FACE IMAGE ANALYSIS FOR SOFT BIOMETRIC CLASSIFICATION

Ph.D. THESIS

by

PATEL BHAVIK RAJNIKANT

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247667 (INDIA)
DECEMBER, 2016
FACE IMAGE ANALYSIS FOR SOFT BIOMETRIC CLASSIFICATION

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

PATEL BHAVIK RAJNIKANT

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247667 (INDIA)
DECEMBER, 2016
CANDIDATE’S DECLARATION

I hereby certify that the work which is being presented in this thesis entitled “FACE IMAGE ANALYSIS FOR SOFT BIOMETRIC CLASSIFICATION” in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Electrical Engineering of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2013 to December, 2016 under the supervision of Dr. R. P. Maheshwari, Professor, Department of Electrical Engineering and Dr. Balasubramanian Raman, Associate Professor, Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

(PATEL BHAVIK RAJNIKANT)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

(R. P. Maheshwari) (Balasubramanian Raman)
Supervisor Supervisor

The Ph.D. Viva-Voce Examination of Mr. Patel Bhavik Rajnikant, Research Scholar, has been held on ________________.

Chairman, SRC Signature of External Examiner

This is to certify that the student has made all the corrections in the thesis.

Signature of Supervisors Head of the Department

Dated: ______________
Facial images have been widely investigated for biometric authentication. Apart from the identity of the user, the human face also carries supplementary information (e.g. age, gender, and ethnicity) which can be used to improve the performance of biometric systems. Such attributes are known as soft biometric attributes since they cannot distinguish the identity by themselves. This thesis concentrates on two soft biometric attributes: gender and kinship.

Soft biometric classification of facial images is a challenging task due to variations in illumination, pose, expression, and resolution. Therefore, it is desirable to design robust facial representations which can deal with most of the challenges if not all. With this aim, this thesis proposes several facial representation algorithms for gender classification and kinship verification. In particular, the thesis focuses on the extension of local binary patterns (LBP) since they have been shown to be highly successful in various applications such as texture classification, face recognition, and facial expression analysis. Further, the choice of LBP is also motivated by its computational simplicity and illumination invariance properties.

The first part of the thesis deals with gender classification problem. Since the LBP operates on raw pixel intensities, initially the focus is made on the design of intensity-domain feature extraction algorithm. The first contribution of this dissertation is a face descriptor namely multi-quantized local binary pattern (MQLBP). The performance of the proposed descriptor is evaluated on several face datasets. Further, noisy images have also been used in experiments to evaluate the robustness against noise. The results of the experiments clearly demonstrated three-fold advantages of the proposed method: higher discrimination power, better noise tolerance capability, and robustness against age variations.

The second contribution of this thesis is compass local binary pattern (CoLBP). To improve the performance of basic LBP, the proposed descriptor operates in the gradient domain and it effectively combines the properties of LBP and Kirsch compass masks. To evaluate the robustness of proposed descriptor, the experiments are performed on real-world facial images. On comparing the performance with several state-of-the-art approaches, the results illustrate that the proposed method provides a good balance between a classification performance and computational cost.

After dealing with facial photographs, this thesis concentrates on gender classification of facial sketches which has got little attention in the literature. Facial sketches are widely used in forensic investigations. Hence, gender classification of sketch images can be useful in heterogeneous face recognition system which matches facial photographs with sketches. The
third contribution of this thesis is to study the applicability of different LBP variants on gender classification of facial sketches. The experiments on two sketch datasets clearly indicated the success of LBP features and its variants for this modality also.

The second part of the thesis deals with kinship verification task which aims to determine whether the persons in a given pair of images are related to each other or not. Since it is a pair matching problem, the role of distance metric used in matching is very important. The thesis, therefore, concentrates on metric learning methods for kinship verification. Unlike existing approaches which learn single global distance metric, this thesis proposes the idea of learning multiple local distance metrics to improve discrimination capability. To this end, the fourth contribution of the thesis is block-based neighborhood repulsed metric learning (BNRML) framework. This framework is generic and can be easily used with any local feature and is also suitable for information fusion. Experiments on two datasets demonstrate the superior performance of the proposed method over basic neighborhood repulsed metric learning (NRML) method.

Inspired by the success of periocular biometrics, this thesis addresses kinship verification of periocular features. The fifth contribution of this work is an evaluation of the effectiveness of periocular features for kinship verification. The experimental results show the encouraging performance of periocular features. Further, fusion scheme of left and right periocular region is also proposed to improve the overall performance.
ACKNOWLEDGEMENTS

A wide range of people has supported me towards this thesis. First of all, I express my sincere gratitude towards my supervisors Dr. R. P. Maheshwari, Professor, Department of Electrical Engineering and Dr. Balasubramanian Raman, Associate Professor, Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, for their constant encouragement and motivation throughout the duration of this research work. Regular discussions with them on every aspect of the research work motivated me to give my best. I am also thankful to my research committee members Dr. P. Sumathi, Prof. R. C. Mittal, Prof. R. S. Anand and Prof. Vinod Kumar, for their constructive suggestions during several meetings held for this research work. I would also like to thank Dr. Pramod Aggarwal, Professor & Dean Academics, and Dr. S. P. Srivastava, Professor & Head, Department of Electrical Engineering for their continuous support. I also acknowledge the support extended by all laboratory and administrative staff, especially Mr. Mohan Singh, Mr. Rishabh Verma, Mr. Jogeshwar Prasad, Mr. Amir Ahmad and Mr. Veer Chand.

I am also indebted to my friends and fellow lab mates who helped me in many ways. Special mention to Dr. Bhargav Vyas, Dr. Arvind Yadav, Dr. Roshan Kumar, Dr. Sachin Singh, Dr. N. Biradar, Dr. KSH Milan Singh, Dr. Om Hari Gupta, Dr. Jitendra Kumar, Mr. Haresh Sabhadia, Mr. Arun Balodi, Mr. Yogesh Sariya, Mr. Harikrishna and Mr. Jayendra Kumar. I also thank Dr. Manisha Verma, Dr. Himanshu Agarwal, Dr. Tasneem Ahmed, Dr. Pushpendra Kumar, Mr. Arun Pundir, Mr. Deepak Murugan, and Mr. Vivek Raj. Many thanks to Mr. Jatin Patel, Mr. Yogesh Makwana, and Mr. Kunal Bhatt.

I would also like to thank our family friends Dr. Mehul Gor, Dr. Mitesh Panchal, Dr. Dilip Jani, Mr. Bhavin Shah, Dr. Pankaj Gohil, and Mr. Amit Joshi.

I take this opportunity to thank Dr. C. L. Patel, Chairman Charutar Vidya Mandal, Dr. R. K. Jain, Principal and Dr. V. K. Thakar, Head, Electronics and Communication Engineering Department, A. D. Patel Institute of Technology, New Vallabh Vidya Nagar, Gujarat for sponsoring me to join the Ph.D. program at Indian Institute of Technology Roorkee. I also thank my all colleagues for their moral support.

I sincerely acknowledge the financial support provided by QIP Center, Indian Institute of Technology Roorkee during the research work.

I wholeheartedly thank my parents, in-laws and other family members for their constant support. Very special thanks to my wife for her unconditional support without which this would not have been possible. I also thank my daughters for their love and happiness they
brought in my life. Finally, I am very thankful to God who gave me blessing and wisdom to carry out this research work.

(Patel Bhavik Rajnikant)
## CONTENTS

ABSTRACT ........................................................................................................... I
ACKNOWLEDGEMENTS ................................................................................... III
CONTENTS .......................................................................................................... V
LIST OF FIGURES ............................................................................................... IX
LIST OF TABLES ................................................................................................. XI
LIST OF ABBREVIATIONS ................................................................................... XIII

CHAPTER 1: INTRODUCTION ........................................................................... 1
  1.1 Motivation ........................................................................................................ 1
  1.2 Overview of pattern recognition system .......................................................... 2
  1.3 Role of feature extraction .................................................................................. 5
  1.4 Research contributions .................................................................................... 7
  1.5 Thesis outline .................................................................................................. 8

CHAPTER 2: LITERATURE REVIEW .............................................................. 11
  2.1 Gender classification ....................................................................................... 11
      2.1.1 Face localization, face alignment and pre-processing .............................. 11
      2.1.2 Feature extraction .................................................................................. 13
          2.1.2.1 Global methods ............................................................................ 13
          2.1.2.2 Local methods ........................................................................... 14
          2.1.2.3 Dimensionality reduction and feature selection ....................... 20
      2.1.3 Pattern classification ............................................................................... 23
      2.1.4 Fusion techniques .................................................................................. 26
  2.2 Kinship verification ......................................................................................... 28
      2.2.1 Feature extraction ................................................................................. 28
      2.2.2 Metric learning .................................................................................... 30
  2.3 Summary ........................................................................................................ 31

CHAPTER 3: MULTI-QUANTIZED LOCAL BINARY PATTERNS FOR GENDER CLASSIFICATION ......................................................... 33
  3.1 Introduction ..................................................................................................... 33
  3.2 Existing local descriptors ................................................................................. 34
      3.2.1 Local binary pattern (LBP) .................................................................. 34
      3.2.2 Local ternary pattern (LTP) ................................................................. 36
      3.2.3 Completed local binary pattern (CLBP) ............................................. 36
3.2.4 Extended local binary pattern (ELBP) ......................................................... 37
3.3 Multi-quantized local binary pattern (MQLBP) ........................................... 37
3.4 Face description using MQLBP ................................................................. 42
3.5 Experimental setup ................................................................................. 43
  3.5.1 Datasets .............................................................................................. 43
  3.5.2 Evaluation protocol ............................................................................ 44
  3.5.3 Parameter setting ................................................................................ 46
3.6 Experimental results ................................................................................ 46
3.7 Effect of number of levels ......................................................................... 50
3.8 Effect of age variations ............................................................................ 51
3.9 Experiments on noisy data ....................................................................... 52
3.10 Comparison of computation time ............................................................. 55
3.11 Summary .................................................................................................. 55

