Asymmetric Region Local Binary Patterns for Face Image Analysis

Shrinivasa Naika C. L.

Asymmetric Region Local Binary Patterns for Face Image Analysis

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

by

Shrinivasa Naika C. L.

Under the supervision of

Dr. Pradip K. Das

Department of Computer Science and Engineering

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

Guwahati 781039, India

March 2014

Dedicated to

My Parents

Whose blessings and love made my path to success

Declaration

I certify that

a. The work contained in this thesis is original and has been done by myself under the general supervision of my supervisor.

b. The work has not been submitted to any other Institute for any degree or diploma.

c. Whenever I have used materials (data, theoretical analysis, results, etc.) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

d. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Place: IIT Guwahati
Date:

Shrinivasa Naika C. L.
Research Scholar
Department of Computer Science and Engineering,
Indian Institute of Technology Guwahati,
Guwahati, Assam, INDIA 781039

Certificate

This is to certify that the thesis entitled "**Asymmetric Region Local Binary Patterns for Face Image Analysis**" being submitted by **Shrinivasa Naika C. L.** to the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, is a record of bona fide research work under my supervision and is worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute.

Place: IIT Guwahati
Date:

Dr. Pradip K. Das
Department of Computer Science and Engineering,
Indian Institute of Technology Guwahati,
Guwahati, Assam, INDIA 781039

DECLARATION ON PLAGIARISM

I, Shrinivasa Naika C. L., (Roll No. 09610110), understand that plagiarism is defined as any one or the combination of the following: Uncredited verbatim copying of individual sentences, paragraphs or illustrations (such as graphs, diagram, etc.) from any source, published or unpublished, including the Internet.

Uncredited improper paraphrasing of pages or paragraphs (changing a few words or phrases, or rearranging, the original sentence order). Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what. (Source: IEEE, The Institute, Dec. 2004)

I have to the best of my abilities ensured that all the concepts, ideas, text, expressions, graphs, diagrams, etc., which are not the outcome of my research have been credited to those who own the same. Elaborate sentences used verbatim from published work have been clearly identified and quoted. I also affirm that no part knowledge and understanding and take compete responsibility if any complaint arises.

I am fully aware that my thesis supervisor(s) are not be in a position to check for any possible instance of plagiarism within this submitted thesis.

Place: IIT Guwahati

Date:

Shrinivasa Naika C. L.

Acknowledgements

"Don't look back when you are moving towards success! But don't forget to look back after reaching your goal!"

I deem it a privilege to be a part of teaching fraternity of an esteem organization *University B.D.T. College of Engineering, Davanagere* which was a constituent college of *Kuvempu University, Shimoga* in the early stage of my venture into research. Later, it became a constituent college of *VTU Belgaum*. I vow my deep sense of gratitude to both the establishments for blessing me with an opportunity to undertake and successfully complete one of my life-time dreams of research.

I am equally grateful to Indian Institute of Technology Guwahati (IITG). This esteemed organization has been a highly resourceful workshop for my ambitious research. It is my pleasure to thank MHRD, Govt. of India for providing financial support during my PhD under the QIP Programme.

I submit wholehearted reverence to my highly respected guide Dr. Pradip K. Das for his immeasurable blessings in the form of continuous guidance and support. My heartfelt reverence is also due to honorable Prof. Shivashankar B. Nair for his invaluable guidance and encouragement in the initial stages of my research. His orientation on ethics and basics of research provided a strong foundation to my further ventures in research. I am indebted to these two respectful scholars for guiding me in accomplishing my research.

I am greatly indebted to highly insightful experts of the Doctoral Committee Dr. Pinaki Mitra and Dr. V. Vijaya Saradhi. Their timely suggestions provided me the clarity and confidence in the successful completion of my research. Their constructive feedback helped me to bring comprehensiveness to my work.

I would love to express my gratefulness from the bottom of my heart to admirable Prof. Chitralekha Mahanta, Dept. of EEE, IIT Guwahati, for her exemplary patience in reading and providing critical reviews of my reports

which has contributed to the quality of this research work.
I go wordless when it comes to outline my gratitude to reverent Dr. G. Sajith, Dr. P. Bhaduri and Prof. Shivashankar B. Nair, as HODs in the Dept. of Computer Science and Engineering during my research work. It is my honour to be grateful to their invaluable support all along the path of my research.
I feel immensely proud to acknowledge the showers of facilities and kindhearted support by Dr. Suresh Chandra Mohan, Dr. Abdul Buden and Dr. Shivaprasad B. Dandagi, respected Principals of B. D. T College of Engineering, Davanagere. I am deeply grateful for Computer Lab and all necessary administrative backup extended by them.
I am equally indebted to Dr. Shreedhara K. S, HOD, Dept. of Computer Science, B. D. T College of Engineering, Davanagere. He is the one who drew my attention towards research and helped me with all related details to take up research under the QIP scheme. Besides he initiated me into this research venture and has been helping me all along in terms of both academic and administrative aspects.
I would be failing in my duty if I do not acknowledge the services and support of the Staff of Dept. of Computer Science and Engineering, IITG for providing access to valuable resources and extending all necessary support for the successful completion of my research work.
I go more humble while I am to thank my friends Guruji, Keshav, Ramesha, Pravati, Mamata, Shashishekhar Jha, Shirshendu Das, Mayank Agarwal and Nilkant Sahu for their volunteered unselfish extension of helping hands at all tough times during this research. I am greatly indebted to all such great friends. Besides I can never forget to share a word of thanks to all my friends and ever well-wishers. They have been a source of inspiration and helped me whenever I needed their services and support.
My special thanks are also due to Mr. Girish G N, M. Tech. Student of B. D. T College of Engineering, Davanagere for his assistance in conducting and consolidating the results of a number of experiments related to the study.
It is needless to say that this sort of research venture would never be possible without the blessings and cooperation of all my family members: My loving parents Sri. Laxman Naik and Gangi Bai, my in-laws Sri. Kubendra Naik D. and Gowra Bai, Sri. Dr. Muralidhara P .D and Jayashree K, Dr. Vinary Kumar K, my beloved wife Roopashree K and children Ruchitha S & Monisha

S. They have always been on my side to shoulder all my responsibilities with a smile. I always remember their unconditional support which made me work with more zeal and dedication.

At last, I would love to thank everyone who is directly or indirectly responsible for the successful completion of my research work. Thanks to all for making my effort a wonderful experience. . .

Place: IIT Guwahati

Date: **Shrinivasa Naika C. L.**

Abstract

This thesis explores feature extraction techniques based on Local Binary Patterns (LBP) for Automatic Face Image Analysis (AFIA). Two tasks of AFIA were considered: Face Recognition (FR) and Facial Expression Recognition (FER) for which a face representation is derived which could be employed in both FR and FER systems. These tasks share many common components and challenges and they need to be addressed for the better performance of the system. Moreover, there exist limited number of LBP based operators which can be deployed in both FR and FER systems. Existing LBP based operators are not well studied for higher scales though scalable operators exist.

This thesis proposes two Asymmetric Region LBP (ARLBP) and Extended LBP (EARLBP) operators, Modified Convolution (MC) operation and a heuristic localization method. These operators along with other proposed techniques were experimented considering different settings viz. with or without localization and registration errors, person dependent or independent, operator scales, number of grids (hence the size) and number of available LBP based codes. The FR system was configured in verification mode using Eigen face approach derived on these LBP based histogram feature vectors. The FER system was configured for multi-class facial expression classification mode using Support Vector Machine (SVM). In FR the proposed techniques were used in person dependent, with localization and registration error settings using the MC method.

Experiments were carried out to illustrate the importance of increase in the number of LBP codes along with different scales and grid sizes and found that the MC increases the available LBP codes enabling the chance of evaluating the operator properties at higher scales. The FR system performance largely depends on the available LBP codes per image rather than scale of the operator and the number of grids. The proposed ARLBP operator is better when compared to EARLBP as it is sensitive to averaging effect of boundary pixels of the operator's sub-regions at larger scales. The performance on several benchmark databases suggests that these operators could be used in FR. The FER system was derived in person dependent and independent, with or without registration and with localization error settings using the MC method. Face image size and number of grids (hence the size) are kept constant to explore the effect of the scale of the operator with different conditions on the system recognition rate.

Experiments on facial expression databases reveals that maximum recognition rate would be obtained for the scale in which *width* is larger than *height* of the operator. Both proposed operators are sensitive to registration errors. However, these operators could be applied in automatic FERS. Further observing the performance of these operators it could be concluded that both operators along with the MC would be useful in deriving AFIA which integrates FR and FER systems to negate the variations of individual systems and enhance the overall integrated system performance.

Contents

List of Tables **xvii**

List of Figures **xix**

Abbreviation **xxi**

1 Introduction **1**

1.1 Context 1

1.2 Challenges 3

1.2.1 Illumination 4

1.2.2 Pose 4

1.2.3 Facial Expressions 5

1.3 Automatic Face Image Analysis 6

1.4 Face Recognition System (FRS) 6

1.4.1 Components of FRS 7

1.5 Principal Component Analysis (PCA) 11

1.5.1 Similarity Measures 13

1.5.2 FR Performance Measures 15

1.6 Facial Expression Recognition System (FERS) 16

1.6.1 Types of Facial Expression 16

1.6.2 Components of FERS 19

1.7 Intersections of the Framework 27

1.8 Contributions 28

1.9 Overview of the Thesis 29

1.10 Chapter Summary 29

2 Literature Survey **31**

2.1 Face Recognition . . . 31

2.1.1 Survey on LBP based Operators in FR . . . 32

2.2 Facial Expression Recognition . . . 34

2.2.1 Survey on LBP based Operators in FER . . . 36

2.3 Intersection of Feature Extraction methods . . . 41

2.4 Chapter Summary . . . 42

3 Feature Extraction and Representations **43**

3.1 Introduction . . . 43

3.2 Local Binary Pattern based Operators . . . 44

3.2.1 Convolution Method . . . 46

3.2.2 Traditional Face Representation . . . 47

3.3 Proposed Face Representation . . . 50

3.3.1 Modified Convolution (MC) Method . . . 50

3.3.2 Asymmetric Region LBP (ARLBP) . . . 52

3.3.3 Extended Asymmetric Region-LBP (EARLBP) . . . 53

3.4 Chapter Summary . . . 54

4 Face Verification **55**

4.1 Introduction . . . 55

4.2 Why Face Verification? . . . 56

4.3 Description of the Databases used . . . 56

4.4 Experimental Process . . . 59

4.4.1 Experimental Design . . . 60

4.5 Experiments and Results . . . 61

4.5.1 Case 1: Effect of Modified Convolution (MC) . . . 61

4.5.2 Case 2: Face Verification. . . . 63

4.6 Chapter Summary . . . 70

5 Facial Expression Recognition **71**

5.1 Introduction . . . 71

5.1.1 Localization and Registration Errors . . . 72

5.1.2 Face Localization Method . . . 74

5.1.3 Description of the Databases used . . . 76

5.2 Facial Expression Recognition using Asymmetric Region based LBPs . . . 78

5.2.1 Experimental Process 78
5.2.2 Experimental Design 81
5.3 Facial Expression Recognition using ARLBP 81
5.3.1 Case 1: Person dependent with registration errors 82
5.3.2 Case 2: Person independent with registration errors 84
5.3.3 Case 3: Person independent without registration errors 87
5.4 Facial Expression Recognition using EARLBP 88
5.4.1 Case 2: Person independent with registration errors 89
5.4.2 Case 3: Person independent without registration errors 91
5.5 Chapter Summary 94

6 Conclusions and Scope for Future Work **95**
6.1 Summary of Contributions 95
6.2 Scope for Future Work 97

List of Tables

2.1 List of studies made in facial expression recognition using LBP based operators. . . . 38

3.1 Comparison of LBP based operators. . . . 49

4.1 Maximum Verification Rate for different Operators. . . . 69

5.1 Facial expression recognition rates (%) for SVM 7-class Facial expression on JAFFE database with RBF kernel for different ARLBP operator sizes in person dependent condition. . . . 82
5.2 Facial expression recognition rates (%) for SVM 7-class Facial expression on FGNET database with RBF kernel for different ARLBP operator sizes in person dependent condition. . . . 83
5.3 Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on JAFFE database and ARLBP in person independent settings. . . 85
5.4 Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on JAFFE database and ARLBP in person independent settings. . . 85
5.5 Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and ARLBP in person independent settings. . . . 86
5.6 Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and ARLBP in person independent settings. . . . 86
5.7 Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and ARLBP in person independent and without registration errors settings. . . . 88
5.8 Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and ARLBP in person independent and without registration errors settings. . . . 89

5.9 Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on JAFFE database and EARLBP in person independent settings. . 90
5.10 Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on JAFFE database and EARLBP in person independent settings. . 90
5.11 Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and EARLBP in person independent settings. . . 91
5.12 Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and EARLBP in person independent settings. . . 92
5.13 Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and ARLBP in person independent and without registration errors settings. 93
5.14 Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and EARLBP in person independent and without registration errors settings. 93
5.15 Maximum Recognition Rates (%) of ARLBP and EARLBP with and without registration errors for image size 64×64. 94

List of Figures

1.1 Face under fixed view with varying illumination (ambient) (adapted from [1]) 4
1.2 Faces with constant illumination but varying poses (adapted from [1]) . . . 5
1.3 Faces from JAFFE [2] with fixed pose and variations due to facial expressions. 5
1.4 Components of generic face recognition system. 7
1.5 Valance-Arousal dimensional space. 17
1.6 Common Action Units [3]. 18
1.7 Components of Facial Expression Recognition System. 19
1.8 Classifier Training and Testing phases. 22
1.9 Image database is divided into training and testing sets to derive person independent and person dependent classifier. 27
1.10 Intersection between challenges and components of FR and FER systems. . 28

2.1 Taxonomy of local descriptors [4] used in face recognition. 33
2.2 Intersection of LBP based operators for face image analysis. 42

3.1 Example of 3×3 sized basic LBP operator [5] 45
3.2 Example of ELBP operator [6] The circular neighborhood (P, R) of the size (8, 1), (16, 2) and (24, 3). 45
3.3 Example of MBLBP [7] with 3×3 sub-regions of the size 3×2. 46
3.4 Convolution method using LBP based operators. 47
3.5 LBP based face representation [6]. 48
3.6 Modified convolution method [8]: (a) 5×5 ARLBP thresholding when region of the operator is out of bound with the image boundary. (b) 5×5 ARLBP thresholding when region of the operator partially overlaps with the image boundary. 51
3.7 Example of 5×5 sized, Asymmetric Region-LBP [8] 52
3.8 Example of 5×5 sized EARLBP [9] 52

4.1 Sample images from the JAFFE database. 57
4.2 Examples of face images from the ORL database. 58
4.3 Examples of face images from the INDIAN database. 58
4.4 Face recognition process. 60
4.5 Number of available codes for different operators with varying (a) scales of the operator (b) number of grids 62
4.6 VR for 1% FAR with MBLBP for INDIAN database. 65
4.7 VR for 1% FAR with MBLBP for JAFFE database. 65
4.8 VR for 1% FAR with MBLBP for ORL database. 65
4.9 VR for 1% FAR with ARLBP for INDIAN database. 66
4.10 VR for 1% FAR with ARLBP for JAFFE database. 66
4.11 VR for 1% FAR with ARLBP for ORL database. 66
4.12 VR for 1% FAR with EARLBP for INDIAN database. 67
4.13 VR for 1% FAR with EARLBP for JAFFE database. 67
4.14 VR for 1% FAR with EARLBP for ORL database. 67

5.1 Localization Error: (a) Face Localization using Viola & Jones method for 3 iterations, (b) Different Face representations. 73
5.2 Registration Error: (a) Frontal face with inconsistency in pose. (b) After correction of pose [10] 74
5.3 (a) Facial features localization using Viola and Jones method for a JAFFE [2] input face image, (b) The cropped image using Face localization method. 76
5.4 Data samples from CK+ database. 77
5.5 Data samples from FGNET database. 78
5.6 Facial Expression Recognition process. 80

Abbreviation

AFIA	Automatic Face Image Analysis
AFA	Automatic Face Analysis
PCA	Principal Component Analysis
FRS	Face Recognition System
FER	Facial Expression Recognition
FERS	Facial Expression Recognition System
VR	Verification Rate
MahCos	Mahalanobis Cosine Distance
LBP	Local Binary Pattern
LBP^{u2}	Uniform Local Binary Pattern
$LBP_{(P,R)}$	Circular Local Binary Pattern
$LBP^{u2}_{(P,R)}$	Uniform Circular Local Binary Pattern
ILBP	Improved Local Binary Pattern
MBLBP	Multi-scale Local Binary Pattern
GRAB	General Region Assigned to Binary
LTP	Local Ternary Pattern
DLTP	Derivative Local Ternary Pattern
TPLBP	Three Patch Local Binary Pattern
FPLBP	Four Patch Local Binary Pattern
CS-LBP	Center Symmetric Local Binary Pattern
V-LBP	Volume Local Binary Pattern
WLD	Weber Local Descriptor
LGBP	Local Gabor Binary Pattern
MHLVP	Multi-resolution Histograms of Local Variation Pattern
LGBPHS	Local Gabor Binary Pattern Histogram Sequence
MULGBP	Multi-resolution Uniform Local Gabor Binary Pattern

ABBREVIATION

EGV-LBP	Effective Gabor Volume Local Binary Pattern
LBPTOP	Three Orthogonal Planes Local Binary Pattern
GV-LBPTOP	Gabor Volume Three Orthogonal Planes Local Binary Pattern
MGCP	Multi-band Gradient Component Pattern
LDP	Local Derivative Pattern
ELGBP	Exploit phase information Gabor Binary Pattern
HGPP	Histogram of Gabor Phase Pattern
LGPDP	Local Gabor Phase Difference Pattern
LGXP	Local Gabor XOR Pattern
MBP	Monogenic Binary Pattern
CBP	Centralized Binary Pattern
LDP	Local Derivative Pattern
LDPv	Local Derivative Pattern variance
CLBP	Compound Local Binary Pattern
VTB	Vertical Time Backward
PLBP	Pyramid Local Binary Pattern
ARLBP	Asymmetric Region Local Binary Pattern
EARLBP	Extended Asymmetric Region Local Binary Pattern
MC	Modified Convolution

Chapter 1

Introduction

1.1 Context

Today's world is experiencing a digital revolution assisted by the rapid growth of technology, computing power and availability of plethoric data. The conglomeration of technology, processing power and data influence one another for their advancements. For example, digital camera and low cost mobile phones are creating a myriad collection of images and videos, converting text based information to multimedia. In the process of this, multimedia data would need different technological solutions and processing infrastructure to efficiently manage, access and process to provide various services to the world. Processing the images automatically to extract semantic information has gained great importance in research due to abundant multimedia data. There is a need of good automatic image analysis and pattern recognition techniques which can organize, index, annotate and retrieve images or videos to extract semantic information more intelligently and efficiently. In particular, face images or videos contain a great deal of information when compared to other multimedia images. For example, looking at the face image of a person, if we know the identity, a lot more information can be retrieved about him. Further, using face images we can determine age, gender, ethnicity and geographical location. In addition to this, by using facial expressions we could approximate the emotion one feels.

Thus, automatic face image processing or analysis has greater importance due to its applications in various domains of computer vision such as Human Computer Interaction (HCI), Human Robot Interaction (HRI), visual video surveillance, biometric identification and animation. Face image processing is a task of transforming the face image into another image. This task could be used as preprocessing step for face image analysis which involves transforming face image into high-level semantic reasoning. Face image analysis may include face detection and facial feature localization, face tracking and pose estimation, face and facial expression recognition, face modeling and animation. The task of localization of face and its features is an important task since it is used as a preprocessing step for remaining tasks of face image analysis. Deriving an automatic image analysis system would require several tasks mentioned above and one must choose tasks among these depending upon the application at hand. For instance, analysis of facial expressions would require face and feature localization task for still image or face tracking for a video sequence.

This thesis considers face recognition and facial expression recognition tasks of face image analysis. Each task can be considered individually as face image analysis since it involves transforming face image and interpreting with high level abstraction. In face recognition, face is identified or verified whereas, in facial expression recognition, face is interpreted and labeled with one emotion out of *neutral, happy, sad, surprise, angry, disgust* and *fear*. Deriving face image analysis by integrating face and facial expression tasks would assist schizophrenic [11] and patients with double dissociation patterns [12] to recognize face with facial expression or vice versa. The integrated framework may have a greater impact in application domain rather than considering as separate tasks. For instance, in commercial face recognition, system performance would be reduced in the presence of a smile. To overcome the shortcoming, external security protocol (*ISO/IEC 19794*) is enforced which specifies constraints on scene and photographic acquisition process. The integrated face analysis system may be designed to render recognized facial expression and transform expressed face to neutral face which in turn is identified or verified by face recognition task. To corroborate the integration of two tasks of face image analysis, Neuropsychology and Neuroimaging literatures suggest that there exists an overlap of activation patterns of the brain parts to various face processing tasks [13,14]. Further, both face recognition and facial expression recognition tasks share common challenges that are addressed by each such as, pose, illumination and scale. In addition to this, face recognition task needs to address facial expressions also. Besides, there exists an intersection between the subtasks of each task viz: face and facial feature localization, normalization, feature extraction and classification.
In designing an automatic face image analysis system by integrating face recognition and facial expression recognition, the first subtask is to detect face(s) in the given input image and localize facial features (nose, eyes and mouth). The localized face or facial features need to be normalized. The most typical normalizations are image alignment and illumination normalization. Next subtask is to extract features from the normalized

faces. Feature extraction is the process of transforming high dimensional pixel data to higher level abstracts. The high level abstracts include line, edges, shapes, texture, color and geometric displacements. In this process, most discriminative features may be selected without affecting the performance of the system. In addition to this, the extracted feature may be subjected to dimensionality reduction to remove redundant information using linear and non-linear projections. These features are fed as input to a classification subtask. Classification is the process of making a decision to assign a label to the input features. The labels may include age, gender, ethnicity, facial expression or identity of a person. The subtasks of face recognition and facial expression may precede one another or they can operate individually. Pipelining of the these tasks of face image analysis would reduce the number of samples needed to be labeled which would increase the speed of classification. For example, facial expression recognition could filter out images with all emotion class except *neutral* faces to feed into the face recognition system. In the process of integration, one must pay attention to individual task and its subtasks. Among all the subtasks of face and facial expression recognition, feature extraction and classification subtasks play an important role in determining the performance of the overall system. To derive robust automatic face image analysis system, one must design feature extraction techniques that are robust to pose, illumination, scale and occlusions and choose the best classifier algorithm depending upon the application.
This thesis aims at deriving robust feature extraction techniques based on Local Binary Pattern (LBP) [15] which extracts texture features from input images. These techniques are employed to evaluate the performance of each face and facial expression recognition tasks. Further, it is envisaged that features extracted by these techniques would be used in automatic face image analysis which integrates face and facial expression recognition tasks to enhance the performance and to address the difficulties faced by individual tasks. The throughput of the system would be increased if we use the same feature extraction techniques for both the tasks by employing different classifiers.

1.2 Challenges

There are some inherent difficulties that face image analysis needs to be addressed due to the non-rigid property of the face and image acquisition process. These difficulties lead to increase in *intra-class* variance when compared with *inter-class* variation which affects the performance of the face and facial expression recognition tasks. One can think of many ways to remove these difficulties, like imposing restrictions similar to (*ISO/IEC 19794*) protocol to design efficient face analysis system, designing feature extraction techniques which are robust to these difficulties and combining individual robust systems to complement one another.

This thesis considers designing of robust feature extraction techniques supported with minimal normalization of face images.

Following Sections summarize the important challenges that influence the performance of the face image analysis in the real world context.

1.2.1 Illumination

Variations in strength and angle of the incident light on the face would induce significant changes in the image as shown in the Figure 1.1. These changes are attributed to ambient and point (focused) source of the light. Ambient light source would induce less *intra-class* variations in the image that can be handled with ease when compared to pointed source like mugshot images. Many approaches have been proposed by extracting illumination variation features and some methods incorporate extra preprocessing such as normalization to negate the effect before feature extraction and classification.

Figure 1.1: Face under fixed view with varying illumination (ambient) (adapted from [1])

1.2.2 Pose

The variance in head pose introduces another level of difficulty which impacts the performance of the face image analysis. There are two types of head rotations i.e., *in-plane* and *out-of-plane*. Many systems assume constrained *frontal* face images which can be processed without normalizing the image. On contrary, out-of-plane rotations would introduce additional computations for face alignment, pose estimation and face registration. Further, some systems would be restricted to operate on a particular pose or multi-view face image. Figure 1.2 shows an example of fixed illumination but pose variation. It can be observed from the figure that, pose causes occlusions of some parts of the face image which make it harder to contain such variations as occluded parts cannot be estimated or reconstructed.

Figure 1.2: Faces with constant illumination but varying poses (adapted from [1])

1.2.3 Facial Expressions

The face is a non-rigid object due to variations introduced by the deformations of the facial features such as muscles, eyes, mouth, burrows and furrows. Figure 1.3 shows different facial deformations which are percepted as facial expressions. We envisage that integration of face and facial expression recognition tasks would handle these variations and enhance the performance of face image analysis.

There are other variations such as age, ethnicity, make-up, skin color and facial hair (mustache, beard) etc. The face and facial recognition task's performance is affected by these challenges. One must pay attention to minimize these effects. The intersection of challenges and functionalities of face and facial expression recognition subtasks are corroborating the integration of these tasks into a single face image analysis system. **This thesis considers exploring feature extraction techniques which could be employed to each face and facial expression recognition task and evaluate their efficacy in representing faces in the presence of the difficulties without incurring additional computation costs.**

Figure 1.3: Faces from JAFFE [2] with fixed pose and variations due to facial expressions.

1.3 Automatic Face Image Analysis

Automatic Face Image Analysis (AFIA) or Automatic Face Analysis (AFA) is a wide topic of research which includes different areas such as image processing, pattern recognition and machine learning. AFA can be loosely defined as the process of transforming face image into high level interpretation. It has great impact in computer vision applications like biometrics, video surveillance human-machine interaction, video conferencing and face image indexing and retrieval [16]. AFA may include face and facial expression recognition, classification of age, gender and gaze, face detection and tracking facial features, pose estimation, etc. AFA may contain only one task or several tasks integrated to handle a particular application. For example, face recognition task alone may be regarded as FA or in combination with age, gender,etc.

This thesis considers AFA which includes face recognition and facial expression recognition using Asymmetric Region Local Binary Pattern for face representation. It assumes an AFA which contains only one task either face recognition or facial expression recognition. The reason is, the proposed Asymmetric Region LBP needs to be employed to derive facial representations and study individual task's behavior and performance to integrate both tasks in one AFA system in future.

Here onwards face recognition and facial expression recognition tasks are referred as system for the sake of convenience. Subsequent sections describe the interactions between different components of face recognition and facial expression recognition systems.

1.4 Face Recognition System (FRS)

Face is a source of variety of information and hosts many sensory (eyes, ears, tongue and nose) organs of human beings. Face may be used to identify and categorize human beings depending upon age, gender, gaze, color, ethnicity and geographical region. Further, it reveals the details related to health and emotions. It is one of the important tools of human communication apart from gestures. Identity of a person would enhance communication further, as humans tend to personalize the content, behavior and the context of the interaction. Automatic recognition of face has resulted in several applications such as computer animation to render identity on avatar, surveillance and law enforcement, smart cards, entertainment, etc. [17].

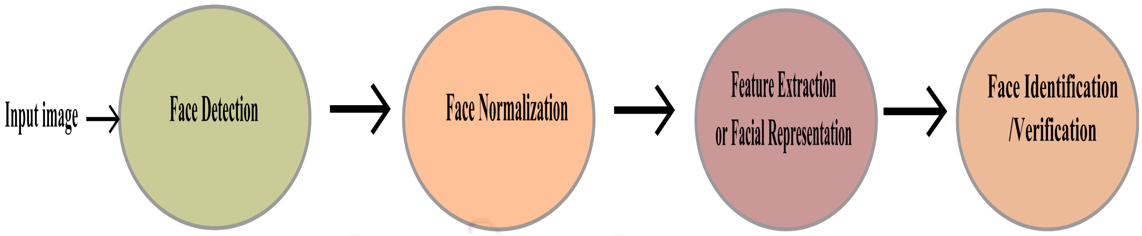

Figure 1.4: Components of generic face recognition system.

1.4.1 Components of FRS

There are four main components of automatic FRS namely, face detection, face normalization, feature representations and face recognition as shown Figure 1.4. There is no strict demarcation between each component i.e., depending upon the application these components may overlap. For instance, face normalization would overlap with face detection or with feature extraction component. Face normalization might be named as preprocessing process consisting of face alignment, illumination normalization, estimating pose, etc. Face detection and face normalization steps could be ignored if FRS is assumed to be non-automatic. In such case face localization normalization are done manually off-line to create training data.

Face Detection

Face detection may be defined as the process of determining whether face is present or not in a given image, if present, then the location of the face is notified. Face can be detected in color or single intensity, single image or video sequence (single image/frame taken at a time). Another term closely related to face detection is face localization. It may be defined as determining position of single face in a given image. The difference is that in face localization it is assumed that only one face is present in the input image. But in the face detection, presence of face either one or many is not known prior. Facial feature detection is to detect the presence and location of facial features like eyes, eyebrow, nose, nostrils, mouth, lips and ears. Face detection is a visual task which human beings do effortlessly in real time, regardless of illumination, orientation and camera distance, but for a computer or machines this task is not easy. For machines face detection is a challenging task due to variations [18] introduced by pose, facial expressions, scale, occlusion, illumination and sensor settings. To automate face detection task two main components; feature extraction method and Classifier have to be derived such that they are robust to the variations. Face detection system may be categorized depending upon the type of features extracted from the face and they are explained below.

- **Knowledge-based methods:** In this method typical facial features (eyes, nose, nose, mouth, etc) are localized and then rules which relate or group the features by geometrical relationships are applied. For example, position of each feature may be represented by relative distance between them or eyes will appear symmetric in position compared to nose and mouth, etc. One thing to be noted in this method is after localizing the features of the face by exploiting general basic-level features of face like color of the skin, edges, contour, etc, then the rule that represents knowledge about the face is used for verification to reduce false detection [18]. The difficulty in this approach is to translate the human knowledge into rules. If rules are generic false detection will increase, if rules are strict face may be missed in the frame. Another challenge is, this approach requires all facial features or at least eyes, nose, mouth, prior and hence this approach can be applied only to faces of frontal view.

- **Feature invariant methods:** The underlying assumption that humans detect face effortlessly in variations in pose; illumination, etc. There may be some invariant features or properties independent of all variabilities. Several methods have been proposed, using edge maps [19], texture [20], skin color [21], shape information or tracking heads and faces in image sequence are also developed [22–25]. These methods suffer from the same challenges in terms of extraction of features from the image, as these features vary with the illumination.

- **Template Matching:** A template (face pattern) is derived manually or parameterized using function. Then similarity or correlation is computed between image and template to check the existence of face. Disadvantage of this method is that it is not easy to design good generic template which can detect face in all poses and scales.

