [96]. Our method is proved to be efficient to facial expression and illumination variations when it achieves perfect accuracy (100%) on Fc set and only misses two images on Fb set. But the most impressive results are on time-lapse variation probe sets: Dup I and Dup II, where LPOG WPCA gains 97.4% and 97.0% RRs, respectively, while the best published ones in the literature are only 96.3% and 94.4%. These numbers surely confirm the robustness of LPOG WPCA against time-lapse variation, which is widely known as one of the most difficult challenges associated to FR. From the best of our knowledge, our system, with its average accuracy of 98.8%, establishes the best results upon FERET database.

Additionally, it can be readily seen that, without using any preprocessing method for illumination normalization, PLOG WPCA attains very impressive results:

6.3. Local Patterns of Gradients for Face recognition



Figure 6.13: Two Fb probe images (a), their wrongly assigned ones (b) and their correct gallery ones (c).

- Perfect recognition performance on Fc set, in which the probe images are strongly affected by illumination variations.
- Comparable with the best reported system in the literature. Indeed, the proposed system has the same overall RR with SLF-RKR [127] while surpassing the rest ones. Since the effect of lighting changes is elusive and omnipresent, it is not isolated in Fc but impairs all four probe sets. Hence, the high RRs of LPOG WPCA on the whole frontal FERET database are big evidences for its efficiency against variable lighting.

From above remarks, we conclude that LPOG is an illumination invariant feature extraction method.

Besides, in comparison with PLPQMC method presented in chapter 5, the results from tables 6.10 and 6.7 agree with each other as in both databases, LPOG outperforms PLPQMC on all probe sets regardless the usage of retinal filter for illumination normalization or not.

As the proposed method only misses two probe images of the Fb set, it is tempting to shed light on the cause of that failure. To this end, we present together in Figs. 6.13 (a-c) two missed Fb probe images (a), their wrongly recognized ones (b) and their expected galleries (c). Surprisingly, three images of each case seem to share one identical identity. It is worthwhile indicating that we determine whether one probe image is correctly recognized to have the same identity with one assigned gallery

image based on their annotated identities, which are the first five characters in the images' file names and can be seen in Figs. 6.13 (a-c) (the file name of each image is the text below it). By this, images *00277fb010_940422* and *00368fb010_940422* (Figs. 6.13 a), having identities *00277* and *00368* respectively, were incorrectly identified since they were assigned to images *00185fa010_940128* and *00463fa010_940422* (Figs. 6.13 b), whose identities are *00185* and *00463*, correspondingly. Their true and expected gallery images, in order, are the ones named as *00277fa010_940422* and *00368fa010_940422*. However, it is probably that they belong to two identical individuals and the proposed system was correct. To clarify this, the original versions of images in Figs. 6.13 are displayed in Figs 6.14. From that figure, we can conclude that the three images *00277fb010_940422*, *00185fa010_940128* and *00277fa010_940422* belong to one person. The same conclusion is drawn for images *00368fb010_940422*, *00463fa010_940422* and *00368fa010_940422*. Hence, our system was turned out to be right and its RR on Fb set of FERET is indeed **100%**. To the best of our understanding, *this is the first time* a FR system reaches such perfect accuracy. Nevertheless, as the finding about these two cases of Fb set above has never been detailed in any study on the FERET database and all the earlier experiments were conducted upon the same datasets, we do not change the reported RRs on the Fa set of every system presented in this thesis.

Non-frontal FERET image sets

Table 6.11: Rank-1 RRs comparison between ELBP, EBLBP, LPQ and LPOG (WPCA) on b-series of FERET database

	ELBP	BELBP	LPQ	LPOG
-40°-bh	81.5	90.5	91.5	95.0
-25°-bg	99.5	100	100	100
-15°-bf	99.5	100	100	100
+15°-be	100	100	100	100
+25°-bd	99.3	100	100	100
+40°-bc	81.5	92.0	92.5	95.5
Avg	93.6	97.1	97.3	98.4

It is easily noticed that the numbers in table 6.11 are consistent with those from table 6.7 and 6.9, with respect to accuracy performance. These results, in addition to prove the efficiency of LPOG WPCA framework, help to confirm again the effectiveness of the LPOG method, when dealing with pose variations challenge. Besides, the RRs are enhanced from ELBP to BELBP as expected. These enhancements are more marked when the pose angles are larger (from $\pm 15^\circ$ to $\pm 40^\circ$).

From comparison results provided in table 6.12, we can conclude that LPOG WPCA is robust against pose variation since it actually outperforms other contemporary com-

6.3. Local Patterns of Gradients for Face recognition

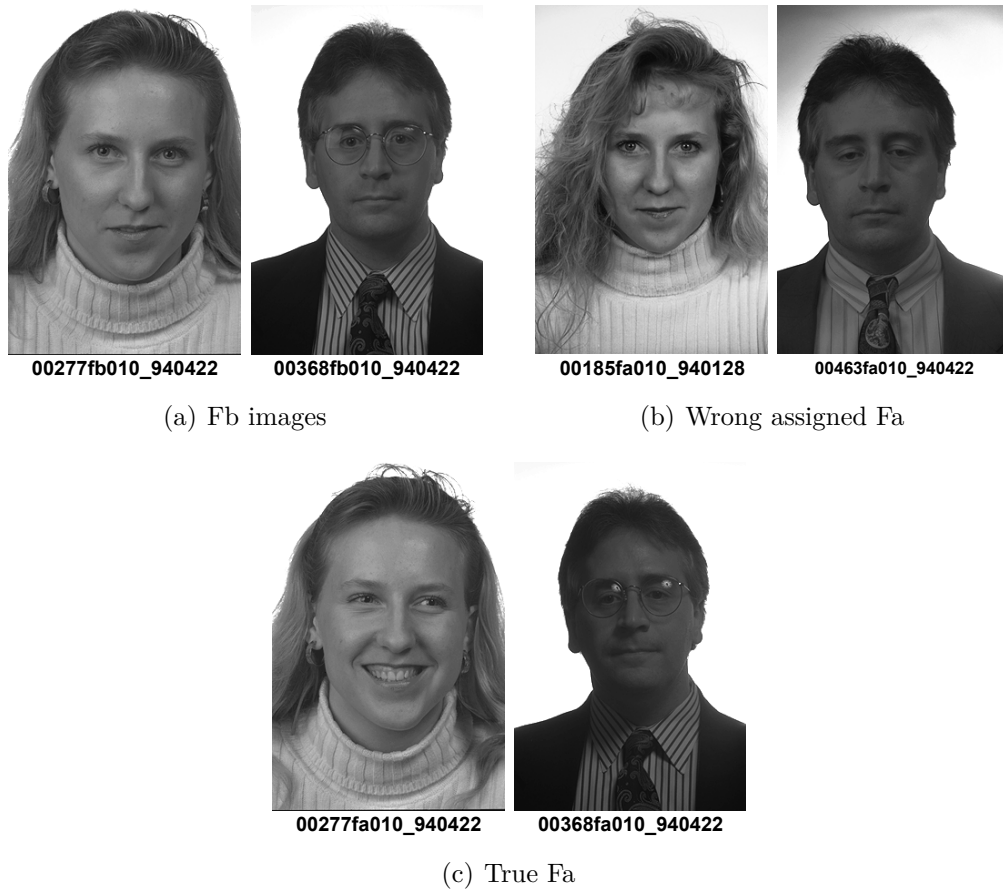


Figure 6.14: Original versions of images in Figs 6.13.

petitors. Admittedly, this is an outstanding capability of our system when taking into account the fact that it is only a general framework for FR while almost others (in table 6.12), except SLF-RKR [127], LSED [121], RFC [20], and GLHO [100], are dedicated ones with special tactics for tackling the problem of pose changes. Besides, our method is the only one that attains perfect RR (100%) under minor changes in head pose (within $\pm 25^\circ$). This is the best known result so far. Furthermore, leading systems, such as DWFF [86], MRF [48], 3D Pose Norm [6], CPN [31] and PAF [131], require a lot of landmark points of the face images or 3D information to synthesize a face model or for extracting facial features to proceed the recognition while our system only need two eyes' coordinates. Additionally, all the top six systems (in table 6.12) are Gabor wavelets based approaches, which are proved to be slower than our (see table 6.16). With its average success rate of 98.4% on b-series images, LPOG WPCA achieves an overall average RR of 98.6% on both frontal and non-frontal probe sets.