CHAPTER 4: COMPASS LOCAL BINARY PATTERNS FOR GENDER CLASSIFICATION ..........................................................57
4.1 Introduction .............................................................................................. 57
4.2 Existing local patterns .............................................................................. 60
4.3 Proposed method ..................................................................................... 61
4.4 Experimental setup .................................................................................. 65
  4.4.1 Evaluation protocol and parameter settings ......................................... 67
4.5 Results on color FERET database ............................................................ 68
4.6 Results on LFW database ........................................................................ 71
4.7 Results on Adience database ................................................................. 73
4.8 Comparison with other state-of-the-art results ....................................... 77
4.9 Computational complexity ....................................................................... 78
4.10 Comparison among components of CoLBP ........................................... 80
4.11 Comparison of CoLBP and MQLBP ....................................................... 81
4.12 Summary .................................................................................................. 82

CHAPTER 5: GENDER CLASSIFICATION OF FACIAL SKETCHES .......... 83
5.1 Introduction .............................................................................................. 83
5.2 Datasets .................................................................................................... 83
5.3 Experimental results ................................................................................ 86
  5.3.1 Results on CUFSF dataset ................................................................. 86
  5.3.2 Results on CUFS dataset ................................................................. 87
CHAPTER 6: KINSHIP VERIFICATION USING FACE AND PERIOCUAR IMAGES

6.1 Introduction ................................................................. 93
6.2 Periocular biometrics......................................................... 95
6.3 Background of NRML ...................................................... 96
6.4 Proposed method.............................................................. 98
   6.4.1 Training stage ......................................................... 98
      6.4.1.1 Feature extraction............................................... 98
      6.4.1.2 BNRML metric learning ....................................... 99
      6.4.1.3 Block weight learning ....................................... 99
   6.4.2 Testing stage .......................................................... 100
6.5 Information fusion of left and right periocular regions ...................... 101
6.6 Experimental setup ....................................................... 101
   6.6.1 Datasets ............................................................... 102
   6.6.2 Evaluation protocol and parameter settings ......................... 102
6.7 Results on full face images ............................................... 103
   6.7.1 Effect of block weighting ........................................ 107
   6.7.2 Comparison of different LBP variants ............................ 108
6.8 Performance evaluation on periocular images ................................ 110
   6.8.1 Comparison with different LBP variants .......................... 113
   6.8.2 Comparison with full face images ................................ 116
   6.8.3 Comparison with previous work .................................. 118
6.9 Summary ........................................................................ 118

CHAPTER 7: CONCLUSION AND FUTURE WORK

7.1 Conclusion .................................................................. 119
7.2 Future work ............................................................... 121

PUBLICATIONS FROM THE WORK ........................................ 123
BIBLIOGRAPHY ................................................................. 125
APPENDIX A ..................................................................... 145
LIST OF FIGURES

Fig. 1.1 Research problems addressed in this thesis: (a) Gender recognition, (b) Kinship verification ................................................................. 3
Fig. 1.2 Block diagram of pattern classification of facial images .................................................. 3
Fig. 1.3 Methodology of local feature extraction ......................................................................... 6
Fig. 3.1 Encoding schemes of (a) LBP, (b) LTP, (c) 2-level MQLBP and (d) 3-level MQLBP .................................................................................. 34
Fig. 3.2 Example of LBP calculation ............................................................................................. 35
Fig. 3.3 Calculation of LBP, LTP and MQLBP operators ............................................................. 39
Fig. 3.4 Comparison of discrimination ability of CLBP and MQLBP ......................................... 40
Fig. 3.5 Comparison of discrimination ability of VAR and MQLBP ........................................... 41
Fig. 3.6 Flowchart of the proposed MQLBP algorithm ................................................................. 42
Fig. 3.7 Examples of cropped facial images from different datasets: (a) FERET, (b) PAL, (c) CASIA, and (d) FEI ................................................................. 44
Fig. 3.8 Performance gain of MQLBP over other feature types .................................................. 49
Fig. 3.9 Examples of misclassified images .................................................................................... 50
Fig. 3.10 Comparison of MQLBP performance for different number of levels ......................... 51
Fig. 3.11 Performance of various descriptors for different age groups on PAL dataset .............. 51
Fig. 3.12 Examples of noisy images with different SNR: (a) 5 dB (b) 10 dB (c) 15 dB (d) 20 dB .......................................................................................................... 52
Fig. 3.13 Classification performance in the presence of noise on FERET database ................. 53
Fig. 3.14 Classification performance in the presence of noise on PAL database ...................... 53
Fig. 3.15 Classification performance in the presence of noise on CASIA database ................. 54
Fig. 3.16 Classification performance in the presence of noise on FEI database ...................... 54
Fig. 4.1 LBP-based facial representation pipeline ........................................................................ 57
Fig. 4.2 Kirsch compass masks .................................................................................................. 60
Fig. 4.3 Coding scheme of existing local patterns ........................................................................ 61
Fig. 4.4 The framework of the proposed method .......................................................................... 62
Fig. 4.5 The face image and corresponding compass LBP images ............................................. 63
Fig. 4.6 Correlation analysis of compass LBP images .............................................................. 65
Fig. 4.7 Examples of cropped face images from three datasets .................................................. 67
Fig. 4.8 ROC curves for color FERET ....................................................................................... 70
Fig. 4.9 ROC curves for LFW dataset ........................................................................................ 73
Fig. 4.10 ROC curves for Adience database ................................................................................. 75
Fig. 4.11 Performance gain by CoLBP over other feature types ........................................76
Fig. 4.12 Examples of misclassifications from all the datasets ........................................76
Fig. 4.13 Performance gain of CoLBP over individual components .................................81
Fig. 4.14 Comparison of classification performance of CoLBP and MQLBP .....................82
Fig. 5.1 Example images from sketch datasets: (a) CUFSF, (b) CUFS ............................84
Fig. 5.2 ROC curves for CUFSF dataset ........................................................................86
Fig. 5.3 ROC curves for CUFS dataset ...........................................................................88
Fig. 5.4 Performance gain by CoLBP over other feature types ........................................89
Fig. 5.5 Illustration of misclassified samples ....................................................................89
Fig. 5.6 Difference in classification rates of male and female subjects.............................90
Fig. 6.1 Processing flow of the proposed method during training stage .........................97
Fig. 6.2 Processing flow of the proposed method during testing stage .........................100
Fig. 6.3 Information fusion of left and right periocular regions using BNRML framework .................................................................101
Fig. 6.4 Sample images from the two datasets: (a) KinFaceW-I, (b) KinFaceW-II ..........102
Fig. 6.5 ROC curves for NRML and BNRML method on KinFaceW-I dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D ........................................................................105
Fig. 6.6 ROC curves for NRML and BNRML method on KinFaceW-II dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D ........................................................................106
Fig. 6.7 Performance gain by BNRML over NRML .........................................................107
Fig. 6.8 Illustration of some misclassifications in kinship verification ............................107
Fig. 6.9 Example image pairs with kinship relation for two datasets: (a) KinFaceW-I and (b) KinFaceW-II ........................................................................110
Fig. 6.10 Verification rates of periocular region on KinFaceW-I dataset .........................112
Fig. 6.11 Verification rates of periocular region on KinFaceW-II dataset .......................112
Fig. 6.12 ROC curves of periocular region on KinFaceW-I dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D ........................................................................113
Fig. 6.13 ROC curves of periocular region on KinFaceW-II dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D ........................................................................113
Fig. 6.14 Comparison of kinship verification rate by region on KinFaceW-I dataset ......116
Fig. 6.15 Comparison of kinship verification rate by region on KinFaceW-II dataset ......116
Table 3.1 Comparison of the proposed approach with the work in [141] .......................... 41
Table 3.2 Summary of images used in experiments .......................................................... 45
Table 3.3 Classification rates (%) on FERET dataset ...................................................... 47
Table 3.4 Classification rates (%) on PAL dataset .......................................................... 48
Table 3.5 Classification rates (%) on CASIA dataset ...................................................... 48
Table 3.6 Classification rates (%) on FEI dataset ......................................................... 49
Table 3.7 Comparison of computation time of different features ................................... 55
Table 4.1 Comparison of different local feature extraction methods .............................. 66
Table 4.2 Summary of images used in experiments ......................................................... 67
Table 4.3 Classification rates (%) of color FERET database ........................................ 69
Table 4.4 Classification rates of LFW database ............................................................. 72
Table 4.5 Classification rates of Adience database ......................................................... 74
Table 4.6 Performance comparison of the proposed method with other representative approaches .......................................................... 77
Table 4.7 Comparison of computational complexity ..................................................... 79
Table 4.8 Overall classification rates (%) of different components of CoLBP .................. 80
Table 5.1 Summary of images used in experiments ......................................................... 85
Table 5.2 Classification rates of CUFSF database ........................................................ 85
Table 5.3 Classification rates of CUFS database ........................................................... 87
Table 6.1 Summary of images used in the experiments .................................................. 103
Table 6.2 Comparison of mean verification accuracy (%) of NRML and BNRML on KinFaceW-I dataset ......................................................................................... 104
Table 6.3 Comparison of mean verification accuracy (%) of NRML and BNRML on KinFaceW-II dataset ......................................................................................... 104
Table 6.4 Effect of block weighting on mean verification accuracy (%) for KinFaceW-I dataset ........................................................................................................ 108
Table 6.5 Effect of block weighting on mean verification accuracy (%) for KinFaceW-II dataset ........................................................................................................ 108
Table 6.6 Parameters of different features for BNRML framework ............................... 109
Table 6.7 Mean verification accuracy (%) of different features under BNRML framework on KinFaceW-I ......................................................................................... 109
Table 6.8 Mean verification accuracy (%) of different features under BNRML framework on KinFaceW-II ......................................................................................... 109
Table 6.9  Mean verification accuracy (%) of periocular region on KinFaceW-I dataset .... 111
Table 6.10 Mean verification accuracy (%) of periocular region on KinFaceW-II dataset ... 111
Table 6.11 Parameters of different features for BNRML framework ........................ 115
Table 6.12 Mean verification accuracy (%) of different periocular features on KinFaceW-I
dataset ........................................................................................................................... 115
Table 6.13 Mean verification accuracy (%) of different periocular features on KinFaceW-II
dataset ........................................................................................................................... 115
Table 6.14 Comparison of verification accuracy (%) with existing methods on KinFaceW-I
dataset ........................................................................................................................... 117
Table 6.15 Comparison of verification accuracy (%) with existing methods on KinFaceW-II
dataset ........................................................................................................................... 117
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAM</td>
<td>Active appearance model</td>
</tr>
<tr>
<td>ASM</td>
<td>Active shape models</td>
</tr>
<tr>
<td>BOG</td>
<td>Bag-of-words</td>
</tr>
<tr>
<td>CCA</td>
<td>Curvilinear component analysis</td>
</tr>
<tr>
<td>CLBP</td>
<td>Completed local binary pattern</td>
</tr>
<tr>
<td>CoLBP</td>
<td>Compass local binary pattern</td>
</tr>
<tr>
<td>CSML</td>
<td>Cosine similarity metric learning</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete cosine transform</td>
</tr>
<tr>
<td>DMML</td>
<td>Discriminative multimetric learning</td>
</tr>
<tr>
<td>DTP</td>
<td>Directional ternary pattern</td>
</tr>
<tr>
<td>ELBP</td>
<td>Extended local binary pattern</td>
</tr>
<tr>
<td>ELDiP</td>
<td>Eight local directional pattern</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>Enhanced local directional pattern</td>
</tr>
<tr>
<td>ESL</td>
<td>Ensemble similarity learning</td>
</tr>
<tr>
<td>GDF</td>
<td>Geometrical distance feature</td>
</tr>
<tr>
<td>GGOP</td>
<td>Gabor-based gradient orientation pyramid</td>
</tr>
<tr>
<td>GLD</td>
<td>Gray level difference</td>
</tr>
<tr>
<td>HGPP</td>
<td>Namely histogram of Gabor phase pattern</td>
</tr>
<tr>
<td>HOG</td>
<td>Histogram of oriented gradient</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent component analysis</td>
</tr>
<tr>
<td>IDP</td>
<td>Interlaced derivative pattern</td>
</tr>
<tr>
<td>KNN</td>
<td>$k$-nearest neighbor</td>
</tr>
<tr>
<td>LBP</td>
<td>Local binary pattern</td>
</tr>
<tr>
<td>LBPH</td>
<td>Local binary pattern histogram</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear discriminant analysis</td>
</tr>
<tr>
<td>LDiP</td>
<td>Local directional pattern</td>
</tr>
<tr>
<td>LDN</td>
<td>Local directional pattern number</td>
</tr>
<tr>
<td>LE</td>
<td>Learning-based</td>
</tr>
<tr>
<td>LFW</td>
<td>Labeled faces in the wild</td>
</tr>
<tr>
<td>LGBPHS</td>
<td>Local Gabor binary pattern histogram sequence</td>
</tr>
<tr>
<td>LGXP</td>
<td>Local Gabor xor pattern</td>
</tr>
<tr>
<td>LMNN</td>
<td>Large margin nearest neighbor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>LTP</td>
<td>Local ternary pattern</td>
</tr>
<tr>
<td>LUT</td>
<td>Look up table</td>
</tr>
<tr>
<td>LVQ</td>
<td>Learning vector quantization</td>
</tr>
<tr>
<td>M³-SVM</td>
<td>Min-max modular support vector machine</td>
</tr>
<tr>
<td>MBP</td>
<td>Monogenic binary pattern</td>
</tr>
<tr>
<td>MCT</td>
<td>Modified census transform</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>MMI</td>
<td>Maximization of mutual information</td>
</tr>
<tr>
<td>MQLBP</td>
<td>Multi-quantized local binary pattern</td>
</tr>
<tr>
<td>NAHD</td>
<td>Normalized absolute histogram difference</td>
</tr>
<tr>
<td>NCA</td>
<td>Neighborhood component analysis</td>
</tr>
<tr>
<td>NN</td>
<td>Neural network</td>
</tr>
<tr>
<td>NPOS</td>
<td>Normalized position</td>
</tr>
<tr>
<td>NRML</td>
<td>Neighborhood repulsed metric learning</td>
</tr>
<tr>
<td>OCI</td>
<td>Object class invariant</td>
</tr>
<tr>
<td>OSL-A</td>
<td>Online similarity learning with average strategy</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PDFL</td>
<td>Prototype-based discriminative feature learning</td>
</tr>
<tr>
<td>PHOG</td>
<td>Pyramid histogram of oriented gradients</td>
</tr>
<tr>
<td>PPBTF</td>
<td>Pixel-pattern-based texture feature</td>
</tr>
<tr>
<td>PR</td>
<td>Pattern recognition</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial basis function</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale invariant feature transform</td>
</tr>
<tr>
<td>SPL E</td>
<td>Spatial pyramid learning-based</td>
</tr>
<tr>
<td>SSRW</td>
<td>Self-similarity representation of Weber face</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>SVMAC</td>
<td>Support vector machine with automatic confidence</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