- **Appearance/Image based methods:** In this method the template (generic representation of face) is learned from the images supplied during off-line training in contrast to template matching method. Therefore training is time consuming but the detection speed is faster and even some time real-time detection may be achieved [26]. These methods have some difficulties that they need a lot of face and non-face images. As learned result is not easily interpreted it is hard to adjust to other applications. In spite of these disadvantages appearance based methods are mostly preferred state-of-the art Viola and Jones [27] method for face detection.

In this thesis faces are detected and cropped automatically using an automated tool [28].

Face Normalization

The challenges that a practical face recognition system needs to address include appearance variations due to lighting, pose, scale, facial expression and occlusion. These variations would increase the *intra-class* variations when compared to *inter-class* variations. These variations affect the system performance such that it may not be appropriate in real world applications. Face normalization is a generic process in which one or more methods like illumination, face alignment, etc are used to remove or reduce these variations to increase the system efficiency. By applying image processing techniques the input face image is transformed into face image with less or no variations. Variations in illumination such as shadows, underexposure and overexposure make face recognition intractable problem. Various approaches have been proposed by several researchers [29–34]. A good review of different approaches to handle the variations are illustrated in [35] and a comparative study on these methods are described in [36]. Varying pose i.e., *in-plane* and *out-of-plane* rotations of the subject's face image would cause greater effect. These rotations need to be negated in the real world application to exploit non-intrusive advantage of face biometrics. Several works have been proposed to handle the pose variations [37–42]. Zhang *et.al* [43] reviews pose invariant techniques used to derive face recognition system. Facial expressions are variations caused by deformation of facial features and muscles. These variations would make face non-rigid object that would harder to model the face in the face recognition systems. Several attempts have been made [44–46] and face recognition systems are derived invariant to facial expressions. In [47] authors have proposed a 3D morphable model to fit facial expression which can be used to neutralize the expression and render a neutral face to recognize.
This thesis does not consider any illumination equalization as proposed feature extraction techniques are robust to monotonic illumination changes. The face recognition system is constrained to frontal pose faces with facial expressions inherent in most of the benchmark databases. The face images were cropped using an automated tool [28] and aligned depending upon the eye coordinates.

Face Representation or Feature Extraction

Feature extraction is a process of transforming the input data into reduced set of features which represent the original data without any redundancy. The redundant information in the input data affects the performance of the classification. The classification algorithm would overfit the training samples poorly generalizing the new test data sample. This may be due to larger dimension of the input data. Hence feature extraction techniques play an important role in image analysis. These feature extraction techniques need to be robust to difficulties/challenges that image analysis system would face in real world application. The feature extraction techniques can be categorized into two categories viz: holistic (Global) and Local (patch) based on the type of image considered to extract features. Holistic feature extraction techniques would extract features from whole image considering each pixel from the image. On the contrary, local feature extraction techniques use regions, patches or neighborhood and landmarks or fiducial points to extract the representative features from the image. For many years holistic approaches such of Principal Component Analysis (PCA) [48], Linear Discriminant Analysis (LDA) [49] and variants of these methods like Marginal Fisher Analysis (MFA) were introduced into face recognition due to their fair performance. However, holistic methods are sensitive to variations of pose, facial expression and scale [10]. These methods use training data to obtained subspace solution which may be biased towards training data. The computation complexity increases as the number of data samples increase. Many subspace learning methods used for face recognition are presented [50] comprehensively. The local feature extraction methods are not sensitive to noise and invariant to small variations in pose, scale, illumination and facial expressions. These methods have an ability to perform better than the holistic methods. Some of the representative local feature methods like Elastic Bunch Graph Matching (EBGM) [51] and Gabor wavelets [52] are extensively used in face recognition for better performance. These local descriptors get affected by the process of manual or automatically annotated landmarks due to localization and alignment errors. However, many texture descriptors have been applied successfully for face image analysis [15] using alternative methods to negate these errors. Local Binary Pattern (LBP) is first texture extraction technique used to derive face representations in face image analysis. LBP based representations have grossed greater attention of researchers in pattern recognition due to less computation cost and possibility of applying in real world applications and a brief survey on the applications of LBP in face recognition can be found in [53].
This thesis proposes a novel LBP based Asymmetric Region based LBP for face representation.

Face Identification/Verification

Face recognition can be configured into two modes of operations: Verification and Identification. Identification is the process of finding out a person's identity/label/name by matching a face pattern (probe/test image) with the person's gallery/trained face features. In the identification mode of operation, the system is trained with the face images obtained during enrollment/registration process of several persons. For each of the persons, a face template or model is learned in the training stage. A probe image is matched against every known model, yielding either a score or a distance describing the similarity between the probe image and the label of the person. A label is assigned by the system which is most similar to the template. In the verification case, a person's identity is claimed apriori. The claim is verified by comparing with the person's individual template. Similar to identification, it is checked for similarity between probe image and the template.
Classification is the task of transforming the input face representation into high level description. The description may be identity of the subject (identification) or it may be affirmation of the subject identity (verification) . Depending upon the output of classification the face recognition system is configured as identification or verification system. For the task of classification of face, there are different approaches in which holistic and local features are used as input for these classifiers. The holistic approach like PCA, LDA, Fisher Discriminant Analysis (FDA) known as Fisherfaces [54], self organizing map and convolution network [55] are used. Further, Probabilistic Decision Based Neural Network (PDBNN) [56] are used as classifier in face recognition. These classification methods can be configured to use local features such as Gabor wavelets and LBP based face representations. Zhang *et.al* [43] have reviewed different face recognition methods along with pros and cons across poses.
In this thesis FRS is configured in verification mode and PCA is used as base line method and it is also used to reduce the high dimensionality of feature histogram method.

1.5 Principal Component Analysis (PCA)

PCA is also called as Hotelling Transform [57]. It is linear projection which transforms the input image space (training images) to orthogonal basis vectors and each basis vector corresponds to the direction of greatest variance of the input image space. PCA is useful to reduce the dimensionality and to contain the statistical covariance in the transformed samples.

The application of PCA in face recognition is called as Eigenspace projection. Eigenspace are nothing but eigenvectors of the covariance matrix derived from the training set images containing all subject images that have been enrolled in face verification system (FRS configured in verification mode). The eigenvectors with non-zero eigenvalues form an orthonormal basis in the image dimensional space. Each image of size $N \times M$ is stored in a vector size NM as:

$$I^j = \left[I^j_{(0,0)} \; \ldots \; I^j_{(N-1,M-1)} \right]^T \tag{1.1}$$

The images are subtracted by the mean image to center each image vector to mean of the whole training or testing image.

$$\overline{I}^j = I^j - m, \text{ where } m = \frac{1}{P} \sum_{j=1}^{P} I^j \tag{1.2}$$

These mean centered vectors are stacked side-by-side, to create the matrix of size $NM \times P$ where P is number of images.

$$\overline{M} = \left[\, \overline{I}^1 \mid \overline{I}^2 \mid \ldots \mid \overline{I}^P \, \right] \tag{1.3}$$

The matrix $\overline{M}$ is multiplied by its transpose to solve for eigenvalues and eigenvectors.

$$\Gamma = \overline{M} \, \overline{M}^T \tag{1.4}$$

The covariance matrix Γ can have P eigenvectors and corresponding non-zero eigenvalues. The eigenvectors are sorted in ascending order of its eigenvalues. Depending upon the amount of dimensionality reduction required the eigenvectors with larger eigenvalues are selected to represent the variance space of the training image. Following steps are followed to use PCA for face verification.

1. **Create Eigenspace:**
   The following steps are followed to create eigenspace of the training images.

   a. **Center data:** Each subject's images are centered subtracting the mean image from these images as in equation 1.2.

   b. **Create Matrix $\overline{M}$:** Each centered image is combined as in equation 1.3.

   c. **Calculate Covariance matrix:** The matrix's $\overline{M}$ transpose is multiplied by the matrix $\overline{M}$ to calculate the covariance matrix.

$$\Gamma' = \overline{M}^T \, \overline{M} \tag{1.5}$$

d. **Compute the eigenvalues and eigenvectors of Γ':**

$$\Gamma' V' = \varPhi' \, V' \tag{1.6}$$

$\varPhi'$ is eigenvalues and V' is corresponding eigenvectors

e. **Compute the eigenvectors of $\overline{M}$ $\overline{M}$:** Multiply the data matrix $\overline{M}$ by the eigenvectors.

$$E = \overline{M} V' \tag{1.7}$$

Divide each eigenvector by its norm.

$$e_i = \frac{e_i}{\|e_i\|} \tag{1.8}$$

f. **Order eigenvectors:** Order the eigenvectors in ascending order and depending upon their eigenvalues and select $P-1$ eigenvectors and stacked side-by-side which forms eigenspace E of the training images.

$$E = [e_1 \mid e_2 \; \ldots \; e_P] \tag{1.9}$$

2. **Project training images:** Each image of each person is centered and projected on the eigenspace by taking dot product between transpose of eigenspace with centered training image.

$$\tilde{I}^j = E^T \, \overline{I}^j \tag{1.10}$$

3. **Project Test image:** The test images of corresponding subject are centered and projected as in equation 1.10. The projected test images are matched with projected training image set (face template) of each subject using different similarity or distance measures. A label of face template is assigned to the test image to which it is similar. A detailed explanation of the PCA projection used in the face image domain is illustrated with examples in [58].

This thesis considers Eigenspace method for feature histograms extracted for FR in verification mode.

1.5.1 Similarity Measures

Face recognition system can be configured in two modes of operations viz. verification and identification. Given the features the face recognition system needs to take decision using similarity measures in order to either verify or identify. Similarity metrics would measure how a given new input features closely represent a learned or enrolled template in the system. Formally decision procedure using similarity measures in verification mode is as follows: given the new feature vector x with claimed identify I_i of an enrolled subject-template $\bar{y}_i$, where, $1 \leq i \leq N$ and N is the number of enrolled subjects, determine the correctness of the identity claim by making decision into one of two classes C_1 or C_2 [59]:

$$(x, I_i) \in \begin{cases} C_1, & \text{if } s(x, \bar{y}_i) \geq \theta, \ i = 1, 2, ..., \text{N}, \\ C_2 & \text{otherwise}, \end{cases} \tag{1.11}$$

where C_1 denotes the class of the genuine identity, C_2 the class of illegitimate identity, In case of identification the decision is taken to assign a class label for the new input feature vector with best match class of enrolled subject templates. This can be formalized as follows: given a new feature vector x and database of enrolled N subject templates $\bar{y}_1, \bar{y}_2, ..., \bar{y}_N$ with identities (class) $I_1, I_2, ..., I_N$, then the decision is taken to find the best match class for the new feature vector, that is,

$$(x) \in \begin{cases} I_i, & \text{if } s(x, \bar{y}_i) = \max s(x, \bar{y}_j) \geq \theta, \ 1 \leq j \leq N, \\ I_{N+1} & \text{otherwise}, \end{cases} \tag{1.12}$$

I_{N+1} stands for the case, where appropriate identity is not assigned for the new feature vector x. The function $s(.,.)$ is any distance metrics like Euclidean, City Block, Cosine and Mahalanobis Cosine distance. The θ represents the experimental threshold. Given two vector x and $\bar{y}_i$ where, x denotes face-feature vector, $\bar{y}_i$ denotes the mean feature vector of the system clients I_i, $(i = 1, 2 \ldots, N)$, C^{-1} is the inverse co-variance matrix and T denote transpose of the vector, each distance can be formalized as similar to [59, 60] follows:

Euclidean distance $D(x, \bar{y}_i)$:

$$D(x, \bar{y}_i) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{1.13}$$

City Block distance $CDB(x, \bar{y}_i)$:

$$CDB(x, \bar{y}_i) = \sum_{i=1}^{N} |x_i - \bar{y}_i| \tag{1.14}$$

Cosine distance $CosDis(x, \bar{y}_i)$:

$$CosDis(x, \bar{y}_i) = 1 - \frac{x.\bar{y}_i}{\|x\| \, \|\bar{y}_i\|} \tag{1.15}$$

$\|.\|$ is norm of the vector which is equal to $\sqrt{\sum_{i=1}^{N}(.)^2}$.
Mahalanobis Cosine Distance $MhCos(x, \bar{y}_i)$ is the cosine of the angle between the vectors. Formally it can be expressed as:

$$MhCos(x, \bar{y}_i) = 1 - \frac{x^T C^{-1} \bar{y}_i}{\|x^T C^{-1} x\| \, \|\bar{y}_i^T C^{-1} \bar{y}_i\|} \tag{1.16}$$

This thesis considers the above distance measures for face verification.

1.5.2 FR Performance Measures

The performance measures are metrics that measure different types of errors a system can make. These errors are used to measure the performance of the FRS. Following types of measure of errors are normally used in FR domain.

1. *False Acceptance Rate*(FAR): An empirical probability at which the system incorrectly accepts the claimed identity but the sample is that of a impostor (different subject). Particularly, it is falsely accepted number of samples divided by total impostor samples.

2. *False Reject Rate* (FRR): Estimate of the probability at which the system rejects the claimed identity but the sample is of actual subject (genuine user). That is number of rejected claim samples (from test set) divided by total number of client samples.

3. *True Acceptance Rate* (TAR): This metrics is also refers to Verification Rate (VR) and it is defined as 1-FRR.

4. *Equal Error Rate*(EER): Point in which FAR and FRR is equal.

All these parameters may be used to determine FR system performance after testing is performed. FAR and FRR are threshold dependent while EER is threshold independent. If EER is lower then system's performance is said to be better. The value of FAR and FRR ranges from zero to one. But either FAR or FRR cannot be used to determine system acceptance in real world deployment though FAR and FRR are minimum at zero.
In this thesis both FAR, FRR and VR are used to measure the system level performance.

1.6 Facial Expression Recognition System (FERS)

Though the study of facial expression is dated back to 18^{th} century, the work of Ekman suggesting universalism between emotions and facial expressions across different cultures of human beings [61] gave a thrust to the research of facial expression analysis. The emotion felt by the human beings are displayed using facial expression which is a result of spatial deformation and movements of muscles under the facial skin. Facial expressions may or may not communicate the internal emotions felt by the individual but they serve as non verbal visual cue which regulate emotions of others in interaction. Several studies have proposed to measure and understand the relation between emotion and facial expressions. There are three methods used to measure the emotion of humans [62]. The first methods is *Neuro-physiological* which measures body responses like brain activity, pulse rate, blood pressure and skin conductance. These measures are analyzed and categorized into emotion labels. Second is *Observer* method in which the visual appearances of the facial expressions are percepted and labeled using classifiers. The third method is *Self-report* in which emotion is labeled by analyzing the outcome of a questionnaire presented to the subject who's emotion to be measured. Deriving a robust FERS incorporating these methods would be a challenging task. In the first method it is difficult to measure emotions by implanting costly sensors on the body restricting mobility of the subject. It is also difficult to handle the sensors without expertise and mapping data unambiguously to a specific emotion label. In *Observer* method it is harder to make context dependent interpretation of facial expressions from sensory data. The sensory data is highly subject dependent and it is not unique across different subjects and the facial expression can be faked i.e, posing complementary facial expression compared to actual emotion. In case of Self-report method, the interpretation of emotions by one subject may be different from the other subject interpretations, which makes it difficult to compare and comprehend. Human beings interpret and label facial expressions effortlessly using *Observer* method. It is relatively easy to automate *Observer* method when compared to other two methods. Hence, many FERS have developed to automatically analyze facial expressions from static images and video sequences when compared to other methods.

1.6.1 Types of Facial Expression

Facial expressions can be classified into posed and spontaneous depending upon the amount of time the expressions are displayed [63]. Posed expressions are intensionally elicited by the selected subject during image acquisition process. These subjects know the context and are instructed about the type of expression one needs to pose in controlled environment. Spontaneous expression are elicited by external stimuli like screening of drama, video, jokes and situations or context which subjects are not aware of or out of their control. Posed expressions are mostly in frontal or profile, percepted easily and compared to spontaneous expressions. Most of the time the real feeling or emotions are reflected by spontaneous expression which lasts for only a few seconds making it hard to percept and recognize automatically.

These facial expressions are categorized into three types like discrete, rule based action

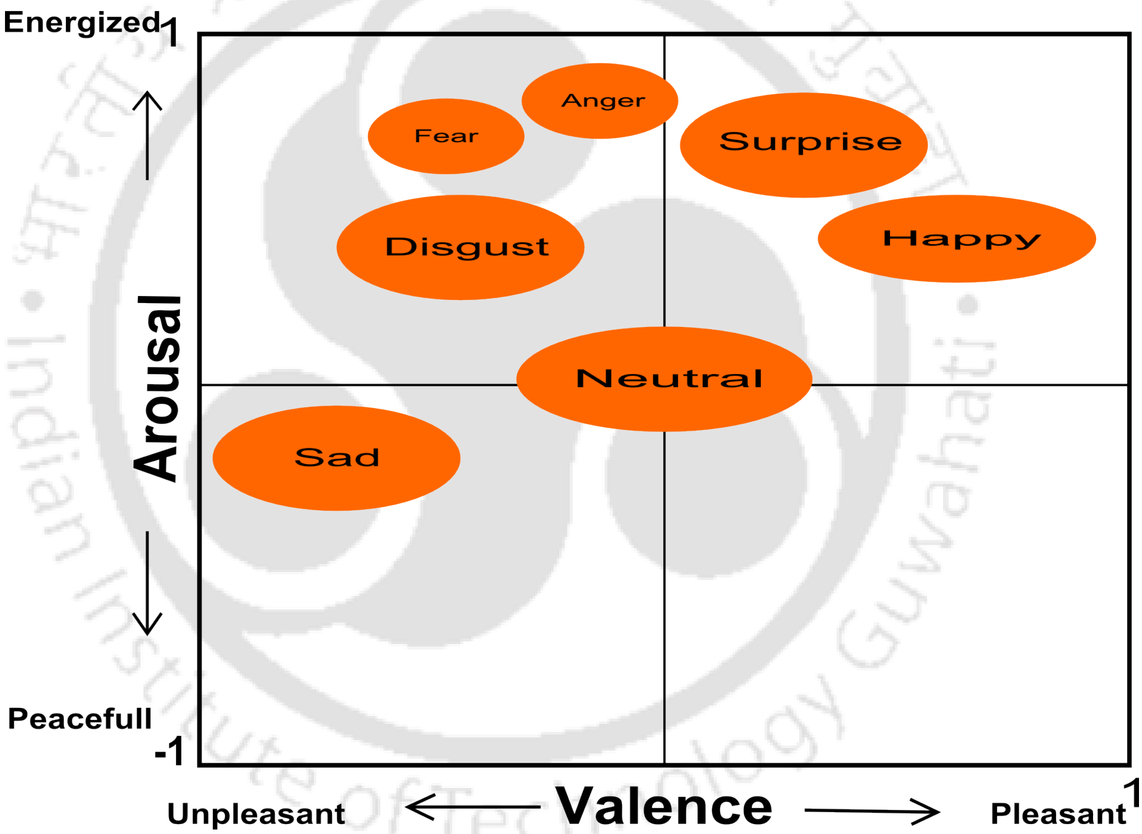

Figure 1.5: Valance-Arousal dimensional space.

units (AUs) and dimensional space depending upon their representation (class/label). AUs are labels for deformation of a particular muscle in the region of face. In discrete representation each emotion is given a label such as *neutral, happy, disgust, fear, sad, surprise and anger* which are considered as basic emotions. Each facial expression is represented as combination of deformation of facial muscles termed as Action Units (AUs). Ekman and Friesen coded all the facial muscle movements called FACS (Facial Action

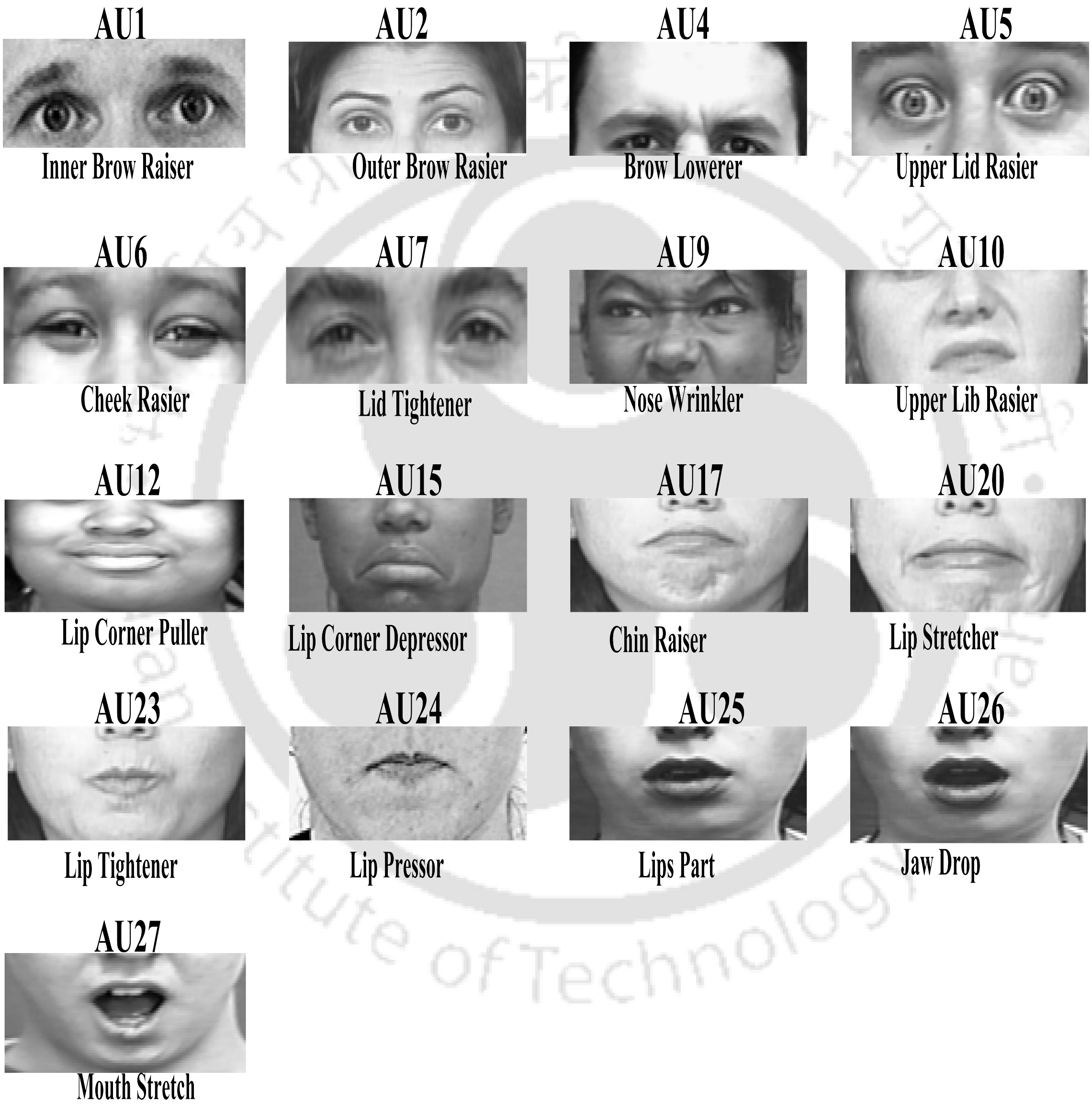

Figure 1.6: Common Action Units [3].

Coding System) [3]. FACS consists of 44 AUs (Action Units) for description of facial actions with regard to their location as well as their intensity. Facial expression can be described as combinations of FAUs (Facial Action Units). Out of 44 AUs 17 AUs (1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 17, 20, 23, 24, 25 and 26) as shown in Figure 1.6 are called primary AUs as their combinations can derive all basic emotions. Combination of AUs 6, 12, 15, 17 *Happy* expression can be derived, *Surprise* expression is represented using AUs 1, 2, 5, 25 and 27. *Fear* is formed by the combination of AUs 1, 4, 7 and 20 , AUs 4, 5, 15 and 17 make expression *Anger*, *Sad* is derived by the integration of AUs 1, 2, 4, 15 and 17 and *Disgust* is derived by composition of 1, 4, 15 and 17. In dimensional category, expressions are represented using 3 or 2 dimensional real valued space like Valence-Arousal-Dominance (VAD) [64] and Valence-Arousal [65] respectively. The valence represents either pleasantness (positive) or unpleasant (negative) and arousal describes the level of activation of the emotion while Dominance illustrates the extent of attention. Most of the researchers consider Valance and Arousal dimensions as shown in Figure 1.5 as they can represent most of the emotions.
In this thesis FERS is designed to process posed near frontal facial expression images and classify using discrete representation.

1.6.2 Components of FERS

A FERS consists of three important components viz: Face detection, Facial representation or Feature Extraction methods and Classification as shown in the Figure 1.7. The last two components largely determine the performance of the FERS, the first one is considered as preprocessing step of FERS. But the quality of the result obtained in preprocessing step can have ripple effects on other two components. Early works in FER is summarized by Samal *et.al* [66], Pantic *et.al* [67], Fasel *et.al* [68] and recent advancements are surveyed by Zeng *et.al* [69] and Huang *et.al* [70]. Next sections describe each component of FERS.
This thesis uses *Observer* method of measuring Facial expression.

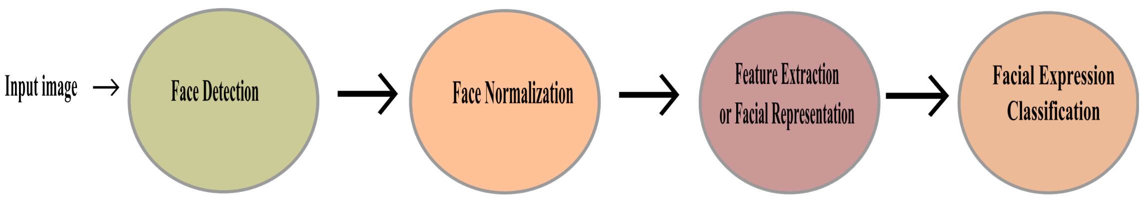

Figure 1.7: Components of Facial Expression Recognition System.

Face Detection

The work of Viola and Jones has made face detection practically feasible for real world applications in camera, mobile phones and photo organizer *Picasa*. Zhang *et.al* [71] have surveyed different boosting based face detection schemes which are de-facto standard of face detection in real-world applications. The survey provides comprehensive details of different feature types used for different boosting learning algorithms. The paper concludes giving suggestions such as considering context, development of new adaptation in learning algorithm and new feature types.
In this thesis the Viola and Jones face detector available in OpenCV library [72] is used for facial expression recognition system design.

Face Normalization

Face normalization is the process of removing inconsistency in the face image size, illumination and position of face in the image. To analyze the facial expression input image needs to align to similar size, refine all images to uniform illumination, correct pose to frontal or near frontal to contain more face region and crop the face image at the same position from all images since facial expressions are more spatial dependent. The localized face of facial feature in the face detection component suffers from inconsistency in position of localization inherent in the face detection component. It is observed that Viola and Jones face localization method [72] produced different positions of face in input image if the algorithm is repeatedly used to detect the face with same image. This inconsistency is termed as localization error. Since facial expressions are spatially sensitive these errors need to be negated before features are extracted or robust feature extraction method should be used.
A heuristic localization method is proposed to negate the localization errors and no illumination correction was done as LBP based representations are robust to monotonic illumination.

Facial Expression Representation

Facial Expression Recognition (FER) deals with classification of movements and deformation of facial features in to abstract classes that are based on visual information. *Emotion Recognition* deals with interpretation of facial expression with context, feeling of the subject and environment around. In 1971, Ekman and Friesen [61] postulated six universal facial expressions such as *Happy*, *Sad*, *Anger*, *Disgust*, *Surprise* and *Fear*. These prototypic displays are termed as basic emotions. The use of the term *emotion* recognition refers to FER unless stated specifically here onwards. The FERS proposed in literatures employ the *Observer* method in order to interpret the emotion. It is also evident from [62] that most of the studies have considered the said method when compared to other two methods of emotion measurement.

In this thesis *Observer* method is automated because of non-intrusiveness and the image acquisition process is real time in nature.

Facial expressions are generated by deformation of facial muscles, which lead to deformation of facial features such as eyebrow, nostrils, mouth, eyes, lips, and skin texture revealed by wrinkles and bulges. These muscular activities can last for 5 seconds to 20 milli seconds. Each basic emotion can be expressed depending upon the location of facial action, intensity and dynamics. Spatial information (location)is important since facial expressions are results of contraction of facial muscles in specific regions of the face. Each expression is expressed using particular muscles, at particular location of face. The intensity of the facial expression is defined as the measure of the extent of muscle contractions monitored in a particular location for a particular expression. It is measured by determining the geometric deformation of facial features or the density of wrinkles appearing in certain face region [68]. Measuring absolute intensity is difficult since variation in inter-personality, across facial actions without referring to neutral face of a given subject. Intensity measurement of spontaneous expression is difficult since it lasts for a few seconds than posed facial expressions. The dynamics of facial expression could be understood by using three temporal parameters namely, *onset* (attack), *apex* (sustain), *offset* (relaxation) and its dynamics (temporal evolution) can be illustrated. There are two main methodologies of measuring the dynamics, *judgement-based* method which directly associates specific facial patterns with mental activities and *sign vehicle-base* method which represents facial actions in a coded way, prior to eventual interpretation attempts. It is evident from the above discussions that the feature representation (feature extraction) methods have to address location, intensity and dynamics of the facial expressions. These methods can be categorized according to whether they extract motion or deformation of facial features, whether they act locally or holistically [68]. These methods try to extract the prominent motion or deformation features from video sequences of facial expression using difference image method, optical flow, etc. These methods can be applied considering the whole input image or a patch/region wise. Some methods use predefined model to fit on fiducial points around prominent deformation regions on the face and the displacement of these points are classified.

The thesis considers still Image/appearance based local feature extraction methods LBP based ARLBP and ERLBP for analysis of facial expression recognition.

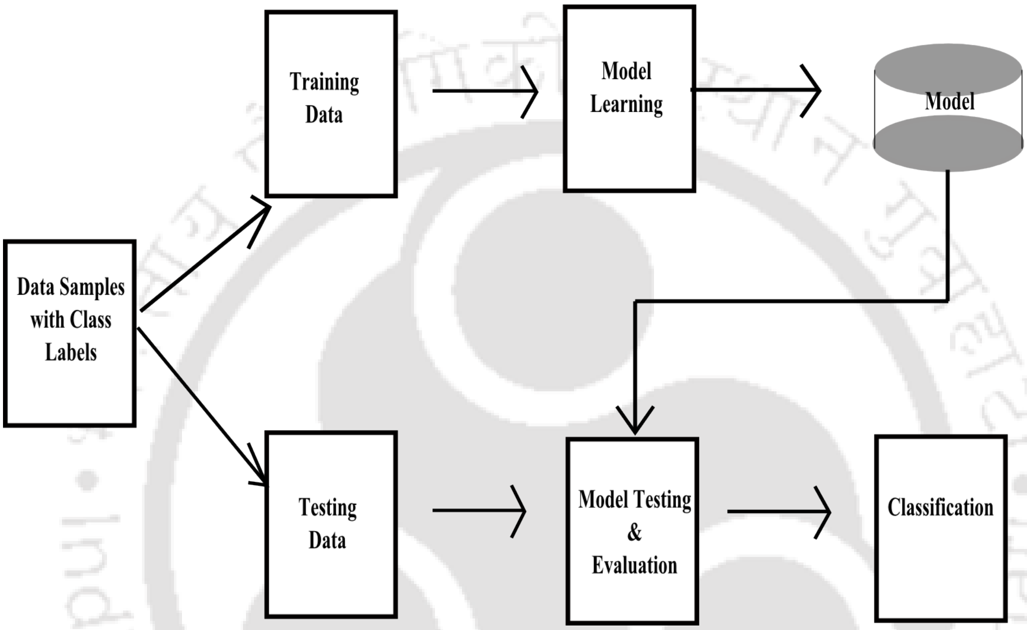

Figure 1.8: Classifier Training and Testing phases.