Once again, and in consistence with comparison results in tables 6.8 (on AR database) and 6.10 (on frontal FERET dataset), one can observe in table 6.12 that even when no preprocessing method is used to handle illumination changes, the proposed system still

Table 6.12: Rank-1 RRs comparison of LPOG WPCA with other leading systems on FERET b-series.

Method	-40° bh	-25° bg	-15° bf	+15° be	+25° bd	+40° bc	Avg
SLF-RKR [127]	N/A ¹	55.0	100	96.0	57.0	N/A ¹	N/A ¹
LSED [121] * ²	78.0	84.0	88.0	89.0	88.0	83.0	85.0
CCA [64] * ²	81.0	91.0	92.0	94.0	89.0	80.0	87.8
PAN [39]	81.5	93.0	97.0	98.5	91.5	78.5	90.0
RFC [20]	84.2	90.2	94.0	93.2	92.5	89.5	90.6
ADMCLS [105]	85.0	94.0	96.0	95.0	94.0	82.0	91.0
LMG [94]	N/A ¹	91.5	98.0	98.5	93.5	N/A ¹	N/A ¹
MRH [7] * ²	87.0	94.0	98.0	99.0	96.0	74.0	91.3
GLOH [100] ** ³	81.1	94.5	100	100	94.5	81.1	91.9
DWFF [86]	87.5	98.0	100	99.0	98.5	82.4	94.2
MRF [48]	91.0	97.3	98.0	98.5	96.5	91.5	95.5
3D Pose Norm [6]	90.5	98.0	98.5	97.5	97.0	91.9	95.6
CPN [31]	94.5	98.0	98.5	99.0	98.5	97.0	97.6
LPOG WPCA NP	93.5	100			99.5	94.0	97.8
PLPQMC	95.0	100			99.5	95.0	98.3
LPOG WPCA	95.0	100			95.5	98.4	
PAF [131]	98.0	98.5	99.25		98.5	98.0	98.6

¹ N/A: Not available result.

² *: The RRs of the method are estimated from plotted figures.

³ **: The RRs on $\pm 25^\circ$ and $\pm 40^\circ$ subsets are average results.

offers a great performance on pose varied images as there is only one method, PAF [131], whose RRs surpass what it gains. Thus, in combination, those results evidently confirm the robustness of the LPOG facial representation under illumination variations.

6.3.3.3 Results on SCface database

In this Section, to investigate the performance of LPOG under video surveillance context, we perform two experiments on SCface database and report our results with two distinct training sets: the frontal Fa set of FERET database [96] (denoted as **Our-F**) like in PCA [42] and PLPQMC-F, and frontal mug-shot images (denoted as **Our-S**) like in DSR [140], ELBP(h+v) [89] and PLPQMC-S.

As can be seen in table 6.13 and table 6.14, the LPOG WPCA framework significantly outperforms other methods on all probe sets in both experiments. It has average performance improvements of 8.1% and 4.4% with DayTime and NightTime experiments

6.3. Local Patterns of Gradients for Face recognition

Table 6.13: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the DayTime protocol [42]

Probe set	PCA	DSR	ELBP(h+v)	PLPQMC-F	Our-F	PLPQMC-S	Our-S
cam1_1	2.3	N/A ¹	43.1	64.6	62.3	70.0	69.2
cam1_2	7.7		56.2	63.9	71.5	68.5	73.1
cam1_3	5.4		45.4	42.3	45.4	43.9	47.7
cam2_1	3.1		36.9	46.9	47.7	54.6	57.7
cam2_2	7.7		50.8	57.7	56.2	65.4	66.2
cam2_3	3.9		42.3	37.7	43.1	41.5	48.5
cam3_1	1.5		34.6	42.3	36.2	49.2	49.2
cam3_2	3.9		46.9	59.2	59.2	64.6	63.1
cam3_3	7.7		51.5	43.1	49.2	49.2	54.6
cam4_1	0.7		32.3	46.9	39.2	47.7	43.9
cam4_2	3.9		50.0	68.5	66.9	72.3	75.4
cam4_3	8.5		50.8	46.2	52.3	51.5	58.5
cam5_1	1.5		36.2	54.6	43.1	55.4	53.9
cam5_2	7.7		32.3	51.5	53.9	56.2	52.3
cam5_3	5.4		31.5	31.5	36.2	40.0	38.5
Average	4.7		20.2	42.7	50.5	50.8	55.3

¹ N/A: Not available result

Table 6.14: Rank-1 RRs (%) comparison with other state-of-the-art results on SCface database using the NightTime protocol [42]

Camera/Distance	PCA[42]	ELBP(h+v)	PLPQMC-F	Our-F	PLPQMC-S	Our-S
cam6_1	1.5	9.2	10.0	13.1	14.6	13.1
cam6_2	3.1	15.4	22.3	22.3	24.6	23.9
cam6_3	3.9	25.4	25.4	26.2	25.4	31.5
cam7_1	0.7	13.1	15.4	17.7	15.4	17.7
cam7_2	5.4	13.1	15.4	17.7	22.3	20.0
cam7_3	4.6	13.9	20.8	19.2	22.3	19.2
Average	3.2	15.0	18.2	19.4	20.8	20.9

(with Fa training set), respectively, in comparison with ELBP(v+h) [89], the best reported results in the literature. Amongst probe images at three distances, the ones of distance 2 (2.6m) bring the highest RRs while those of distance 1 (4.2m) have the lowest accuracies. Since camera 1 has the best resolution [42], its accuracy is higher. One can also notice the accuracy discrepancies between DayTime and NightTime probe sets (i.e., 50.8% versus 19.4% on average RR). These tremendous performance drops are caused by the extreme effect of IR night vision mode’s illumination conditions on the probe images of NightTime test. The illumination variations, therefore, will continue being a substantial challenge to FR systems. Another conclusion we can draw from

these results is that the accuracy rates of LPOG WPCA are higher (50.8% versus 56.8% in DayTime tests and 19.4% versus 20.9% in NightTime tests on average RR) when using images of test objects for training than a totally different image set (the frontal Fa set of FERET database).

Specifically, when comparing the overall RRs between LPOG and PLPQMC, LPOG is the winner as it surpasses PLPQMC in both protocols and with two training sets: the Fa set from the FERET database and the mugshot set of the SCface database. This victory of the proposed method is consistent with the comparison results upon the AR and FERET databases in the previous Sections.

Although LPOG WPCA’s results on SCface database are interesting, they are far below the ones upon AR, FERET databases previously reported (in tables (6.7-6.12)). In our opinion, the striking contrast between gallery images, which are high quality frontal mug-shot taken under controlled lighting condition, and probe images, that are low resolution ones (small in size and very poor in quality) acquired under uncontrolled conditions (with variations due to blur, pose, illumination, distance) by varying quality surveillance cameras, is the reason for those massive performance gaps. These are also the proofs that FR is far from being a completed research, at least in video surveillance context. Besides, this suggests a pressing need of more attention from scientists on FR in video surveillance systems with more powerful methods.

6.3.3.4 Parameters setting

Table 6.15: Details of divided sub-regions and window size used with LPOG

	AR	FERET	SCface
Window size (M)	11	11	9
BELBP sub-regions	7x9	9x9	3x3
LPQ sub-regions	8x9	10x10	4x4

There are several parameters in LPOG method: the BELBP radiuses (horizontal and vertical), its number of neighborhood pixels, the window size (M) and ρ associated to LPQ, and the number of divided sub-regions for each local patterns image. All the values of these parameters (except the window size and the divided sub-regions) are selected through empirical experiments and fixed when testing with all databases.

We use two symmetric pairs of radiuses, (5, 3) and (3, 5), and 8 neighborhood pixels in BELBP to build LPOG feature vectors. LPQ’s parameter ρ is assigned as 0.89. Details of sub-regions and window size used with BELBP and LPQ upon different databases are shown in table 6.15. One can notice that on SCface database, because of their quite small images’ sizes (48x48 resolution), we just use 3×3 and 4×4 sub rectangles for BELBP and LPQ, respectively.