This chapter describes the importance of soft biometric attributes and need for their classification. The components of typical pattern classification systems are briefly discussed to provide a background. The role of feature extraction in classification task is clearly explained. Research contributions of the thesis are also highlighted. Finally, the chapter ends by outlining the structure of rest of the thesis.

1.1 Motivation

Biometrics is the science of recognizing individuals in an automated manner based on their physiological or behavioral characteristics. It has received growing attention in recent years in our security aware society. The traditional unimodal biometric systems rely on a single characteristic such as the face, iris, fingerprint, voice, and gait. Such unimodal systems have been successfully employed in various applications ranging from personal computer login to criminal identification. However, these unimodal systems are confronted with the following challenges [1]:

- Performance degradation under noisy input data
- Upper bound on identification accuracy
- Lack of universality

The aforesaid problems motivated several biometric researchers to look for alternative solutions. Towards this, Jain et al. [1] introduced the idea of using supplementary information (e.g. age, gender, height, weight, hair color) regarding the user. They coined the term, soft biometrics to refer to such supplementary information. Similar to biometric traits, soft biometric traits also represent physical or behavioral human attributes. However, unlike biometric traits, the soft biometric attributes cannot distinguish the identity of an individual by themselves. Nevertheless, they can be used to improve the performance of the biometric system. For example, using face-based automatic gender classification as a front-end of face recognition system; the effective search space can be reduced to achieve faster recognition output. Further, recognition accuracy can also be improved by training separate classifiers for each gender category [2, 3]. Since its introduction, the research in soft biometrics is ever increasing and many novel soft biometric traits such as clothing [4] and facial marks [5] are being proposed from time to time.

Humans naturally use soft biometric traits for recognition and description. People often use phrases like “the man with blue eyes”, “the young lady with brown hair” and “the tall boy with red shirt”. These phrases contain several soft biometric traits (shown using underlined text) including gender, eye color, age, hair color, height, and clothing. Primary biometric traits
cannot be expressed by such discrete labels. Thus soft biometric traits have great potential to bridge this semantic gap in order to enhance the performance of biometric systems. Further, soft biometric traits offer several other advantages as mentioned below [6]:

- Low computational cost
- Less user cooperation
- Suitability for long distance identification
- Deducible from primary biometric traits

With aforesaid motivations, this thesis addresses the soft biometric classification of face images. The face modality is chosen since face images can be captured in non intrusive manner from long distance and thus requires less user cooperation. In particular, the thesis focuses on two soft biometric attributes: gender and kinship.

Automatic gender recognition of human facial images is a binary classification task aiming at identifying the correct gender from the given input image. It is one the most fundamental task for humans since many social activities depend on the perception of correct gender. As mentioned earlier, automatic gender recognition can be used to improve the performance of face biometric systems. Automatic gender recognition can also be integrated into many other potential applications like intelligent human-computer interface, automatic image annotation, demographic data collection, surveillance and access control, content-based image retrieval and targeted advertising for retail business intelligence. Moreover, it can also be used to foster other face analysis tasks like automatic facial expression recognition [7] and emotion recognition [8].

The other major task of interest is kinship verification which aims to determine whether the two persons share a common blood relation. Kinship verification is emerging soft biometric problem which has gained substantial attention in the last five years. It has interesting and valuable applications in photo album management, forensic science, social media analysis and missing child/parent search. The two research problems addressed in this thesis are summarized in Fig. 1.1.

1.2 Overview of pattern recognition system

Automatic gender recognition and kinship verification are typical pattern recognition (PR) tasks. The aim of pattern recognition systems is to classify sensed objects into predefined classes. In our case, the sensed objects are facial images which need to be classified into two classes (male/female or kin/non-kin) for both the problems of interest. The
components of typical pattern recognition system are depicted using the block diagram in Fig. 1.2. The function of each block is briefly described as follows:

![Male or female?](Image source: FERET dataset [9])

![Are they related?](Image Source: KinFaceW-I dataset [10])

Fig. 1.1 Research problems addressed in this thesis: (a) Gender recognition, (b) Kinship verification

![Block diagram of pattern classification of facial images](image)

Fig. 1.2 Block diagram of pattern classification of facial images
• **Image database:** The processing starts with getting an input image from the database. The face image shown in Fig. 1.2 is taken from PAL database [11].

• **Face localization:** It aims at identifying the face region and segregates it from the background. This step is crucial as any error in face localization can provide misleading results at the expense of huge computational cost involved in performing the following tasks such as face alignment, feature extraction, and pattern classification.

• **Face alignment:** It is used to position the face into a canonical pose so that consistent features can be extracted from the localized face image. This is carried out by detecting a fixed set of landmarks such as eyes, nose, and mouth. The face image is then normalized such that the detected landmarks appear at the same location within a spatial coordinate system of an image.

• **Pre-processing:** For better classification results, the input images are generally pre-processed to reduce the effects of illumination. Histogram equalization is a commonly used technique for this purpose. Pre-processing may also include a reduction in the size of the image to lower the computational burden.

• **Feature extraction:** It acquires relevant data representation to extract discriminative information. Templates created from extracted features are stored in the system database during the training (enrollment) phase for the purpose of matching when the test sample is presented during identification phase.