Classifier

A Classifier is a model derived from input data or feature vector to assign class label for each input feature vector. Formally, given a set of input items, $X = \{x_i\}, 1 \leq i \leq n$ and a set of labels/classes, $Y = \{y_i\}$ and training data $T = \{(x_i, y_i)|y_i$ is the label/class for $x_i\}$, *Classifier* maps X to Y, $f(T, x) = y$. Classification is a process of assigning a class label to unseen input data items. There are two phases to derive a *Classifier*; training and testing phases as shown in the Figure 1.8. Before training phase the input data items with known label or class are divided into two or more sets, called training and test data sets. In Training phase, a classifier model is first fed training data in which each item is labeled with the correct class or label and learning algorithm is used to derive classifier model. Once the model is derived for a particular input data the performance of the model needs to be evaluated before it is used for classification. In testing phase the model performance is evaluated to test how best it can classify the data item in test set with known labels. The classifier performance largely depends on the parameters of the model and the characteristics of the input data. There is no generalized classifier that performs best on all data sets. Empirical tests need to be performed to find the characteristics of data that determine classifier performance and find the best classifier by comparing performance of different classifiers. Several classifiers are used in FER domain such as, Neural Networks [73,74], Bayesian Network [75], Dynamic Bayesian Networks [76], Hidden Markov Models [75], Rule-based classifiers [77], Support Vector Machines (SVM) [9,78–81] Instance-based classifier (KNN) [82] and Sparse Representation [83]. Among these SVM is profusely used as classifier in image/appearance based FER in which LBP based feature representation of facial expression is employed due to better performance in the domain. **In this thesis SVM with Radial Basis Function and Liner kernels are used for FERS.**

Support Vector Machines

Support Vector Machine (SVM) [84] are binary supervised classifiers in which they use pre labeled training data to derive a learning model and are used to classify the unseen data items. For concrete comprehension of SVM, the discussion is divided into four parts [85] viz: (i) Separating hyperplanes (ii) The maximum-margin hyperplane (iii) The soft-margin (iv) The kernel function. Given training data set $D = \{(x_i, y_i)\}^k$ $i = 1$, k is the number of data samples $x_i \in \Re^n$ is data item and $y_i \in \{-1, +1\}$ is class label of x_i data item. SVM finds an hyperplane to separate these data samples into respective classes -1 and +1. The hyperplane in the form of $\vec{w}\ \vec{x} + b = 0$ where, $\vec{w}$ is weights to be learned over data items, $\vec{x}$ is data item and b is bias. There can be many hyperplanes exists to classify the data items but the best classifier is one which chooses the best hyperplane that generalizes over unseen data. The selection of best hyperplane is to maximizes the margin of separation between two data classes -1 and +1. Maximizing the margin is a problem of constrained optimization that can be solved by Lagrange method. Thus w can be solved as [84]:

$$w = \sum_{i=1}^{k} \alpha_i y_i x_i, \tag{1.17}$$

where the coefficients α_i is a Lagrange multiplier. A non-zero α_i indicates that x_i associated with α_i is a support vector. An unseen data item x_{unseen} can be classified by the decision function:

$$f(x_{unseen}) = sign\left(\sum_{i=1}^{k} \alpha_i y_i (x_i, x_{unseen}) + b\right), \tag{1.18}$$

where positive output means the x_{unseen} belongs to positive or negative class. There may be several instances that SVM cannot classify the data samples. Introduction of soft-margin to the SVM allows it to handle misclassification errors i.e, outliers or noisy measurements or the data slightly non-linear which are near or within the margin of the hyperplane. By introducing soft-margin parameters user can control a number of misclassification and these parameters specify a trade-off between hyperplane violation and the size of the margin. In real world data samples are not linearly separable, therefore SVM uses kernel functions to project the data samples from lower dimensions to higher dimension and finds the hyperplane that can separate the data samples in two classes. In the experiments of FERS two kernels are used viz: Radial Basis Function (RBF) shown in the Equation 1.19 and Equation 1.20.

$$k(x, y) = exp(-\gamma\|x - y\|^2), \tag{1.19}$$

$$k(x, y) = x^T y, \tag{1.20}$$

And the decision function in Equation 1.18 can be rewritten as:

$$f(x_{unseen}) = sign\left(\sum_{i=1}^{k} \alpha_i y_i k(x_i, x_{unseen}) + b\right), \tag{1.21}$$

The intuition behind the kernel functions is if the data samples are transformed from non-separable lower dimensions into higher dimensions such that data samples are linearly separable. This intuition does not always hold due to increase in dimensions of data samples and it is difficult to select best separating hyperplane from myriad hyperplanes making difficult for learning algorithms to converge and generalize. One must choose the best kernel function and soft-margin parameters to suit the data samples. It is difficult and time consuming to tune these parameters. Nevertheless, there is a technique to tune SVM parameters available in the literature Cross-Validation. This technique is a statistical method to evaluate the learning algorithms and tune the parameters of learning algorithm.

Cross-Validation

Cross-Validation (CV) is a statistical technique used to evaluate the performance and tune the parameters of a learning algorithm. The data samples are divided in to training and testing sets and these sets have to cross-over in successive rounds such that each data point is used as testing sample. In each CV round Recognition Rate (RR) is used measure the performance. The RR is defined as total number of testing samples correctly classified divided by the total number of the testing samples.

Though, CV is used to estimate performance however, main goal of CV is to validate or check generalizing ability of the learning model. The learning model should not overfit the data i.e., the model performs best on the training and testing data but performs worst for unseen data. There are three types of CV such as, Hold out method, *k-fold* cross-validation and Leave-One-Out CV (LOOCV). Hold out method is simplest of CV in which the data set is randomly sampled into two sets training and testing sets. A model is derived using training set and the performance is validated using the test set. *k-fold* cross-validation is common type of CV type used in machine learning domain. In *k-fold* CV data is segmented into k equal size and k iterations of training using $k-1$ segments as training set and one segment as testing set. LOOCV is special type of *k-fold* CV in which k is equal to n the number of data samples. The performance is measured using common metrics such as precision, recall, F-score and accuracy. Accuracy is commonly used as a measure of performance in classification task. CV can be used to tune the parameters using grid search as in OpenCV library [72]. With respective to SVM kernels there are two parameters that one needs to pay attention i.e, Cost parameter C and (γ) gamma parameter. The parameter C relates to the cost function which gives weight to the misclassified data samples and gamma to the kernel. If C is large, learning model will choose a smaller-margin hyperplane so as to enable it to classify every data sample correctly. Conversely, a very small value of C causes the learning model have larger-margin separating hyperplane, which may lead to more misclassification.

In this thesis grid search method from OpenCV library is used to tune the parameters of the kernels and SVM.

Multi-class Classification

SVM is a binary classifier which classifies only two valued label data samples restricting the use of SVM for classifying data samples with many class labels. For example, numeral classification has 10 labels and facial expression data sets may have 7 or more class labels. Multi-class classification problem is to predict a label from n labels for each data instance. The multi-class classification problem is reduced to binary classification problem. The main idea is to reduce the multi-class classification problem into multiple binary classification problem and using of majority voting principle to predict the label of unseen data sample. There are two basic reduction methods, one-against-rest (one-against-all) and one-against-one (all-against-all) by which a multi-class classification is realized using binary classification. In one-against-rest strategy, single classifier is trained to separate one class from the rest. To predict the class of the new unseen data sample, the class label is assigned where classifier decision function is largest. Another approach is one-against-one builds $k(k-1)/2$ binary classifiers, trained to separate each pair of classes against each other, and uses a majority voting scheme (max-win strategy) to determine the class label of the new unseen data sample.

In this thesis a multi-class SVM with RBF and Linear kernels is used for classification of Facial expressions.

Person Dependent and Independent FER

Emotion theories can be categorized in two types viz: cognitive appraisal and somatic factors [62]. The cognitive theory says that the emotion is regulated by external stimuli evaluation in a particular context. The second theory says that the bodily response causes emotional reactions but not to the external stimuli. Emotion can be represented by using two approaches; discrete and continuous. There are six basic emotions (*happiness, sadness, anger, fear, disgust* and *surprise*) that are universal across cross-culture [61]. Different emotions would be derived as a combination of these basic emotions. FERS must be derived in person independent settings to recognize these universal facial expressions. Such an FERS would recognize facial expressions irrespective of subjects age, gender, culture, race, etc. To incorporate this requirement in FERS the classifier is derived by training features extracted from face images from training samples which does not contain any overlap in testing samples with respect to a subject (person). In person dependent setting overlap exists in training and testing samples. For instance Figure 1.9 shows how the face database is divided to get training and test sets so that person independent or person dependent classifier would be trained.

It can be observed that in case of person independent there is no data sample of one person in both the sets. However, there are data samples of one or more persons (P1, P2, P3) facial expression is present in both the sets. A multi-class classifier could be derived using these training sets which may classify the facial expressions. The obtained classifier is evaluated using unseen testing set. The implication of the setting is the classifier obtained in person independent is not biased towards recognizing person but person's facial expressions. This setting would test the ability of the feature extraction method (face representation) representing features which are relevant to the recognition of facial

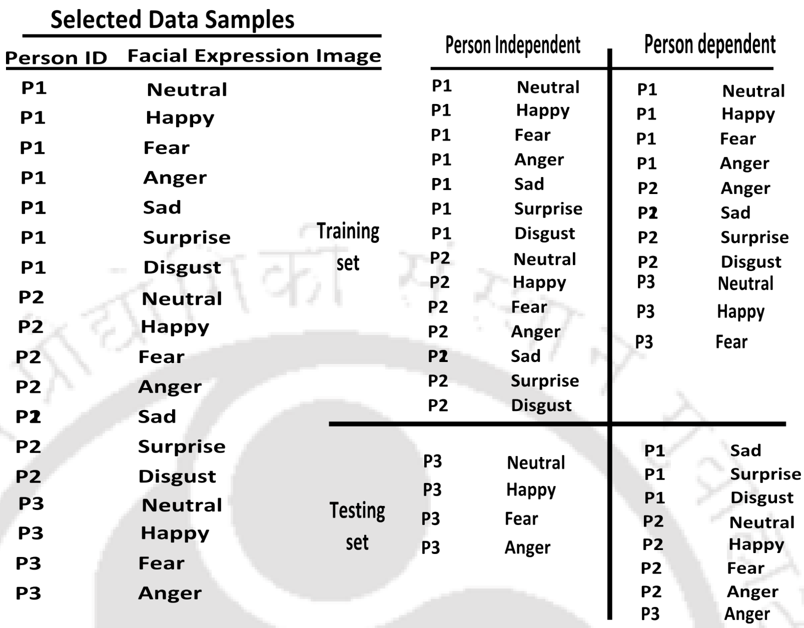

Figure 1.9: Image database is divided into training and testing sets to derive person independent and person dependent classifier.

expressions independent of person's face. In case of the person dependent setting the obtained classifier would be biased as it had seen the face related features while training. A universal generic recognition of facial expression system that is independent of person age, gender, cultural appearance etc, which represents real world scenario could be derived in person independent setting.

In this thesis classifier is trained in person dependent and independent scenarios.

1.7 Intersections of the Framework

There are many commonalities between the components of the face recognition and facial expression recognition system and both the systems suffer from common challenges. The Figure 1.10 (a) shows the intersection in difficulties needs to be addressed by each system while Figure 1.10 (b) shows the intersection of the components of the both systems. In FR and FER literature systems are derived addressing the issues like pose, illumination, occlusion and facial expression separately with different feature extraction techniques and classifiers. This thesis ideate to propose common feature extraction techniques that would be applied in both FR and FER so as to integrate these systems together to enhance

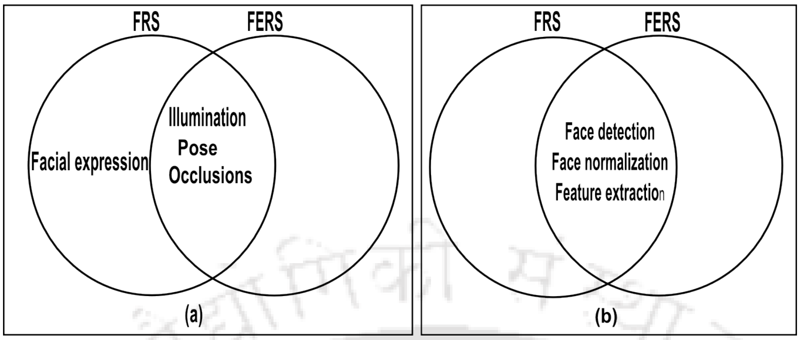

Figure 1.10: Intersection between challenges and components of FR and FER systems.

the performance of resultant system. However, before integrating FR and FER, both the systems need to throughly experimented to understand the behavior and measure individual performance of feature extraction techniques applied to each system. **This thesis uses a convention to mention the size of an image, scale of an LBP based operators and grids in terms of** $width \times height$

1.8 Contributions

The goals pursued in this work prominently consider deriving novel feature extraction techniques that could be used for face representations and evaluate in the context of face recognition and facial expression recognition in the presence of localization and registration errors. The thesis focuses on these following objectives:

- Proposing an heuristic based localization method to remove the inconsistency of the face detection subtask.
- Proposing novel Asymmetric Region based LBPs to extract texture features from the face images.
- Proposing and using a Modified Convolution method to derive a face representation which enhances the available labeled image codes to derive a concatenated histogram.
- Evaluating the proposed feature extraction techniques in the context of face recognition.
- Using feature extraction methods for evaluating the performance of facial expression recognition in the presence of localization and registration errors.

1.9 Overview of the Thesis

The rest of the thesis is organized as follows:

Chapter 2 outlines the survey of various Local Binary Pattern based features used to derive facial representation in the context of face and facial expression recognition.

Chapter 3 introduces different basic LBP based features along with the proposed novel Asymmetric Region LBP features. Modified convolution technique is described to obtain enhanced facial representation.

Chapter 4 and 5 describe face and facial expression recognition experiments and databases that are used for the same with different experiment settings. Finally, **Chapter 6** summarizes the work and findings along with suggestions for future research.

1.10 Chapter Summary

This Chapter discuss various aspects, challenges and components of Face Image Analysis along with the description about face and facial expression recognition components and interaction between the components. Further, the Chapter outlines the commonality between the challenges and components of face image analysis that can be integrated. Successful integration of FR and FER systems to design a face image analysis would serve in various real world applications. Different researchers have designed FR and FER systems using different approaches to handle pose, illumination, facial expression and occlusions with different feature extraction techniques to derive face representations. These works do not exploit the commonalities prevailing among challenges and components of these systems. Thus there is a need to pay attention developing new feature extraction techniques which are robust and can be applied in both FR and FER systems as they form important components capable of influencing the performance.In addition, the knowledge of prevalent feature extraction techniques would enhance the possibilities of proposing new techniques. Besides, it outlines the contributions of the thesis and organization the thesis. A literature survey of different feature extraction methods used in face and facial expression recognition systems follows in the next chapter.

Chapter 2

Literature Survey

This chapter illustrates different feature extraction methods used in FRS and FERS. It is well known that FR and FER systems can be categorized into two types depending upon the type of input image processed by the feature extraction methods. The system is said to be global if feature extraction method processes whole (global/holistic) input image or it is said as local if it processes input image part wise (local/region). Moreover, it is also known that systems which incorporate local feature extraction techniques would give better performance than holistic ones. The survey illustrates both local and global feature extraction methods used to derive FR and FER systems. The survey pays more attention to local texture based feature extraction techniques as they are well established due to their better performance compared to their counter parts in both FR and FER domains.

2.1 Face Recognition

Global feature extraction methods have made considerable contributions in FR and have established a favorable position. Though, PCA, LDA artificial neural networks [55] and ICA [86] are used extensively, often they are criticized for their inability to handle variations in pose, illumination and occlusion. They suffer from the challenges of handling large amounts of training data due to the need of high computations. Further, to exploit the dimension reduction capabilities and reduce the affects of variation, some researchers have employed these global methods to extract features from patches (local) of the input image i.e., modular PCA [87]. These methods require more preprocessing techniques to make them robust against variations. Recently local appearance based texture extraction techniques LBP and its variants have attained a lot of attention of researchers in face image analysis due to their ability to represent face efficiently, robust to variations and less computation cost. These methods enable to obtain a global spatially enhanced face representation processing at pixel level using local patches for each pixel.

2.1.1 Survey on LBP based Operators in FR

The work of Ahonen *et.al* [15] introduced a texture based feature extraction technique in face image analysis. The authors ideated the application of LBP and its derivatives such as $LBP_{(P,R)}$ a circular LBP and Uniform LBP LBP^{u2} [88] with the idea that the face is made of micro structures like, spot, line, edges, etc. Since its inception a number of LBP operators are proposed and used in different application areas such as face image analysis, image and video retrieval, visual inspection, motion analysis, biomedical and aerial image analysis and remote sensing [70]. In [89] an Improved LBP (ILBP) is used to detect the face under Bayesian framework with multivariable Gaussian model of face and non-face. A Multi-scale Block LBP (MBLBP) [7,90] is used for face detection and face recognition. The authors have used Adaboost learning algorithm to select the most effective uniform features to construct the classifier. The authors claim several advantages like robustness, encoding of micro and macro patterns of the face and computationally efficient by the use of integral image to calculate average. Tan *et.al* [91] have proposed Local Ternary Pattern (LTP) to derive face representation for face recognition under difficult lighting conditions. A face verification framework is experimented combining Gabor wavelets features with LTP using score level fusion. In [92] the authors have used Differential LTP (DLTP) for face recognition which is robust to noise and illumination. They have conducted experiments with linear and nonlinear dimensionality reduction and subspace learning techniques for multispectral face recognition.

Wolf*et.al* [93] have used Three Patch (TPLBP) and Four Patch LBP (FPLBP) for face identification and verification. The authors have evaluated similarity learning method performance in comparison with descriptor based methods and found that descriptor methods would out perform similarity based methods if appropriate learning method is chosen. Center-Symmetric LBP (CS-LBP) is proposed in [94] and the authors have illustrated experimentally that CS-LBP is better than other region based descriptors. Although variants of LBP operators are proposed, many works have proposed to use Gabor filtered images to derive LBP histograms to enhance the performance of FRS. Volume based Gabor Binary Patterns (V-LGBP) is presented in [95]. This operator encodes index of the Gabor filter together with spatial dimensions. Another operator which encodes three dimensions is proposed in [96]. This operator is further extended and called Three Orthogonal Planes LBP (LBP-TOP) to represent dynamic texture considering 2D image

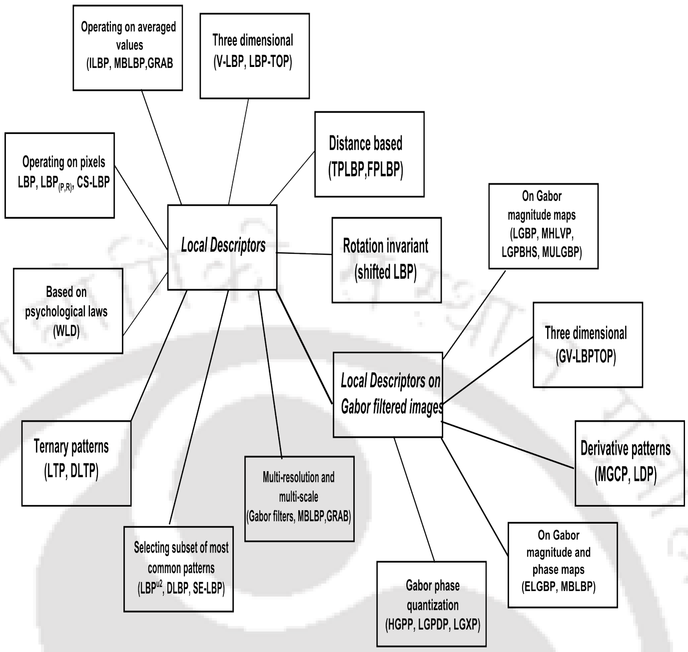

Figure 2.1: Taxonomy of local descriptors [4] used in face recognition.

with temporal dimension. This method may be useful for FR using video sequences. Many Gabor based operator in combination with LBP operator which exploit Gabor phase apart from magnitude have shown greater potential in FR. Multi-resolution Histograms of Local Variation Patterns (MHLVP) [97] is one among such operators and another operator Local Gabor Binary Pattern Histogram Sequence (LGBPHS) [98] has shown good results in FR. Multi-resolution Uniform Local Gabor Binary Pattern (MULGBP) operator is proposed in [99]. It produces long feature vectors and empirical test is not conducted for multi-resolution analysis on Gabor filtered images. Gabor Volume Three Orthogonal Planes LBP (GV-LBPTOP) [100] considers to code scale or orientation of the Gabor filter as third dimension apart from 2D image. An operator Effective GV-LBP (EGV-LBP) [100] was proposed to reduce the computational complexity. In [101] Multi-brand Gradient Component Pattern (MGCP) is proposed which exploits the multi-frequency bands of Gabor filter response of the face image to encode energy variations of Gabor magnitude. Local Derivative Pattern (LDP) is proposed in [102] which encodes directional patterns on variations of local derivatives. LDP is capable of extracting higher order spatial information within the local regions when compared with LBP. Enhanced LGBP (ELGBP) is proposed in [103] which combines Gabor phase and magnitude together for face description. This framework proved that Gabor phase information is also important in face recognition. Histogram of Gabor Phase Pattern (HGPP) [104] proposed for FR which is a combination of Local Gabor Phase Pattern (LGPP) and Global Gabor Phase Pattern (GGPP). HGPP is a combination of spatial histogram derived from LGPP and GGPP. Local Gabor Phase Difference pattern (LGPDP) [105] takes into account the relationships of difference of Gabor phase information in local region pixels. Those features encode the difference between Gabor phase rather than phase information making it robust to local variations in Gabor phase information in the local region. Local Gabor XOR Pattern(LGXP) is proposed in [106]. It encodes Gabor phase information using XOR pattern operator. The LXGP is the fusion of LXP and Gabor phase information. Experiments using block-based Fisher linear discriminant has established the efficiency of the LXGP operator. A Monogenic Binary Pattern is proposed in [107] which encodes monogenic magnitude using uniform LBP and orientations are encoded as quadrant bits. Bereta *et.al* [4] have described various LBP based operators applied in face recognition and a comparative analysis of these operators with their respective advantages and disadvantages are tabulated. A taxonomy of the operators is derived as shown in the Figure 2.1.

2.2 Facial Expression Recognition

Facial expressions are highly spatially dependent as each facial expression is a result of deformation of muscles in a particular region of the face and also facial features like mouth, eyes and nose. FER system should have the ability to extract these deformations from still images or video sequences. Depending upon the type of input image considered for feature extraction FER system can be divided into holistic method which considers whole image at once and local method which considers regions/patches that the image is divided into. FER system can also be categorized depending upon the type of information extracted like motion, deformation of facial feature and displacement of fiducial points from full or regions of the still or video sequence images. Motion extraction methods focus to extract the facial changes from the video sequence of facial expression. These methods could be applied holistically i.e., whole facial expression or locally considering regions that divide the facial expression images. There are three prominent methods of motion extraction viz: *difference-image*, *feature point tracking* and *dense optical flow*. In *difference image* method features are extracted by taking difference of intensities of each pixel in each frame from the intensities in the neutral frame video sequence of particular subject's facial expression. To increase robustness with regard to variations, frame normalization is necessary to align the reference frame with the test faces. In [108] B. Fasel*et.al* have used holistic image difference to extract the motion of facial changes. Choudhury and Pentland [109] have used Motion field histograms derived from the difference of image from consecutive image frames. The motion features are extracted in terms of displacement of points placed manually in the first frame on facial features likely to produce maximum deformations for example, eyebrow, lips, chin and cheeks in *feature point tracking* method. A feature point may be a window of size $w \times h$ where, w is width and h is height of the window. Feature points are placed in high contrast areas on the facial features so that it is easy to approximate its probable position in the next frame of facial expression sequence. Though this method performs better in terms of computation cost when compared to the other methods but it ignores other activities of facial features except tracking of the feature points. Tian *et.al* [110] have proposed a multistate parametric templates for facial features such as mouth, eyes, eyebrows, cheeks and furrows. These models are placed on the first frame of the facial expression and displacement of the parameters are tracked using color and shape distributions. The classification of AUs is performed by using rule base depending upon the parameters of the templates. The *dense optical flow* estimates the amount of velocity that each pixel in the first frame will attain in the next incoming frame. There are many variations in which holistic and local feature extraction methods are applied to extract directions of the gradients and magnitude. All the methods of motion extraction are sensitive to motion discontinuities.

The deformation of facial features is the result of the changes in shapes and texture of the facial features. The deformation extraction methods would extract these changes in order to represent the deformations. These methods can be categorized into two parts viz: *Image/ appearance based* and *Model-based* [68]. Before extracting the features, feature points are placed on the deformation features such as lips, eyebrows, etc. manually and a model is derived. Active Appearance Model (AAM) proposed by Cootes *et.al* [111] is a statistical model of gray-level appearance and shape of the deformable object with the ability to generalize to any object. This model has two stages: training and tracking. While training, 122 feature landmarks are manually placed on training database then the model parameters are estimated. Using this model the deformation of facial feature can be obtained and classified. Lanitis *et.al* [112] have proposed a method of interpreting facial image using AAM. In [75] Cohen *et.al* have proposed a wire frame model to extract motion features by placing 16 surface patches on face region manually where deformation is likely in the first frame of the facial expression sequence to extract the displacements of these patches in the video. The wire frame model is based on piece wise *Bézier* volume.

Further, these wire frame displacements are classified to specific facial expressions using Gaussian Tree-Augmented Naive Bayes as a classifier. Kostia *et.al* have derived a model to extract texture features using a subspace method based on Discriminant Non-negative Matrix Factorization (DNMF). The motion information is extracted by Candide model and the fusion of both texture and motion features of facial expression are classified using SVM (Support Vector Machine) [113]. These methods are not employed in automatic FERS since they result in prodigious computational cost and localization errors introduced by manual annotations of the feature points to extract the motion and deformation features.

Image/appearance based methods processing locally have attracted many researchers due to their ability to perform better in automatic real time applications with less computation complexity. Though many local feature extraction techniques such as, Gabor wavelets [80, 114, 115], Curvelets [116] and subspace learning methods [117–120] are used in analyzing facial expressions applied for static images locally and holistically they incur high computational cost. Gu *et.al* have proposed a framework in which local multiscale Gabor wavelets features of facial images were extracted from different regions across the face image and radial grid which imitates Human Visual Cortex is used to represent the features. Local Binary Patterns (LBP) [5] and Haar features [121] are also used in facial expression representation. Though the LBP operator was originally proposed for analysis of texture [5], it has been used in different applications. An overview of LBP methodologies applied to different applications are available online [122]. LBP based operators are applied in static images and video sequences of facial expressions. It is used in real time applications due to its low computation cost in deriving features from the facial expression images [123].

2.2.1 Survey on LBP based Operators in FER

Much progress has been made in the research of LBP operators and its application to various domains such as face detection, face recognition, demographic classification and other related applications. A good review of existing work on LBP based operators applied to different domains can be found in [70]. Some of the studies in which LBP based operators are applied for facial expression recognition are briefly reviewed below.

Table 2.1 depicts the studies made in facial expression recognition using LBP based operators arranged in chronological order of year of publication with different parameters used for feature representation and classification. We can categorize different works listed in Table 2.1 into three groups depending upon the type of features used as input to the classifier.

(a) LBP based Features : In this group the LBP based operators and their variants are used to derive feature histogram for facial representation and these features are used to train the classifier without processing further. Feng [124] have proposed a two-stage classification method. In the first stage LBP based operator is used to classify the expressions by Chi- Square Statistics dissimilarity measure is used to match the test samples and averaged templates derived from LBP feature vector for each class of expressions. Then, multi-templates are derived in a similar way which are used to classify the expressions from the first stage using the same dissimilarity measure. Shan *et al.* [79] has used LBP features to analyze low resolution images and the trained model was tested on smart meeting database. Local Ternary Patterns (LTP) [135] and Multi-scale LBP feature histograms [126]are extracted by scaling the rectangular regions over the facial expression images to train a SVM classifier. The input histogram is selected by using the Adaboost learning algorithm [136]. Gritti *et al.* [128] describe facial expression recognition system using LBP and LTP by extracting feature histogram from overlapped rectangular grids from the input image. Further the authors studied the LBP-Overlap and LTP-Overlap in the presence of registration errors which are simulated by adding perturbations using Gaussian noise to the eye distance. Experiments were carried out for different perturbations for eye distance. It was concluded that LBP-Overlap perform better than LBP. Shoyaib *et al.* [129] have proposed weighted LBP histogram and derived features from the image using overlapped rectangular blocks and Adaboost learning algorithm. A multi-class SVM was used to train and classify the input facial expression respectively. Kabir *et al.* [131] have derived input features using Local Directional Pattern Variance (LDPv) which encodes the contrast information of an image. These features are used to classify facial expressions by SVM and Template matching methods. In [74] the authors have proposed a method for recognizing facial expressions by Centroid Neural Network as classifier which is trained using LBP feature histograms. Moore *et al.* [80] have used two sequential stages to study the effect of pose of the face in different range of views of the face to classify the facial expressions. They have used LBP based operators like LGBP (Local Gabor Binary Patterns), Multi-scale LBP and gradient magnitude LBP to derive features to train a multi-class SVM. Ahmed *et al.* [133] have proposed CLBP (Compound-LBP) features for recognizing facial expressions by training a SVM classifier.

Table 2.1: List of studies made in facial expression recognition using LBP based operators.