6.3. Local Patterns of Gradients for Face recognition

6.3.3.5 Computational cost

For examining computational cost of the LPOG method, its Matlab implementation is benchmarked by the test upon the Fa image set of the FERET database, whose description is in Section 4.1.4.5 of chapter 4. The obtained results (total required time, extraction time and speed-in terms of number of images per second) are compared with those of initial step of Gabor wavelets (just generating Gabor wavelets components at 5 scales and 8 orientations) based approaches, and some advanced feature extraction algorithms such as MBC-A, MBC-O [128], PLPQMC (one of our propositions presented in chapter 5 of this dissertation) as well as LPQ [5] and BELBP.

Table 6.16: Computation time of LPOG in comparison with other feature extraction methods

Method	Image size	Time (seconds)	Extraction time (milliseconds)	Images/second
LPQ	128×128	5.45	4.56	219
BELBP	128×128	7.4	6.19	162
PLPQMC	128×128	27.91	23.34	43
LPOG	128×128	27.02	22.59	44
MBC-A [128] ^{*1}	150×130	30.54	25.54	39
MBC-O [128] ^{*1}	150×130	87.00	72.74	14
Gabor wavelets	88×80	96.23	80.46	12

¹ *: We used the Matlab code provided by the author.

The comparison results from table 6.16 apparently indicate the fast speed of the LPOG method, despite the fact that it requires about twice the total time of BELBP and LPQ. This is because LPOG operates on two gradient images while BELBP and LPQ only work with one intensity image of the same size. Our method is around 3 times faster than MBC-O [128]. As MBC-F [128] is the fusion approach which needs MBC-A, MBC-O and MBC-P (which extracts LXP [128] features from phase component of monogenic signals and has the same speed as MBC-O), LPOG is obviously faster than this method. One can observe that the initial step of Gabor based methods is really slow since it could only process 12 images in a second, even though this operation works on quite smaller images of 80×88 resolution. More impressively, LPOG is faster than any feature extraction algorithms in advanced systems, whose average RRs on FERET database (in table 6.10) are at least 96.9%, since they are Gabor wavelets based (EPFDA [104], FLPGMP [107], G-LQP [52], GSF [123], GOM [16] and SLF-RKR [127]), or Monogenic based (MBC-F [128]), which are all slower than our method. Also, since its first step is just generating two directional gradient images whilst PLPQMC needs to produce 6 DBCs for doing the feature extraction process, LPOG is slightly faster PLPQMC. Having the processing speed of 44 images per second, LPOG is at least 3.5 times faster

than any Gabor wavelets based feature extraction methods and it is definitely capable of being used in real-world applications, including video surveillance context.

6.3.4 Conclusions

Motivated by the results of visual perception researches ([61], [56], [49]), and the benefits of BELBP and LPQ operators in local texture patterns encoding, we presented the Local Patterns of Gradients (LPOG), a novel feature extraction method for FR. Firstly, a new variant of ELBP is proposed and named as Block-wise ELBP. Through comparative experiments' results upon AR and FERET databases, BELBP is proved to be more efficient than ELBP. Then, LPOG method is formed by exploiting both BELBP and LPQ operators on gradient images to generate LPOG images. Each LPOG image is next partitioned into non-overlapped sub-regions from which we calculate the histogram sequences. The concatenations of histogram sequences from each image are then incorporated to constitute the LPOG feature vector of the input image. Equipped with LPOG for feature extraction, we have proposed a novel, unified, single sample per person FR framework called Local Patterns of Gradients Whitened Principal Components Analysis (LPOG WPCA) by using WPCA for dimension reduction, k-NN and weighted angle-based distance function for classification.

Extensive experiments conducted on three large scale public face databases prove the efficiency and effectiveness of the proposed method. Comparison results upon these databases strongly show that LPOG WPCA convincingly outperforms other contemporary systems. The method achieves stable, consistent and outstanding performance with respect to a diversity of challenges such as single sample per person, illumination, facial expressions, occlusion, time-lapse, pose variations and low resolution probe images. In more details, our system gains, for the first time ever, 99.3% average recognition rate on AR database (without time-lapse variation) using just one sample per person for training and gallery sets, 98.6% overall accuracy on frontal and non-frontal probe sets of FERET database. Meanwhile, it also outperforms other methods and attains good performance on SCface database, whose probe images are extremely challenging. To the best of our knowledge, these are *the best results* in the FR literature.

In addition to the robustness in feature extraction, one of the most prominent properties of the LPOG method is its fast processing speed. LPOG is faster than many advanced feature extraction methods, which are usually based on Gabor wavelets. With its processing speed up to 44 images per second (provided by an un-optimized Matlab implementation), LPOG is certainly able to be used in any real-world application.

As a summary, we highlight here some representative characteristics of LPOG method:

1. LPOG extracts local texture patterns from gradient images, this makes it contains

6.4. Conclusions

many discriminative features for FR and provides it with illumination invariant capability. These attributes come both from the fact that LPOG works on gradient images instead of intensity images, and from the capacity of BELBP and LPQ operators.

2. LPOG is strong to blurred images, a characteristic derived from LPQ. This accounts for the high recognition rates it achieves on SCface database.
3. LPOG is robust against many challenging factors such as: facial expression, occlusion, pose, time-lapse variations and low resolution images.
4. Illumination insensitivity is a great property of LPOG description. This is mainly stemmed from the fact that LPOG features are extracted from gradient images where the illumination effects are highly reduced.
5. The method is computational efficiency and can be used in real life scenarios.

Notwithstanding the standout performance we have shown in this paper, the results also confirm that FR remains an open problem, especially under the effects of the hardest challenges such as time-lapse, pose, illumination variations, and low resolution probe images.

6.4 Conclusions

In this chapter we have focused on feature extraction based on gradient images and have presented two novel methods referred as Elliptical Patterns of Oriented Edge Magnitude Images (EPOEM) and Local Patterns of Gradients (LPOG).

With the aim to enhance the discriminatory features of POEM, a description based on LBP, an horizontal ELBP descriptor is proposed to be used in EPOEM to extract local patterns from oriented edge magnitudes images of different orientations. As expected, EPOEM has been demonstrated to be more effective than POEM under various FR issues. In addition to that, the efficiency of the method was also verified when it attained very promising results upon AR, FERET and SCface databases. Besides, however, the existence of a weakness of EPOEM is figured out. To rectify this drawback, several suggestions have been given.

Inspired by the advantages of the gradient images (horizontal and vertical) over magnitude and orientation components, as well as the intensity appearance of a face image, and by introducing Block-wised ELBP (BELBP), a new variant of ELBP, then employing it together with LPQ to capture local features from gradient images, we proposed Local Patterns of Gradients, a robust yet low computation cost feature extraction method for

face recognition. In applying LPOG for FR, the *WPCA based* framework is employed to build a new SSPP FR system, in which the *weighted angle based* distance function is used to measure the similarities between probe images and gallery ones. Numerous experiments on three public databases (AR, FERET and SCface) have been undertaken and the results show that the proposed method outperforms almost all other existing systems and is robust against a variety of challenges. Impressively, LPOG is an *illumination invariant* facial representation as it gained amazing RRs without relying on any illumination normalization technique. Additionally, due to its low complexity, the method is viable to be implemented in real world applications.

In comparison with other methods in chapter 4, which are intensity-based features, and the one based on Monogenic components and PLPQ in chapter 5 (PLPQMC), our facial representations described within this chapter are gradient images based features. In fact, EPOEM was proposed just after the appearance of horizontal ELBP as a proof of the superiority of ELBP over LBP when applying in a multi-resolution method. Afterwards, the PLPQMC was introduced, and then the LPOG came, for addressing the aim of more sophisticated descriptions. As empirically shown, LPOG has achieved the highest results and in our standpoint, it has met our objectives with this thesis the most. Nevertheless, we do not slavishly believe that these results are the best a FR system can gain. Instead, we keep churning out new, more advanced methods in the future work.