• **Pattern classification:** Classification involves comparison of features extracted from the test sample with templates stored in the system database. Based on scores obtained from the comparison, the output of pattern classification block indicates the estimated class of the input test image.

It is important to note that kinship verification problem requires an image pair instead of a single image at the input. During feature extraction step, some distance measure (e.g. Euclidean distance) is computed between two feature vectors representing two faces. This distance vector is then fed to the classifier. More details about the kinship verification task are covered in Chapter 6.

Face localization and alignment have been extensively studied in literature and they will not be addressed in this work. The focus of this research work is to explore more discriminative features so as to improve the overall system performance. Thus, in this work, it is assumed that the given image database contains localized and aligned faces. This
assumption is also supported by several research works which have either provided both localized and aligned faces [10] or they give the coordinates of detected faces [12].

1.3 Role of feature extraction

Humans can easily and accurately extract facial attributes such as gender and kinship with just a glimpse of facial appearance. The growing role of computers in our life motivates the researchers to mimic the machines with the same capabilities. However, several parameters make such tasks complicated for machines. Some of the major challenges are briefly described as follows:

- **Illumination**: Facial images captured under different ambient lighting exhibit changes in illumination. These illumination variations make the classification task difficult since it often causes intra-class distance to exceed inter-class distance.
- **Pose**: The 3D structure of human face causes different types of variations in head pose. The major challenge is the out-of-plane rotations. The pose angles larger than ±45° severely degrades the classification performance.
- **Image quality**: The growing use of surveillance cameras raises additional issues like low image resolution, blur, and noise. Such factors make it really challenging to achieve high classification rates.
- **Occlusion**: The use of sunglasses, hats, and masks are common among people. Such accessories cause the faces to be only partially visible, making it difficult to extract the desired information reliably.
- **Facial attributes**: Human face contains information about various attributes including identity, gender, age, ethnicity, and emotion. Variations in these attributes cause large appearance changes. For example, kinship verification of a father-daughter pair involves age and gender differences between the subjects. Large appearance gap between them due to such factors often increases misclassification rate.

Since face images can be captured non-invasively and without the cooperation of users, above mentioned factors are often present in groups rather than appearing singly. Thus, the task of soft biometric attribute extraction gets even more challenging.

Among various processing stages shown in Fig. 1.2, feature extraction stage is perhaps the most critical stage. In order to perform classification, it is essential to represent the available data appropriately. If the features are not discriminative enough, even the best classifier cannot provide decent classification performance. The feature extraction algorithms
should be able to capture inter-class variations and at the same time, minimize the effect of intra-class variations. Further, the features should be robust enough to handle various challenges described above. Hence, extraction of meaningful features is a fundamental issue in computer vision and facial attribute extraction tasks.

Feature extraction methods can be broadly classified into two categories: Global methods and local methods. Global methods (or holistic methods) use whole face images for extracting features without employing any knowledge about the input object. Classical examples of global feature extraction methods are raw pixel intensity, principal component analysis (PCA) [13] and linear discriminant analysis (LDA) [14]. Due to their simplicity, global methods have been widely used for face representation in biometric and soft biometric classification. However, holistic methods are very sensitive to various types of image variations described earlier. Any small change in local face part can influence the whole feature description. Hence, the performance of holistic methods degrades severely in unconstrained settings. Local methods can surmount these difficulties as they can effectively quantify the information extracted from various local regions. Due to their promising performance, local features have been successfully used in various applications including content-based image retrieval [15-18], medical image analysis [19], texture classification [20] and biometric identification [21-25].

![Methodology of local feature extraction](image)

**Fig. 1.3** Methodology of local feature extraction

The basic principle of local feature extraction method is simple as depicted in Fig. 1.3. The input image is first divided into blocks and from each block, the features are extracted locally. Finally, the features of various blocks are fused together to form an image descriptor. Since features are extracted locally, any change in the local region of the image will affect the
representation of corresponding block only. The representation of other blocks will not be altered. Hence, local features provide more robustness as compared to global features.

1.4 Research contributions

This thesis concentrates on two soft biometric attributes: gender and kinship. Based on the preceding discussion in Section 1.3, the thesis focuses on local features. The major goal of the thesis is to explore local features for soft biometric classification of 2D face images in an unconstrained environment. The existing approaches have been reviewed and analyzed. An attempt is made to enhance the classification performance of the two soft biometric traits by overcoming the limitations of existing approaches. The main contributions of this work are highlighted as follows:

- The thesis proposes multi-quantized local binary patterns (MQLBP) for facial gender classification task. Local binary patterns (LBP) have gained a lot of attention in computer vision and face biometrics in particular. However, the basic LBP employs binary quantization of difference vectors generated from the neighboring pixels within the window of interest. The proposed MQLBP generalizes the concept of LBP and uses multi-level quantization scheme to achieve more discriminative power and better noise tolerance. The proposed method is evaluated on four publicly available datasets and it showed improved gender classification performance across all the datasets.

- The LBP method operates directly on raw pixel intensities. To enhance the performance, the thesis proposes a compass local binary pattern (CoLBP) framework. The CoLBP operates in the gradient domain by employing Kirsch compass masks to achieve robustness against the unconstrained environment. The proposed descriptor is evaluated on three benchmark datasets containing real-world face images captured in an unconstrained manner. In comparison with state-of-the-art approaches, the proposed CoLBP method provided a good balance between gender classification performance and computational complexity.

- Gender classification of facial sketches has gained little attention in spite of useful forensic applications. This thesis addresses this issue and comprehensively evaluates the performance of various local descriptors on sketch modality. The experimental results clearly demonstrated the potential of existing LBP-based texture features in classifying the gender of facial sketches.
• Most of the existing approaches have used full face images to verify kinship between a pair of images. Periocular region (the region surrounding eyes) has received increasing interest in recent years. Periocular features have been successfully applied to person identification tasks. Based on this motivation, this thesis proposes the idea of automatic kinship verification using periocular features. Various local descriptors were extracted from both the left and right eye regions and their effectiveness in automatic kinship verification task was thoroughly evaluated using two publicly available datasets. The results were very promising. Further, the information fusion from left and right periocular features was also performed to achieve performance gain. It is important to remember that the periocular region has already demonstrated its potential for face recognition, ethnicity classification, and gender recognition task. Our experiments extend this knowledge and demonstrate the suitability of the periocular region for automatic kinship verification.

• This work also proposes a block-based neighborhood repulsed metric learning (BNRML) framework, an extension of NRML, to yield more discriminative power for kinship verification task. Unlike global distance metric learned by NRML method, the proposed method learns multiple local distance metrics from different blocks of the images represented by local descriptors. Further, it is known that various blocks of the image have different discrimination capability. Hence, to reflect this diversity in discrimination power of different blocks, weighted score-level fusion scheme is used to obtain a similarity score between a pair of images from a parent and a child. The proposed BNRML framework is generic and can be used with any local descriptor.

1.5 Thesis outline

The rest of the thesis is structured as follows:

Chapter 2 reviews the related work in gender classification and kinship verification. The existing methods have been systematically categorized and summarized. The chapter ends with an outline of major findings which form the basis for the work presented in the following chapters.

In Chapter 3, the proposed multi-quantized local binary pattern is introduced. The technical aspects of the proposed method are compared with local binary pattern and some of its variants. The gender classification performance of the proposed method is evaluated on four datasets. Further, noise tolerance capability of various methods is also evaluated.
The gradient-domain feature descriptor namely compass local binary pattern is proposed in Chapter 4. Further, the gender classification performance of the descriptor is evaluated on challenging real-world face datasets. The chapter also compares computational complexity of the proposed descriptor with several state-of-the-art methods.

Most of the works have focused on gender classification of facial photographs. However, facial sketches are widely used in forensic applications. Hence, it is an interesting problem to identify the gender of facial sketches. Chapter 5 focuses on sketch modality and evaluates the performance of several local descriptors on two facial sketch datasets.

Chapter 6 describes kinship verification problem. Unlike existing approaches, it addresses the kinship verification of image pairs using periocular features. Further, it also proposes block-based neighborhood repulsed metric learning (BNRML) framework which is an extension of the NRML method. Experiments carried out on two kinship datasets are presented and the performance of periocular features is compared with full face features.

Finally, Chapter 7 presents conclusions and outcomes of the thesis. The perspectives regarding future scope are also discussed.
CHAPTER 2: LITERATURE REVIEW

This chapter reviews the existing approaches for gender classification and kinship verification. Literature regarding each task is discussed separately. Various existing methods have been categorized systematically. At the end of the chapter, a summary of major findings is also reported.

The research work presented in this thesis is focused on two soft biometric classification tasks: Gender classification and kinship verification. This chapter provides a systematic review of the literature related to these tasks. The chapter begins with the review of gender classification techniques. Next, literature on kinship verification is reviewed. Finally, the chapter ends by summarizing the major findings which paved the path for research work presented in this thesis.

2.1 Gender classification

The major steps of pattern classification pipeline were discussed in Fig. 1.2. Although the major focus of this thesis is feature extraction block, for the sake of completeness, other blocks (face localization, face alignment, and pattern classification) are also briefly reviewed here. Thus, the literature on gender classification is reviewed systematically by breaking down the task into major blocks and reporting the relevant works for each block.

2.1.1 Face localization, face alignment and pre-processing

The first important task of automatic gender classification system is the face localization task which detects and localizes the face present (if any) in a given input image. The most widely used algorithm for this task is by Viola and Jones [26]. To obtain real-time detection of faces, they used integral image representation providing an efficient computation of Haar-like features. Further, Adaboost-based learning algorithm was employed not only for classification but also for selecting a minimal number of discriminatory features. Then they introduced a cascade combination of classifiers to improve both the processing speed and detection accuracy.

The popularity of Viola-Jones face detector in face analysis task is because of the following advantages.