Type of Operator	DB	NoS	NoE	PI	Image Size	Dynamic	Grids	LoFH	MoFL	Classifier	Measure	RR(%)	YoPub
LBP^U [124]	JAFFE	193	7	Y	128 × 150	N	80	NA	PoE	Coarse-to-fine	9-CV	77	2004
LBP^U + LP [78]	JAFFE	213	7	NS	128 × 150	N	80	20480	M	SVM Linear	10-CV	93.8	2005
LBP^U [79]	CK	1280	7	NS	110 × 150	N	42	2478	M	SVM-RBF	10- CV	92.1	2005
ALBP+Tsallis Entropy [125]	JAFFE	213	7	N	64 × 64	N	10	488	M	SVM-RBF	NA	94.59	2006
BLBP + SLPP [82]	CK	1280	6	N	110 × 150	N	NA		M	K-NN	Euclidean distance	94.59	2006
Multi-scale LBP^U [126]	CK	1280	7	NS	110 × 150	N	42	19824	M	AdaSVM-RBF	10-CV	93.1	2008
CBP +IMED [127]	JAFFE	196	7	NS	64 × 64	N	12	384	M	K-NN	Chi-square	94.76	2008
	CK	1820										94.86	
LBP^U-Overlap [128]	CK	1240	7	Y	108 × 147	N	42	8437	M	SVM-Linear	10- CV	92.9	2008
LBP^U + SLLE [117]	JAFFE	213	7	Y	64 × 64	N	NA	256	M	SVM-RBF	LOSO	71.5	2009
$LBP^U_{(P,R)}$ [129]	CK	NA	6	NS	110 × 150	N	72	4248	M	SVM-RBF	10-CV	98.69	2010
LDP [130]	CK	1224	6	Y	110 × 150	N	42	2352	M	SVM-RBF	LOSO	96.4	2010
LDP_V [131]	CK	1224	6	NS	110 × 150	N	42	NA	M	SVM-RBF	7-CV	96.7	2010
		1632	7									93.1	
LBP^U [132]	JAFEE	210	7	Y	64 × 64	N	16	4096	M	SRC	10-CV	62.9	2010
LBP + Curvelet [116]	JAFFE	213	7	N	NA	N	NA	NA	M	KNN	Chi-Square	93.69	2010
	CK	NA										90.33	
LBP^U [74]	JAFFE	213	7	N	130 × 150	N	42	2478	A	CNN	Chi-square	91.43	2010
LBP^U+KDIsomap [118]	JAFFE	213	7	NS	110 × 150	N	42	2478	M	SVM-RBF	10-CV	81.59	2011
	CK	1409										94.88	
Multi-scale LBP^U [80]	BU3DFE	48000	7	Y	80 × 110	N	64	30208	A	SVM	10-CV	65.02	2011
	Multi-PIE	4200			100 × 100							76.7	
LGBP [80]	BU3DFE	48000	7	Y	80 × 110	N	64	30208	A	SVM	10-CV	67.96	2011
	Multi-PIE	4200			100 × 100							82.1	
LBP^U + Gabor [114]	JAFFE	210	7	Y	NA	N	NA	NA	M	SVM-Linear	LOSO	77.62	2011
CLBP [133]	CK	1224	6	NS	110 × 150	N	42	2478	M	SVM-RBF	10-CV	94.2	2011
LBP^U + DCT [134]	JAFFE	213	7	Y	64 × 64	N	NA	NA	IP	SVM-RBF	LOSO	74.87	2011
LBP + Gabor [115]	JAFFE	213	7	N	NA	N	42	10206	PoE	Ensemble of SVM	10-CV	96.2	2011
	CK	1281	7									99.4	
LBP^U + SRC [83]	CK	470	7	NS	110 × 150	N	42	2478	PoE	SVM-Linear	10- CV	95.24	2012
LBP^U+VTB [123]	CK	348 seq	7	NS	64 × 64	N	10	2752	A	SVM-Poly	10-CV	97.2	2012
						Y						93.68	
ARLBP [8]	JAFFE	213	7	N	64 × 64	N	16	4096	A	SVM-RBF	10-CV	95.7	2012
	FGNET	1009										79.46	
EARLBP [9]	JAFFE	213	7	Y	64 × 64	N	16	4096	A	SVM-RBF	10-CV	60.35	2012
	CK+	1236								SVM-Linear		83.09	
LBP^U + PCA [119]	Personal	50	7	NS	24 × 24	N	NS	NA	M	SVM-RBF	NS	93.75	2012
LBP^U [120]	JAFFE	213	7	N	110 × 150	N	42	2478	A	DKLLE	Kernel ED	84.06	2012
	CK	1409										95.85	
PLBP [81]	CK+	309 seq	6	Y	NS	Y	5	295	A	2-NN	ED	96.7	2013

Legend: **ToP**= Type of Operator, **DB**= Database, **NoE**=Number of Expressions considered, **NoS**= Number of Sample images considered, seq= Sequence, **PI**=Person Independent, **N**=No, **Y**=Yes, **NS**=Not Specified, **LoFH**= Length of Feature Histogram, **NA**=Not Applicable, **MoFL**= Method of Face Localization, **M**=Manual, **A**=Automatic, **PoE**= Position of Eyes, **IP**=Interal Projections, **CNN**= Centroid Neural Network, **RR**= Recognition Rate, **LOSO**= Leave one subject out, **K-CV**=K fold Cross-Validation, **ED**= Euclidean Distance, **YoPbu**=Year of Publication

In [81] authors have proposed a framework which may be deployed in real time application and Pyramid LBP (PLBP) for facial expression recognition. They have extracted feature histograms from salient regions of the face and used a classifier to divide the facial expression into two groups depending on their salient facial regions. The other two classifiers are used to classify the features from respective groups. For example, feature histograms extracted from the mouth region is used to classify *Happy, Surprise* and *Sad* while to classify *Anger, Disgust* and *Fear* the feature histograms were extracted from mouth and the eye regions.

(b) LBP with other Features: Facial expression classification is challenging because of intra class variability i.e., Happy expression image of the same person will be different due to influence of different artifacts such as illumination, pose, scale and noise. To address these variabilities different features need to be extracted from the images like texture, geometric shapes, edges and their orientations to represent facial expressions. The LBP based features represent local texture of the image ignoring other structural information. In order to derive robust FERs many studies have considered to fuse global and local features. Liao *et al.* [125] have proposed fusion of the texture features and global appearance features that could increase the performance of FERs. The first set is obtained by using extended local binary patterns (ELBP) in both intensity and gradient maps and computing the Tsallis entropy of the Gabor filtered responses. The global features are obtained by performing nullspace based linear discriminant analysis on the training face images. Fu *et al.* [127] have embedded IMage Euclidean Distance (IMED) with Centralized Binary Pattern (CBP). The facial expression images are transformed using IMED, then CBP features are extracted by dividing the transformed images into grids of different sizes. Saha *et al.* [116] have used digital curvelets transforms to the face image at a specific scale and orientations. Local binary patterns are extracted from the selected curvelet sub-bands to form the descriptive feature set of expressions. Bafandehkar *et al.* [114] have used Successive Mean Quantization Transform (SMQT) features to detect facial salient regions of the face like two eyes and the mouth. They have obtained Gabor features at different scales and orientations. The final LBP histogram was derived by encoding the Gabor images. Moore *et al.* [80] have derived several feature sets which are combinations of LBP, variations of LBPs and Gabor features to study the effect of intrinsic pose in facial expressions at different views and resolutions. Guo *et al.* [134] have derived two sets of feature combination from two layers of images. The first set of features of the image is derived from salient regions by LBP and DCT (Discrete Cosine Transform) coefficients and the other set is derived by down sampling the salient regions of the images. Ji *et* *al.* [123] have proposed two features viz. Vertical Time-Backward (VTB) and moments which extracts general shape changes and motion texture details. These features represent spatiotemporal behavior of the facial expression or the dynamic deformation of facial features. Feng *et al.* [78] have used LBP features along with the Linear programming technique to classify facial expressions.

(c) LBP with Sub-space Learning: High dimensionality and redundancy in input data may reduce the classifier performance due to the curse of dimensionality. Sub-space learning is a method of transforming input features of high dimensionality into low dimensionality feature subspace such that the ratio of the between-class distance to the within-class distance is maximized in order to reduce the effect of the curse of dimensionality. In this group, the authors have considered to reduce redundancy and dimensionality of the input feature space to increase the facial expression recognition rate. First, the LBP based feature histograms are extracted from input images and maps the data set in the high dimensional space is mapped to the lower dimensional space such that certain properties like, global geometric shapes and neighborhood are preserved. Shan *et al.* [82] have investigated linear subspace methods such as Locality Preserving Projections (LPP), Orthogonal Neighborhood Preserving Projections (ONPP) and Supervised LPP (SLPP) for facial expression analysis. The authors have experimented using Boosted LBP (BLBP) features which are LBP based histograms selected by Adaboost algorithms and found that Supervised LPP (SLPP) is superior in supervised methods, while ONPP performs best in unsupervised methods. Ying *et al.* [117] have used LBP histogram features mapped into manifold space using locally linear embedding (LLE) and Supervised LLE (SLLE) method to train a SVM classifier. Zhao *et al.* [118] have proposed a non-linear kernel discriminant isometric mapping (KDIsomap). It is used to perform non-linear dimensionality reduction on the extracted local binary patterns (LBP) facial features and produce low-dimensional discriminant data representations and classify the facial expressions. Zhang *et al.* [83] have used Compressive Sensing (CS) theory to construct a Sparse Representation Classifier (SRC) which is trained with LBP feature histogram. The authors have concluded that SRC performs better when compared to other classifiers. Luo *et al.* [119] have investigated dimensionality reduction using principle component analysis on LBP feature histogram from whole face and from mouth region and have inferred that this would give better performance using SVM as classifier. Zhao *et al.* [120] have proposed discriminant kernel locally linear embedding (DKLLE) for facial expression recognition using LBP feature histogram. Further, the authors have compared the results obtained with other subspace learning methods to conclude that DKLLE performs better than other

learning methods.
It is apparent from Table 2.1 that most of the studies used CK and JAFFE databases while a few other used BU3DFE [137], FGNET [138] and Multi-PIE [139]. Some of the researchers have concentrated on recognizing person independent facial expressions in which the input database is divided into training and testing sets such that there is no overlap of facial images in both sets. The image and grid size having *width* less than *height* are considered by most of the works with a few exceptions of square image and grid sizes. The length of the feature vector depends on the number of grids and number of bins in the histograms of each grid. Prior to recognition of facial expression, face has to be localized in the input image. Face localization methods has to be consistent because LBP based operator feature extractions are sensitive to face localization errors discussed in Chapter 5. Majority of the work has localized the face manually, which puts restrictions on automatic FERS. Support Vector Machine (SVM) is deployed as a classifier in LBP based FERS, as it is suitable for real time applications, curse of dimensionality and the skewness in the number of available facial expressions per class in the database do not affect the performance adversely. Many research groups have considered subspace learning for facial expression recognition. Many of these subspace studies have not specified explicitly about person independent recognition which is one of the important criteria for ideal FERS.
It can be observed from Table 2.1 that FERS performance is higher on CK, CK+ and JAFFE databases with person dependent and manually localized expression faces. The work [123] satisfies most of the criteria of ideal FERS except person independent criterion. Many of the FERS described in the literature does not possess the ideal functionalities except [80] and [81]. These systems can be used in real time applications by automatic localization of face and feature extraction for person independent facial expression recognition with a reduced feature vector length.

2.3 Intersection of Feature Extraction methods

The texture feature extraction based on LBP to derive face representation for FR and FER system is shown in Figure 2.2. There are more operators proposed for FR when compared to FER. Moreover, Gabor based operators are considered to be better in FR but it seems these operators are not good at representing facial expressions and to our best knowledge none of these operators are used in FERS. However, there exists a group of operators which can be used in both the systems. One could derive a face image analysis system

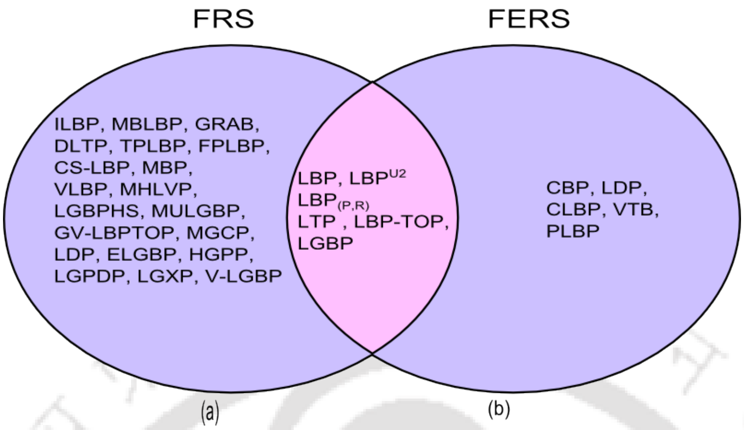

Figure 2.2: Intersection of LBP based operators for face image analysis.

by integrating FE and FER using any single operator to obtain single face representation instead of separate representations. This idea would reduce computation cost and time to respond to the input image by the integrated system.

The existing common operators do have disadvantages apart from the advantage of less computation cost except LGBP that would encourage to deploy in real world applications [4]. The LBP based operators in the intersection are not flexible to represent macro patterns of the face as they increases the length of the feature vector which reduces the performance to the system due to the curse of dimensionality. LGBP and LBP-TOP are computationally more demanding though it is flexible to contain macro patterns and they also increase the length of the resultant feature vector. Hence there is a scope to develop new operators which will be useful for both FR and FER systems.

2.4 Chapter Summary

As discussed in this Chapter several FR and FER systems employ different feature extraction techniques. However, a few studies have considered deriving face representations that can be used in both systems. There are many parameters that affect the performance such as, input image size, grid size, poses of the face images, type of matching metrics used to take a decision, type of classifier used, etc. However, robust feature representation methods need to be derived to minimize these variations which could cater to a single face representation for both FR and FER systems. The following Chapter addresses different local feature extraction and representations.

Chapter 3

Feature Extraction and Representations

3.1 Introduction

Feature extraction or feature representation is a special form of dimensionality reduction method. Human beings tend to store only prominent distinguishable (without redundancy) features of objects which are used to detect, describe and retrieve that object in later stage instead of storing full object in the memory. With minimal information humans are able to process complex operations with respect to the object. Similarly an image contains redundant data in it and it is rational to remove redundant data in order to reduce the storage and processing cost. The process of extracting prominent discriminative features from an image is called *feature extraction* or *feature representation*. The prominent features could be color, texture, edges, corners, blobs, shapes and motion. The features extracted must be good enough to produce the result similar to the result when full input data is processed. The performance of a system depends largely on feature extraction techniques as decide what and how discriminating information is extracted from the image. In face image analysis tasks FR and FER need to address common challenges and contain common components deriving a robust feature extraction technique that could be important to increase the performance when both the tasks are integrated. The feature extraction techniques are categorized as holistic and local depending upon whether the methods operate on whole image or in parts to extract the features. Recently local feature extraction methods have proved to be efficient in FR and FER systems. LBP based texture extraction methods have gained much attention and established a firm position in the face

image analysis domain due to less computational cost and the ability of their use in real time applications. LBP operators are applied in different domains such as face detection, face recognition, demographic classification and other related applications. In [70] authors provide a review of recent works of LBP based operators applied to different domains.
LBP based operators are applied in static images and video sequences of facial expressions. They are used in real time applications due to their low computation cost in deriving features from the facial expression images [123]. Further, very few LBP methods are applicable both in FR and FER systems but these methods have shortcomings which could motivate to propose new methods that would be used in FR and FER systems. However, these proposed methods must be evaluated in independent FR and FER system to understand the behavior and ability so as to apply in the integrated face image analysis to handle common challenges of individual system.

3.2 Local Binary Pattern based Operators

The LBP based operators are computationally efficient and can be applied in real time applications. However, they have a few drawbacks. For example, the basic LBP [5] shown in Figure 3.1. The LBP resultant code can be expressed for a pixel (x_c, y_c), of the face image in decimal form as follows:

$$LBP(x_c, y_c) = \sum_{i=0}^{7} S(p_i - p_c)2^i \tag{3.1}$$

where, p_i and p_c are pixel gray values of surrounding ($i = 0 \ldots 7$) and central pixel value respectively and the function S(x) is defined as:

$$S(x) = \begin{cases} 0 & x < 0; \\ 1 & x \geq 1, \end{cases} \tag{3.2}$$

The LBP is not scalable and hence cannot represent macro patterns of the face efficiently. Moreover it is not robust to noise. To mitigate the shortcomings of LBP, Extended-LBP (ELBP) [6] was proposed which has good discriminative ability when compared to the basic LBP operator.

Figure 3.2 shows an example of ELBP operators which label the pixels of the image by considering center pixel intensity and thresholding from its circular neighborhood pixel intensities. The circular neighborhood of the operator is defined by two parameters *(P, R)* where P is number of pixels to be considered on the circle of radius R around the central

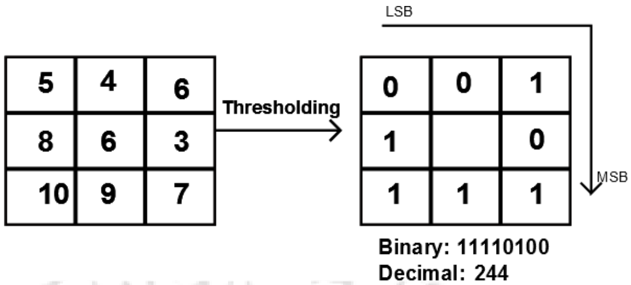

Figure 3.1: Example of 3×3 sized basic LBP operator [5]

pixel to be labeled. As R increases P also increases. As a result the range of the value of the code for pixel (2^P) also increases. It has been shown that a subset of 2^P binary patterns can be used to represent texture of images which are called *uniform* patterns and denoted as $LBP^{U2}_{(P,R)}$. A *uniform* pattern is a LBP code in which it contains two bitwise transitions from 0 to 1 or 1 to 0 when circularly shifted. For example, 00000100, 00000000 are *uniform* patterns since they contain 2 and 0 transitions, whereas 00101101 and 00101011 contain 5 transitions respectively and they are called *non-uniform* patterns. Ahonen *et.al* divided 2^P patterns into two types *uniform* and *non-uniform* patterns. The *non-uniform* patterns can be collected in one bin and *uniform* patterns would be assigned one separate bin for each type of *uniform* patterns (58 types) that would result in 59 bins.

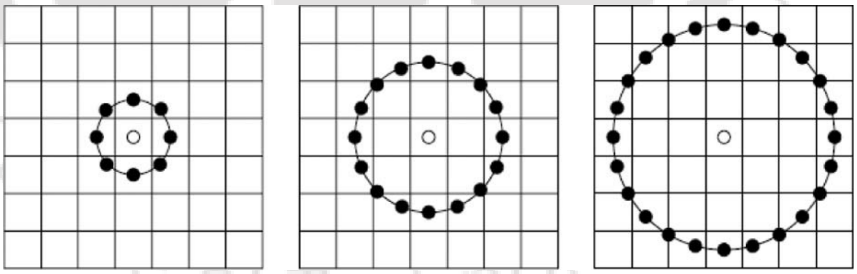

Figure 3.2: Example of ELBP operator [6] The circular neighborhood (P, R) of the size (8, 1), (16, 2) and (24, 3).

If we consider a grid of the image labeled with ELBP, it can contain pixel values in the range $0 \ldots (2^P - 1)$. The histogram derived on this grid will contain 2^P number of bins. Thus the feature dimensionality of the concatenated histogram is dependent on the number of grids that the image is divided and P. The length of the feature vector will increase if the number of grids or P increases. Hence the idea of *uniform* patterns is used to reduce the histogram length.

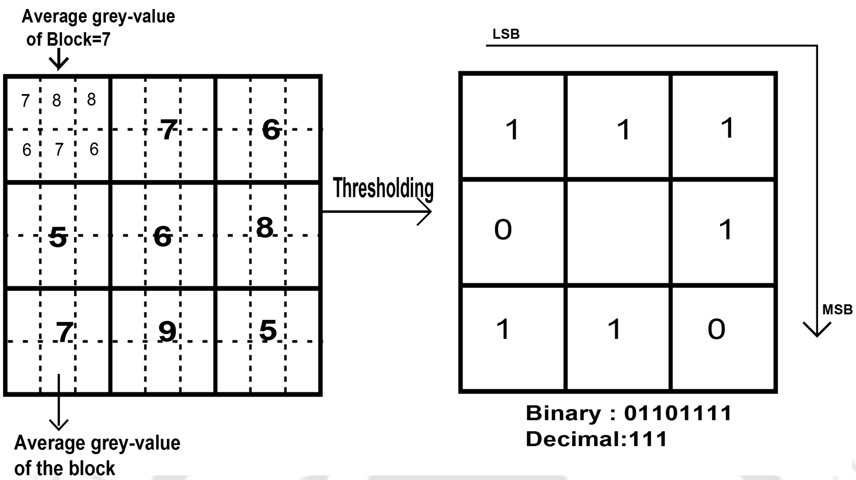

Figure 3.3: Example of MBLBP [7] with 3×3 sub-regions of the size 3×2.

Zhang *et al.* [7] proposed Multi-Block LBP (MBLBP) for face detection applications. This operator has discriminative property and it is robust to noise as it considers average values of surrounding pixels. Figure 3.3 illustrates labeling of pixel of the image using 3×3 sub-regions of size 3×2. MBLBP operator labels pixel of the image by considering average value of intensities from neighboring sub-regions and thresholding with average intensity of central sub-region. It is scalable and the dimensionality of the feature histogram is constant as the size of subregions increase. MBLBP is computationally efficient since the average gray-pixel value of a region is computed using integral image or summed-area tables [140].

3.2.1 Convolution Method

Convolution is a local operation where a window of some size would be used in order to perform some operation and assign a resultant to the central pixel considering the pixel's neighborhood. Figure 3.4 (a) shows the process of convolution in which 3×3 LBP operator is scanned row wise on an image of size 13×14. Figure 3.4 (b) shows the resultant 12×13 labeled pixel image.

The well known convolution method in LBP based feature extraction has a constraint that the operator boundary should not cross the image boundary. Ahonen *et al.* [15] had suggested that the constraint would decrease the influence of boundary pixels which would enhance the discriminative ability of the operator. However, there is a loss in the number of available codes which are used in deriving a histogram from each grid. The statistical

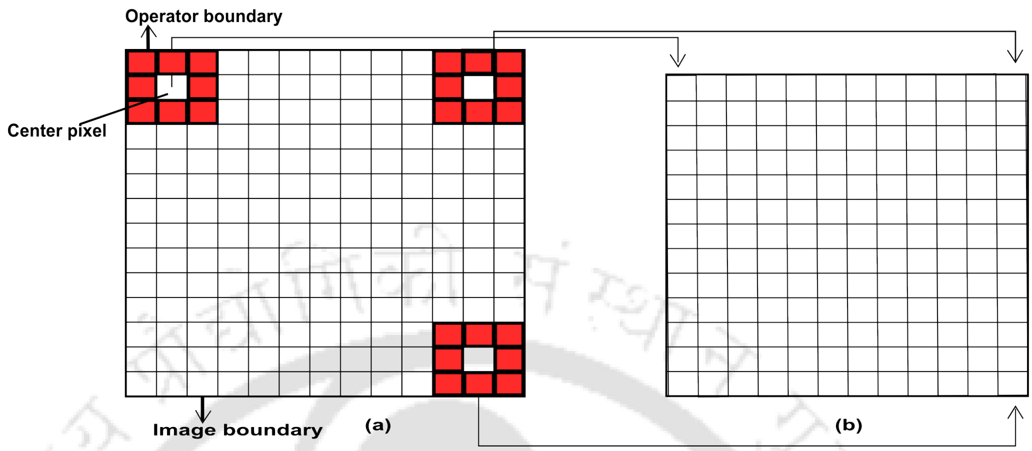

Figure 3.4: Convolution method using LBP based operators.

nature of the histogram is dependent on the number of bins and frequency of the occurrence of the bin values. The decrease in the number of LBP codes would affect the histogram and inturn the performance of the face image analysis system.

3.2.2 Traditional Face Representation

The well known LBP based face representation is illustrated in the Figure 3.5. LBP histogram could be derived from whole face image by convolving with LBP based operators. However, such histogram would encode only the texture patterns without spatial information. This is due to the fact that histogram does not preserve the spatial information but frequency of occurrence of LBP codes. To incorporate spatial information the face image is divided into fixed size grid of subregions and from each subregion a histogram is derived after convolution with a LBP based operator. For whole image a face representation is obtained by concatenating each region histogram to get feature concatenated histogram vector.

This is a popularly known and widely used face representation or feature extraction method for face image analysis. The face representation is a very important component of any face image analysis task as it largely determines system performance. It is widely known that the property of face representation is dependent on convolution method, type of mask or operator used and number of grids that the face image is divided. The feature extraction using LBP based operators are not convolved from boundary pixels of the image to reduce the influence of boundary pixels on the feature histogram. For example, convolution of ELBP operator starts at pixel (R, R) of the image where R is the radius

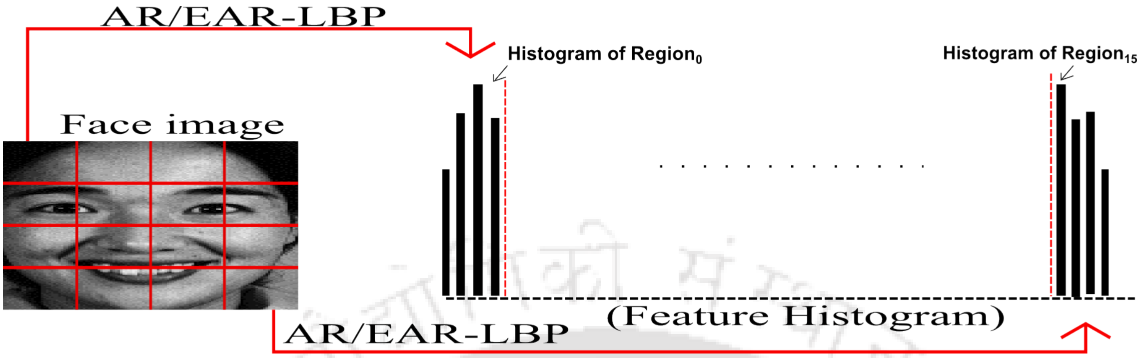

Figure 3.5: LBP based face representation [6].

of the operator. As the radius R increases, the number of ELBP codes decreases which may affect the number of available labeled codes. The available labeled code per image can be approximated as $(n - R) \times (m - R) \times$ (number of grids), where n and m are width and height of the grid respectively and $R < n, m$. As the radius R increases, the feature histogram may contain bins with less frequency of gray level values which may affect the statistical property of the histogram which in turn reduces the performance of face image analysis system. In case of MBLBP, the number of local binary codes per grid is approximated as $(\alpha - (3s - 1)) \times (\beta - (3s - 1)$ where, α and β are width and height of the grid, s is a variable used to represent the condition $m = n = s$ where, m and n is width and height of single region of the operator respectively then operator's size is $3s \times 3s$. As s increases the number of local binary codes decrease. Moreover, for a particular scale, if the operator contains similar sized regions as the scale of MBLBP operator increases the number of labeled codes per grid decrease. The size of the operator when $n \neq m$ is $(2n + 1) \times (2m + 1)$ where, n and m are width and height of the grid, which limit the analysis for odd numbered scales. Further, not all cropped image size and grid size combinations are possible in the case of non-overlapped grid, because the cropped image size should be divisible by the grid size. Hence, the size of the operator also influences the number of labeled codes per image, size of cropped image and the grid size.

There is a well known proposition in the literature of LBP based face analysis *"The performance of the FR or FER systems would increase, as the number of grids increases"* [15,88]. However, choice of the number of grids (and thus size) would depend on application and also the type of normalization applied on the image. It is evident that the size of the grids should not be too small as the histogram does not represent spatial information properly [4]. It is possible that the proposition might have been derived by experimenting with small scale LBP based operators. Although many researchers have proposed scalable

Table 3.1: Comparison of LBP based operators.

Operators	Advantages	Disadvantages
LBP	• Tolerance to monotonic gray-scale illumination changes. • Computationally efficient. • Good descriptor of micro-patterns such as spot, line end, edges and corners.	• Cannot capture macro-patterns efficiently. • Sensitive to noise. • Operator is not scalable.
ELBP	• Captures micro-patterns at higher scales. • Tolerance to monotonic gray-scale illumination changes. • Operator is scalable. • Rotation invariant.	• As neighborhood size increases the size of feature histogram and computation cost increase [80]. • Number of local binary codes per image is $(\alpha - R) \times (\beta - R)$, as R increases number of local binary code will decrease. Where R is radius of the operator, α and β are width and height of the image respectively.
MBLBP	• Captures micro and macro patterns. • Operator is scalable without increasing the feature histogram. • Robust to noise as it considers average of the region to encode the patterns.	• As the neighborhood increases, texture gets averaged and discriminative information is lost [80]. • As the scale of the operator increases number of available code decreases. • Not used for facial expression recognition.

LBP based operators to the best of investigator's knowledge none of the studies in FR or FER consider experimenting with large scales of the operator. This may be due to the fact that larger scales of (for example, ELBP) would produce feature histograms with larger lengths which would affect the performance due to the curse of dimensionality. This would be motivation to investigate whether the proposition holds for face representations derived using larger scales of the operators. To study the effect one must propose new LBP based operators such that the scale of the operator is independent of feature length to negate the effect of curse of dimensionality. MBLBP is one such operator which produces constant feature length histograms with increased scale of the operator. Further, it could be argued why smaller grids lead to statistically unreliable histograms [4]. A comparison of different LBP based operators which are within the scope of this research would corroborate the motivation to propose new convolution method and LBP based operators which are used in both FR and FER systems.

It is evident from the Table 3.1 that basic LBP operators have a few drawbacks.

Moreover, a single operator is not capable of describing images with respect to different scales of the operator, reduction in length of the feature vector, its discriminative ability and robustness. Thus, before applying LBP features in an application, it requires to choose an operator by evaluating its properties. Further, LBP features lack generalization across databases. Its performance depends on different factors as listed in the first row except the last two columns of the Table 2.1. These shortcoming need to be addressed and some advantages need to be exploited.

3.3 Proposed Face Representation

To investigate the proposition discussed above, it seems it is not possible with the traditional face representation and present LBP based operators as they are less explored. A Modified convolution technique and two Asymmetric Region based operators which incorporate advantages of the above discussed LBP variants operators are proposed. If a face representation would be derived using the proposed Modified convolution instead of traditional convolution method with any LBP variants to increase the available codes. From here onwards the term code(s) refers to the decimal code that is derived using LBP based operator by convolving face images. Experiments with larger scales of LBP operator would be possible to derive face representation like in Figure 3.5. Further, it is possible to explore whether the proposition holds for larger scales.

3.3.1 Modified Convolution (MC) Method

The traditional convolution method labels the pixels of a face image using LBP based operators ignoring the boundary pixels. However, the scale of the operator and the size of the grid that an image is divided may increase the number of pixels that are ignored. The scale of operator and smaller grids reduces the available codes to derive the feature histogram for the face image as the grids and scales are varied. The histogram statistical property is largely determined by the number of bins and frequency of the occurrence of values of the bins. The decrease in available codes would greatly influence the statistical property. Hence, the idea is to increase the available LBP codes to derive the histogram and to study the effect of the idea on the number of grids, size of the grids and size of the operators. Further, it may facilitate to verify the proposition for larger scales of the operators.