Chapter 7

Conclusions and Future work

7.1 Conclusions

With the main focus on feature extraction, the task being at the heart of every Face recognition system, this doctoral dissertation has proposed several methods targeted to robust facial representations against a *diversity of challenging factors*, including variations of illumination, expression, occlusion, time-lapse and pose, and unconstrained low resolution probe images. Another important aspect is aimed at *low computational cost* solutions that are feasible to apply in real life applications.

To address the above objectives, while how to devise such a method is not available in words in the literature, we have explored into the insights of knowledge gained over years from Image processing, Visual perception and Face recognition fields. It is worth noting that merely mimicking how human beings recognize each others based on their faces, or copying what has been done, does not lead to a right and sufficient approach, and is not the way we have followed. However, by analyzing pros and cons of the currently used methods and distilling thoughtful observations learnt from the results of those study areas into detailed steps, new methods were proposed.

Before enumerating what has been made in this thesis, we wish to recount in brief, and in order of time, the underlying concepts of all the propositions we have done. Our first finding is that facial information is oriented in horizontal direction and important facial features, like the eyes and mouth, are in shapes of horizontal ellipse. This is the motivation for the horizontal Elliptical Local Binary Pattern (ELBP) descriptor, which is further applied to construct the Elliptical Patterns of Oriented Edge Magnitudes (EPOEM) description. Afterwards, the complementary property of a symmetric pair of a horizontal and a vertical ELBP is discovered. Then, advantages of Monogenic filter's directional bandpass components and Local Phase of Quantization (LPQ) are

found and exploited in the Patch based LPQ of Monogenic components (PLPQMC) method. Lastly, the Local Patterns of Gradients (LPOG) facial representation is inspired from the fact that horizontal and vertical gradient images contain a large amount of meaningful features while being unaffected by variable lighting, and from the efficiency of Block-wised ELBP, a new variant of ELBP, and LPQ descriptors.

Amongst those methods, PLPQMC and LPOG are the most preminent ones for which we have strived. But it should be underlined that they are not a matter of luck and the details of their methodologies did not come *instantly* or *by coincidence*. Instead, they are logical results of a long-term continuous progress of attacking FR challenges based on elementary descriptors, like ELBP and LPQ, while pursuing some *common key principles* derived from analysis of existing systems and our results (in chapters 2 and 4): 1) the local features extracted by elementary methods on intensity image are not adequate for a robust facial representation. By some manner, the extraction process must be effectively carried out on multiple components generated from an input image and it is better if they are less affected by illumination variations. 2) The larger amount of facial information these components retain, the more robust features the resulting vector has. 3) Though it will be a long period of time before we can see a perfect FR system, however, there are always some ways to improve the efficiency of a feature extraction method. For this, we did not stop, because the solutions did not reach the perfection stage. One can raise a question like that: Why did you continue when EPOEM and PLPQMC, at some level of accuracy and in some certain circumstances, were enough? Fortunately, their results are good with controlled conditions, but under uncontrolled scenarios, i.e. video surveillance context, poor performance is the best they gained. Spontaneously, we have struggled to find new and more efficient feature extraction strategies with the hope that, someday, the newly discovered things are sufficient to propose a novel method, because we believe, that day will come.

The list below are some of the main contributions of the present thesis:

- Despite the existence of many variants, LBP is still widely used as an elementary descriptor for building up numerous advanced facial feature extraction methods. This is probably changed with the advent of ELBP, our novel method detailed in chapter 4, due to its substantial performance superiority over LBP, especially under challenging obstacles, while having the same complexity and computing cost (horizontal ELBP) or does not affect the speed too much (when the combination of horizontal and vertical ELBP is used). Based on oriented characteristics of facial information and critical features, such as eyes and mouth, we showed that horizontal ELBP, using a horizontal ellipse sample for computing ELBP patterns, is more relevant and efficient for FR than LBP, which is originally designed for texture classification problem. Further, realizing that a pair of symmetric ELBPs (horizontal and vertical) are complementary descriptors, they are so combined to capture more useful features from face images. Delivering

7.1. Conclusions

promising results when plugging in the *WPCA based* framework and owing a rapid processing speed, ELBP is potential for extracting local features in more sophisticated methods and using in other face analysis related tasks, for instance face detection and gender classification.

- Through extensive experiments under various conditions, LPQ, a phase based local descriptor whose details are in Section 4.2.2 of chapter 4, was found much more robust than a lot of Gabor wavelets, LBP and its variants based methods. The merits of LPQ are blur tolerant, uniform illumination invariant and fast speed. Being elementary methods and proceeding the feature extraction step on intensity image, which is largely affected by lighting changes and has noise, it is understandable that the performance of ELBP and LPQ can not competing with the best approaches in the literature. However, importantly, these basic descriptors have their places, they are cornerstones in founding more robust descriptions. This was evidently illustrated in chapters 5 and 6 where they were exploited to constitute PLPQMC, EPOEM and LPOG, three powerful facial representations in this thesis.

- Description in chapter 5 are details of PLPQMC, a novel feature extraction method for FR. Based on Patch based LPQ, a new variant of LPQ, and the Directional Bandpass components generated by Monogenic filter, the descriptor has many valuable attributes, e.g. tolerant to blurred images, high discriminant features and *illumination insensitivity*. Experimental results showed that our method significantly outperforms almost all other state-of-the-art systems when coping with a large spectrum of difficulties. Even *without the usage of any illumination normalization technique*, its accuracy performance is comparable to the leading rivals. Regarding to the computational aspect, the proposed method is *faster* than all other advanced algorithms based on Gabor wavelets and Monogenic filter.

- In the first part of chapter 6, a new variant of POEM, referred as EPOEM, was presented. To improve the discriminatory power of local patterns extracted from oriented edge magnitudes images in POEM, the horizontal EBP descriptor was applied instead of LBP one. As expected, EPOEM was empirically shown its superiority over POEM and gained very encouraging results in comparison with state-of-the-art systems.

- To mining local patterns hidden inside gradient images, which serve as the supply of facial features rather than the intensity appearance, Block-wised ELBP (BELBP), a refined variant of ELBP, is applied together with LPQ directly on them. By this way, the Local Patterns of Gradients, an elite facial feature extraction method is formulated with many desirable characteristics. Chiefs among these are *blur insensitivity* and *illumination invariant*. They are the fruition of the methodology in which the LPOG vector is built. It contains two kinds of local features: the micro textures from BELBP and the phase based textures from LPQ. While BELBP features contribute to the general discriminative ability of the method, LPQ ones provide it with blur tolerant attribute.

Since the extraction process is performed on gradient images, which carry a number of helpful visual features, e.g. local contrast, strong edges, discontinuities properties, and especially are less influenced by illumination variations, the resulting representation is thus more discriminant and robust to lighting changes. The robustness of the method was convincingly validated when it substantially outperformed all other considered systems and established *the best results on all three tested databases*. Moreover, having a fast processing speed, LPOG is capable of applied in real time scenarios, even under video surveillance context.

- For the first time in the **FR** literature, the **complete and highest** results upon **SCface**, a very challenging database with *unconstrained* and low resolution probe images captured by surveillance cameras, is reported by ELBP(h+v) WPCA framework in chapter 4. These are then considerably improved by PLPQMC WPCA and LPOG WPCA in chapter 5 and 6, respectively.
- For evaluating one feature extraction methods or other related techniques, for instance face detection, face alignment and illumination normalization, the *Template matching* and *WPCA based* frameworks in chapter 3 are good paradigms to use. They can be adapted without difficulties to meet such kinds of requirement.

7.2 Future work

In spite of excellent results delivered by our propositions in this thesis, there are many unanswered questions. Face recognition is like a never ending story where the perfection is hard to come by, particularly under uncontrolled conditions or in video surveillance context.

For future works, there are two sorts of task we wish to do: (1) improving the proposed methods by remedying some weaknesses they have or adding some boosting steps, and (2) further exploring the capabilities of them in face analysis related tasks. Hereafter is the list of some potential directions that could be pursued.