- Computationally simple feature extraction method.
- The algorithm is very fast and suitable for real-time applications.
- It is fairly robust and provides good detection accuracy.
- Good open source codes are easily available.
Face localization provides a coarse estimation of the face region. For better performance of the overall system, face alignment step may be required to position the face into a canonical pose. This step aims to detect a fixed set of landmarks such as eyes, nose, and mouth. The face image is then normalized such that the detected landmarks appear at the same location within a spatial coordinate system of an image.

In gender classification literature, both the manual and automatic face alignment methods have been used. Among the methods using manual alignments, some researchers [27-29] use eye center coordinates only while another group of researchers [30] uses other additional landmarks like mouth corners and eye corners.

Among various automatic face alignment approaches, appearance model based approaches have been very popular in face processing literature. Active shape models (ASM) were used in [4, 31, 32] used to locate facial landmarks in gender classification task. Active shape models are considered less reliable as they utilize limited information extracted from areas around detected landmarks only. Active appearance model (AAM) which is an extension of ASM, overcome this limitation by incorporating all the information within the image region. AAM is one of the most successful methods of face alignment and some representative works using AAM for gender classification are found in [33, 34]. Nicely drafted technical details on AAM along with excellent review are provided in [35].

Makinen and Raisamo [12] critically evaluated the effect of automatic face alignment on gender classification accuracy. After extensive experimentation, they indicated that automatic alignment methods did not improve classification accuracy. Thus, automatic face alignment methods were found to be less reliable as compared to manual face alignment which showed improvements in classification rate.

High computational cost without apparent performance improvement led several recent works [36-40] to avoid the use of automatic face alignment step. In spite of this, they reported competitive classification performance. The noteworthy work by Mayo and Zhang [36] transferred the face alignment problem from the testing phase to the training phase of the classifier. They expanded the training dataset by duplicating multiple copies of the original images obtained through artificial rotations and translations. Using this approach, they demonstrated that face alignment step is not essential and robustness to misalignment can be achieved by proper choice of feature set and classifier. Chu et al. [41, 42] also generated a set of randomly cropped images from the detected face. However, in contrast to using individual unaligned face images, they used sets of different unaligned images and represented them as linear subspaces. The other recent work by Bekios-Calfa et al. [40] used a pose-aware
classification methodology. Once faces are detected, they clustered them according to pose information extracted from eye coordinates. These clustered faces were then used to train the classifier to simultaneously predict the pose and gender both for avoiding face alignment step.

Irrespective of whether face alignment is used or not, some other pre-processing steps are usually required before feature extraction step. The most commonly used step is illumination normalization using histogram equalization. Image resizing is also generally used to reduce the number of pixels to be processed. However, there exists diversity in the choice of resized image. For example, Moghaddam and Yang [43] used image size as small as 21×12 while Shan [44] used 127×91 pixel images.

2.1.2 Feature extraction

Perhaps the most important stage of any pattern classification task is a feature extraction step as its selection makes a huge impact on the performance of the system. The feature extraction algorithms should be able to capture inter-class variations and at the same time, minimize the effect of intra-class variations. Thus, effective feature extraction is a very challenging task and it remains an active area of research. Feature extraction methods can be broadly classified into two categories: global methods and local methods. This section describes a detailed review of feature extraction methods along with various dimensionality reduction and feature selection methods.

2.1.2.1 Global methods

Global methods (or holistic methods) use whole face images for extracting features without employing any knowledge about the input object. Some of the very early works [45-48] on gender classification showed that it is possible to identify gender using pixel values of pre-processed whole face images. However, these works used a limited-size database raising questions about generalization ability of pixel-based global features to large databases. Later, the famous work of Gutta et al. [49] validated the generalization ability by using a subset of FERET database [9] consisting of 3006 images. Contrary to using pixel intensities directly, Baluja and Rowley [27] presented the use of pixel-based binary features obtained by different types of pixel comparison operators.

Global methods have the benefit of being simple since they do not discard any information. Hence they are easy to compute [50]. However, such techniques demand intensive computations and suffer from significant performance degradation under local variations like illumination and pose [51]. To reduce the influence of illumination, Lu et al.
proposed a new feature namely, pixel-pattern-based texture feature (PPBTF) extracted by transforming a gray-scale image into a pattern-map using edges and lines. They claimed that PPBTF is insensitive to lighting-variation as the local pattern matching removes the intensity of the original image. For validating their claim, they also conducted experiments under 24 different illumination conditions using a subset of illumination dataset of PIE [52]. However, as they used only 8 subjects in their experiment, the generalization of illumination invariance to a larger dataset cannot be ensured.

Several studies focused on transform domain global feature extraction using discrete cosine transform (DCT). The major motivations for using DCT are twofold. First is the information packing ability of DCT that helps to reduce the number of features by truncating higher frequency components of DCT [53]. Second is the less computational complexity of DCT resulting in faster processing speed compared to other feature reduction methods like principal component analysis (PCA) and linear discriminant analysis (LDA) [54]. Because of these advantages, DCT is widely used in image compression. However, DCT was first used by Pan et al. [55] for face recognition application. Encouraged by the outcome of [55], Majid et al. [56] used truncated DCT features for gender classification problem. The choice of most significant DCT features was carried out based on the magnitude of their variance. Mozaffari et al. [57] and Mirza et al. [58] also used DCT based global feature in addition to other local features which were fused together to improve the performance.

DCT based global features are also sensitive to illumination variations. To overcome this difficulty, Chen et al. [59] proposed a novel method for normalizing illumination variation by discarding an appropriate number of DCT coefficients in the logarithm domain. They claimed that their model-free method is very fast and can be used in a real-time face recognition system.

2.1.2.2 Local methods

As mentioned earlier, the major disadvantage of the global methods is their sensitivity to variations in pose and illumination. Moreover, they give equal importance to each part of the face. In contrast, research shows that various parts of face carry a different amount of discrimination power pertaining to face recognition [60]. Local methods can surmount these difficulties as they can effectively quantify the features extracted from various local regions which are obtained by splitting the whole face into smaller sub-images.
A. Local binary patterns and its variants

Since its introduction as texture feature by Ojala et al. [61], local binary patterns (LBP) have gained increasing attention in several computer vision applications particularly in facial representation tasks such as face detection [62], face recognition [63], anti-spoofing solution [64], facial expression analysis [65], image retrieval [66] and image quality measurement [67]. A comprehensive review of LBP and its applications to the analysis of facial images is found in [68].

Researchers have also investigated the use of LBP for extracting local appearance-based features with regard to gender classification problem. Sun et al. [69] demonstrated the gender discrimination ability of the local features extracted from conventional LBP operator by dividing a facial image using fixed sub windows. Further, they improved gender classification accuracy by using movable and scalable sub windows which capture more detailed information than that obtained with fixed-size sub windows. However, they evaluated the effectiveness of LBP features on FERET database only in contrast to three different databases used by Yang and Ai [30]. They used FERET database, PIE database, and their personal Chinese snapshot database and showed the effectiveness of LBP features in gender classification task. Previous two approaches used frontal face images for their experiments. In contrast, multi-view gender discrimination ability of LBP features was demonstrated by Lian and Lu [70]. Particularly, they showed that regional and global descriptions captured by LBP method instigated it to perform well on faces with pose variations as much as 30 degrees on the CAS-PEAL face database [71].

All of the above studies used databases which contain face images captured under relatively controlled conditions with limited variations in lighting, pose, and facial expression. In contrast, Shan [44] investigated a more challenging problem of gender classification of real-life faces using a more recent database namely labeled faces in the wild (LFW) database [72]. Real-life faces contain significant appearance variations in pose, illumination, expression, occlusion etc. Unlike previous methods, they urged that all the bins in LBP histogram (LBPH) do not carry useful information pertaining to gender discrimination task. In view of this, they proposed a compact and efficient LBP feature extraction method which learns discriminative LBP histogram bins using Adaboost. Based on comparison with pixel intensity based global method, they showed performance improvement using boosted LBP features.
The popularity of LBP method in computer vision community is due to several advantages which include the following: 1) It is a non-parametric method. 2) Invariance to monotonic illumination transformations. 3) Computational simplicity leading to very fast feature extraction. However, LBP has some disadvantages like sensitivity to non-monotonic illumination changes and random noise. Further, it relies only on the first-order derivative in local neighbor and discards directional information which may limit discriminative information [73]. To overcome these problems, basic LBP method has undergone several modifications. Some relevant modifications applied to gender classification task are discussed here.

To extend the robustness against non-monotonic illumination variation and random noise, Jabid et al. [29] introduced Local Directional Pattern (LDiP) by computing the edge response values in different directions in contrast to simple pixel intensity changes computed by LBP. Further, they urged about the imbalanced importance of the edge response values in different directions and hence they proposed to choose fixed number of most prominent directions for generating LDiP code. Based on experiments on FERET database, they showed superior gender classification performance of LDiP operator as compared to LBP. However, LDiP has limitations which include following: 1) It relies excessively on a number of most prominent edge directions and 2) it generates unstable codes in uniform and smooth areas. To overcome these limitations, Tan and Triggs [74] proposed local ternary pattern (LTP) as an extension of LBP by providing 3-valued code to improve robustness in uniform and smooth areas influenced by noise. They evaluated LTP method on face recognition problem. In contrast, Ahmed et al. [75] investigated gender classification problem using directional ternary pattern (DTP) which is a combination of LTP and LDiP method. Specifically, they encoded local texture by allocating a ternary code to each pixel by computing the edge response values in different directions. Experiments on FERET database showed improved gender classification performance as compared to LBP, LDiP and LTP.

To obtain more discriminative information, Zhang et al. [76] proposed a high-order local pattern operator namely local derivative pattern (LDP) and applied it to face recognition task showing consistently superior performance when evaluated on FERET, CAS-PEAL, CMU-PIE, Extended Yale B [77] and FRGC [78] databases. However, local derivative pattern operator increases the computational burden and the processing time as it encodes derivatives along four directions producing a 32-bit wide representation in contrast to 8-bit wide representation used by LBP. To address this issue, Shobeirinejad and Gao [73] proposed a novel feature known as interlaced derivative patterns (IDP) which preserve an important
property of capturing more detailed discriminative information and substantially decreases computational complexity of local derivative patterns. Based on experiments on FRGC ver.2.0 database, they also demonstrated better gender classification accuracy than LBP and local derivative pattern.