Modified convolution technique is proposed to code every pixel without ignoring the boundary pixels in the face image and also grids to derive the face representation for face

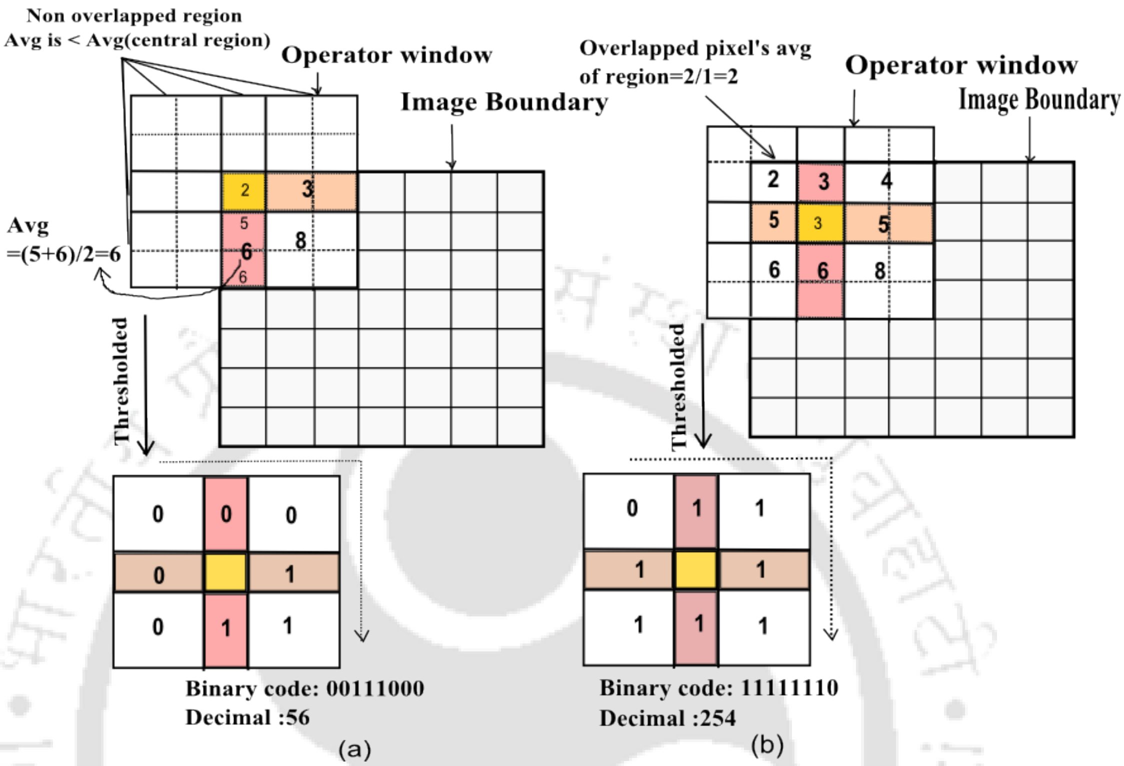

Figure 3.6: Modified convolution method [8]: (a) 5×5 ARLBP thresholding when region of the operator is out of bound with the image boundary. (b) 5×5 ARLBP thresholding when region of the operator partially overlaps with the image boundary.

image analysis. This technique increases the available labeled codes per image and the size of the operator does not influence the size of the cropped image and the grid size since the boundary constraint is relaxed. Figure 3.6 shows the modified convolution method. In Figure 3.6 (a) a 5×5 ARLBP operator is placed at pixel (0, 0) of the image. It can be observed that top three and left two regions are out of bound with the image boundary. The average of gray pixel values for these sub-regions is considered as zero and pixel is (0, 0) coded as decimal 56. In Figure 3.6 (b) a 5×5 ARLBP operator is placed at pixel (1, 1) of the image. In this case, the average for these sub-regions of the operator is calculated considering only the overlapped pixels of that region with image grid and pixel (1, 1) coded as decimal 254. The bold numerals represent average values of the respective image pixels overlapped with operator sub-regions.

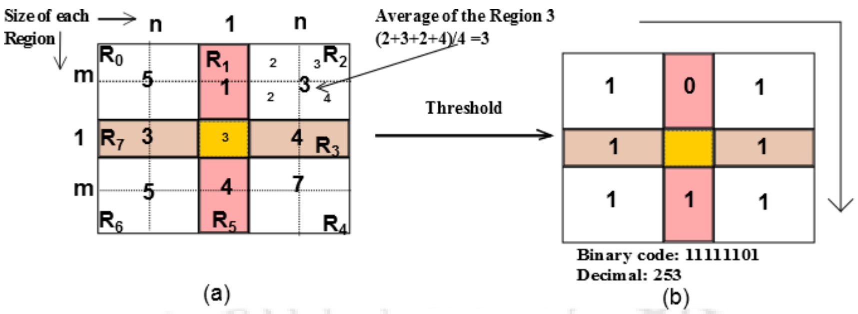

Figure 3.7: Example of 5 × 5 sized, Asymmetric Region-LBP [8]

3.3.2 Asymmetric Region LBP (ARLBP)

Among the operators discussed earlier, a single operator alone does not address the issues of discriminative ability, dimensionality of the feature histogram and scalability. To address these issues using a single operator, ARLBP was proposed. The operator is motivated from MBLBP. The ARLBP operator considers average intensities of sub-regions of different sizes around a pixel to be labeled. ARLBP is scalable which can capture dominant feature at larger scales. The length of feature histogram obtained is constant and it is equal to the number of grids the image is divided multiplied by 2^P where, P is the number of neighborhood regions of the operator. ARLBP consists of different sized regions which reduce the loss of texture information since average of pixel intensities are calculated for different sizes. The average gray-pixel value of ARLBP regions is calculated using summed-area tables, making it computationally efficient.

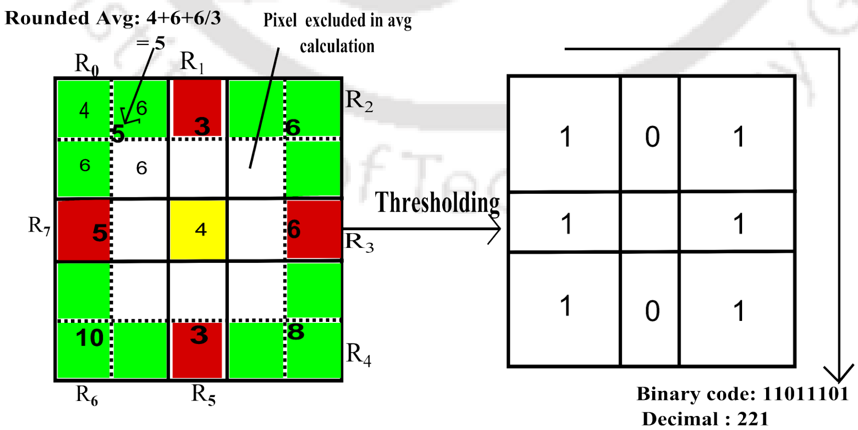

Figure 3.8: Example of 5 × 5 sized EARLBP [9]

Figure 3.7 shows ARLBP operator consisting of nine regions, R_i (i= 0...7) and central region (not labeled). The size of the regions R_0, R_2, R_4 and R_6 varies in both horizontal and vertical directions. Whereas, the sizes of the region R_1 and R_5 vary in vertical and R_3, R_7 in horizontal directions. The size of the central region is invariant to changes to the sizes of the sub-region and size of the operator. It consists of 4-$n \times m$ (R_0, R_2, R_4 and R_6), 2- $1 \times m$ (R_1 and R_5), 2-$n \times 1$ (R_3 and R_7) and 1- 1×1 (central) rectangular sized regions and their size variations are not similar as MBLBP, hence the name Asymmetric Region, where n and m are width and height of the sub-regions. The size of the operator can be generalized to $(2n + 1) \times (2m + 1)$. With $n = 1$ and $m = 1$ ARLBP will be a 3×3 basic LBP operator. Figure 3.7 shows how a pixel of an image is coded using ARLBP of size 5×5, as decimal 221. Each region's average is calculated as shown in the region with bold numerals except for central region for illustration.

Formally, ARLBP resultant code can be expressed for a pixel (x_c, y_c), of the face image in decimal form as follows:

$$ARLBP(x_c, y_c) = \sum_{i=0}^{7} S(avg_i - avg_c)2^i \tag{3.3}$$

Where, avg_i and avg_c are average gray values of surrounding region i, ($i = 0 \ldots 7$) and central region respectively and the function S(x) is defined as:

$$S(x) \;=\; \begin{cases} 0 & x < 0; \\ 1 & x \geq 1, \end{cases} \tag{3.4}$$

The ARLBP is coding micro patterns at smaller scales as it is similar to LBP operator at the size of 3×3, it also codes macro patterns at larger scales as it is scalable without increasing the feature histogram. Robust to noise as it considers averaging of pixel intensities to label and threshold the pixel values.

3.3.3 Extended Asymmetric Region-LBP (EARLBP)

ARLBP has the disadvantage that averaging of sub-region pixels intensity value reduces its discriminative ability as neighborhood sub-region sizes increase. In order to preserve the discriminative ability of the operator the effect of averaging needs to be reduced. The effect of averaging can be reduced by considering the boundary pixels of each sub-regions. With this idea the EARLBP was proposed which inherits the properties of ARLBP and enhances the discriminative ability by considering the outer boundary pixels of the sub-regions. EARLBP is computationally efficient since the average of outer boundary gray-pixel value of a region is computed using the integral image as in [8]. EARLBP operator consists nine sub-regions as shown in the Figure 3.8, R_i ($i = 0 \ldots 7$) and central region (not labeled). These sub-regions have the properties of ARLBP [8] except that the boundary pixels of each region is considered in order to calculate average of each region. Figure 3.8 also shows how a pixel of an image is coded using EARLBP of size 5×5, as decimal 221. The rounded average of each region is shown within the region with bold numerals except for the central region for illustration. The non boundary pixels are in white color (example pixel value 3 in region R_0).

Given a center pixel (x_c, y_c) of the face image, the resultant EARLBP code can be formally expressed in decimal form as:

$$EARLBP(x_c, y_c) = \sum_{i=0}^{7} s(bavg_i - bavg_c)2^i \tag{3.5}$$

Where, $bavg_i$ and $bavg_c$ are rounded (to higher value) average gray values of the outer boundary pixels of surrounding region i, ($i = 0 \ldots 7$) and the central region respectively. The function $S(x)$ is similar to Equation 3.4. EARLBP is similar to the ARLBP and it is also scalable and capable of representing macro and micro patterns.

3.4 Chapter Summary

This Chapter illustrated the importance of feature extraction and representation methods. Further, it describes LBP based operators and the traditional face representation methods and their shortcomings. It may be noted that larger size operators and their effects on the performance of FR and FER system have not been explored well. The same is also true regarding the effect of grid size, number of grids and number of LBP codes. To cater to the study, new operators along with a Modified Convolution method has been illustrated. In the subsequent Chapters these techniques are explored in FR and FER system individually to test their efficacy.

Chapter 4

Face Verification

4.1 Introduction

Face recognition is one of the challenging areas of research to the difficulties that need to be addressed by a system to produce reasonable performance. A successful system would find useful in applications like security, surveillance, authentication, etc. The FRS may be configured in two modes viz: face verification and face identification. The configuration would depend on the application that the system is deployed. To derive a good system one needs to pay attention to the design of the components such as feature extraction and classifier which would be invariants to variation. Although myriad global and local feature extraction techniques are proposed, the local methods have established a noticeable position in the domain. It seems texture based LBP operators have been used by the researchers due to their computational simplicity and their ability to be applied in various domains. In fact these features could be applied in both FR and FER systems. These LBP based face representations are sensitive to different parameters like size of the image, size and number of the girds that a face image is divided, size of the operators, type of classifier, etc. Experiments were conducted to illustrate the need for the Modified Convolution (MC) and proposed face representation to derive an FR system configured in face verification mode. PCA is used to reduce the dimensionality of face representation and face templates of individual subject during the training and Mahalanobis Cosine distance is used as matching measure for verification.

4.2 Why Face Verification?

The configuration of FRS depends on the application that it is designed for and deployed. These configurations may be interpreted with a different perspective. Face Verification is process of confirming or rejecting the claimed identity of the face image. Whereas, face identification is processing of retrieval of identity of given face image. It is hard to tell which task is superior as it depends on the application. For instance, if two persons are about to start a conversation they perform face identification using the descriptions they have about each other and retrieve labels/names. If both the parties successfully retrieve the names of each other then they feel comfortable and the interaction would be interesting and personalized. On the contrary, if both the parties conduct face verification of a common friend the conversation would be same as before. In fact, face verification would be used to extend the knowledge base of known persons i.e., enrollment in FRS. Face verification is more important when providing access based services, for example accessing of the airport, banking, biometrics, etc. The face authentication system puts constraints on the image acquisition process and environment (i.e,*ISO/IEC 19794*) to enhance the performance as it is unable to handle the variations (see Section 1.2). It is apparent that FRS in verification configuration would need to address the facial expression difficulty which is complementary to FERS (See Figure 1.10). This thesis envisages an integrated AFA system in which FER linearly precedes FR. This could be achieved if a common feature extraction method is proposed which can be applied to derive FR and FER systems. The FR configured to verify the neutral face which is rendered to remove facial expressions by the FERS could be use successfully to authenticate face with facial expressions. FERS should be carefully designed by deriving and evaluating a feature extraction technique robust to variations like pose, illuminations and occlusion. This would warrant deriving an FR which is configured in verification mode. Hence, evaluating the proposed face representation in face verification mode is considered.

4.3 Description of the Databases used

One needs to be aware of the nature of the data samples for better understanding of variations like pose, illumination, occlusion, etc. that it contains. New methods may be derived to deal with these variations. Three databases are used to conduct face verification experiments. The description of the databases used in these experiments is as follows:
The JAFFE database [2] consists of 213 TIFF images sized 256×256 posed by 10 Japanese

models. Each subject has posed 7 facial expressions (6 basic expressions and 1 neutral) photographed with constant illumination and uniform background. An example set of images is shown in Figure 4.1. The database is used to check the ability of FRS invariance to facial expression inherent in the database. The Olivetti Research Laboratory (ORL)

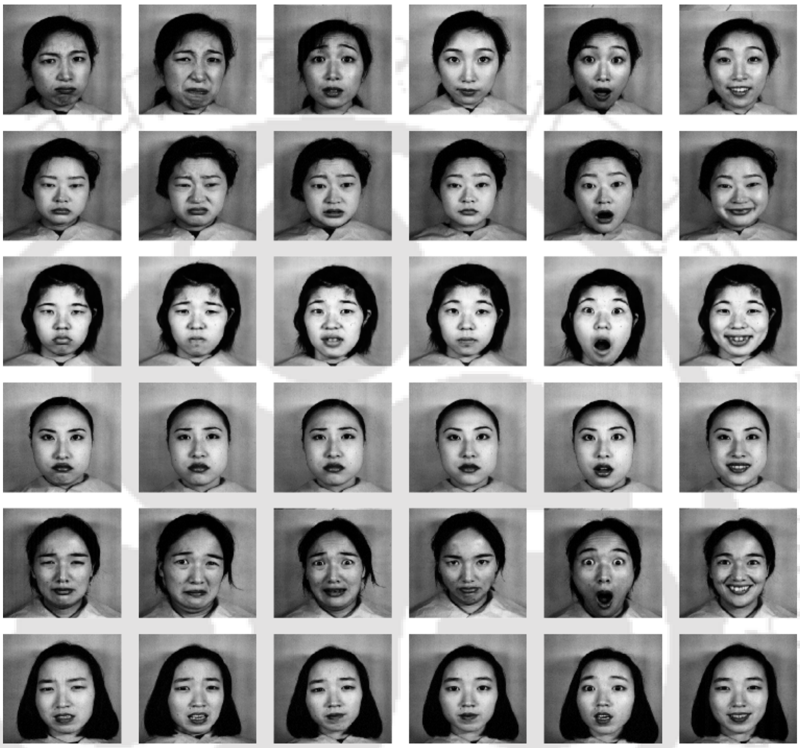

Figure 4.1: Sample images from the JAFFE database.

database [141] contains 40 distinct subjects. Each subject has 10 images in different poses. The images are taken by introducing some variation like time, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses) but with constant background. The faces are mostly frontal position with tolerance for some side movement. Moreover, there is also some variation in the scale upto about 10 percent. The images are gray scale and with a resolution of 92×112 pixels. This database is very challenging as it contains all the variations that an FRS has to address.

The INDIAN face database [142] consists of 40 subject images with different poses. These poses are looking front, looking left, looking right, looking up, looking up towards left, looking up towards right and looking down. The images are blended with four facial expressions - neutral, smile, laughter and sad/disgust. These images were taken at constant laboratory conditions with uniform lighting and background condition. The database does

Figure 4.2: Examples of face images from the ORL database.

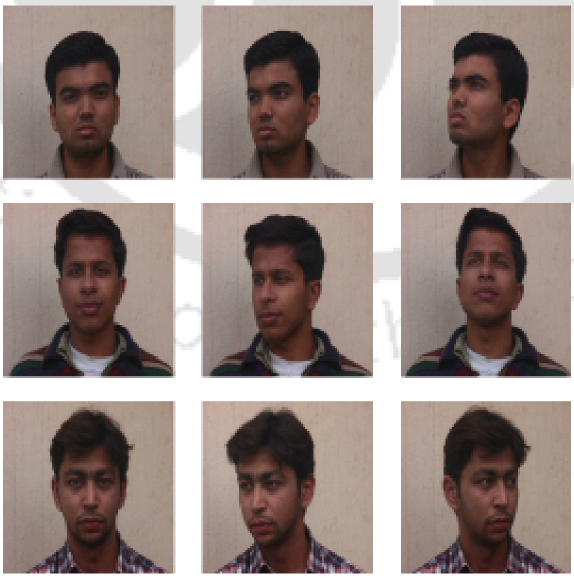

Figure 4.3: Examples of face images from the INDIAN database.

not introduce variations with respect to scale. The files are in JPEG format. The size of each image is 480 × 640 pixels, with 256 gray levels per pixel.
These databases have different variations such as pose, illumination, occlusion and facial expressions. The INDIAN database contains facial expression with large variations in pose. The JAFFE has variations in facial expressions with frontal pose. ORL database contains all the variations and it is very challenging database when compared to other two databases. The use of these three image databases for FR would help us substantiate the generalizing ability of the proposed operators across databases.

4.4 Experimental Process

The experimental process is divided into two phases: training and testing phases. Figure 4.4 shows the different component for testing mode after deriving face templates for every subject for face verification. The training and testing phases are illustrated in the following steps:
i. Data Preparation: In this step face images are selected manually from the benchmark database to construct training and testing data samples from which face templates are derived. In FR domain the label/ID of the image is important as it serve as class label in verification and identification modes. For each subject some of the images would be included into training set and the remaining in the testing set. The face sampling is called person dependent since both training and test data sets include images from a particular subject(person).
ii. Training: In this step a face template per subject using training data set may be derived by extracting features from these images. For each image of each subject face is localized using FaceTool [28]. The localized face may be normalized to remove the effect of illumination. The face is resized and divided into number of grids of different sizes to construct a histogram from each grid on the pixels labeled using the proposed operator and modified convolution or traditional convolution methods. A feature histogram is obtained for each image by concatenating each grid histograms as shown in the Figure 3.5. All feature histograms of training set are projected using PCA linear subspace method to reduce the dimensionality as discussed in Chapter 1. The face template is derived by projecting each person's images on to covariance matrix of reduced PCA projection.
iii. Testing: This step is called verification or matching. For each test set, image face may be localized and normalized. Features are extracted and a concatenated feature histogram is derived. The feature histogram is projected on covariance matrix of reduced PCA

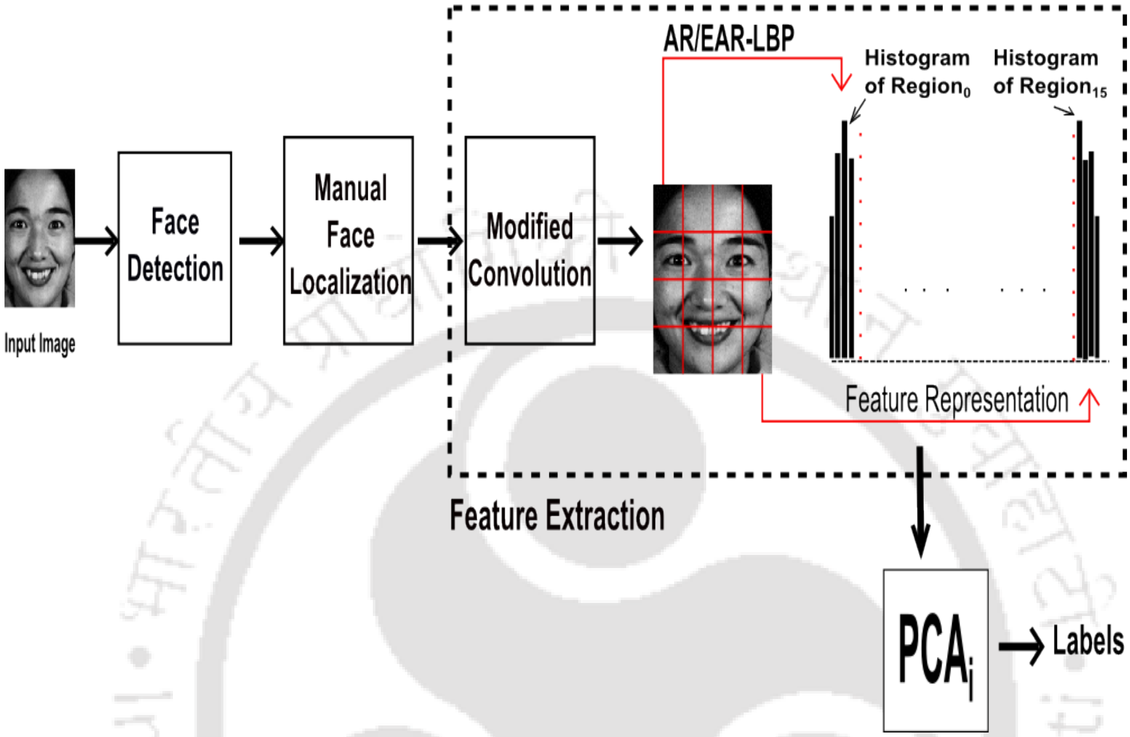

Figure 4.4: Face recognition process.

projection and matching is performed using different distance or similarity measures and the test image is assigned a label same as the label of the closest matched face template. The detailed description of FR with PCA subspace learning is illustrated with examples in [58].

4.4.1 Experimental Design

The following experiment cases are designed to investigate the effect of ignoring boundary pixels by traditional convolution technique and suitability of proposed operators in FR domain.

1. **Case 1: Effect of Modified Convolution(MC)**
   This case is designed to study the effect of different parameters like operator scale, number of grids an image is divided into and the number of available labeled codes. Further, MBLBP is used in experiment to verify the effect of MC on number of available codes.

2. **Case 2: Face Verification**

In this case the ability of the proposed operators ARLBP and EARLBP using the proposed MC are evaluated in the face verification mode of FRS. MBLBP with MC is experimented to justify that increase in available code increases the performance of the FRS. The experiments were conducted in the presence of localization errors and person dependent settings.

4.5 Experiments and Results

The experiments were conducted using the FRS in which different components are integrated as shown in the Figure 4.4. Two stages are followed to construct the FRS configured in face verification mode. In first stage, the face template is derived following the steps illustrated in Section 1.5. Prior to the training steps the training image and testing image sets must be constructed i.e, data preparation. Face templates are obtained in training step then the testing is performed using Mahalanobis Cosine Distance (MahCos) and Verification Rate (VR) is determined. Though the experiments were conducted for different similarity measures like Euclidean, City Block and Cosine distances, VR with MahCos was better when compared to other distance measures in [143] and hence MahCos is considered for experiments.

4.5.1 Case 1: Effect of Modified Convolution (MC)

It is well known in the LBP based AFA that the performance of FR and FER systems would get influenced by a number of different parameters. Specifically, the performance depends on the number of grids, scales of the operator and the resolution (size) of the image. A proposition exists in LBP based FR and FER that "the performance of FR or FER system would increase when number of grids increase". However, the survey suggests that this proposition might have drawn by experimenting with limited combination of parameters. These experiments does not consider to study the effect of number of grids and scale of the operators as experiments were not conducted with larger scale operators. It is evident that due to smaller grid size (hence large number of grids) it would produce unreliable histogram [4]. On the contrary, less number of grids (hence larger grid size) would lead to histograms without proper spatial information embedded in it. Hence, there exists a crux point (threshold) between number of grids and grid size that would optimize performance of the FR or FER systems. The scale of the operator would also influence the performance along with grid size or number of grids. It is difficult to assess the influence

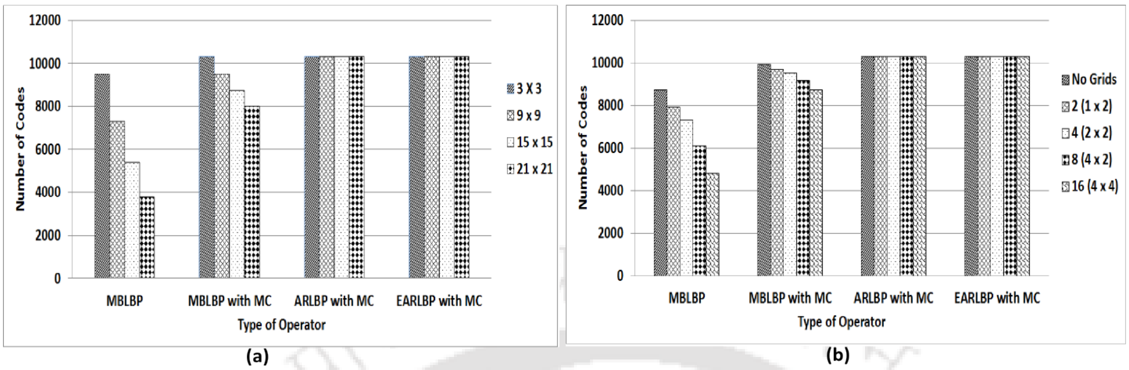

Figure 4.5: Number of available codes for different operators with varying (a) scales of the operator (b) number of grids .

of these parameters on the performance of the systems as they are inter-dependent. However, one parameter stands out i. e., number of available codes which are used to derive feature histograms. The scale of the operator, size of the grid or number of grids and convolution method affect the available codes which in turn affect the statistical property of the histogram features. It is obvious that the statistical property could be affected by either changing the number of bins or width of the bins or the frequency of occurrence of the bin values.

An investigation to study the effect of available codes would help to understand the proposition for larger scales of the operator and the inter dependence between different parameters. To corroborate the effect of other parameters on available codes an experiments was conducted in which available codes are counted and plotted. In the experiment an ORL image of size 92×112 was considered having 10304 pixels. Different types of operators are used to convolve with traditional and MC methods on the ORL image with varying grids and scales of the operators. The scale of the operator was fixed at 9×9 and the number of grids are varied as 1 (92×112), 2 (92×56), 4 (46×56), 8 (46×28) and 16 (23×28). The size of each grid is mentioned within bracket in respective places. Available codes for different types of operators e.g, MBLBP with traditional convolution are counted and a barchart is plotted as shown in Figure 4.5 (a). Next, the number of grids were set at 4 each with size 46×56 and the scale of the operator is varied as 3×3, 9×9, 15×15 and 21×21. The available codes were counted for different types of operators as before and a barchart is plotted as shown in Figure 4.5 (b). It is observed from Figures 4.5 (a) and (b) that for MBLBP operator with traditional convolution method the available codes would rapidly decrease as the number of grids and the scale of the operator increase. This decrease in the available codes is attributed to the pixels that are ignored when traditional convolution method is used. The rate at which the number of pixels are ignored are higher in case of scale variation than in the number of grids. Girish *et al.* [143] conducted the experiment using MBLBP with traditional face representation and have concluded that increase in the scale of the operator and number of grids upto certain extent would increase the Verification Rate (VR). But further increase in the same would decrease the VR. This may be due to the fact that the available codes would decrease rapidly with increase in the scale of the operator and the number of grids. The preposition could be re-framed as "the performance of FR or FER system would increase when the available codes increase". The proposition would fail for larger scales and more number of grids. There is a trade-off exists between available number of codes and number of grids. The available codes per grid would affect the statistical property of the histogram and the number of grid would affect the spatial information embedded in to the concatenated histogram feature vector. It would be interesting to study the effect of spatial information i.e., number of grids and scale of the operator with constant available codes per image. To achieve the constant available codes or coding every pixel of the whole image, the MC is proposed as illustrated in Section 3.3.1. In Figure 4.5 (a) and (b) for MBLBP operator with MC method it is observed that the available codes are decreasing but the rate at which it decreases is less when compared to the MBLBP without MC. However, for ARLBP and EARLBP with MC the available codes are constant per image as the scale of the operator or size of the grid changes. The decrease in the codes with respect to MBLBP with MC is due to the effect of the size of the central region of the MBLBP. As the size of the central region increases the number of available codes decrease. This is due to the fact that the available codes are still dependent on the scale of the central region. On the contrary, ARLBP and EARLBP central region scale is constant though the scale of operator is dynamic. This property of asymmetric region based operator with MC produces constant number of codes per image without ignoring the boundary pixels. When the available codes are constant per image, the effect of number of grids and scale of the operator on FR or FER could be explored.

4.5.2 Case 2: Face Verification.

There are two possibilities to study the effect of the grid and scale: Increase the number of available codes by reducing the ignored boundary pixels or code every pixel using MC so as to achieve constant codes per image irrespective of scale and number of the grids.

MBLBP, ARLBP and EARLBP with MC are experimented to explore the implication of the two possibilities.

Data samples must be prepared before conducting experiments. The training and testing set needs to prepared a priori. For each database image samples were selected to build the respective training and testing sets as follows:

In case of ORL database, for each of the 40 subjects 3 images were selected randomly and included into the training set and the remaining 7 images were included in the testing set. From JAFFE, 3 images were selected with *Sad*, *Angry* and *Fear* facial expression and 7 containing 6 basic expressions (*Happy, Sad, Anger, Disgust, Surprise* and *Fear*) are added to the training and testing set for all 10 subjects respectively. The training data set of INDIAN database was built by adding 3 face images of front pose and around 45 °C right and left pose respectively for each of the 30 subjects out of 40 subjects. In testing set 8 images of different poses of the same 30 persons were included. The second stage of FRS consists training and testing processes. The training process precedes with feature extraction and deriving face representation as discussed in Section 3.3. For each image included in training set of the respective databases, the face is localized automatically using FaceTool [28]. The localization errors are not removed but all images are normalized to common eye co-ordinates using this tool. Illumination normalization was not applied on any database. It was assumed that Asymmetric Region operators are robust to monotonic increasing illumination. To extract the features and to derive face representation, each pre-processed image from INDIAN database training set is resized to 270×200 and divided into 15 grids with size of 90×40 and 25 grids sized 54×40. The JAFFE database testing set localized face images resized to 150×110 and 15 grids of size 50×22 and 25 grids of size 30×22 are divided. In case of ORL localized and pre-processed training images were resized to 112×92 and 4 grids with size 56×46 and 16 grids each sized 28×23 was divided respectively. For all the grid faces of training images in respective databases MBLBP, ARLBP and EARLBP operators were used to convolve each grid and for each scale using MC to obtain concatenated histograms. The face templates were obtained using the method illustrated in Section 1.5. for the test image database to derive feature histogram (i.e., face representation). For each database test feature histograms were matched and verification results for different number of scales of the respective operator are obtained. For each grid size, different scales of operators like 3×3, 6×6, 9×9, 12×12, 15×15, 18×18, 21×21, 24×24, 27×27 and 30×30 are experimented. The VR at 1% FAR is plotted in respective figures.