- As the upper part of the face (above the nose) contains a larger amount of important facial features than the lower part, it is expectable that this can be a suggestion to apply for enhancing the discriminant of local descriptors like ELBP and LPQ in which the final vector is a combination of histogram sequences computed from evenly divided sub-regions. We believe that an asymmetric dividing strategy should be better.
- In order to reduce LPQ vector size, a statistic algorithm or a learning method from machine learning study can be exploited.
- Whilst Directional Bandpass components are better than other ones generated by

7.2. Future work

Monogenic filter, greater performance might be achieved if a fusion tactic (at feature level or score level) is applied with the join of ELBP.

- Gradient images in horizontal and vertical directions, which were proved to be good source of facial features in LPOG method, are in fact the first order gradients, then some new questions are arisen. What about the second order gradient images? Do they play any part in the feature extraction context? Can they be combined with their first order counterparts? And if the answers are positive, then how to do that effectively?
- With a variety of local feature extraction methods presented in this thesis, we wish to have a fusion method to combine all kinds of features from them to found an unified and robust facial representation.
- In order to make what we have proposed more feasible for real life situations, an automatic face detection method, such as Active Appearance Model [23] and its variants [81], should be applied.
- One sound way to improve the recognition performance of FR frameworks examined in this dissertation is the use of an efficient face alignment technique, rather than the simple face cropping algorithm we adopted. For this, works in [50, 51] are interesting to proceed with.
- As pointed out in chapters 2 (Section 2.2.6 about super resolution approaches) and 6 (Section 6.2), one of the major issues with low resolution probe images captured under video surveillance context is the contrast in image quality, particularly the blur level, between gallery images and probe ones. Hence, finding a relevant method to make gallery images more blurred before feeding them into feature extraction stage is a worthwhile path to take.
- Further exploring proposed methods for face recognition related tasks, such as facial expressions recognition, gender classification, age estimation, face verification, face anti-spoofing, forensic sketches-mugshots matching, and video-based face recognition.
- There is a haunting question that is: while LBP, LPQ, Gabor wavelets approaches are first successfully introduced for the texture classification problem and then are brought into FR to exhibit their efficiency, is it possible to apply what we have proposed, ELBP(h+v), EPOEM, PLPQMC and LPOG, back into such kind of task? This will be great if the answer is yes.

Bibliography

- [1] A. Pentland, T. Starner, N. Etcoff, A. Masoiu, O. Oliyide, and M. Turk. Experiments with eigenfaces. In *Looking at People Workshop, IJCAI 93*, pages 1–6, August 1993. [68](#)
- [2] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):721–732, 1997. [36](#), [122](#)
- [3] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. *ECCV*, pages 469–481, 2004. [29](#), [40](#), [41](#), [42](#), [44](#), [45](#), [47](#), [53](#), [66](#), [76](#), [86](#), [89](#), [93](#), [100](#), [137](#)
- [4] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2037–2041, 2006. [77](#), [78](#)
- [5] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkilä. Recognition of blurred faces using local phase quantization. In *ICPR*, pages 1–4, 2008. [44](#), [45](#), [46](#), [89](#), [139](#), [153](#)
- [6] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. Fully automatic pose-invariant face recognition via 3D pose normalization. In *ICCV*, pages 937–944, 2011. [85](#), [97](#), [98](#), [100](#), [116](#), [130](#), [131](#), [149](#), [150](#)
- [7] A. Asthana, C. Sanderson, T. Gedeon, and R. Goecke. Learning-based face synthesis for pose-robust recognition from single image. *BMVC09*, 4(5), 2009. [85](#), [97](#), [116](#), [130](#), [150](#)
- [8] E. Bailly-Baillié, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariétoz, J. Matas, K. Messer, V. Popovici, and F. Porée. The BANCA database and evaluation protocol. In *Audio-and Video-Based Biometric Person Authentication*, pages 625–638. Springer, 2003. [52](#)
- [9] M. R. Banham and A. K. Katsaggelos. Digital image restoration. *Signal Processing Magazine, IEEE*, 14(2):24–41, 1997. [89](#)

- [10] C. Baptiste, R. Sami, and C. Liming. 3D-aided face recognition robust to expression and pose variations. In *IEEE CVPR*, pages 1–8, June 2014. [113](#), [114](#), [127](#), [128](#), [144](#)
- [11] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski. Face recognition by independent component analysis. *IEEE Trans. Neural Netw.*, 13(6):1450–1464, November 2002. [37](#)
- [12] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, 1997. [28](#), [29](#), [36](#)
- [13] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of sift features for face authentication. In *CVPR Workshop*, 2006. [52](#)
- [14] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):1042–1052, October 1993. [75](#), [77](#)
- [15] H. Cevikalp, M. Neamtu, and M. Wilkes. Discriminative common vector method with kernels. *IEEE Trans. Neural Netw.*, 17(6):1550–1565, November 2006. [48](#)
- [16] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, and T. Tan. Gabor ordinal measures for face recognition. *IEEE Trans. Inf. Forensics Security*, 9(1):14–26, January 2014. [46](#), [49](#), [115](#), [146](#), [153](#)
- [17] C. H. Chan, J. Kittler, N. Poh, T. Ahonen, and M. Pietikäinen. (Multiscale) Local Phase Quantisation histogram discriminant analysis with score normalisation for robust face recognition. In *ICCV Workshops*, pages 633–640, 2009. [46](#), [96](#), [115](#), [116](#), [129](#), [146](#)
- [18] C. H. Chan, M. A. Tahir, J. Kittler, and M. Pietikäinen. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35:1164–1177, may 2013. [46](#)
- [19] H. F. Chen, P. N. Belhumeur, and D. W. Jacobs. In search of illumination invariants. In *IEEE CVPR*, volume 1, pages 254–261, 2000. 00304. [122](#)
- [20] S. Chen, C. Sanderson, S. Sun, and B. C. Lovell. Representative feature chain for single gallery image face recognition. In *ICPR*, pages 1–4, 2008. [85](#), [97](#), [116](#), [130](#), [149](#), [150](#)
- [21] W. Chen and Y. Gao. Recognizing partially occluded faces from a single sample per class using string-based matching. In *ECCV*, pages 496–509, 2010. [82](#), [95](#), [113](#), [128](#), [144](#)

Bibliography

- [22] J. Choi, W.R. Schwartz, H. Guo, and L.S. Davis. A complementary local feature descriptor for face identification. In *WACV*, pages 121–128, January 2012. [42](#), [96](#), [115](#), [129](#), [146](#)
- [23] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV*, pages 484–498. Springer, 1998. [161](#)
- [24] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. [29](#)
- [25] S. C. Dakin and R. J. Watt. Biological "bar codes" in human faces. *Journal of Vision*, 9(4), April 2009. [77](#)
- [26] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, 2005. [52](#), [122](#)
- [27] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Optical Society of America, Journal, A: Optics and Image Science*, 2:1160–1169, 1985. [29](#), [37](#)
- [28] J. G. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust., Speech, Signal Process.*, 36(7):1169–1179, 1988. [37](#)
- [29] W. Deng, J. Hu, and J. Guo. Gabor-eigen-whiten-cosine: a robust scheme for face recognition. *AMFG*, pages 336–349, 2005. [39](#), [68](#), [96](#), [129](#)
- [30] W. Deng, J. Hu, and J. Guo. Extended SRC: undersampled face recognition via intraclass variant dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(9):1864–1870, September 2012. [51](#), [52](#), [96](#), [115](#), [129](#), [145](#), [146](#)
- [31] L. Ding, X. Ding, and C. Fang. Continuous pose normalization for pose-robust face recognition. *Signal Processing Letters*, 19(11):721–724, November 2012. [85](#), [97](#), [98](#), [100](#), [116](#), [117](#), [130](#), [131](#), [149](#), [150](#)
- [32] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Trans. Signal Process.*, 49(12):3136–3144, December 2001. [50](#), [106](#)
- [33] X. Feng, M. Pietikäinen, and A. Hadid. Facial expression recognition based on local binary patterns. *Pattern Recognition and Image Analysis*, 17(4):592–598, December 2007. [41](#)
- [34] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987. [105](#)