The original LBP operator encodes local texture around each pixel by using conventional 3x3 neighborhood window. The limitation of this approach is its inability of capturing major features with large-scale structures. To deal with this issue, Ojala et al. [61] generalized the original LBP operator by defining a local neighborhood as a set of \( P \) evenly spaced sampling points on a circle with radius \( R \). This extended LBP operator is denoted as \( \text{LBP}_{(P, R)} \) which consists of \( 2^P \) bits for each pixel. However, a minor increase in the value of \( P \) results in a rapid explosion (due to geometric progression) of feature dimensionality which lowers the computational efficiency. Fang et al. [79] investigated the effect of high-dimensional LBP features on gender classification accuracy and showed that such high-dimensional features could not evidently improve the performance as compared to low-dimensional LBP features. Further, they proposed to split conventional 3x3 neighborhood into two subsets by considering 4-neighbors and diagonal neighbors. Fusion of low density LBP features extracted from these two subsets drastically reduced feature dimension without lowering the gender classification accuracy.

B. Scale invariant feature transform (SIFT)

Scale invariant feature transform (SIFT) was introduced by Lowe [80] for robust image matching. Since then, it has been widely used in diverse applications like image retrieval, object recognition, video tracking, 3D modeling and various biometric applications. The wide acceptability of SIFT is due to following advantages: 1) Invariance to scale, rotation and illumination changes, 2) robustness to noise and affine distortion. Further, these invariant properties of SIFT effectively eliminate the need for pre-processing steps like face alignment, image resizing and histogram equalization for gender classification problem [81].

However, SIFT results in large feature dimension. Hence, Ke and Sukthankar [82] used PCA to reduce the feature dimension and using so-called enhanced PCA-SIFT they improved SIFT in terms of feature length and robustness to image deformations. This PCA-SIFT method was applied to gender classification by Yiding and Ning [83]. However, to enable gender discrimination ability, they further modified the basic PCA-SIFT method by computing PCA-SIFT projection separately for male and female thus avoiding the same projection matrix to obtain the descriptors. By evaluating their approach on three different
datasets of FERET, CAS-PEAL and BUAA-IRIP they showed better classification accuracy of their approach as compared to other feature extraction methods like PCA, LBP, and Gabor.

Wang et al. [84] addressed the difficulties of SIFT when applied to facial gender identification problem. Specifically, they mentioned the failure of SIFT-based key point detection due to missing texture and low resolution of face images. Further, they also mentioned that SIFT returns a variable number of sparse key points at different locations for different faces. Due to this, sparse SIFT features are not suitable for common machine learning techniques like support vector machine (SVM). To overcome these difficulties, they employed a dense SIFT descriptor approach by extracting local feature descriptors at regularly spaced grid points as opposed to selective interest points used in original SIFT. To evaluate their approach, they used a hybrid face database consisting face images from color FERET, CAS-PEAL, Yale and a private I2R database. Using this hybrid dataset, they showed better performance of dense SIFT as compared to another local descriptor namely shape context [85]. To further improve, very recently El-Din et al. [86] used an extended active shape model before applying SIFT to extract local feature descriptors at specific facial landmarks in contrast to regularly spaced grid points used in [84]. The added advantage of their landmarks-SIFT approach is its ability to effectively integrate shape and texture information. The shape information captured by extended active shape model is implicitly fused with texture information captured by SIFT. Using FERET, LFW and UB KinFace [87] databases, they showed superior performance of Landmarks-SIFT approach over dense SIFT on all three datasets.

For the very challenging task of gender classification from arbitrary viewpoints and under occlusion, Toews and Arbel [88] proposed a unique framework covering all the three tasks of face detection, localization, and gender classification. In this framework, they used SIFT features and quantified them probabilistically to learn an appearance-based object class invariant (OCI) face model which was used for face detection and localization. Using FERET database, they compared geometry-free bag-of-words (BOW) model with their proposed SIFT-based OCI model and showed superior face localization performance of the later. Further, they showed that, with reference to their common framework, this superior face localization facilitated the task of gender classification by showing improved performance.

C. Gabor wavelets

The two-dimensional Gabor wavelets based filter bank framework was introduced by Daugman [89] to understand the neural activity in brain's visual cortex. Based on the
experiments, they showed Gabor functions as the best fit to the impulse response of a simple cell in visual cortex. The two most desirable properties of Gabor wavelets are orientation selectivity and optimized spatial-frequency resolution. Some of the very early works exploited these properties for texture analysis [90]. Later, Lades et al. [91] applied 2D Gabor wavelets for the first time in face recognition application. Since then, 2D Gabor wavelets have attracted an enormous interest in face biometrics community at the global level. This success has led to review articles [92, 93] devoted to Gabor-based face recognition. Further, Gabor wavelets have also been used in image indexing and image retrieval applications [94].

The use of Gabor wavelets for gender classification task was pioneered by Wiskott et al. [95]. They used Gabor wavelets to create a local template for labeling the nodes in the framework of elastic graph matching used to represent faces. The similar approach combined with PCA-based eigenface algorithm was used by Lyons et al. [96] for automatic classification of faces by gender, race, and expression. Grosso et al. [97] used nodes of the uniform 8×8 grid to extract Gabor features. Using FRGC database, they compared pixel intensity and LBP features with Gabor features and found later to be more effective against variations in illumination and facial expressions. In contrast to graph-based methods used in aforementioned works, Lian et al. [98] used retina sampling method to extract facial feature points and then Gabor wavelet was used to extract frequency and orientation of those feature points. In order to tackle the issue of high-dimensional feature vectors, aforementioned works used a few selected discrete points in faces to extract local features using 2D Gabor wavelets. In contrast to this, several works use whole face image to extract Gabor features and then use downsampling and dimensionality reduction algorithm to reduce feature vector length. Strictly speaking, these approaches are said to be global feature extraction methods, however, these are described here to maintain continuity. Leng and Wang [99] used such holistic Gabor features followed by Adaboost feature selection method. Using the similar approach, Wang et al. [100] compared the performance of Gabor filters with SIFT descriptor on the same hybrid dataset as used in [84]. They demonstrated the superiority of Gabor features by reporting fewer error rates for all three kinds of Adaboost namely Real Adaboost, Gentle Adaboost and Modest Adaboost. Yan [101] demonstrated the superior robustness of Gabor transform based features to pixel-based image representation in terms of illumination variation, expression variation, and different face ornaments. Further, they also showed that as the size of face images gets larger, Gabor-based method gives higher recognition rate at the cost of increased training time. Recently, Rai and Khanna [102, 103] applied Gabor filters with six orientations on approximation face sub-images obtained using two-level wavelet decomposition. Further,
to reduce the high dimension of the real Gabor space thus obtained, they employed (2D)\(^2\) principal component analysis.

Gabor wavelets offer several advantages like invariance to translation, rotation, and scale. Moreover, they are also found to be robust against illumination variation and image noise [104]. The major disadvantage of Gabor methods is that they demand heavy computational load [105, 106].

2.1.2.3 **Dimensionality reduction and feature selection**

Feature extraction methods described earlier often generate a large number of features making classification task difficult. To mitigate this problem, dimensionality reduction techniques are commonly employed. These techniques transform a high-dimensional data into lower dimensional subspace by retaining the most meaningful features to facilitate the classification task with added advantage of reduced computational resources. Some of the relevant techniques applied to gender classification task are described below.

After the pioneering work of Turk and Pentland [13], principal component analysis (PCA) has found applications in many different areas including scene modeling [107] and face image analysis. Some very early works [45, 48] on gender classification also utilized PCA for feature extraction. However, these works used a limited size database. Balci and Atalay [108] were the first to evaluate gender classification using PCA on FERET dataset. Further, they analyzed the role of eigenvectors and found that only a subset of eigenvectors corresponding to higher eigenvalues is important for gender discrimination. Other works using PCA for reducing feature length includes [79, 109, 110]

The major advantages of PCA include: 1) less memory requirement since only a few components are retained for creating the training set, 2) low noise sensitivity as PCA retains high variance components and discards small variations. However, PCA has several limitations which are as follows: 1) PCA works only on linear data, 2) It considers second order statistics only and 3) PCA is highly data dependent and hence cannot capture any simple invariance which is not contained in training data [111].

To overcome the assumption of linear data, Buchala et al. [112] used a nonlinear projection method known as curvilinear component analysis (CCA) and showed that CCA can reduce data to its intrinsic dimension which indicates the minimum number of free variables needed to represent the data without substantial information loss. Moreover, they showed that CCA can achieve classification accuracy comparable to PCA with much fewer components than PCA.
PCA-based methods also suffer from computational difficulty encountered while obtaining accurate covariance matrix. This difficulty arises due to high-dimensional image vector obtained by transforming 2D image matrix into 1D data vector. To reduce this computational cost, Yang et al. [113] proposed a two-dimensional principal component analysis (2DPCA) which extracts features directly from image matrices in contrast to 1D vectors used by conventional PCA. Based on extensive experiments on ORL [114], AR [115] and Yale face databases, they demonstrated superior face recognition accuracy using 2DPCA than PCA. Further, they also showed that 2DPCA performs much better under variations in illumination and expressions. Moreover, 2DPCA was shown to be much faster than PCA. However, the only disadvantage of 2DPCA is that it needs more coefficients than PCA for image representation. Based on these promising results on face recognition, more recently Lu and Shi [116] and Bui et al. [117] investigated the performance of 2DPCA approach on gender classification problem and showed improved recognition rate in comparison to PCA.