The MBLBP with MC is considered here to compare the effect of increasing the

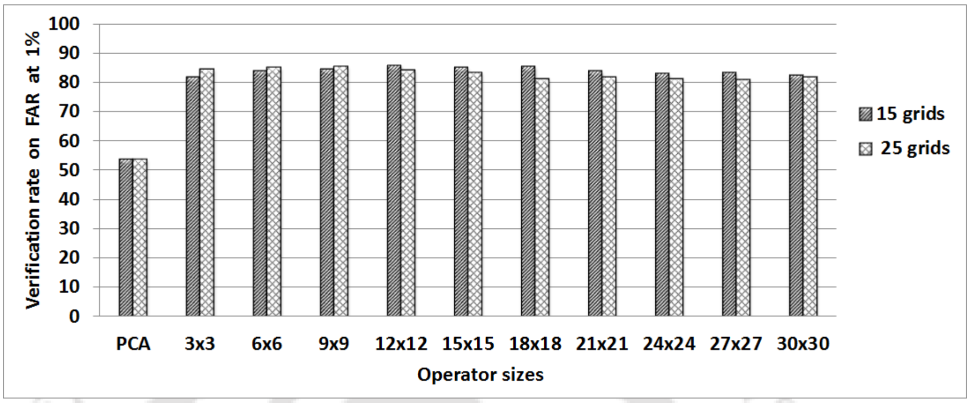

Figure 4.6: VR for 1% FAR with MBLBP for INDIAN database.

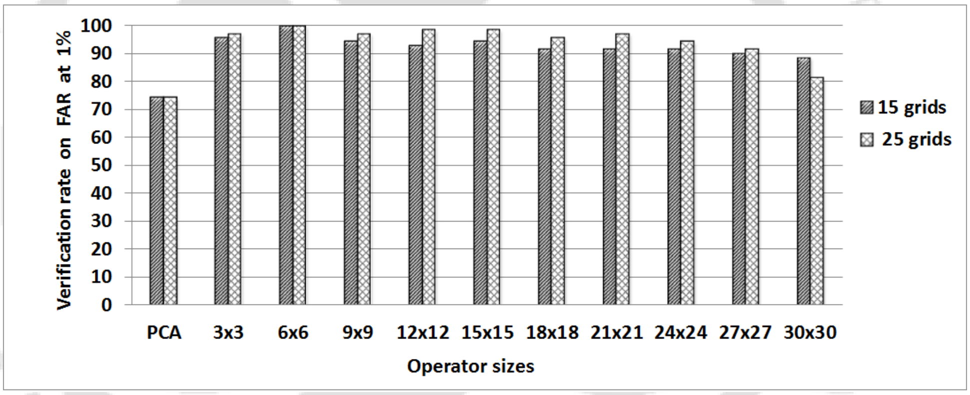

Figure 4.7: VR for 1% FAR with MBLBP for JAFFE database.

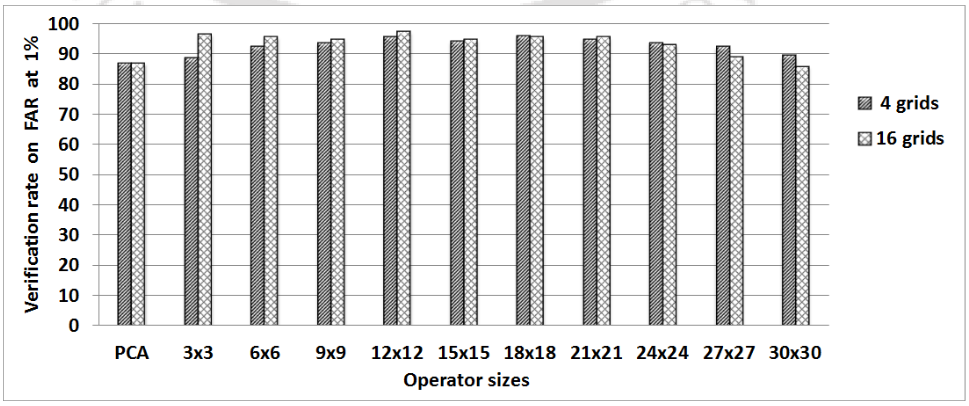

Figure 4.8: VR for 1% FAR with MBLBP for ORL database.

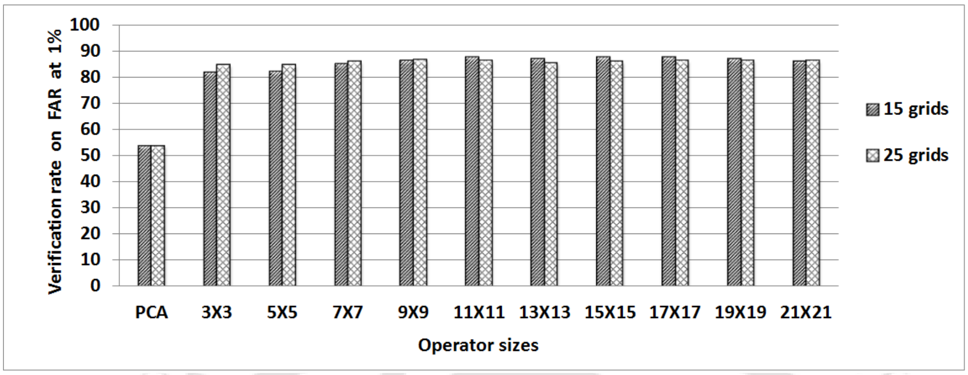

Figure 4.9: VR for 1% FAR with ARLBP for INDIAN database.

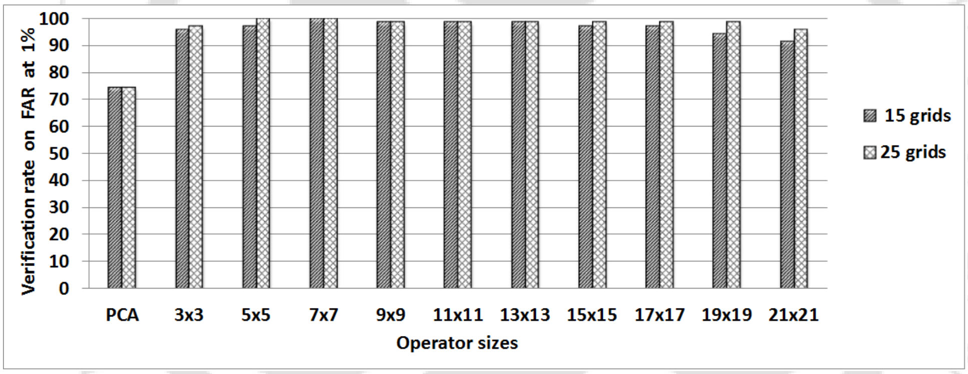

Figure 4.10: VR for 1% FAR with ARLBP for JAFFE database.

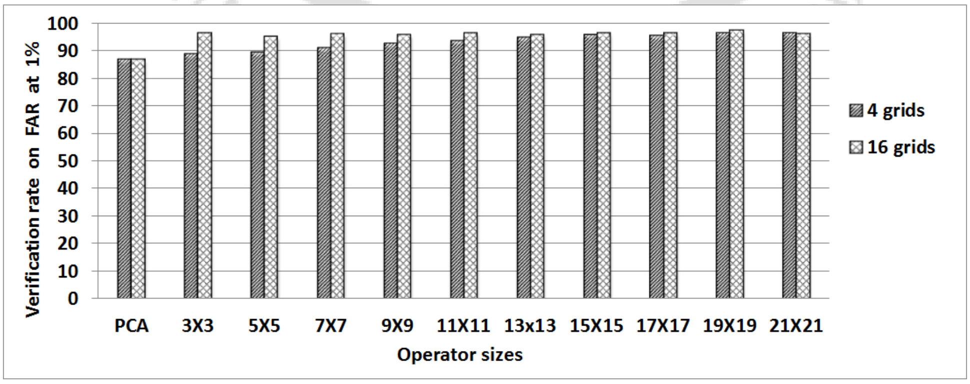

Figure 4.11: VR for 1% FAR with ARLBP for ORL database.

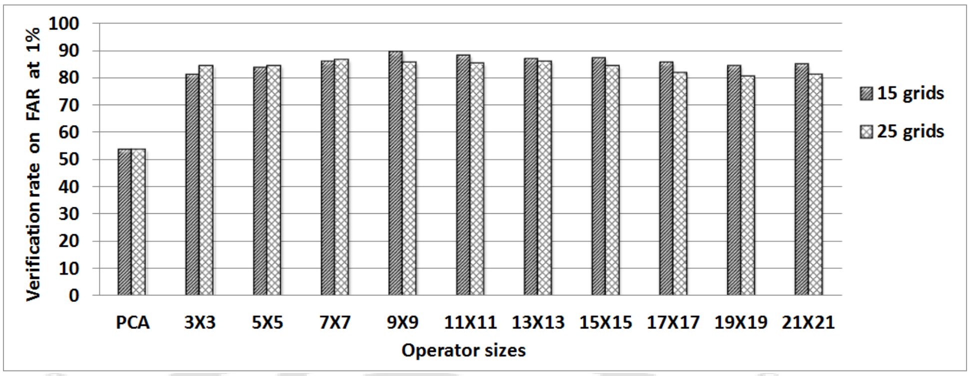

Figure 4.12: VR for 1% FAR with EARLBP for INDIAN database.

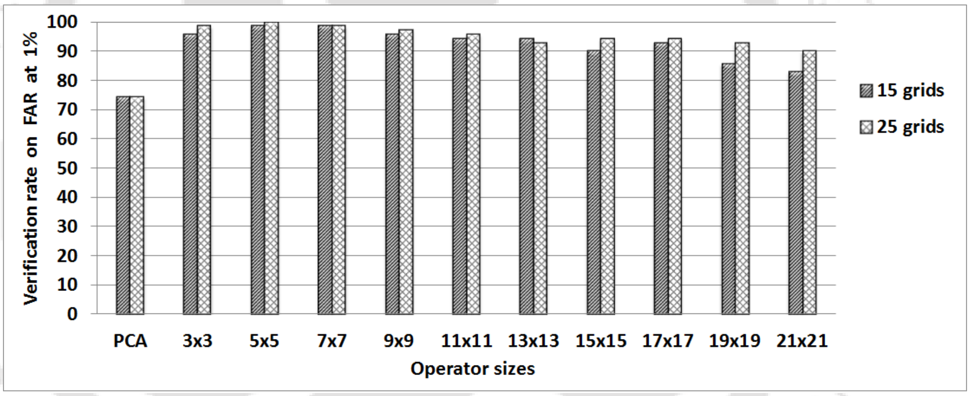

Figure 4.13: VR for 1% FAR with EARLBP for JAFFE database.

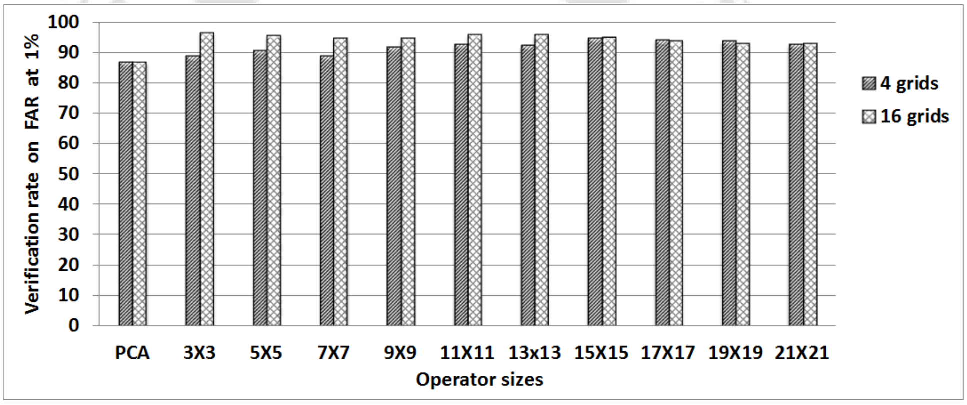

Figure 4.14: VR for 1% FAR with EARLBP for ORL database.

available codes with MBLBP with the traditional convolution method. Moreover, it can also be used to compare with ARLBP and EARLBP operators which code all the image pixels irrespective of scales of the operator and grids. The Figures 4.6, 4.7 and 4.8 show the VR at 1% FAR for different scales of the MBLBP operator for two cases of grid variations for the three databases. It can be observed that in case of 15 grids in INDIAN and JAFFE databases, VR increases with an increase in scale of the MBLBP operator to some extent, and further increase in the scale of the operator leads to decrease in VR. This may be due to two factors like viz: decrease in the available codes as the scale of the operator increases. Another could be the decrease in discriminative ability of the operator due to texture averaging. If the number of grids is increased to 25 (hence decrease in grid size) the VR increases for smaller scales of the operators, as the operator scale increases further the VR decreases. This is due to the fact that the averaging texture property and the number of available codes per grid affect the VR though the available codes per image or across the concatenated histogram is constant. In ORL database similar trend is observed in the case of 4 grid and 16 grids.

The Figures 4.9, 4.10 and 4.11 show the VR at 1% FAR for ARLBP on INDIAN, JAFFE and ORL databases respectively with different operator scales and grid sizes. In 15 grids case, the VR on INDIAN and JAFFE databases increase as the scale of the operator increases but gets saturated when the scale of the operator is increased further. For 25 grids, the VR increases with the enhancement of number of grids but saturates as the scale of the operator increases. In case of the ORL database for 4 grids, VR increases as the scale of the operator increases and then saturates. When 16 grids are considered, the VR is high and as the scale of the operator increases there is no further improvement. The trend "increase in scale increases the VR and increase in number of grids increases VR" is largely visible for three databases.

Figure 4.12, 4.13 and 4.14 show the VR with 1% FAR for different database with different scales and grids for EARLBP operator. For 15 grids, the VR increases with increase in the scale of the operator on INDIAN database. But the VR decreases as increase in the scale of the operator further (observed from the scale 13×13 onwards). Similarly, when the number of grids was increased to 25, the VR followed the same trend for smaller scales of the operator. But further increase in the operator scales decreased the VR. It may be observed that increase in scales would decrease the VR (see from scale 15×15 onwards) which is contrary to ARLBP operator. This may be caused due to the averaging of boundary pixel values of sub-regions of the operator without considering the average of all pixels' gray values in sub-region. In case of JAFFE and ORL database similar trend was observed. However, with respect to JAFFE database (observed scales from (15×15) onwards) the VR in the case 25 grids was higher when compared to the VR in case of 15 grids. This trend was not observed in other two databases. This may be the fact that the EARLBP is sensitive to poses and occlusions present in the ORL and INDIAN databases. On the contrary, the ARLBP is able to perform better than EARLBP at higher scales of the operator. In all the three databases the global PCA representation VR is less than the proposed local texture based operators.

It is possible to analyze the VR of the FR system with respect to the parameters like number of grids (and hence size of the grid) and the scale of the operator. It is difficult to correlate the performance with respect to these parameters as they are inter-related. The variations in the number of grids (and size) affect the available number of codes in the grid and in turn the total number of codes in the whole face representation. The number of available codes per grid changes as the number of grids (hence the size of the grid) changes. The only parameter that gets affected is available codes. Hence this parameter was chosen for the study.

Table 4.1: Maximum Verification Rate for different Operators.

Type of the Operator	**Database used**	**Number of Grids**	**Scale of the Operator**	**Maximum VR in (%) at 1% FAR**
MBLBP	JAFFE	15, 25	6×6	100
ARLBP	JAFFE	15, 25	7×7	100
EARLBP	JAFFE	25	5×5	100
MBLBP	ORL	16	12×12	97.14
ARLBP	ORL	16	19×19	97.14
EARLBP	ORL	16	3×3	96.43
MBLBP	INDIAN	15	12×12	85.83
ARLBP	INDIAN	15	15×15	87.5
EARLBP	INDIAN	15	9×9	89.58

Table 4.1 shows the maximum VR at 1% FAR for different operators. It is observed that performance of all three operators was best on JAFFE followed by ORL and INDIAN databases with different scales and grids. MBLBP with MC performs better when

compared to Grish*et al.* [143] as the authors have reported the VR of 79.23%, 98.57% and 93.21% on INDIAN, JAFFE and ORL database respectively. It is evident that increase in the available codes would increase system VR. Asymmetric Region operator perform better on INDIAN database when compared with MBLBP. The INDIAN database is difficult due to its nature of variations in pose and facial expressions. From the table it is observed that for an application one must choose type of operator, number of grids and scale of the operator for a particular database as the results are different for various settings across the databases. However, if the average of maximum VR over three database is considered, the Asymmetric Region operators performs better than MBLBP which suggests that the boundary pixels would contribute to the performance of the system. The experiments considered only scales with equal width and height of the operators. Experiments with different width and height could also be considered.

4.6 Chapter Summary

This chapter illustrates the importance, experimental design and procedure of deriving an FR system in verification mode. It also describes the need of an MC and its effect on different parameters such as grids and scale of the operator. It details the experiments along with results using different settings. It also presents the maximum VRs for each database using different operators. In the next Chapter these proposed techniques are evaluated in the FER domain and their performance is discussed.

Chapter 5

Facial Expression Recognition

5.1 Introduction

There are two components of FERS that influence its performance i, e., feature extraction technique and classifier. It is difficult to derive a generic feature extraction technique which extracts the features from images of different domains. Nonetheless, proposing a generic feature extraction technique for overlapping domains could be tried with regard to FR and FER. The FR and FER would take similar input images and they have common challenges, components and hence they are overlapped domains. It is observed that the proposed face representation performance is better and comparable with state-of-art FR system. These representations should be explored in case of FERS as well. The face representation would be applied in FERS in different conditions to test the ability of the proposed operators. The conditions are person dependent or independent settings and presence or absence of registration and localization errors. Moreover, the experiments consider to explore only the effect of size (scale) of the operator but not the size of the grid and the number of grids as in the case of FRS discussed in previous chapter. Although, many LBP based scalable operators were proposed, studies have experimented more with varying scales which is revealed in related FER literature. With respect to the classifier, Support Vector Machine (SVM) [113] is used for facial expression recognition since it can be applied in real time applications. It is not affected by the skewness of the class labels and its training time depends on the number of samples rather than the number of features. Training of SVM classifier usually gives a global solution if the optimal parameters are known [144]. Moreover, it is widely used in the LBP literature for its decent classification performance.

5.1.1 Localization and Registration Errors

Facial expressions are the result of deformation of muscles in a particular region of the face. Facial expressions are spatially dependent with respect to the face. Position of the prominent facial regions like face, eyes, mouth, nose and eyebrows must be known prior to extraction of features. Further, the positional information of facial features can be used by the preprocessing algorithms such as normalization process in which the alignment of face images with respect to eye coordinates [80, 145] can be done. Prior to localizing (knowing) the positions of facial feature, the face has to be localized in the given image automatically and that would lead to designing of robust real time FERS without any preprocessing step of face alignment. Face can be localized manually or automatically in the given input static image or video sequence. It is rather easy to localize the face manually but it is more challenging to automate this process. Automatic face detection problem may be addressed easily assuming deterministically single face with constrained background in the given image. In constant background low level features like color of the skin [146], edges and motion [147] can be used to localize the face. In unconstrained background Viola and Jones methods is used to locate face automatically in static images or video sequences. It is one of the best and extensively used face localization method available in the face detection literature. OpenCV library contains the implementations of several automatic localization which can be used to localize face, eyes and mouth. But these methods suffer from inconsistency of localization i.e., for the same input face image, the detecting algorithm gives different locations for each iteration. For example, Figure 5.1 (a) shows that inconsistency of localization leading to different representations (b) in the face feature space. To decrease the effect of localization error on facial expression feature extraction methods a heuristic localization method is used. Registration error is defined as the inconsistency (image of same person looks different in various conditions) in pose, scale and illumination of the images of the localized face. For example, the Figure 5.2 shows before and after pose alignment. As the thesis considers databases with frontal facial expressions, alignment of pose across facial expression is not considered. The alignment process may increase the computation cost and it may also reduce the real time ability of FERS. There exists a trade off between pre-processing or post-processing operations and real time efficiency of FERS. Introduction of more pre/post-processing computations FERS would take more time for recognition and give better recognition with reduction in real time ability. It would be feasible to derive FERS with reasonably good performance without compromising with its real time ability. The idea is to derive feature extraction methods which are robust to these errors. LBP based operators represent textures effi-

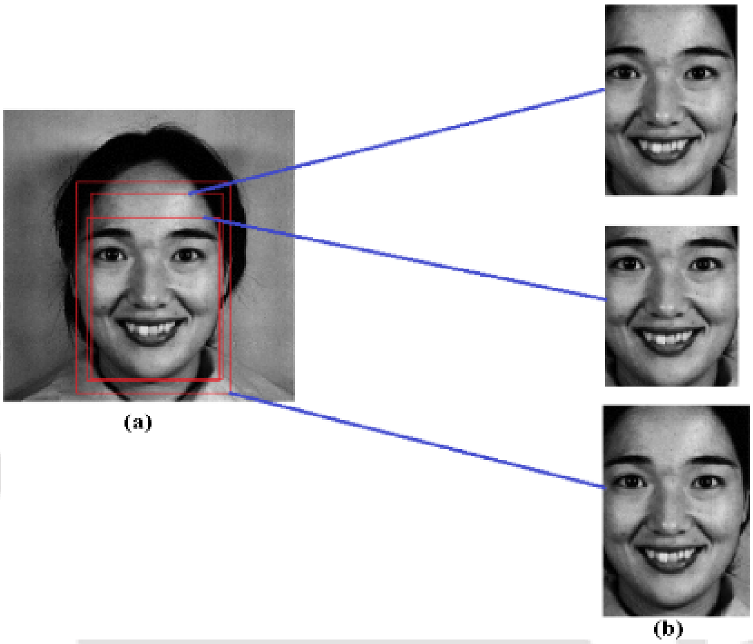

Figure 5.1: Localization Error: (a) Face Localization using Viola & Jones method for 3 iterations, (b) Different Face representations.

ciently, they are robust to monotonic illumination, less computation cost and they can be deployed in real time face analysis applications. The survey has revealed that there are less number of studies in which LBP based operators are applied in presence of registration errors in automatic real time FERS.

To study the suitability of a LBP operator in the presence of face registration errors, Gritte *et al.* [128] have simulated the error by adding Gaussian noise in the range of 0-7% to the distance between eyes. Testing and training sets are derived from the CK database and it was found that the performance using LBP is 77% while in the case of LBP-Overlap the performance was increased by 3-6%. Khan *et. al* [81] have proposed a framework for real time facial expression recognition using Pyramid of LBP (PLBP). The feature histogram vector is a concatenated histogram created by extracting a uniform pattern histogram dividing the stimuli into finer spatial sub-regions by iteratively doubling the number of divisions in each dimension of the stimuli. The framework uses multiple classifiers and maximum recognition rate of 96.7% is reported using 2-Nearest Neighbor (2NN) on CK+ database. Further the framework was tested for generalization to other databases. A maximum recognition rate of 81.9% was obtained when CK+ database was used as training set and FGNET FEED [138] database is considered as testing sample. To

Figure 5.2: Registration Error: (a) Frontal face with inconsistency in pose. (b) After correction of pose [10]

the best of our knowledge most of the existing works on facial expression analysis using LBP based operators except [81] are not able to achieve the ideal characteristics of automatic facial feature extraction and recognition system due to manual face localization (localization errors) and facial expression image alignment (registration) errors. To mitigate these errors most of the LBP based works follow one of the following techniques: (i) the face cropped manually and geometric transforms are applied to align the images with respect to the position of the eyes to reduce the registration errors [80,145] which restricts the automatic feature extraction. (ii) Overlapping grids [145] are used to derive the facial expression representation which in-turn increase the feature histogram length. (iii) Multi-scale [80] and multi-level operators are used which increase the feature histogram length as well. For automatic facial feature extraction there is a need to derive good face localization methods or derive new LBP based operators which are intrinsically robust to registration errors and achieve good generalization capabilities across databases.

5.1.2 Face Localization Method

Facial expression is the outcome of deformation of several facial muscles at a particular portion of a face. For example, lower part of the face muscles will be predominately involved in producing *Happy* and the mouth and eye brows are used to express *Surprise* facial expressions respectively. Hence, feature extraction methods have to extract deformations to recognize the facial expressions from salient regions of the face. Prior to the recognition of facial expressions face should be localized in the given input image. In the context of automatic feature extraction, the face localization and registration of the images should be consistent in order to derive the features of the facial expression. For this purpose Viola and Jones face and eye detectors available in OpenCV 2.3 [72] library are used to localize the face as it is the most cited, considered to be the fastest and most accurate pattern recognition method for face detection [148]. It was observed that face and eye detectors produce false positives. The eye detectors (left and right) are used to reduce false positives of the face detector. The combination of face and eye detector can reduce false positives effectively but they can introduce localization errors in facial features detection i.e. if the same image is given as input to localization of the face region, detector gives different locations for each detection iteration. Localization errors of the face are critical for the performance of the FERS using features based on LBP operators which are sensitive to face localization errors and also registration errors [80]. Hence, most of the studies in the literature (see Table 2.1) pre-process the input images manually to remove localization errors and image normalization to remove registration errors [79, 124, 126, 129, 133]. This inhibits to derive automatic FERS. ARLBP and EARLBP features are extracted automatically by localizing the face using Viola and Jones detectors. The localization errors are mitigated using a proposed heuristic localization method to enable automatic feature extraction.

$$d = [x_{\mathrm{RE}} + (\alpha_{\mathrm{RE}}/2)] - [x_{\mathrm{LE}} + (\alpha_{\mathrm{LE}}/2)] \tag{5.1}$$

$$x_{\Gamma'} = \begin{cases} x_{\mathrm{LE}}, & \text{if } y_{\mathrm{RE}} \geqslant y_{\mathrm{LE}} \\ x_{\mathrm{RE}} - d, & \text{otherwise} \end{cases} \tag{5.2}$$

$$y_{\Gamma'} = \begin{cases} y_{\mathrm{LE}} - d/2, & \text{if } y_{\mathrm{RE}} \geqslant y_{\mathrm{LE}} \\ y_{\mathrm{RE}} - d/2, & \text{otherwise} \end{cases} \tag{5.3}$$

$$\alpha_{\Gamma'} = \alpha_{\mathrm{LE}}/2 + \alpha_{\mathrm{RE}}/2 \tag{5.4}$$

$$\beta_{\Gamma'} = \begin{cases} y_{\mathrm{RE}} - y_{\Gamma} + (2 * d) - k, & \text{if } y_{\mathrm{RE}} \geqslant y_{\mathrm{LE}} \\ y_{\mathrm{LE}} - y_{\Gamma} + (2 * d) - k, & \text{otherwise} \end{cases} \tag{5.5}$$

Figure 5.3 (a) shows three rectangles as output of the Viola and Jones face detector. Out of the three rectangles one is face rectangle $\Gamma(x_{\Gamma},\ y_{\Gamma},\ \alpha_{\Gamma},\ \beta_{\Gamma})$ and two are eyes rectangles, LE $(x_{\mathrm{LE}},\ y_{\mathrm{LE}},\ \alpha_{\mathrm{LE}},\ \beta_{\mathrm{LE}})$ for left eye (LE) and RE$(x_{\mathrm{RE}},\ y_{\mathrm{RE}},\ \alpha_{\mathrm{RE}},\ \beta_{\mathrm{RE}})$ for right eye (RE). Γ' is a face rectangle whose parameters can be obtained using equations

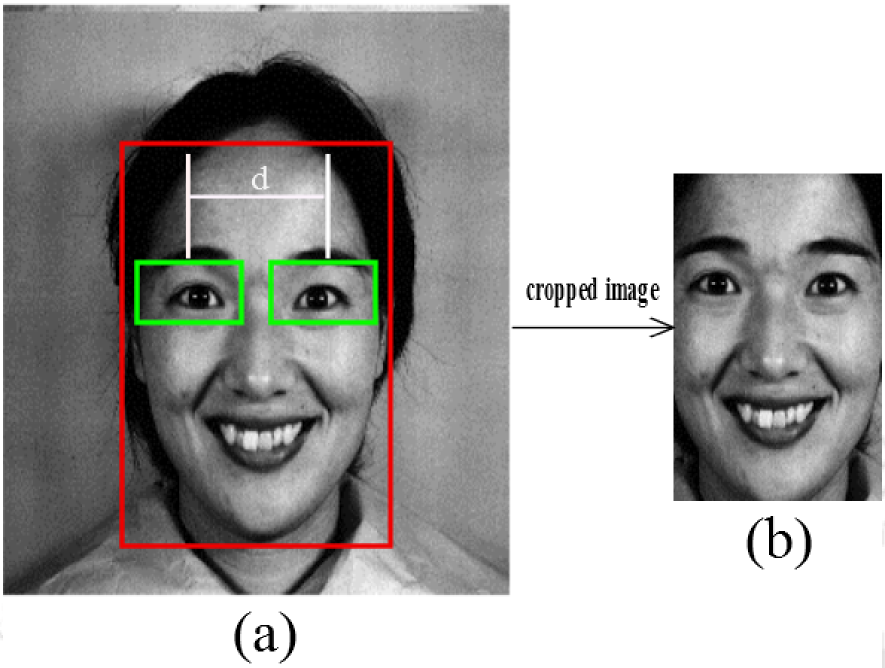

Figure 5.3: (a) Facial features localization using Viola and Jones method for a JAFFE [2] input face image, (b) The cropped image using Face localization method.

5.1, 5.2, 5.3 , 5.4 and 5.5. The x and y are co-ordinates measured from top left corner of the input image and d is the approximate distance between two eyes. *Left* and *Right* directions are with respective to the observer point of view, k is an empirical constant. α, β and Γ represent *width, height* and *Face* rectangle of the image respectively. Figure 5.3 (b) shows the localized face image obtained using Γ' rectangle.

5.1.3 Description of the Databases used

To derive feature extraction or face representation techniques it is important to understand the properties of the database and the protocol followed during the image acquisition process. Moreover, it will enable to judge feature extraction method's robustness to different types of variations present in the database. For facial expression recognition three databases were used and their description is as follows:
The CK+ database consists of facial sequences of 123 adults aged in the range of 18-50 years of which 69% are females, 81% Euro-American, 13% African-American and 6% other groups. The subjects were instructed to display the prototypic facial expression from neutral to the target expression. These image sequences are digitized to 640 $\times$ 490

pixels with 8-bit representation. The database consists of 593 sequences with peak frame FACS (Facial Action Coding System) coded and 327 sequences out of 593 peak frames are emotion coded (*Neutral, Happy, Sad, Surprise, Fear, Disgust, Angry* and *Contempt*) using Active Appearance Model and Support Vector Machine. A sample images from this database is shown in Figure 5.4.