- [35] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7):179–188, 1936. 36, 39
- [36] I. H. Fraser, G. L. Craig, and D. M. Parker. Reaction time measures of feature saliency in schematic faces. *Perception*, 19(5):661–673, 1990. 30
- [37] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995. 42
- [38] D. Gabor. Theory of communication. *J. Inst. Elect. Eng.*, 93:429–457, 1946. 37, 105
- [39] H. Gao, H. Ekenel, and R. Stiefelhagen. Pose normalization for local appearance-based face recognition. *Advances in Biometrics*, pages 32–41, 2009. 85, 97, 116, 130, 150
- [40] V. Goffaux and S.C. Dakin. Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, 1, 2010. 43, 75, 77
- [41] R. Gottumukkal and V. K. Asari. An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, 25(4):429–436, March 2004. 30
- [42] M. Grgic, K. Delac, and S. Grgic. Sface – surveillance cameras face database. *Multimedia Tools and Applications*, 51(3):863–879, October 2009. 15, 16, 17, 35, 56, 59, 60, 64, 85, 86, 98, 99, 117, 118, 131, 132, 134, 140, 150, 151
- [43] R. Gross, I. Matthews, and S. Baker. Fisher light-fields for face recognition across pose and illumination. In L. V. Gool, editor, *Pattern Recognition*, number 2449 in LNCS, pages 481–489. Springer Berlin Heidelberg, January 2002. 36
- [44] A. Hadid, M. Pietikäinen, and T. Ahonen. A discriminative feature space for detecting and recognizing faces. In *CVPR*, volume 2, pages 797–804, 2004. 41
- [45] N. D. Haig. Exploring recognition with interchanged facial features. *Perception*, 15(3):235–247, 1986. 75, 81
- [46] M. Heikkilä, M. Pietikäinen, and C. Schmid. Description of interest regions with center-symmetric local binary patterns. *Computer Vision, Graphics and Image Processing*, pages 58–69, 2006. 42
- [47] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding*, 91(1-2):6–21, 2003. 28

- [48] H. T. Ho and R. Chellappa. Pose-invariant face recognition using markov random fields. *IEEE Trans. Image Process.*, 22(4):1573–1584, 2013. [85](#), [97](#), [100](#), [116](#), [130](#), [131](#), [149](#), [150](#)
- [49] I. P. Howard and B. J. Rogers. *Binocular Vision and Stereopsis*. Oxford University Press, February 1996. [122](#), [134](#), [137](#), [154](#)
- [50] G. B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *IEEE ICCV*, pages 1–8, 2007. [161](#)
- [51] G. B. Huang, M. A. Mattar, H. Lee, and E. Learned-Miller. Learning to align from scratch. In *NIPS*, pages 773–781, 2012. [161](#)
- [52] S. U. Hussain, T. Napoleon, and F. Jurie. Face recognition using local quantized patterns. In *BMVC*, September 2012. [46](#), [49](#), [96](#), [100](#), [115](#), [129](#), [130](#), [146](#), [153](#)
- [53] A. K. Jain, B. Klare, and U. Park. Face recognition: Some challenges in forensics. In *FG*, pages 726–733, 2011. [26](#)
- [54] (A.P.) James. Pixel-level decisions based robust face image recognition. In Milos Oravec, editor, *Face Recognition*. InTech, April 2010. [82](#), [95](#), [113](#), [114](#), [128](#), [144](#)
- [55] H. Jin, Q. Liu, H. Lu, and X. Tong. Face detection using improved LBP under bayesian framework. In *Proc. Image and Graphics*, pages 306–309, 2004. [42](#)
- [56] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97, March 1981. [122](#), [134](#), [137](#), [154](#)
- [57] P. Karthigayani and S. Sridhar. A novel approach for face recognition and age estimation using local binary pattern, discriminative approach using two layered back propagation network. In *Proc. Trendz in Information Sciences and Computing (TISC)*, pages 11–16, 2011. [41](#)
- [58] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(6):1127–1133, 2010. [53](#)
- [59] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(1):103–108, 1990. [35](#)
- [60] P. Kovesi. What Are Log-Gabor Filters and Why Are They Good?, 2014. Available online at <http://www.csse.uwa.edu.au/~pk/research/matlabfns/PhaseCongruency/Docs/convexpl.html>. Last viewed in May, 2014. [105](#)

- [61] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical society of America*, 61(1):1–11, 1971. [122](#), [134](#), [137](#), [154](#)
- [62] T.S. Lee. Image representation using 2D gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–971, 1996. [37](#)
- [63] Z. Lei, T. Ahonen, M. Pietikainen, and S. Z. Li. Local frequency descriptor for low-resolution face recognition. In *FG*, pages 161–166. IEEE, 2011. [45](#)
- [64] A. Li, S. Shan, X. Chen, and W. Gao. Maximizing intra-individual correlations for face recognition across pose differences. In *CVPR*, pages 605–611, 2009. [85](#), [97](#), [116](#), [130](#), [150](#)
- [65] S. Z. Li and A. K. Jain, editors. *Handbook of Face Recognition*. Springer, 2nd edition. edition, August 2011. [25](#), [26](#), [28](#), [105](#), [122](#)
- [66] X.-X. Li, D.-Q. Dai, X.-F. Zhang, and C.-X. Ren. Structured sparse error coding for face recognition with occlusion. *IEEE Trans. Image Process.*, 22(5):1889–1900, May 2013. [52](#), [145](#)
- [67] Z. Li, J. Imai, and M. Kaneko. Robust face recognition using block-based bag of words. In *ICPR*, pages 1285–1288, August 2010. [113](#), [127](#), [128](#), [144](#)
- [68] H. C. Lian and B. L. Lu. Multi-view gender classification using local binary patterns and support vector machines. *Advances in Neural Networks*, pages 202–209, 2006. [41](#)
- [69] S. Liao and A. Chung. Face recognition with salient local gradient orientation binary patterns. In *ICIP*, pages 3317–3320, 2009. [123](#)
- [70] S. Liao and A.C.S. Chung. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In *ACCV*, pages 672–679, Berlin, Heidelberg, 2007. Springer-Verlag. [42](#), [75](#), [83](#)
- [71] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Li. Learning multi-scale block local binary patterns for face recognition. *Advances in Biometrics*, pages 828–837, 2007. [42](#)
- [72] C. Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):572–581, 2004. [39](#)
- [73] C. Liu and H. Wechsler. Enhanced fisher linear discriminant models for face recognition. In *ICPR*, volume 2, pages 1368–1372, 1998. [39](#)

- [74] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.*, 11(4):467–476, 2002. [39](#)
- [75] J. Liu, S. Chen, Z. H. Zhou, and X. Tan. Single image subspace for face recognition. *AMFG*, pages 205–219, 2007. [82](#), [95](#), [113](#), [128](#), [144](#)
- [76] N. Liu, J. Lai, and H. Qiu. Robust face recognition by sparse local features from a single image under occlusion. In *ICIG*, pages 500–505, August 2011. [82](#), [113](#), [114](#), [127](#), [128](#), [144](#)
- [77] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004. [52](#), [122](#)
- [78] J. Lu, Y. P. Tan, and G. Wang. Discriminative multi-manifold analysis for face recognition from a single training sample per person. In *ICCV*, pages 1943–1950, 2011. [82](#), [95](#), [96](#), [113](#), [128](#), [129](#), [144](#)
- [79] J. Maatta, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis. In *IJCB*, pages 1–7, 2011. [41](#)
- [80] A. M. Martinez and R. Benavente. The AR face database. *CVC Technical Report*, 24, 1998. [48](#), [52](#), [56](#), [57](#), [134](#), [140](#)
- [81] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004. [161](#)
- [82] D. Maturana, D. Mery, and A. Soto. Learning discriminative local binary patterns for face recognition. In *FG*, pages 470–475, 2011. [43](#), [95](#), [96](#), [129](#)
- [83] E. Meyers and L. Wolf. Using biologically inspired features for face processing. *Int Journal of Computer Vision*, 76:93–104, 2008. [52](#)
- [84] R. Min, A. Hadid, and J.L. Dugelay. Improving the recognition of faces occluded by facial accessories. In *AFGR*, 03 2011. [82](#), [95](#), [113](#), [128](#), [144](#)
- [85] H. Moon and P. J. Phillips. Computational and performance aspects of PCA-based face-recognition algorithms. *Perception*, 30(3):303–321, 2001. [68](#)
- [86] E. A. Mostafa and A. A. Farag. Dynamic weighting of facial features for automatic pose-invariant face recognition. In *9th Conference on Computer and Robot Vision (CRV)*, pages 411–416, 2012. [85](#), [97](#), [100](#), [116](#), [130](#), [149](#), [150](#)
- [87] K. Naka and W. AH Rushton. S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):587–599, 1966. [63](#)