Independent component analysis (ICA) is another popular dimension reduction technique which takes into account higher-order statistics rather than just relying on second order statistics as used in PCA. The goal of ICA is to find independent components in contrast to uncorrelated components found by PCA. Thus, ICA put more stringent requirement and it is a generalization of PCA [118]. Based on the assumption of a higher-order relationship between pixels in the facial image, Bartlett et al. [119] showed superior face recognition performance using ICA representation in comparison to PCA. Inspired by these results, Jain et al. [120] applied ICA to gender classification problem and reported a better accuracy using a subset of FERET dataset. In contrast to the fixed-point algorithm of ICA used in [120], recently Kumari and Majhi [121] used information maximization approach of ICA to extract features for face images to identify gender. Although their results were inferior to those presented in [120], these two works cannot be directly compared as they used different classifiers and a different number of images in training and test set for classification. Further, Kumari and Majhi used just 50 independent components in contrast to 200 components used in [120]. Wilhelm et al. [122] compared the active appearance model (AAM) based feature extraction method with face images described by ICA. Among other classification tasks, they found a better performance of AAM for gender classification when used with classifiers like multi-layer perceptron, radial basis function, and nearest neighbor. However, they reported just 20 ms of processing time needed by ICA in contrast to 2800 ms required by AAM suggesting better suitability of ICA method for real-time application. Further, gender
classification accuracy was slightly higher using ICA when applied to generalized learning vector quantization network for classification.

Feature selection methods aim to find the optimal subset of the input feature set by discarding features with low discrimination power [123]. The purpose of the remaining part of this section is to review relevant feature selection techniques.

As mentioned earlier, Balci et al. [108] showed that only a subset of eigenvectors carries more gender discrimination ability. This point was further verified in [124] which used genetic algorithms to select a reliable subset of eigen-features computed using PCA. In particular, they showed that the best eigenvectors selected by the genetic algorithm were able to capture the strong gender discriminatory information although identity-related information was obscured. Using four different classifiers on their private dataset, they showed that the gender classification accuracy was significantly improved for all the classifiers by employing a genetic algorithm for feature selection. Unlike PCA-based feature extraction, Wang and Mu [125] used a genetic algorithm to select the subset of features derived from ICA. Using a CAS-PEAL database, which is larger than that used by [124], they also reported a significant improvement in classification performance using a much lesser number of features.

Adaboost algorithm was originally proposed by Freund and Schapire [126] to boost the classification by cascading a collection of weak classifiers. However, several researchers used it as a feature selection method for improving gender classification accuracy. For example, Xu et al. [33] used Adaboost to select strong global features. They adopted Haar-like rectangle features as weak classifiers which were effectively combined by Adaboost to form a small set of gender discriminative features. Further, Adaboost-based feature selection significantly reduces execution time making it very suitable for real-time gender classification system. This real-time ability of Adaboost was demonstrated by Lu et al. [28] by reporting computation time of 1.8 ms for extracting best 100 features selected by Adaboost during the testing step. Previous two works used Adaboost for selection of global features. In contrast, Shan [44] used Adaboost to select the discriminative LBP features at multiple scales and showed improved gender recognition rate with feature dimension reduced from 2478 features (standard LBP) to 500 features (boosted LBP).

Recently Tapia and Perez [127] investigated the mutual information based feature selection method for gender classification. Using the feature-level fusion of features selected by different measures for mutual information, they reported the best results till date on three different datasets: FERET and UND database containing face images captured under controlled conditions and LFW database containing unconstrained face images. Further, they
showed that mutual information based feature selection method significantly reduced (70% - 90% reduction depending on the size of the image) the feature length making real-time gender classification feasible. Moreover, they also demonstrated the robustness of their feature selection method by reporting improved accuracy on AR face database which includes occlusions like sunglasses and scarves. Other interesting work of Jun et al. [128] used an iterative approach based on maximization of mutual information (MMI) to select compact LBP (CLBP) codes. The reduced code length helped not only in real-time performance but also improved gender classification accuracy as compared to LBP, uniform LBP and modified census transform (MCT) when evaluated on a large dataset containing images from web and POSTECH face database (PF07) [129].

2.1.3 Pattern classification

The final stage of an automatic gender classification system is the design of a classifier which aims to acquire knowledge from extracted features in order to assign the gender label to the given test samples. As mentioned by Jain et al. [123], the selection of a classifier is a difficult task and it depends on either availability or the knowledge of the user. Hence, researchers have used several classifiers for automatic gender classification. The most commonly used classifiers are a neural network (NN), Adaboost and support vector machine (SVM) which are discussed below.

Very early works on gender classification used neural networks. Golomb et al. [45] employed a two-stage fully connected three-layer back-propagation neural network for gender identification of 30×30 face images. The first stage was used for image compression and the second neural network was designed for gender classification. In particular, they showed the importance of compression stage by indicating poor generalization ability to identify new faces when it was not employed. However, the limitation of their compression stage was that they used computationally intensive nonlinear function for that task. Abdi et al. [48] overcome this limitation by using a linear compression technique of PCA and demonstrated the significant reduction in the size of the neural network. Further, they found gender classification problem as linearly separable and claimed that the choice of a particular classifier network i.e. a perceptron or radial basis function (RBF) network, does not critically affect the gender classification performance. Later, Tamura et al. [47] demonstrated the potential of neural networks in identifying gender from images with resolution as low as 8×8 or 8×6.
Recently, support vector machine has gained a lot of interest in computer vision community. It has also been used widely for gender classification problem. Moghaddam and Yang [43] were the first to use SVM classifier in gender classification problem. By comprehensive evaluation on low-resolution (21×12 pixels) face images, they showed superior performance of SVM in comparison with several other classifiers like RBF network, quadratic Bayesian, Fisher linear discriminant and nearest-neighbor classifier. Andreu et al. [109] further verified the similar results on XM2VTS [130] database and FERET database by using PCA before classification. Using face images projected on ICA subspace, Jain et al. [131] showed superior performance of SVM as compared to the nearest neighbor classifier and linear discriminant analysis (LDA) classifier when tested on FERET database. In contrast to these linear subspace methods, Buchala et al. [112] used nonlinear dimensionality reduction technique of CCA before performing classification. By comparing the performance of SVM and multilayer perceptron (MLP) classifier, they also found SVM to be superior when evaluated on two different datasets. Several other works employing SVM may be found in [28], [29], [79], [33], [127].

One interesting work worth mentioning is that of Du et al. [132] as they used SVM in their two-level decision tree (DT) framework using ethnicity-specific gender classification. Using SVM and LDA classifiers, they compared their decision tree based approach with a conventional approach which does not take ethnicity information into account for gender classification. For evaluation of their approach, they used hybrid dataset constructed from FERET and CAS-PEAL face database. The results of decision tree approach showed much better improvement with LDA. In contrast, SVM showed just marginal improvement which indicates SVM is more robust in comparison with LDA with respect to ethnicity variations. However, their work involved only two ethnicities and hence it demands further work to verify such robustness of SVM in multi-ethnic environments.

The major advantages of SVM are: 1) robust performance due to good generalization ability and 2) flexibility provided by kernel functions to deal with a large set of features. The major disadvantage of SVM is the high computational complexity which increases training and testing time.

To reduce training time, Lu et al. [133] introduced a min-max modular SVM (M³-SVM) which divides the training data into several subsets each associated with independent SVM and then carries out parallel training of SVMs. Later, Lian et al. [98] applied M³-SVM to gender classification task and reported better classification accuracy on a challenging dataset containing pose and expression variation and faces with glasses. Further, by using a novel
method of task decomposition incorporating age information, they improved generalization performance of $M^3$-SVM.

Several other variants of SVM have also been applied to gender classification problem. Zafeiriou et al. [134] modified the SVM by recasting the within-class variance minimization problem of Fisher discriminant ratio to construct the optimal separating hyperplane. Using this variant of SVM, they demonstrated performance improvement in gender classification. Leng and Wang [99] focused on a less-addressed issue of generalization ability of gender classifiers in different environments involving large variations in illumination, pose, and facial expressions. They applied fuzzy SVM combined with learning vector quantization (LVQ) to automatically generate fuzzy membership function. By evaluating their approach on three different datasets, they claimed better generalization ability of fuzzy SVM as compared to SVM. Zheng and Lu [135] incorporated automatically calculated label confidence values of training samples into learning framework of SVM and reformulated the corresponding quadratic programming problem. They called this approach as SVM with automatic confidence (SVMAC). To evaluate the performance of SVMAC, they used frontal images from FERET and BCMI database along with CAS-PEAL database containing pose variations. When compared with SVM using pixel intensity and different Gabor and LBP based features, SVMAC achieved higher classification accuracy and lower standard deviation along with improved generalization performance.

Adaboost is the third commonly used classifier for gender classification task. Shakhnarovich et al. [136] employed a threshold Adaboost for a challenging task of real-time gender classification using real-world face images extracted from the unconstrained video. On this difficult dataset, they reported superior performance obtained with Adaboost as compared to SVM. Further, they claimed that Adaboost yielded 1000 times faster results. However, the limitation of threshold Adaboost is longer convergence time resulting in a large number of features. To overcome this limitation, Wu et al. [137] proposed a look up table (LUT) weak classifier based Adaboost method and showed that LUT based Adaboost is better than threshold Adaboost with respect to parameters like the convergence speed and classification accuracy. Similar findings were reported by Baluja and Rowely [27] using experiments conducted on FERET database. In particular, they observed the comparable performance of SVM and Adaboost in terms of classification accuracy. However, Adaboost was found to be 50 times faster than SVM. Further, when tested for sensitivity to rotation, scale and translation offsets of face images, they found slightly better performance with Adaboost classifier as compared to SVM.
Makinen and Raisamo [12] presented a comparative study of the best gender classification approaches with respect to several design factors like recognition rate, classification speed, selection of face detection methods and various normalization methods. For comparing the performance of the classifier, they used multilayer neural network, SVM and threshold Adaboost. After, extensive evaluation on a subset of FERET database, they recommended to use SVM when classification accuracy is of prime importance. On the other hand, when classification speed is the most important issue, then Adaboost should be used while neural network should be preferred to have a good compromise between accuracy and speed.

In contrast to most of the existing approaches relying on high-performance computer systems, Bekios-Calfa et al. [138] recently focused on a gender classification task in limited resource environments like mobile devices and networked computing. They claimed LDA techniques to be a better choice in comparison with SVM and Adaboost, when limited computational resources and small-size training data is available. However, they also showed that when enough training data and high computational resources are available, SVM classifier is the supreme choice.