Figure 5.4: Data samples from CK+ database.

The FGNET [138] is more challenging facial expression database since it captures facial expression as natural as possible when people react to a movie being played. Some samples are shown in the Figure 5.5. The database has a video sequence of 18 individuals (male and female) performing each facial expression three times. Altogether, it consists of 399 sequences of seven basic facial expressions (*Happiness, Anger, Surprise, Disgust, Fear, Sadness* and *Neutral*). These sequences are converted to a size of 320 $\times$ 240 8-bit JPEG static images.
The JAFFE database is described in the Section 4.3 as it is used in FR for experimentations.

Figure 5.5: Data samples from FGNET database.

5.2 Facial Expression Recognition using Asymmetric Region based LBPs

FERS which incorporates LBP based facial representations is constrained by registration errors introduced by the methods used for localization of face and its salient regions. Though the LBP operators have the advantage of lower computation cost, but the performance of the systems that use them depends on many parameters as listed in Table 2.1 from column 2 to 12 and also parameters such as database used (as it may contain frontal or non-frontal faces) different illumination, classifier and face localization methods employed. The following Sections describe the use of proposed ARLBP and EARLBP operators for facial expression recognition in the presence of face localization and registration errors to study the effect of size (scale) of the operator.

5.2.1 Experimental Process

In order to train the classifier, data samples from databases of facial expressions need to be selected to create a training set and testing set following a sampling protocol. The first two/three images per person per expression are selected and included in training set and the remaining images of the same person's expressions were included in testing set. There were 143 images in training set and 70 in the testing set for the JAFFE database. From the FGNET database, six frames from two sequences for each expression per subject are selected and included in training set. In addition, three frames from the remaining sequence per expression per subject form the testing set are also included. Altogether the training and test sets consisted of 684 and 325 images respectively. No particular selection criteria was applied while selecting the images. The training and the test sets of both databases are person-dependent for case 1 which means that there exists an overlap between the training and the testing sets with respect to the person and the facial expression as illustrated in Section 1.6.2.

For Cases 2 and 3, 213 images (30 neutral, 31 happy, 31 sad, 30 surprise, 30 angry, 32 fear and 29 disgust) from the JAFFE database were selected. No particular selection criteria was applied while selecting the images. From CK+ database 309 emotion coded and labeled sequences were considered excluding *Contempt*. A total of 1236 images (309 neutral, 210 happy, 84 sad, 249 surprise, 138 angry, 75 fear and 171 disgust) from each sequence (maximum of three peak frames and minimum one peak frame) were considered such that each frame can be categorized into one of the seven prototypical expressions.

These selected facial image samples are used to build or train the classifier using the training set in person independent or person dependent configuration (see Figure 1.9). There are two stages to derive a classifier; one is training and the other is testing phase. The aim of training stage is to derive a classifier employing a learning algorithm to generate a model on training data samples. The learned or trained classifier model is evaluated using the testing samples to test how well it represents the data samples. Figure 5.6 illustrates the FERS with its different components interacting to recognize the facial expression. Given a face image, the facial features are localized automatically using Viola and Jones face and eye detectors. Since these detectors are inconsistent in localizing the facial features the proposed heuristic localization method is used to re-localize the face. The localized face might be processed if necessary to enhance particular features like illumination correction and face alignment. Asymmetric Regions LBP patterns are obtained by convoluting using modified convolution technique for each grid. A histogram is obtained for each grid which is concatenated with other grids histograms to get the image face representation (feature) as shown in Figure 3.5. These features are then fed to the classifier model to classify as one of the six basic emotion labels including neutral expressions.

For each scale (size) of the ARLBP operator the face in images were localized using

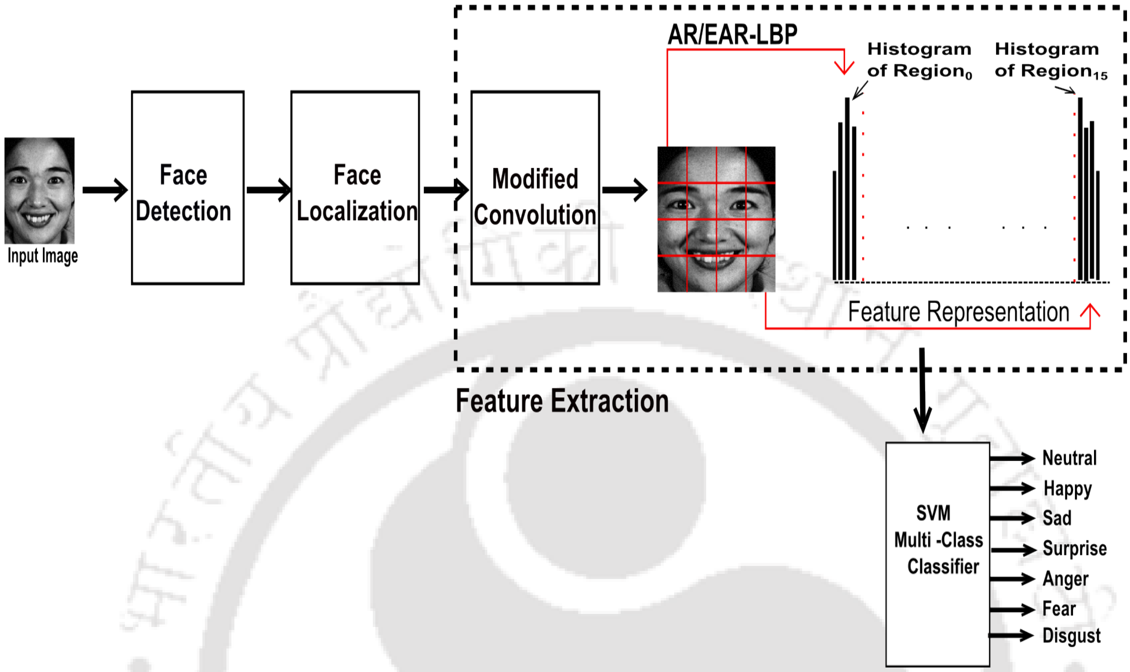

Figure 5.6: Facial Expression Recognition process.

Viola and Jones face and eye detectors [72] in each case illustrated below. First, the face detector was used and further eye detectors (both left and the right eye) were used as verifiers. Face is localized using proposed heuristic method to remove inconsistencies of localization method. The localized face image is cropped and the face image is resized to 64×64. Illumination correction or face alignment process was not considered as LBP based operators robust to monotonic illumination. The cropped 64×64-pixel face image is divided into 16 grids (sub-regions) of size 16×16 pixels. From each grid, ARLBP is convoluted using modified convolution to label each pixel and a 256 bin-length histogram is derived and a concatenated histogram of length 4096 ($16*256$) was obtained as the facial expression image representation. This representation is obtained for each sample image for the training set and a SVM multi-classifier is derived. The parameter during training is set using grid search method available in OpenCV [72]. The above process is followed for image samples in testing set to get facial representation to test the SVM classifier. This process is followed in all the cases considered for experiments. The experiments were conducted for different sizes of the operator. The size of the proposed operators does not affect the histogram length as in other LBP based operators except MBLBP operator. Moreover, the size of the operator does not affect or influence the number of the grid (hence size of the grid) and the available codes per image. This is due to fact that MC technique labels all the image pixels without ignoring the boundary pixels. A heuristic rule used is that the scale of the operator should be less than the grid size in varying the size of the operator in different cases. There are some cases in which different process was followed that will be discussed in the respective cases explicitly.

5.2.2 Experimental Design

Experiments are designed to study the ability of the proposed face representation techniques with different settings. The settings like person dependent or person independent, with or without registration errors. The following cases can be designed and experimented.

- **Case 1: Person dependent FER with registration errors:**
  In this case the FER is configured as person dependent and features are extracted from the face images without any normalization and alignment process . This will test the robustness of the operators to registration errors inherent in databases. This setting is similar to FR which is person dependent recognition with manual face alignments. This case study will also test whether the proposed Asymmetric Region based operator (ARLBP) can be applied in person dependent FER.

- **Case 2: Person independent FER with registration errors:**
  It is well known that LBP based operators are sensitive to registration errors which influence the performance of FER. The person independent setting would test whether, the face representation is robust to variations due to subjects age, gender, culture, etc. in the presence of registration errors without incurring extra processing cost to remove registration errors.

- **Case 3: Person independent FER without registration errors:** This case is designed to compare the performance of the proposed operators obtained in Case 2 so as to judge the performance of these operators.

5.3 Facial Expression Recognition using ARLBP

It is well known that FERS consists of two phases of which one is training and another is testing. In this process it is necessary to build a classifier model before testing stage. ARLBP operator is used to derive face representation using modified convolution and

Table 5.1: Facial expression recognition rates (%) for SVM 7-class Facial expression on JAFFE database with RBF kernel for different ARLBP operator sizes in person dependent condition.

Width \ Height	3	5	7	9	11	13	15
3	84.29	88.57	91.43	88.57	85.71	85.71	88.57
5	91.43	91.43	**95.71**	92.86	91.43	91.43	91.43
7	90	92.86	91.43	91.43	92.86	90	90
9	88.57	94.29	92.86	92.86	92.86	90	88.57
11	94.29	94.29	92.86	92.86	92.86	91.43	91.43
13	**95.71**	92.86	94.29	**95.71**	91.43	92.86	92.86
15	**95.71**	92.86	94.29	**95.71**	92.86	92.86	92.86

classifier is built using the protocol as discussed above. The trained model is used to test using the illustrated experimental process. Different cases are considered to study the behavior and ability of operator in different conditions.

5.3.1 Case 1: Person dependent with registration errors

In this case size of the operator is varied but size of the image, number of grids and size of the grids an image is divided are treated as constant. A multi-class SVM is trained (as discussed above) using facial representation and the parameters[1] of SVM for RBF kernel are set using OpenCV 2.1 grid search method. A 10-fold CV was performed with the SVM multi-class classifier (7-class) and the average recognition rate of 10-fold CV is tabulated in Table 5.1 and Table 5.2. Table 5.1 shows SVM facial expression recognition rate for each scale of the ARLBP operator. For example using 3×3 ($width \times height$) scale operator recognition rate is 84.29%.

From the table it is observed that the recognition rate attains increasing trend when the width is greater than the height of the operator and attains a maximum of 95.71% for 13×3 and 15×3 sizes with an exception for 5×7 size. It is hard to find the reason

[1] For all sizes of the operator, RBF's C=62.5, γ= 0.00001 for JAFFE database and C=20, γ= 0.000003 for FGNET database

Table 5.2: Facial expression recognition rates (%) for SVM 7-class Facial expression on FGNET database with RBF kernel for different ARLBP operator sizes in person dependent condition.

		Height						
		3	**5**	**7**	**9**	**11**	**13**	**15**
Width	**3**	71.6	72.21	74.32	74.32	75.53	75.23	74.62
	5	74.32	75.53	78.55	76.13	77.64	76.74	77.95
	7	69.79	74.02	77.34	76.74	77.34	76.13	76.13
	9	72.51	76.13	76.74	76.44	75.83	77.04	75.53
	11	71.9	74.92	78.55	78.25	76.44	74.92	74.32
	13	74.02	75.23	76.13	77.04	77.95	75.23	73.41
	15	73.41	72.81	75.53	**79.46**	78.25	75.23	74.02

for this exception since the face alignment and variation of texture and misalignments are not controlled. When the width and height of the operator is equal (observed diagonally from top left to right bottom) the recognition rate does not increase. This may be due to averaging of the texture. This suggests that varying in width and height one at a time would increase the performance. However, when the height decreases and as the width increases (observed diagonally from top right to bottom left), the recognition rate exhibits an increasing trend with the exception of 11×5 size.

In Table 5.2 the observations inferred with the JAFFE database were not observed in recognition rates with the FGNET database. This may be because the ARLBP operator is database dependent as two different databases were considered. However, the maximum recognition rate of 79.46% was obtained when the width is greater than the height using the FGNET database for the operator size 15×7. It is unfair to compare with similar work performed using LBP as the recognition rate is dependent on several parameters such as image size, image sub-region size, databases used and classifier parameters. The work reported by Liao *et.al* [90] can however be compared with this work. The authors experimented with 64×64 image size for person dependent facial expression recognition and have been reported a recognition rate of about 94.59% on the JAFFE database using Advanced-LBP with Null space Linear Discriminant Analysis together with Tsallis. The proposed ARLBP operator with Modified convolution technique provides a recognition rate of 95.71% on the same JAFFE database and can be considered a degree better than that reported by Liao.

5.3.2 Case 2: Person independent with registration errors

For all sizes of ARLBP operator, the length of feature histogram was kept constant at 4096 and facial representations are obtained as illustrated above. For each size of the ARLBP operator, the CK+ and JAFFE database images were divided into 10 random parts without overlap of a person's facial expression between parts making the training and testing set person-independent. A multi-class SVM using Linear and Radial Basis Functions (RBF) with different parameters[2] was trained using 9 parts and remaining one part was used as the test set. The training and testing experiment was repeated 10 times to get a 10-fold cross validation recognition rate. For different sizes of the ARLBP operator, average facial recognition rate is tabulated in respective tables.

Experiments conducted varying the size of the operator but not the number of grid and size of the grids. It can be observed from the Table 5.3 and Table 5.4 that there is not much difference in performance for SVM Linear and RBF kernels. The maximum recognition rate of 58.95% was obtained for JAFFE database when the ARLBP operator width is greater than the height. This is evident as the face contains more horizontal features than vertical features and ARLBP represents these to quite some extent. It is difficult to compare to the results obtained with other similar studies as the performance depends on various parameters discussed in Chapter 2. Nevertheless, the work in [117] can be compared and a recognition rate of 71.5% is reported by using sub-space learning technique. The recognition rate of ARLBP is low when compared to these studies because the features are extracted in the presence of localization errors and ARLBP operator averages the texture as the size of the operator increases due to which the performance of the operator decreases. It is inferred from Table 5.3 and 5.4 that as width and height of

[2]For JAFFE database SVM with Linear kernel C=12.5 for all sizes of the ARLBP. For RBF kernel C=312.5 and γ=0.033750 for 3×9, 5×5, 5×13, 7×15, 9×9, 11×13 and 11×15 sizes of the ARLBP. For remaining size of the operator C=12.5 and γ=0.506250 was set. In the case of CK+ database SVM with Linear kernel C=62.5 for 3×7 and for other sizes C=12.5 was set. For SVM with RBF kernel for the size of operator C=312.5 was set for sizes 3×7, 7×9, 11×9, C=62.5 for sizes 3×13, 5×3, 5×11, 7×5, 7×11, 9×15, 11×15, 13×15, 15×9, C=2.5 was set for sizes 11×11, 11×13, 13×5 and for rest of the sizes C=12.5 was set. γ=0.0375 for 3×7, 3×13, 5×3, 5×11, 7×5, 7×9, 7×11, 9×15, 11×9, 11×15, 13×15 was set and for the rest of the sizes of the operator γ=0.0506 was set.

Table 5.3: Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on JAFFE database and ARLBP in person independent settings.

		Height						
		3	**5**	**7**	**9**	**11**	**13**	**15**
Width	**3**	54.73	52	52.12	50.11	50.59	54.40	51.95
	5	55.25	56.24	50.63	53.97	56.28	53.81	55.26
	7	54.37	54.34	56.21	52.36	53.47	55.79	58.14
	9	55.16	55.79	**58.95**	55.09	55.53	54.68	51.04
	11	53.72	56.00	55.23	51.52	55.53	53.15	51.01
	13	56.49	54.20	52.79	50.44	53.92	52.91	51.01
	15	52.83	53.24	54.23	52.71	50.04	50.54	49.07

Table 5.4: Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on JAFFE database and ARLBP in person independent settings.

		Height						
		3	**5**	**7**	**9**	**11**	**13**	**15**
Width	**3**	54.65	50.52	48.13	49.20	47.36	47.25	49.10
	5	56.65	56.18	53.38	49.70	53.25	50.06	50.04
	7	55.25	56.21	54.64	49.54	49.05	50.02	51.41
	9	**58.52**	57.63	58.00	55.09	51.28	51.37	50.58
	11	53.81	56.14	54.25	53.45	56.05	50.37	47.77
	13	53.81	55.65	53.84	55.68	50.58	53.30	47.29
	15	50.99	52.85	54.70	54.66	52.45	50.02	47.62

Table 5.5: Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and ARLBP in person independent settings.

Width \ Height	3	5	7	9	11	13	15
3	79.89	80.00	79.61	79.41	78.30	77.24	76.55
5	81.40	80.36	80.01	81.14	79.31	76.80	75.81
7	80.93	**82.43**	81.02	81.54	79.14	78.47	77.38
9	80.60	80.95	80.43	81.67	79.98	79.70	78.74
11	80.18	81.95	81.58	81.59	80.48	79.76	79.17
13	79.77	81.71	82.36	81.20	79.98	78.52	78.96
15	80.74	80.40	81.45	80.96	79.83	79.20	79.37

Table 5.6: Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and ARLBP in person independent settings.

Width \ Height	3	5	7	9	11	13	15
3	79.39	80.59	80.71	77.30	76.17	79.52	74.85
5	82.94	81.59	80.07	78.68	81.20	76.54	75.92
7	80.24	**83.19**	80.97	81.62	81.06	78.72	77.77
9	80.77	81.21	79.92	79.51	79.88	78.94	82.31
11	80.60	80.58	80.05	82.61	78.63	78.95	82.34
13	80.38	80.52	80.29	78.76	78.63	78.33	81.47
15	79.96	79.90	80.69	82.07	78.18	77.94	78.73

the operator increases, recognition rate attains decreasing trend. In case of CK+ database it is apparent from Table 5.5 and Table 5.6 that the maximum recognition rate is 83.19% and performance decreases as the size of the operator increases due to texture averaging effect of ARLBP operator and localization errors. The performance is superior for sizes where width is greater than height of the operator.

5.3.3 Case 3: Person independent without registration errors

A face image preprocessing tool [149] is used to crop the faces manually. This tool has different features like detection of face, normalization, histogram equalization, cropping frontal face images and saving the cropped images in *XML* format. In order to remove registration errors this tool was used to crop the face images and manually change the co-ordinates of face such that all salient regions of face are included in the cropped image. Then the cropped faces were normalized depending on the eye distance similar to [145] and the images were resized to 64×64 pixels. ARLBP histogram features were derived as described in Figure 3.5. The parameters[3] were set using grid search method available in OpenCV 2.3 for the SVM classifier.

Experiments are carried out on CK+ database and JAFFE, but the recognition rate for JAFFE database is not reported as the performance was below 40%. This may be due to the fact that image normalization process would normalize the texture in low resolution images while pre-processing. None of the works have reported performance on low resolution JAFFE images. It is evident from the Table 2.1 that the recognition rate on JAFFE database is better when either image size is greater than 64×64 or person dependent settings [74, 78, 114, 118, 125, 127, 131]. Nevertheless, few studies report a recognition rate of more than 50% with image size 64×64 and person independent case without considering all images of the database.

Table 5.7 and Table 5.8 shows the performance of SVM with Linear and RBF kernels on the CK+ database. Since maximum results were obtained in the case with registration errors when width is greater than height of the operator. The experiments were conducted considering the size of the operator in which width is more than the height of the operator. The maximum recognition rate of 88.02% was obtained in case of SVM with Linear kernel, which is comparable to recognition rate of 89.9% reported in [145] for an image resolution

[3]For CK+ database and SVM with Linear kernel C=0.05 for 3×3, 5×3, 5×5 and 0.01 was set for rest of the operator sizes. In case of RBF kernel C=12.5 was set for operator sizes 7×7, 9×3, 9×5, 11×3, 11×5, 13×5, 15×5, 15×7 and for rest of the sizes 62.5 was set. The value of γ=0.00001 was set for all the corresponding operator size

Table 5.7: Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and ARLBP in person independent and without registration errors settings.

Width \ Height	3	5	7
3	85.05	85.33	85.51
5	87.27	86.30	85.42
7	87.24	86.72	85.62
9	87.93	87.07	85.87
11	87.05	86.14	86.21
13	**88.02**	86.34	86.57
15	87.73	86.91	87.33

of 55×75.

From above tables and discussions in Sections 5.3.2 and 5.3.3, it may be noted that performance of ARLBP operator get affected by registration errors. The maximum recognition rate was obtained in the case where SVM was trained using Linear Kernel. This may be due to the fact that Linear kernel's performance is better than RBF kernel as feature dimension (4096) is more than the number of samples used to train the SVM [150]. Further, ARLBP is sensitive to registration errors just like any other LBP based operators as performance is higher in the case of without registration errors when compared to images with registration errors.

5.4 Facial Expression Recognition using EARLBP

EARLBP operator is an extended version of ARLBP. It is proposed to reduce the averaging effect on texture which in turn tries to increase the recognition rate of FERS. The averaging effect is reduced by considering the boundary pixels of the regions as described in Section 3.3.3. Experiments were conducted to verify sensitivity to registration errors by reducing the averaging effect which may lead to higher performance and be less sensitive to registration errors. The operator is not experimented with Case 1 setup assuming that the operator would work better in person dependent settings. The settings of Case 2 and

Table 5.8: Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and ARLBP in person independent and without registration errors settings.

		Height		
		3	**5**	**7**
Width	**3**	84.07	85.67	84.77
	5	85.06	86.41	85.27
	7	86.24	86.30	85.94
	9	**86.92**	86.18	84.98
	11	86.40	86.14	85.45
	13	86.57	86.60	85.90
	15	86.60	86.58	86.89

3 are more challenging than Case 1 and hence the experiments were not conducted using Case 1 for EARLBP. The experimental process similar to ARLBP is followed in order to build the training and testing sets to derive the classifier.

5.4.1 Case 2: Person independent with registration errors

ARLBP Case 2 experimental process was used for all sizes of EARLBP operator. Experiments were conducted as described and average facial recognition rate from 10-fold cross validation are tabulated in respective tables. The parameters[4] of SVM for RBF and Linear kernels were set using OpenCV 2.3 grid search method for all sizes of the EARLBP operator for both databases.

Table 5.9 and Table 5.10 show the average recognition rate of SVM facial expression

[4]For JAFFE database SVM with Linear kernel C=12.5 for all sizes of the EARLBP, for RBF kernel C=312.5 and γ=0.033750 for 3×9, 5×5, 5×13, 7×15, 9×9, 11×13 and 11×15 sizes of the EARLBP. For remaining sizes of the operator C=12.5 and γ=0.506250 was set. In the case of CK+ database SVM with Linear kernel C=2.5 for 7×9, 7×11, 11×11, 11×13 and for other sizes C=12.5 was set. For SVM with RBF kernel for the sizes of operator 3×7, 5×3, 5×5, 5×11, 7×3, 11×3, 11×13, 13×3, 13×7, 13×11, 15×11, 15×15, C=312.5 γ=0.033750 and for sizes 5×13, 7×9, 9×5, 9×7, 9×13, 9×15, 13×13, 13×15 C=62.5, γ=0.033750. For 9×9 C=2.5, γ=0.506250 and for other sizes of the operator C=12.5 γ=0.506250 was set.

Table 5.9: Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on JAFFE database and EARLBP in person independent settings.

Width \ Height	3	5	7	9	11	13	15
3	56.68	52.55	56.43	48.83	47.65	50.28	49.80
5	57.70	55.77	55.95	49.85	52.15	56.78	56.70
7	56.31	51.53	55.27	55.93	49.85	51.53	56.27
9	55.81	52.60	51.96	51.57	51.75	52.05	54.26
11	54.02	51.98	49.75	51.19	51.27	49.89	50.42
13	56.84	54.37	47.60	55.38	52.59	50.70	53.87
15	**58.74**	53.96	47.60	54.38	53.61	57.13	53.88

Table 5.10: Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on JAFFE database and EARLBP in person independent settings.

Width \ Height	3	5	7	9	11	13	15
3	55.38	51.13	53.66	48.18	47.23	48.46	48.91
5	**60.35**	55.06	54.89	49.89	49.97	54.74	51.95
7	55.32	52.28	56.67	55.54	50.80	49.36	54.95
9	52.77	54.28	51.30	50.52	49.23	50.69	51.19
11	52.75	55.07	52.06	52.29	52.24	49.51	50.62
13	55.99	54.88	49.94	54.51	53.67	52.60	49.24
15	58.02	53.13	54.34	53.58	51.77	52.93	48.21

Table 5.11: Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and EARLBP in person independent settings.

Width	Height: 3	5	7	9	11	13	15
3	80.07	79.03	79.23	78.79	78.52	77.34	77.20
5	81.30	82.17	80.91	80.60	77.55	78.49	79.00
7	81.89	82.49	80.61	79.16	79.29	78.74	79.81
9	81.17	81.42	80.52	79.37	79.49	78.70	80.70
11	80.18	82.24	79.32	79.61	82.23	80.49	81.30
13	82.30	**83.09**	79.65	79.30	80.52	81.01	81.12
15	80.61	82.20	79.19	79.30	79.38	79.98	80.17

for each size of the EARLBP operator with linear and RBF kernels on the JAFFE database. For example using 3×3 size, average recognition rate was 56.68% and 55.38% for Linear and RBF kernel respectively. One can observe that the recognition rate is considerably higher for size when width is more than the height of the operator. A maximum of 58.74% and 60.35% recognition rate was obtained with Linear and RBF kernels respectively.

From Table 5.11 and Table 5.12, similar observations on CK+ database can be made as obtained in the case of the JAFFE database. Maximum recognition performance of 83.09% and 82.21% were obtained with Linear and RBF kernel respectively. Further, when width and height of the operator is greater than 9, the recognition rate was less compared to the lesser sizes of the operator (observed 4^{th} to 7^{th} columns of all tables). On the contrary, better performance was obtained when the *width* and *height* of the operator was less than 9 (observed 1^{st} to 3^{rd} columns of all tables) for both databases.

5.4.2 Case 3: Person independent without registration errors

In order to remove registration errors we used a tool as described in Case 2 of ARLBP earlier to crop the face and each face was normalized depending on the eye distance similar to [145] and resized the resulting images to 64×64. Next the EARLBP histogram feature is derived. The experiments were carried out on CK+ and JAFFE database, but the

Table 5.12: Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and EARLBP in person independent settings.

Width \ Height	3	5	7	9	11	13	15
3	78.24	79.31	78.13	75.32	76.35	75.56	74.75
5	81.68	81.02	77.93	77.05	78.31	78.46	75.50
7	80.02	79.83	78.97	79.41	77.27	76.72	76.37
9	78.24	**82.21**	80.44	78.35	77.78	78.45	79.81
11	80.42	79.92	78.87	78.26	78.25	80.60	77.04
13	80.85	79.36	79.93	77.88	79.73	80.35	81.40
15	76.43	79.44	77.94	77.56	79.49	78.25	79.80

recognition rate for JAFFE database is not reported as the performance was below 40% but the maximum results obtained are mentioned for comparison purpose in Table 5.15. The lower recognition rate may be due to pre-processing of texture similar to Case 3 of ARLBP. The parameters[5] were set using grid search method available in OpenCV 2.3. Table 5.13 and Table 5.14 show the performance on CK+ database with SVM Linear and RBF kernel respectively. It is evident that the maximum recognition rate was obtained in the case where width of the operator is greater than its height. SVM with Linear kernel yielded higher recognition rate of 87.62% when compared to the performance of 86.76% in case of RBF kernel. Table 5.15 consolidates the maximum recognition rates of ARLBP and EARLBP operators with and without registration errors for image resolution of 64×64. These operators perform better on CK+ database in case of without registration errors when compared to with registration errors. With registration errors, the operators' performances are difficult to correlate for CK+ and JAFFE databases. In case of ARLBP on CK+ database, recognition rate of 83.19% was obtained with RBF kernel which is higher than the performance with Linear kernel but with EARLBP operator on the same database, a recognition rate of 83.09% with Linear kernel was obtained when compared to the result with RBF kernel. On JAFFE database, we obtained a contradicting trend compared to the results obtained using CK+ database in case of both the operators. In

[5]For CK+ database and SVM with Linear kernel, C=0.05 for all operator sizes. In case of RBF kernel C=62.5 and γ=0.00001 was set for all operator sizes.

Table 5.13: Average SVM 7-class Facial expression recognition rates (%) with Linear kernel on CK+ database and ARLBP in person independent and without registration errors settings.

		Height		
		3	**5**	**7**
Width	**3**	85.04	84.77	83.04
	5	86.68	85.64	84.05
	7	87.40	86.16	83.89
	9	**87.62**	86.62	84.31
	11	86.04	85.04	84.89
	13	85.98	85.44	84.19
	15	85.06	85.66	83.29

Table 5.14: Average SVM 7-class Facial expression recognition rates (%) with RBF kernel on CK+ database and EARLBP in person independent and without registration errors settings.

		Height		
		3	**5**	**7**
Width	**3**	84.40	84.89	83.83
	5	85.29	85.62	84.03
	7	**86.76**	86.02	84.17
	9	85.92	85.46	84.39
	11	85.10	85.38	84.91
	13	84.03	85.11	84.61
	15	84.36	84.55	82.81

Table 5.15: Maximum Recognition Rates (%) of ARLBP and EARLBP with and without registration errors for image size 64 × 64.

Database	Type of Operator							
	ARLBP				EARLBP			
	With Registration Errors		Without Registration Errors		With Registration Errors		Without Registration Errors	
	Type of Kernel		Type of Kernel		Type of Kernel		Type of Kernel	
	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF
CK+	82.43	83.19	88.02	86.92	83.09	82.21	87.62	86.76
JAFFE	58.95	58.52	30.01	28.02	58.74	60.35	29.13	27.90

case of ARLBP, recognition rate of 58.95% was obtained with Linear kernel which is higher than the result obtained with RBF kernel while for EARLBP operator, recognition rate of 60.35% was obtained with RBF kernel when compared to Linear kernel performance. This may be due to fact that both operators are sensitive to registration errors and they are specific to the databases. Further, it is observed that maximum recognition rate is obtained for ARLBP operator with the size of 9 × 3 which is higher than the EARLBP operator of the 13 × 5 size. One common behavior of these operators was that irrespective of the settings, maximum recognition rate is obtained when width of the operator is greater or equal to the height of the operator.

5.5 Chapter Summary

This chapter illustrates the process and protocols followed to conduct the experiments to study the behavior of the proposed Asymmetric Region based LBP operators on FERS. Experiments were conducted considering different cases with different scales of the operators. The number of grids (hence their sizes) and size of the images were kept constant. It was found that the performance of the system is affected by the size of the operator when grids and image sizes are kept constant. Further, it may inferred that the proposed operators could be applied to derive automatic FERS which are invariant to localization and registration errors inherent in image databases.

Chapter 6

Conclusions and Scope for Future Work

This thesis proposes LBP based operators along with convolution and a heuristic localization method to derive a facial representation for FR and FER sub-tasks of AFIA system. These face representations were investigated considering different parameters like scale of the operator, number of grids and number of available LBP codes. This chapter summarizes the major contributions of this thesis. It also elucidates the scope for future research works in the concerned field.