- [88] H. Nguyen, L. Bai, and L. Shen. Local gabor binary pattern whitened pca: A novel approach for face recognition from single image per person. *Advances in Biometrics*, pages 269–278, 2009. [83](#), [96](#), [129](#)
- [89] H. T. Nguyen and A. Caplier. Elliptical local binary patterns for face recognition. In *Computer Vision - ACCV 2012 Workshops*, number 7728 in LNCS, pages 85–96. Springer Berlin Heidelberg, January 2013. [60](#), [96](#), [98](#), [99](#), [100](#), [117](#), [129](#), [131](#), [132](#), [134](#), [135](#), [139](#), [150](#), [151](#)
- [90] H.J. Oh, K.M. Lee, and S.U. Lee. Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. *Image and Vision Computing*, 26(11):1515–1523, November 2008. [82](#), [95](#), [113](#), [128](#), [144](#)
- [91] T. Ojala, M. Pietikäinen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002. [40](#)
- [92] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. *Image and Signal Processing*, pages 236–243, 2008. [44](#), [90](#), [91](#), [117](#), [139](#)
- [93] Y. Pang, L. Zhang, M. Li, Z. Liu, and W. Ma. A novel gabor-LDA based face recognition method. *Advances in Multimedia Information Processing-PCM*, pages 352–358, 2005. [39](#)
- [94] C.A. Perez, L.A. Cament, and L.E. Castillo. Methodological improvement on local gabor face recognition based on feature selection and enhanced borda count. *Pattern Recognition*, 44(4):951–963, 2011. [85](#), [96](#), [97](#), [115](#), [116](#), [129](#), [130](#), [146](#), [150](#)
- [95] V. Perlibakas. Distance measures for PCA-based face recognition. *Pattern Recognition Letters*, 25(6):711–724, 2004. [68](#)
- [96] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, 2000. [15](#), [16](#), [17](#), [39](#), [48](#), [56](#), [58](#), [60](#), [83](#), [96](#), [115](#), [129](#), [134](#), [140](#), [145](#), [146](#), [150](#)
- [97] J. Ren, X. Jiang, and J. Yuan. Relaxed local ternary pattern for face recognition. In *ICIP*, pages 3680–3684, Melbourne, VIC, 2013. IEEE. [42](#)
- [98] C. Rosenberger and L. Brun. Similarity-based matching for face authentication. In *ICPR*, pages 1–4, 2008. [52](#)
- [99] J. Sadr, I. Jarudi, and P. Sinha. The role of eyebrows in face recognition. *Perception*, 32(3):285–293, 2003. [30](#)

Bibliography

- [100] M. Saquib, Olaf Hellwich, and Zahid Riaz. Feature extraction and representation for face recognition. In Milos Oravec, editor, *Face Recognition*. InTech, April 2010. [30](#), [85](#), [97](#), [116](#), [130](#), [149](#), [150](#)
- [101] B. Schölkopf, A. Smola, and K.R. Müller. Kernel principal component analysis. *Artificial Neural Networks-ICANN'97*, pages 583–588, 1997. [39](#)
- [102] W. Schwartz, H. Guo, and L. Davis. A robust and scalable approach to face identification. *ECCV*, pages 476–489, 2010. [42](#)
- [103] Á Serrano, I. M. de Diego, C. Conde, and E. Cabello. Recent advances in face biometrics with gabor wavelets: A review. *Pattern Recognition Letters*, 31(5):372–381, 2010. [37](#)
- [104] S. Shan, W. Zhang, Y. Su, X. Chen, and W. Gao. Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition. In *ICPR*, volume 4, pages 606–609, 2006. [46](#), [47](#), [48](#), [49](#), [96](#), [115](#), [129](#), [146](#), [153](#)
- [105] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, and D. W. Jacobs. Robust pose invariant face recognition using coupled latent space discriminant analysis. *Computer Vision and Image Understanding*, 116(11):1095–1110, November 2012. [85](#), [97](#), [116](#), [130](#), [150](#)
- [106] J. W. Shepherd, G. M. Davies, and A. W. Ellis. Studies of cue saliency. In G. Davies, H. Ellis, and J. Shepherd, editors, *Perceiving and Remembering Faces*, pages 105–131. New York: Academic Press, 1981. [43](#), [75](#), [77](#), [81](#)
- [107] X. Shufu, S. Shiguang, C. Xilin, and C. Jie. Fusing local patterns of gabor magnitude and phase for face recognition. *IEEE Trans. Image Process.*, 19(5):1349–1361, May 2010. [46](#), [47](#), [48](#), [49](#), [50](#), [96](#), [100](#), [115](#), [129](#), [130](#), [146](#), [153](#)
- [108] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, November 2006. [75](#)
- [109] D. L. Swets and J. J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):831–836, 1996. [68](#)
- [110] M. A. Tahir, C. H. Chan, J. Kittler, and A. Bouridane. Face recognition using multi-scale local phase quantisation and linear regression classifier. In *ICIP*, pages 765–768, 2011. [46](#)
- [111] Z. Taiping, Y. T. Yuan, F. Bin, S. Zhaowei, and L. Xiaoyu. Face recognition under varying illumination using gradientfaces. *IEEE Trans. Image Process.*, 18(11):2599–2606, 2009. [122](#), [139](#)

- [112] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *AMFG*, volume 4778, pages 168–182. Springer Berlin Heidelberg, January 2007. [42](#), [65](#)
- [113] X. Tan and B. Triggs. Fusing gabor and LBP feature sets for Kernel-Based face recognition. In *AMFG*, volume 4778, pages 235–249. Springer Berlin Heidelberg, 2007. [47](#), [48](#), [83](#), [95](#), [96](#), [115](#), [129](#), [146](#)
- [114] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, January 1991. [28](#), [29](#), [35](#), [36](#), [53](#), [68](#), [69](#)
- [115] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. Subspace learning from image gradient orientations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(12):2454–2466, December 2012. [123](#), [129](#), [139](#)
- [116] N. S. Vu and A. Caplier. Illumination-robust face recognition using retina modeling. In *ICIP*, pages 3289–3292, 2009. [58](#), [62](#), [65](#), [66](#), [67](#), [71](#)
- [117] N. S. Vu and A. Caplier. Face recognition with patterns of oriented edge magnitudes. *ECCV*, pages 313–326, 2010. [43](#), [123](#), [124](#), [139](#)
- [118] N. S. Vu and A. Caplier. Mining patterns of orientations and magnitudes for face recognition. In *IJCB*, pages 1–8, October 2011. [43](#), [95](#), [96](#), [115](#), [129](#), [139](#), [146](#)
- [119] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. In *ICIP*, volume 1, pages 129–132, October 1997. [11](#), [37](#), [39](#)
- [120] L Wolf, T Hassner, and Y Taigman. Descriptor based methods in the wild. In *Faces in Real-Life Images Workshop in ECCV*, October 2008. [68](#)
- [121] Y. Wong, M. T. Harandi, C. Sanderson, and B. C. Lovell. On robust biometric identity verification via sparse encoding of faces: Holistic vs local approaches. In *IJCNN*, pages 1–8, 2012. [85](#), [97](#), [116](#), [130](#), [149](#), [150](#)
- [122] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009. [51](#), [145](#)
- [123] K. Yan, Y. Chen, and D. Zhang. Gabor surface feature for face recognition. In *1st ACPR*, pages 288 –292, November 2011. [47](#), [49](#), [96](#), [100](#), [115](#), [129](#), [130](#), [146](#), [153](#)
- [124] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Process.*, 19(11):2861–2873, 2010. [53](#)