6.1 Summary of Contributions

Deriving a generic feature representation (feature extraction) method which would be applied in FR and FER is challenging due to variations that both the systems need to be addressed. Hence in this thesis, focus was to propose novel local based texture face representation and evaluate them against different parameters like scale of the operator, the number of grids, person independent or dependent, with localization and with or without registration errors and available codes for FRS and FERS domains. The main contributions are elucidated as follows:

- **Facial Verification:** In Chapter 4 two novel Asymmetric Region based texture operator were experimented with different parameters. A MC method which is used to convolve the face image is also used to obtain a constant available codes per face image. The experiments were designed into two cases. The important observations that are made in FR configured in verification mode for person dependent settings are as follows in respective cases:

- **Case 1:** It was observed that the total available codes per image when a face image is convoluted using MC and Asymmetric Region based operators was constant. The MC and the property of the operators enabled to experiment with larger scales in FR. The increase in available codes does increase the performance of the system.
- **Case 2**: The MBLBP operator with MC increases the available codes which increases the performance of the system when compared to traditional convolution method. The performance of the system is reduced due to the decrease in available codes affected by increase in the number of grids and scale of the operator. The number of grids would affect the histogram of each grid. The discriminative ability decreases due to averaging of texture as scale of the operator increases. With respect to ARLBP with MC for three databases, it is found that increase in the scale of the operator and the number of grids do not affect the performance of the system. This is due to the fact that the number of available codes were constant in the histogram. The performance of the system would be maintained as long as the number of available codes do not get affected. The performance of the system is highly dependent on the available codes rather than the scale of the operator and the number of grids. EARLBP is sensitive to boundary pixel averaging of the sub-region at larger scales. Hence, for larger scales of the operator the VR decreases as the scale and number of grids increase although the available number of codes are constant. EARLBP is not robust to the variations like pose, facial expression and occlusions which are inherent in the database. The proposed operators along with MC can be applied to derive FRS with minimum normalization. As the proposed operators are robust to localization and registration errors which are inherent in the databases. It is difficult to select a perfect combination of the parameters to deploy in an application without analyzing the behavior of the operator.

- **Facial Expression Recognition:** In Chapter 5, experiments were conducted in the presence of localization and registration errors and empirically evaluated the suitability of the proposed operators in FER domain. A heuristic localization method was proposed to negate face localization errors so as to facilitate these operators to be applied in automatic FERS. The size of the image and number of grids were considered as constants but the scale of the operators were varied. The following conclusions were drawn:

- ARLBP and EARLBP operator performances depend on the scale of the operator. To get maximum performance the *width* of the operator should be more than the *height*.
- ARLBP and EARLBP are sensitive to registration errors. Maximum recognition rate was obtained without registration errors.
- the proposed operators could be applied to derive automatic FERS which are invariant to localization and registration errors inherent in the databases as there is little difference in recognition rates with and without registration errors.

The proposed operators are capable of deriving face representation in both FRS and FERS domains. Hence these operators would be used in AFIA in which FR and FER system would be integrated linearly to neutralize the effect of facial expressions to enhance the FRS performance.

6.2 Scope for Future Work

Facial Expression Recognition in Video sequence could be considered using the proposed MC, ARLBP and EARLBP to understand the dynamics of the facial expressions. Such systems can be used to enhance the interactions in animations, robotic facial expression rendering, pedagogical systems, etc. Rather than using a few images under a controlled environment, if FR were to use video sequences this could lead to systems which would be useful in analyzing CCTV (Closed-Circuit TeleVision) footages. Integration of both FRS and FERS which could use the same feature extraction techniques would complement each other's challenges and enhance the overall system performance.

References

[1] S. Duffner, *Face image analysis with convolutional neural networks.* GRIN Verlag, 2009. xix, 4, 5

[2] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings* 3^{rd} *IEEE International Conference on Automatic Face and Gesture Recognition*, Apr. 1998, pp. 200–205. xix, xx, 5, 56, 76

[3] "FACS (facial action coding system)," December 2013. [Online]. Available: http://www.cs.cmu.edu/afs/cs/project/face/www/facs.htm xix, 18, 19

[4] M. Bereta, W. Pedrycz, and M. Reformat, "Local descriptors and similarity measures for frontal face recognition: A comparative analysis," *Journal of Visual Communication and Image Representation*, vol. 24, no. 8, pp. 1213 – 1231, 2013. xix, 33, 34, 42, 48, 49, 61

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002. xix, 36, 44, 45

[6] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006. xix, 44, 45, 48

[7] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Li, "Face detection based on multi-block LBP representation," in *Advances in Biometrics*, ser. Lecture Notes in Computer Science, S.-W. Lee and S. Li, Eds. Springer Berlin Heidelberg, 2007, vol. 4642, pp. 11–18. xix, 32, 46

[8] S. Naika C.L, P. K. Das, and S. B. Nair, "Asymmetric region local binary pattern operator for person-dependent facial expression recognition," in *International*

Conference on Computing, Communication and Applications (ICCCA), Feb. 2012, pp. 1–5. xix, 38, 51, 52, 53, 54

[9] S. Naika C.L, S. S. Jha, P. K. Das, and S. B. Nair, "Automatic facial expression recognition using extended AR-LBP," in *Wireless Networks and Computational Intelligence*, ser. Communications in Computer and Information Science, K. Venugopal and L. Patnaik, Eds. Springer Berlin Heidelberg, 2012, vol. 292, pp. 244–252. xix, 23, 38, 52

[10] C. Geng and X. Jiang, "Fully automatic face recognition framework based on local and global features," *Machine vision and applications*, vol. 24, no. 3, pp. 537–549, 2013. xx, 10, 74

[11] W. Gaebel and W. Wölwer, "Facial expression and emotional face recognition in schizophrenia and depression," *European archives of psychiatry and clinical neuroscience*, vol. 242, no. 1, pp. 46–52, 1992. 2

[12] V. Bruce and A. Young, "Understanding face recognition," *British journal of psychology*, vol. 77, no. 3, pp. 305–327, 1986. 2

[13] M. T. Posamentier and H. Abdi, "Processing faces and facial expressions," *Neuropsychology review*, vol. 13, no. 3, pp. 113–143, 2003. 2

[14] A. J. Calder and A. W. Young, "Understanding the recognition of facial identity and facial expression," *Nature Reviews Neuroscience*, vol. 6, no. 8, pp. 641–651, 2005. 2

[15] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer Vision - ECCV 2004*, ser. Lecture Notes in Computer Science, T. Pajdla and J. Matas, Eds. Springer Berlin / Heidelberg, 2004, vol. 3021, pp. 469–481. 3, 10, 32, 46, 48

[16] M. Pietikäinen, *Computer vision using local binary patterns.* Springer, 2011, vol. 40. 6

[17] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003. 6

[18] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002. 7, 8

[19] J. Canny, "A computational approach to edge detection," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, no. 6, pp. 679–698, 1986. 8

[20] M. F. Augusteijn and T. L. Skufca, "Identification of human faces through texture-based feature recognition and neural network technology," in *IEEE International Conference on Neural Networks, 1993.* IEEE, 1993, pp. 392–398. 8

[21] J. M. Chaves-Gonzälez, M. A. Vega-Rodríguez, J. A. Gómez-Pulido, and J. M. Sánchez-Pérez, "Detecting skin in face recognition systems: A colour spaces study," *Digital Signal Processing*, vol. 20, no. 3, pp. 806–823, 2010. 8

[22] H. P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, "Multi-modal system for locating heads and faces," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996.* IEEE, 1996, pp. 88–93. 8

[23] M.-H. Yang and N. Ahuja, "Detecting human faces in color images," in *Proceedings in International Conference on Image Processing, 1998. ICIP 98.*, vol. 1. IEEE, 1998, pp. 127–130. 8

[24] J. Yang and A. Waibel, "A real-time face tracker," in *Proceedings* 3^{rd} *IEEE Workshop on Applications of Computer Vision, WACV'96.* IEEE, 1996, pp. 142–147. 8

[25] S. J. McKenna, Y. Raja, and S. Gong, "Tracking colour objects using adaptive mixture models," *Image and vision computing*, vol. 17, no. 3, pp. 225–231, 1999. 8

[26] W.-K. Tsao, A. J. Lee, Y.-H. Liu, T.-W. Chang, and H.-H. Lin, "A data mining approach to face detection," *Pattern Recognition*, vol. 43, no. 3, pp. 1039–1049, 2010. 8

[27] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on and Pattern Recognition, CVPR 2001.*, vol. 1, 2001, pp. 511–518. 8

[28] "A computational tool," February 2014. [Online]. Available: http://fei.edu.br/~cet/facepreprocessing_tool.zip 9, 59, 64

[29] Y. Wu, yan Jiang, Y. Zhou, W. Li, Z. Lu, and Q. Liao, "Generalized weber-face for illumination-robust face recognition," *Neurocomputing*, 2014. 9

[30] A. Baradarani, Q. J. Wu, and M. Ahmadi, "An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTWT filtering," *Pattern Recognition*, vol. 46, no. 1, pp. 57 – 72, 2013. 9

[31] X. Cao, W. Shen, L. Yu, Y. Wang, J. Yang, and Z. Zhang, "Illumination invariant extraction for face recognition using neighboring wavelet coefficients," *Pattern Recognition*, vol. 45, no. 4, pp. 1299 – 1305, 2012. 9

[32] Z. Lian, M. J. Er, and Y. Liang, "A novel efficient local illumination compensation method based on DCT in logarithm domain," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1725 – 1733, 2012. 9

[33] T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li, and F. Lang, "Multiscale facial structure representation for face recognition under varying illumination," *Pattern Recognition*, vol. 42, no. 2, pp. 251 – 258, Feb. 2009, learning Semantics from Multimedia Content. 9

[34] O. Arandjelovic and R. Cipolla, "A methodology for rapid illumination-invariant face recognition using image processing filters," *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 159 – 171, 2009. 9

[35] X. Zou, J. Kittler, and K. Messer, "Illumination invariant face recognition: A survey," in *First IEEE International Conference on Biometrics: Theory, Applications, and Systems, BTAS 2007.* IEEE, 2007, pp. 1–8. 9

[36] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, no. 6, pp. 1691 – 1699, 2013. 9

[37] M. G. Dehkordi, M. R. Daliri, A. Ayatollahi, and F. Ayatollahi, "Pose invariant face recognition using biological inspired features based on ensemble of classifiers," *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 968 – 976, 2013. 9

[38] J. Wang, J. You, Q. Li, and Y. Xu, "Orthogonal discriminant vector for face recognition across pose," *Pattern Recognition*, vol. 45, no. 12, pp. 4069 – 4079, 2012. 9

[39] 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, vol. 116, no. 11, pp. 1095 – 1110, 2012. 9

[40] A. Li, S. Shan, X. Chen, and W. Gao, "Cross-pose face recognition based on partial least squares," *Pattern Recognition Letters*, vol. 32, no. 15, pp. 1948 – 1955, 2011. 9

[41] S.-I. Choi, C.-H. Choi, and N. Kwak, "Face recognition based on 2^d images under illumination and pose variations," *Pattern Recognition Letters*, vol. 32, no. 4, pp. 561 – 571, 2011. 9

[42] S. R. Arashloo, J. Kittler, and W. J. Christmas, "Pose-invariant face recognition by matching on multi-resolution MRFs linked by supercoupling transform," *Computer Vision and Image Understanding*, vol. 115, no. 7, pp. 1073 – 1083, 2011, special issue on Graph-Based Representations in Computer Vision. 9

[43] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876 – 2896, 2009. 9, 11

[44] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "meshSIFT: Local surface features for 3d face recognition under expression variations and partial data," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158 – 169, 2013. 9

[45] X. Li and F. Da, "Efficient 3D face recognition handling facial expression and hair occlusion," *Image and Vision Computing*, vol. 30, no. 9, pp. 668 – 679, 2012. 9

[46] H.-S. Lee and D. Kim, "Expression-invariant face recognition by facial expression transformations," *Pattern Recognition Letters*, vol. 29, no. 13, pp. 1797 – 1805, 2008. 9

[47] F. B. ter Haar and R. C. Veltkamp, "Expression modeling for expression-invariant face recognition," *Computers & Graphics*, vol. 34, no. 3, pp. 231 – 241, 2010, international Conference on Shape Modelling (SMI) 2010. 9

[48] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990. 10

[49] W. Zhao, R. Chellappa, and P. J. Phillips, *Subspace linear discriminant analysis for face recognition.* Citeseer, 1999. 10

[50] W. Bian and D. Tao, "Face subspace learning," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds. Springer London, 2011, pp. 51–77. 10

[51] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997. 10

[52] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern analysis and applications*, vol. 9, no. 2-3, pp. 273–292, 2006. 10

[53] S. Marcel, Y. Rodriguez, and G. Heusch, "On the recent use of local binary patterns for face authentication," *International Journal on Image and Video Processing Special Issue on Facial Image Processing*, pp. 06–34, 2007. 10

[54] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul 1997. 11

[55] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997. 11, 31

[56] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 114–132, 1997. 11

[57] H. Hotelling, "Analysis of a complex of statistical variables into principal components." *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933. 11

[58] W. S. Yambor, "Analysis of PCA-based and Fisher discriminant-based image recognition algorithms," Master's thesis, Citeseer, 2000. 13, 60

[59] V. Štruc and N. Pavešić, "The complete Gabor-Fisher classifier for robust face recognition," *EURASIP Advances in Signal Processing*, vol. 2010, pp. 1–26, 2010. 14

[60] ——, "Gabor-Based kernel partial-least-squares discrimination features for face recognition," *Informatic (Vilnius)*, vol. 20, no. 1, pp. 151–138, 2009. 14

[61] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, pp. 124–129, 1971. 16, 20, 26

[62] I. Lopatovska and I. Arapakis, "Theories, methods and current research on emotions in library and information science, information retrieval and human-computer interaction," *Information Processing & Management*, vol. 47, no. 4, pp. 575 – 592, 2011. 16, 21, 26

[63] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: An addition to the MMIfacial expression database," in *Proc. Intl Conf. Language Resources and Evaluation, Workshop on EMOTION*, 2010, pp. 65–70. 16

[64] C. E. Osgood, *The measurement of meaning.* University of Illinois Press, 1957, vol. 47. 19

[65] J. A. Russell, M. Lewicka, and T. Niit, "A cross-cultural study of a circumplex model of affect," *Journal of personality and social psychology*, vol. 57, no. 5, pp. 848–856, 1989. 19

[66] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern recognition*, vol. 25, no. 1, pp. 65–77, 1992. 19

[67] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000. 19

[68] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003. 19, 21, 35

[69] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 31, no. 1, pp. 39–58, 2009. 19

[70] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, no. 99, pp. 1 –17, 2011. 19, 32, 36, 44

[71] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research, Tech. Rep., 2010. 20

[72] "OpenCV," October 2011. [Online]. Available: http://opencv.willowgarage.com/wiki/Welcome 20, 25, 75, 80

[73] Y.-l. Tian, "Evaluation of face resolution for expression analysis," in *Conference on Computer Vision and Pattern Recognition Workshop, CVPRW'04.* IEEE, 2004, pp. 82–82. 23

[74] D.-C. Park, H. Thuy, D.-M. Woo, and Y. Lee, "Recognition of facial expression using centroid neural network," in *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Oct. 2010, pp. 480–485. 23, 37, 38, 87

[75] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 160–187, 2003. 23, 35

[76] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 699–714, 2005. 23

[77] M. Pantic and I. Patras, "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, vol. 36, no. 2, pp. 433–449, 2006. 23

[78] X. Feng, M. Pietikäinen, and A. Hadid, "Facial expression recognition based on local binary patterns and linear programming," *Pattern Recognition and Image Analysis*, vol. 15, pp. 546–548, 2005. 23, 38, 40, 87

[79] C. Shan, S. Gong, and P. McOwan, "Recognizing facial expressions at low resolution," in *IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2005.*, Sept. 2005, pp. 330 – 335. 23, 37, 38, 75

[80] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541–558, 2011. 23, 36, 37, 38, 39, 41, 49, 72, 74, 75

[81] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz, “Framework for reliable, real-time facial expression recognition for low resolution images,” *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1159 – 1168, 2013. 23, 38, 39, 41, 73, 74

[82] C. Shan, “A comprehensive empirical study on linear subspace methods for facial expression analysis,” in *Conference on Computer Vision and Pattern Recognition Workshop,* CVPRW, June 2006, p. 153. 23, 38, 40

[83] S. Zhang, X. Zhao, and B. Lei, “Robust facial expression recognition via compressive sensing,” *Sensors*, vol. 12, no. 3, pp. 3747–3761, 2012. 23, 38, 40

[84] V. N. Vapnik, *Statistical learning theory.* John Wiley & Sons Ltd., 1998. 23

[85] W. S. Noble, “What is a support vector machine?” *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006. 23

[86] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002. 31

[87] A. Pentland, B. Moghaddam, and T. Starner, “View-based and modular eigenspaces for face recognition,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR’94.* IEEE, 1994, pp. 84–91. 31

[88] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec 2006. 32, 48

[89] H. Jin, Q. Liu, H. Lu, and X. Tong, “Face detection using improved LBP under Bayesian framework,” in *IEEE First Symposium on Multi-Agent Security and Survivability*, Dec 2004, pp. 306–309. 32

[90] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Li, “Learning multi-scale block local binary patterns for face recognition,” in *Advances in Biometrics*, ser. Lecture Notes in Computer Science, S.-W. Lee and S. Li, Eds. Springer Berlin Heidelberg, 2007, vol. 4642, pp. 828–837. 32, 83

[91] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, June 2010. 32

[92] A. Bendada and M. Akhloufi, "Multispectral face recognition in texture space," in *Canadian Conference on Computer and Robot Vision (CRV)*, May 2010, pp. 101–106. 32

[93] L. Wolf, T. Hassner, Y. Taigman *et al.*, "Descriptor based methods in the wild," in *Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition*, 2008, pp. 1–14. 32

[94] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns," in *Computer Vision, Graphics and Image Processing*. Springer, 2006, pp. 58–69. 32

[95] S. Xie, S. Shan, X. Chen, and W. Gao, "V-LGBP: Volume based local Gabor binary patterns for face representation and recognition," in 19^{th} *International Conference on Pattern Recognition, ICPR 2008*, Dec 2008, pp. 1–4. 32

[96] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, June 2007. 32

[97] W. Zhang, S. Shan, H. Zhang, W. Gao, and X. Chen, "Multi-resolution histograms of local variation patterns (mhlvp) for robust face recognition," in *Audio and Video-Based Biometric Person Authentication*, ser. Lecture Notes in Computer Science, T. Kanade, A. Jain, and N. Ratha, Eds. Springer Berlin Heidelberg, 2005, vol. 3546, pp. 937–944. 33

[98] 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 *Tenth IEEE International Conference on Computer Vision, ICCV 2005.*, vol. 1. IEEE, 2005, pp. 786–791. 33

[99] B. Jun, H.-S. Lee, J. Lee, and D. Kim, "Statistical face image preprocessing and non-statistical face representation for practical face recognition," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2009, pp. 392–397. 33

[100] Z. Lei, S. Liao, R. He, M. Pietikainen, and S. Z. Li, "Gabor volume based local binary pattern for face representation and recognition," in 8^{th} *IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG'08.* IEEE, 2008, pp. 1–6. 33

[101] Y. Guo, J. Chen, G. Zhao, M. Pietikinen, and Z. Xu, "Multi-band Gradient Component Pattern (MGCP): A new statistical feature for face recognition," in *Image Analysis*, ser. Lecture Notes in Computer Science, A.-B. Salberg, J. Hardeberg, and R. Jenssen, Eds. Springer Berlin Heidelberg, 2009, vol. 5575, pp. 229–238. 33

[102] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533–544, Feb 2010. 33

[103] W. Zhang, S. Shan, L. Qing, X. Chen, and W. Gao, "Are Gabor phases really useless for face recognition?" *Pattern Analysis and Applications*, vol. 12, no. 3, pp. 301–307, 2009. 33

[104] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 57–68, Jan 2007. 34

[105] Y. Guo and Z. Xu, "Local Gabor phase difference pattern for face recognition," in 19^{th} *International Conference on Pattern Recognition, ICPR 2008.*, December 2008, pp. 1–4. 34

[106] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of Gabor magnitude and phase for face recognition," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1349–1361, May 2010. 34

[107] M. Yang, D. Zhang, L. Zhang, and D. Zhang, "Monogenic binary pattern (MBP): A novel feature extraction and representation model for face recognition," in 20^{th} *International Conference on Pattern Recognition (ICPR)*, Aug 2010, pp. 2680–2683. 34

[108] B. Fasel and J. Luettin, "Recognition of asymmetric facial action unit activities and intensities," in *Proceedings* 15^{th} *International Conference on Pattern Recognition*, vol. 1. IEEE, 2000, pp. 1100–1103. 35

[109] T. Choudhury and A. Pentland, "Motion field histograms for robust modeling of facial expressions," in *Proceedings* 15^{th} *International Conference on Pattern Recognition*, vol. 2. IEEE, 2000, pp. 929–932. 35

[110] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001. 35

[111] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Computer Vision ECCV98*. Springer, 1998, pp. 484–498. 35

[112] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, 1997. 35

[113] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, September 1995. 36, 71

[114] A. Bafandehkar, M. Nazari, and M. Rahat, "Pictorial structure based keyparts localization for facial expression recognition using Gabor filters and local binary patterns operator," in *International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, October 2011, pp. 429 –434. 36, 38, 39, 87

[115] T. Zavaschi, A. Koerich, and L. Oliveira, "Facial expression recognition using ensemble of classifiers," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 1489–1492. 36, 38

[116] A. Saha and Q. Wu, "Facial expression recognition using curvelet based local binary patterns," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 2470 –2473. 36, 38, 39

[117] Z.-L. Ying, Y.-W. Zhang, and J.-W. Li, "Manifold learning approach to facial expression recognition on local binary pattern features," in *International Conference on Machine Learning and Cybernetics*, vol. 1, July 2009, pp. 405–410. 36, 38, 40, 84

[118] X. Zhao and S. Zhang, "Facial expression recognition based on local binary patterns and kernel discriminant isomap," *Sensors*, vol. 11, no. 10, pp. 9573–9588, 2011. 36, 38, 40, 87

[119] Y. Luo, C.-M. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of PCA and LBP with SVM," *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 17, pp. 2767–2770, 2013. 36, 38, 40

[120] X. Zhao and S. Zhang, "Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding," *EURASIP Journal on Advances in Signal Processing 2012*, vol. 20, 2012. 36, 38, 40

[121] Y. Wang, H. Ai, B. Wu, and C. Huang, "Real time facial expression recognition with adaboost," in *Proceedings of the* 17^{th} *International Conference on Pattern Recognition, ICPR 2004*, vol. 3. IEEE, 2004, pp. 926–929. 36

[122] "LBP bibliography," October. 2011. [Online]. Available: http://www.ee.oulu.fi/mvg/page/lbp_bibliography 36

[123] J. Yi and I. Khalid, "Automatic facial expression recognition based on spatiotemporal descriptors," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1373– 1380, 2012. 36, 38, 40, 41, 44

[124] X. Feng, "Facial expression recognition based on local binary patterns and coarse-to-fine classification," in 4^{th} *International Conference on Computer and Information Technology, CIT '04*, September 2004, pp. 178 – 183. 37, 38, 75

[125] S. Liao, W. Fan, A. Chung, and D.-Y. Yeung, "Facial expression recognition using advanced local binary patterns, Tsallis entropies and global appearance features," in *IEEE International Conference on Image Processing*, Oct. 2006, pp. 665–668. 38, 39, 87

[126] C. Shan and T. Gritti, "Learning discriminative LBP-histogram bins for facial expression recognition," in *Proc. British Machine Vision Conference*, 2008, pp. 1–10. 37, 38, 75

[127] X. Fu and W. Wei, "Centralized binary patterns embedded with image euclidean distance for facial expression recognition," in 4^{th} *International Conference on Natural Computation, ICNC '08.*, vol. 4, Oct. 2008, pp. 115 –119. 38, 39, 87

[128] T. Gritti, C. Shan, V. Jeanne, and R. Braspenning, "Local features based facial expression recognition with face registration errors," in 8^{th} IEEE *International Conference onAutomatic Face Gesture Recognition,* FG *'08.*, Sept. 2008, pp. 1–8. 37, 38, 73

[129] M. Shoyaib, M. Abdullah-Al-Wadud, J. M. Youl, M. Alam, and O. Chae, "Facial expression recognition based on a weighted local binary pattern," in 13^{th}

International Conference on Computer and Information Technology (ICCIT), Dec 2010, pp. 321–324. 37, 38, 75

[130] T. Jabid, M. Kabir, and O. Chae, "Facial expression recognition using local directional pattern (LDP)," in 17^{th} *IEEE International Conference on Image Processing (ICIP)*, Sept. 2010, pp. 1605–1608. 38

[131] M. Kabir, T. Jabid, and O. Chae, "A local directional pattern variance (LDPv) based face descriptor for human facial expression recognition," in 7^{th} *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Sept. 2010, pp. 526–532. 37, 38, 87

[132] M.-W. Huang, Z. wei Wang, and Z.-L. Ying, "A new method for facial expression recognition based on sparse representation plus LBP," in 3^{rd} *International Congress on Image and Signal Processing (*CISP*)*, vol. 4, Oct. 2010, pp. 1750 –1754. 38

[133] F. Ahmed, E. Hossain, A. Bari, and A. Shihavuddin, "Compound local binary pattern (CLBP) for robust facial expression recognition," in *IEEE* 12^{th} *International Symposium on Computational Intelligence and Informatics (CINTI)*, Nov. 2011, pp. 391–395. 37, 38, 75

[134] J.-M. Guo and J.-L. Chen, "Two-layer features extracted from essential areas for facial expression recognition," in *International Conference on System Science and Engineering (ICSSE)*, June 2011, pp. 288–291. 38, 39

[135] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Analysis and Modeling of Faces and Gestures*, ser. Lecture Notes in Computer Science, S. Zhou, W. Zhao, X. Tang, and S. Gong, Eds. Springer Berlin Heidelberg, 2007, vol. 4778, pp. 168–182. 37

[136] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," 1996, pp. 148–156. 37

[137] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D facial expression database for facial behavior research," in 7^{th} *International Conference on Automatic Face and Gesture Recognition,* FGR, April 2006, pp. 211–216. 41

[138] F. Wallhoff, "Facial expressions and emotion database," 2006. [Online]. Available: http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html 41, 73, 77

[139] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," in 8^{th} IEEE *International Conference on Automatic Face Gesture Recognition, FG '08.*, September 2008, pp. 1–8. 41

[140] C. Crow and Franklin, "Summed-area tables for texture mapping," in *Proceedings of the* 11^{th} *Annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '84, 1984, pp. 207–212. 46

[141] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*. IEEE, 1994, pp. 138–142. 57

[142] V. Jain and A. Mukherjee, "The Indian face database," 2002. [Online]. Available: http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/ 57

[143] G. N. Girish, S. Naika C L, and P. K. Das, "Face recognition using MB-LBP and PCA: A comparative study," in *International Conference on Computer Communication and Informatics (ICCCI -2014)*. Coimbatore INDIA: IEEE, January 2014. 61, 63, 70

[144] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005.*, vol. 2. IEEE, 2005, pp. 568–573. 71

[145] C. Shan, S. Gong, and P. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009. 72, 74, 87, 91

[146] A. Geetha, V. Ramalingam, S. Palanivel, and B. Palaniappan, "Facial expression recognition–a real time approach," *Expert Systems with Applications*, vol. 36, no. 1, pp. 303–308, 2009. 72

[147] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer vision and image understanding*, vol. 83, no. 3, pp. 236–274, 2001. 72

[148] M. Kolsch and M. Turk, "Analysis of rotational robustness of hand detection with a Viola-Jones detector," in *Proceedings of the Pattern Recognition,* 17^{th} *International Conference on (ICPR'04)*, ser. ICPR '04, vol. 3. Washington, DC, USA: IEEE Computer Society, 2004, pp. 107–110. 75

[149] C. E. Thomaz and G. A. Giraldi, “A new ranking method for principal components analysis and its application to face image analysis,” *Image and Vision Computing*, vol. 28, no. 6, pp. 902 – 913, 2010. 87

[150] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, “A practical guide to support vector classification,” 2003. 88

Publications Related to Thesis

1. **Communicated**

   (a) Shrinivasa Naika C. L., Pradip K. Das, "Automatic Facial Expression Recognition using Asymmetric Region based Local Binary Patterns in Presence of Facial Registration Errors", **Under revision** Elsevier Optik Journal.

2. **Published in Journals**

   (a) Shrinivasa Naika C. L., Vibhor Nikhra, Shashi Shekhar Jha, Pradip K Das, Shivashankar B Nair, "An Intelligent Face Tracking System for Human-Robot Interaction using Camshift Tracking Algorithm", International Journal of Machine Intelligence, ISSN:0975-2927,E-ISSN:0975-9166, Vol.3, Issue 4, pp. 263-267, Bioinfo Publishers.

   (b) Shashi Shekhar Jha, Shrinivasa Naika C. L. and Shivashankar B Nair,"Driving Robots Using Emotions", Transactions on Computational Science XXI, Special Issue on Innovations in Nature-Inspired Computing and Applications, Springer Lecture Notes in Computer Science Vol. 8160, 2013, pp 64-89.

3. **Published in Conferences**

   (a) Shrinivasa Naika C. L., Pradip K. Das, Shivashankar B Nair, "Asymmetric Region Local Binary Pattern For Person-dependent Facial Expression Recognition", International Conference on Computing, Communication and Applications, ICCCA-2012, 22-24th February, 2012, Dindigul, India pp. 1-5(indexed in IEEE Digital Library).

   (b) **(Best paper Award)**Shrinivasa Naika C. L., Pradip K. Das, Shashi Shekhar Jha, Shivashankar B. Nair, "Automatic Facial Expression Recognition using Extended AR-LBP", Sixth International Conference on Information Processing

(ICIP-2012), August 10th - 12th 2012, Communications in Computer and Information Science, Eds. Venugopal, K. R., Patnaik, L. M., Vol. 292, Part 3, Springer Berlin Heidelberg, pp. 244-252.

(c) Saurabh Prajapati, Shrinivasa Naika C. L., Shashi Shekhar Jha and Shivashankar Nair, "On Rendering Emotions on a Robotic Face", Advances in Robotics, International Conference of the Robotics Society of India, July 4-6,2013 at Pune, India. (To be published in ACM Digital Library).

(d) Girish G N, Shrinivasa Naika C. L., Pradip K. Das, "Face Recognition using MB-LBP and PCA: A Comparative Study", in International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03- 05, 2014, Coimbatore, India (To be published in IEEE Digital Library).

(e) Shashi Shekhar Jha, Shrinivasa Naika C. L., Vibhor Nikhra and Shivanshankar B. Nair, "On Using Emotions in Decision-Making by Situated Robots", Proceedings of the 12th International Conference on Hybrid Intelligent Systems (HIS 2012), 4-7 December, 2012, Pune, India, pp. 372-377,.

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati
Guwahati 781039, India