Bibliography

- [125] J. Yang, D. Zhang, A.F. Frangi, and J. Yang. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):131–137, 2004. [37](#)
- [126] J. Yang, D. Zhang, X. Yong, and J.-Y. Yang. Two-dimensional discriminant transform for face recognition. *Pattern Recognition*, 38(7):1125–1129, July 2005. [37](#)
- [127] M. Yang, L. Zhang, S. C.-K. Shiu, and D. Zhang. Robust kernel representation with statistical local features for face recognition. *IEEE Trans. Neural Netw.*, pages 1–1, 2013. [46](#), [49](#), [85](#), [97](#), [115](#), [116](#), [145](#), [146](#), [147](#), [149](#), [150](#), [153](#)
- [128] M. Yang, L. Zhang, S.C.-K. Shiu, and D. Zhang. Monogenic binary coding: An efficient local feature extraction approach to face recognition. *IEEE Trans. Inf. Forensics Security*, 7(6):1738–1751, December 2012. [50](#), [87](#), [88](#), [96](#), [100](#), [104](#), [109](#), [115](#), [116](#), [119](#), [146](#), [153](#)
- [129] M. Yang, L. Zhang, J. Yang, and D. Zhang. Robust sparse coding for face recognition. In *CVPR*, pages 625–632, 2011. [82](#), [95](#), [113](#), [127](#), [128](#), [144](#)
- [130] M. Yang, L. Zhang, L. Zhang, and D. Zhang. Monogenic binary pattern (mbp): A novel feature extraction and representation model for face recognition. In *ICPR*, pages 2680–2683, 2010. [50](#), [95](#), [96](#), [104](#), [109](#), [115](#), [116](#), [129](#), [146](#)
- [131] D. Yi, Z. Lei, and S.Z. Li. Towards pose robust face recognition. In *IEEE CVPR*, pages 3539–3545, June 2013. 00001. [116](#), [117](#), [149](#), [150](#)
- [132] A. W. Young, D. C. Hay, K. H. McWeeny, B. M. Flude, and A. W. Ellis. Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14(6):737–746, 1985. 00255. [30](#)
- [133] B. Zhang, S. Shan, X. Chen, and W. Gao. Histogram of gabor phase patterns (HGPP): a novel object representation approach for face recognition. *IEEE Trans. Image Process.*, 16(1):57–68, January 2007. [47](#), [48](#), [50](#), [83](#), [96](#), [100](#), [129](#)
- [134] G. Zhang, X. Huang, S. Li, Y. Wang, and X. Wu. Boosting local binary pattern (lbp)-based face recognition. In *Advances in Biometric Person Authentication*, pages 179–186. 2005. [42](#), [43](#)
- [135] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. In *ICCV*, volume 1, pages 786–791, 2005. [47](#), [50](#), [82](#), [83](#), [88](#), [95](#), [96](#), [100](#), [113](#), [128](#), [129](#), [144](#)
- [136] Y.-J. Zhang. *Advances in face image analysis techniques and technologies*. Medical Information Science Reference, Hershey, PA, 2011. [62](#)

- [137] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003. [26](#), [27](#), [28](#), [30](#), [41](#), [105](#), [122](#)
- [138] W. Zhen and Y. Zilu. Facial expression recognition based on local phase quantization and sparse representation. In *ICNC*, pages 222–225, 2012. [52](#)
- [139] J. Zou, Q. Ji, and G. Nagy. A comparative study of local matching approach for face recognition. *IEEE Trans. Image Process.*, 16(10):2617–2628, October 2007. [30](#)
- [140] W.W.W. Zou and P.C. Yuen. Very low resolution face recognition problem. In *BTAS*, pages 1–6, 2010. [53](#), [60](#), [86](#), [98](#), [117](#), [131](#), [150](#)
- [141] W. Zuo, K. Wang, and H. Zhang. Subspace methods for face recognition: Singularity, regularization, and robustness. In *State of the Art in Face Recognition*. I-Tech Education and Publishing, January 2009. [82](#), [95](#), [113](#), [128](#), [144](#)

List of Publications

The contributions presented in this manuscript were published in the following articles:

1. International Conferences

[1] H.-T. Nguyen, N.-S. Vu, and A. Caplier. “How far we can improve micro features based face recognition?”, 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA) 2012, October, Istanbul, Turkey.

[2] N.-S. Vu, H.-T. Nguyen, and A. Caplier. “Multiple patterns of gradient magnitudes for face recognition”, International Conference on Image Processing (ICIP) 2012, October, Orlando, USA.

[3] H.-T. Nguyen and A. Caplier. “Elliptical Local Binary Patterns for Face recognition”, 1st LBP workshop at 11th Asian Conference on Computer Vision (ACCV) 2012, November, Daejeon, Korea.

[4] H.-T. Nguyen and A. Caplier. “Patch based Local phase quantization of Monogenic components for Face recognition”, International Conference on Image Processing (ICIP) 2014, October, Paris, France. *To appear.*

2. Articles under review

[1] H.-T. Nguyen and A. Caplier. “Local Patterns of Gradients (LPOG) for Face recognition”, submitted to IEEE Transactions on Information Forensics and Security.

Contributions à l'extraction de caractéristiques pour la reconnaissance de visages

Résumé – La tâche la plus délicate d'un système de reconnaissance faciale est la phase d'extraction de caractéristiques significatives et discriminantes. Dans le cadre de cette thèse, nous nous sommes focalisés sur cette tâche avec comme objectif l'élaboration d'une représentation de visage robuste aux variations majeures suivantes: variations d'éclairage, de pose, de temps, images de qualité différentes (vidéosurveillance). Par ailleurs, nous avons travaillé également dans une optique de traitement temps réel.

Tout d'abord, en tenant compte des caractéristiques d'orientation des traits principaux du visages (yeux, bouche), une nouvelle variante nommée ELBP de célèbre descripteur LBP a été proposée. Elle s'appuie sur les informations de micro-texture contenues dans une ellipse horizontale. Ensuite, le descripteur EPOEM est construit afin de tenir compte des informations d'orientation des contours. Puis un descripteur nommée PLPQMC qui intègre des informations obtenues par filtrage monogénique dans le descripteur LPQ est proposé. Enfin le descripteur LPOG intégrant des informations de gradient est présenté. Chacun des descripteurs proposés est testé sur les 3 bases d'images AR, FERET et SCface. Il en résulte que les descripteurs PLPQMC et LPOG sont les plus performants et conduisent à des taux de reconnaissance comparables voire supérieur à ceux des meilleurs méthodes de l'état de l'art.

Contributions to facial feature extraction for Face recognition

Abstract – Centered around feature extraction, the core task of any Face recognition system, our objective is devising a robust facial representation against major challenges, such as variations of illumination, pose and time-lapse and low resolution probe images, to name a few. Besides, fast processing speed is another crucial criterion. Towards these ends, several methods have been proposed through out this thesis.

Firstly, based on the orientation characteristics of the facial information and important features, like the eyes and mouth, a novel variant of LBP, referred as ELBP, is designed for encoding micro patterns with the usage of an horizontal ellipse sample. Secondly, ELBP is exploited to extract local features from oriented edge magnitudes images. By this, the Elliptical Patterns of Oriented Edge Magnitudes (EPOEM) description is built. Thirdly, we propose a novel feature extraction method so called Patch based Local Phase Quantization of Monogenic components (PLPQMC). Lastly, a robust facial representation namely Local Patterns of Gradients (LPOG) is developed to capture meaningful features directly from gradient images. Chiefs among these methods are PLPQMC and LPOG as they are per se *illumination invariant* and *blur tolerant*. Impressively, our methods, while offering comparable or almost higher results than that of existing systems, have low computational cost and are thus feasible to deploy in real life applications.

Key words: Robust face recognition, real-time, illumination, facial expressions, occlusion, time-lapse and pose variations, video surveillance, local descriptors, local features, ELBP, Patch based LPQ, Monogenic filter based, EPOEM, LPOG, gradient images based features.