2.1.4 Fusion techniques

To improve classification accuracy, recently many methods employ information fusion approaches which can mainly be carried out at feature-level, score-level or decision-level. Feature-level fusion combines various feature extraction methods by concatenating individual feature vectors to form a single feature vector. While feature-level fusion is applied prior to classification task, the other two methods of information fusion are applied after classification. Among these post-classification methods, score-level fusion aims at combining the matching scores of different classifiers using techniques like weighted sum. In contrast, decision-level fusion combines the final decisions output by individual classifiers using techniques like majority voting. Exhaustive coverage of fusion approaches can be found in [139, 140].

Several works fused appearance-based features with a geometric-based feature for improving the performance of gender classification. Xu et al. [33] used Haar wavelet based strong appearance features selected by Adaboost and fused them with the most significant geometry features derived using active appearance model (AAM). From the 3403 geometry features extracted using 83 landmarks located by AAM, ten most significant features were selected using step-wise discriminant analysis. Further, min-max normalization was used
before fusing both types of features. After evaluating their method on a hybrid dataset containing images from FERET, AR and web picture database, they showed improved gender recognition rate. Mozaffari et al. [57] used DCT (global feature) and LBP (local feature) as appearance features and fused them with geometrical distance feature (GDF). They used AR face database and a private ethnic database consisting of face images of Iranian persons.

In contrast to aforementioned works, several researchers used only appearance-based features for fusion. In two different works, Wang et al. [84, 100] fused dense SIFT descriptors with other features. In [84], they used shape contexts for fusion with dense SIFT, while in [100] they used Gabor features for the same purpose. Adaboost algorithm was applied for feature selection. On evaluating the performance using a hybrid dataset consisting of images from FERET, CAS-PEAL, Yale and a private I2R database, they showed reduced error rates by fusion of features. Ylionas et al. [141] used the fact that LBP is independent of gray scale but vary with rotation. On the other hand, contrast is rotation independent but depends on gray scale. Thus, they fused LBP with local contrast information to complement each other. Using three datasets namely FERET, XM2VTS and FRGC database, they not only showed improved classification accuracy but also demonstrated the relative robustness against illumination changes. Further, they evaluated their fusion approach for generalization ability and suggested to train the system on larger and different datasets to improve generalization performance.

Yang et al. [142] and Mirza et al. [58] used a fusion of global and local appearance based features. The work of [142] used AAM as global features which were fused with LBP features whose dimensionality was reduced using PCA. For evaluation purpose, they used FG-Net face database. In contrast, Mirza et al. [58] used pixel intensity values as global features with PCA-based dimensionality reduction. These global features were fused with local features extracted from LBP and DCT. Their approach was evaluated on FERET database showing improved performance.

Recently Tapia and Perez [127] used a fusion of intensity (pixel value), texture (LBP) and shape (edge histogram) features at different scales. This fusion approach was evaluated on four different databases namely FERET, UND [143], LFW and AR face database. They showed that feature fusion improves gender classification accuracy even for facial images containing occlusions with sunglasses and scarves. Other notable work of Alexandre [37] used the same features; however fusion was applied at decision level using majority voting. They also used three different scales to compute different type of features. This multiscale decision approach was evaluated on FERET and UND database. On comparing the
performance obtained with a fusion of different types of features at a single scale versus fusion of a single type of feature at different scales, they noted greater improvements for the later case. However, the highest performance was achieved using a fusion of different features at various scales. The similar observations were further verified by Tapia and Perez [127].

2.2 Kinship verification

Some group of researchers in the psychology literature [144, 145] demonstrated the human capability of verifying kinship from facial images. Motivated by this finding, several computer vision researchers evaluated automatic kinship verification from a pair of facial images. Thus, unlike gender classification task, kinship verification system is presented with two input images. The goal of the system is to determine if the persons present in an image pair share a common blood relation or not. The steps such as face localization, face alignment, and feature extraction are performed on both the images. However, being a pair matching problem, it is required to find the distance between two feature vectors before classification can be performed. Thus it is very important to find suitable distance metric for better classification performance. Owing to the importance of distance metric, many of kinship verification approaches have focused on learning a novel distance metric to improve the classification performance. Since the literature on face localization and face alignment has been briefly reviewed in Section 2.1, kinship verification approaches have been reviewed on the basis of two categories: Feature extraction and metric learning.

2.2.1 Feature extraction

The issue of automatic kinship verification was first addressed by Fang et al. [146]. They extracted features like skin color, gradient histogram and pixel intensity from localized parts of faces. Then, a difference of two feature vectors (one for each face) was computed to represent facial resemblance between the two faces. They employed k-nearest neighbor classifier (KNN) and SVM for pattern classification.

The work in [146] was evaluated on a relatively small dataset containing 150 pairs. Zhou et al. [147] experimented on a larger dataset containing 800 images. Further, they proposed a novel feature descriptor namely spatial pyramid learning-based (SPLE) descriptor which incorporated both the local and global information from the given face. In contrast to simple difference operator, they utilized normalized absolute histogram difference (NAHD) operator to measure facial resemblance of the given face pair. Using SVM classifier, they reported superior performance of SPLE feature as compared to other feature extraction
methods like PCA, LBP, histogram of oriented gradient (HOG) and learning-based (LE) method. Later Zhou et al. [148] proposed Gabor-based gradient orientation pyramid (GGOP) method for feature extraction which further improved the performance as compared to SPLE. However, they used cosine similarity measure and also proposed a new kernel learning method to weight multiple features so that more complementary information can be extracted.

Kohli et al. [149] proposed self-similarity representation of Weber face (SSRW) as feature extraction method and used Chi-square distance as a measure of facial resemblance. Using UB KinFace database [150] and IIITD kinship database, they reported the performance superior to SPLE features.

Guo and Wang [151] used DAISY descriptor to extract facial familial traits (eyes, nose, and mouth) by comparing pairs of the corresponding facial part. Further, they used probability based likelihood ratio as a similarity measure and proposed a dynamic scheme to stochastically combine familial traits to make a final decision.

In contrast to using hand-crafted low-level features, Yan et al. [152] proposed a learning-based method to extract mid-level features from face images. The method namely, prototype-based discriminative feature learning (PDFL) trained support vector machine (SVM) using face samples from independent dataset without kinship labels. The decision scores of SVM were used to learn discriminative features for kinship verification. Other recent approaches to kinship verification include visual attributes [153], gated autoencoder [154], deep convolutional neural network [155] and graph-based approach [156].

Several works focused on using a combination of features to extract complementary information such that the overall verification rate can be improved. Vieira et al. [157] used geometric, textural and holistic features. To extract geometric features, they used active shape model to detect local facial landmarks. Then they extracted features such as normalized position (NPOS) and normalized length of a set of segments (SEGS) from detected landmark points. Further, they used Gabor filters and local binary pattern based texture features. Holistic features were extracted using PCA. These three types of features were combined using two-stage feature selection algorithm to retain more effective features which were fed to SVM for classification purpose. The work in [158] used intensity features, Gabor filters, SIFT and pyramid histogram of oriented gradients (PHOG) whereas the work in [159] employed Gabor features, LBP and dense SIFT features. The recent work in [160] also proposed the idea of combining global and local features. Global features were extracted using Zernike moments whereas SPLE was used as a local descriptor. These feature vectors were then
concatenated with each other. Further, to reduce feature dimension they introduced a genetic algorithm based feature selection method for kinship verification.

### 2.2.2 Metric learning

Kinship verification is a pair matching problem. Given a pair of input images, the task is to determine if the two persons are related to each other or not based on facial resemblance. Thus the quality of a distance metric in feature space is also equally important as the quality of extracted features. Therefore several researchers focused their attention on metric learning approaches to improve kinship verification performance.

Lu et al. [10] proposed a neighborhood repulsed metric learning (NRML) approach to learn a distance metric which aims to bring positive kinship samples close to each other while negative kinship samples are repelled as far as possible. Based on experiments conducted on two datasets, they demonstrated the superior performance of NRML in comparison with state-of-the-art metric learning approaches including cosine similarity metric learning (CSML), large margin nearest neighbor (LMNN) and neighborhood component analysis (NCA).

In [87], Xia et al. exploited the idea that it is easier to verify kinship between children and parents when the parents are in youth. To demonstrate the idea, they proposed to use transfer subspace learning method which learns an intermediate subspace where the children and old parents are projected close together. The subspace model was then used to make images of parents look younger providing more discriminative information for kinship verification.

Since kinship verification is an extremely challenging task, intraclass samples often have larger distance than interclass samples. Based on this observation, Xu and Shang [161] proposed an online similarity learning with average strategy (OSL-A). The goal of OSL-A is to find a distance metric such that the distance between intraclass samples should be smaller than average distance between some arbitrarily chosen interclass samples.

Aforementioned metric learning methods mainly deal with learning a single metric from feature space. To improve further, several works proposed multimetric learning methods to handle multiple feature representations. The work in [162] proposed discriminative multimetric learning (DMML) to better characterize the complementary information represented by multiple features. Based on maximum likelihood (ML) theory, they formulated DMML method to maximize the probability that the intraclass samples have higher similarity than interclass samples. Hu et al. [163] proposed a large margin multi-metric learning method to jointly learn multiple distance metrics for multiple features. The work in [158] proposed to
separately learn a combined kernel over multiple features for each class. For appropriate choice of kernel type and to make problem size tractable, they also employed sample space clustering and feature space clustering. Thus their ensemble metric learning method combined the advantages of task-specific learning and feature selection. Recently, Zhou et al. [164] proposed to combine merits of similarity learning and ensemble learning methods. Unlike most of the existing approaches which focused on learning Mahalanobis distance metric, they first introduced sparse bilinear similarity function to efficiently model genetic similarity between parent-child pairs. Then they incorporated bilinear similarity function into the ensemble similarity learning (ESL) framework to decode kinship information. Using experiments on benchmark datasets, they demonstrated the improved kinship verification rate of their approach. Recently Chen et al. [165] proposed to learn multiple linear distance metrics from various facial patches. They adopted canonical correlation analysis (CCA) to learn multi-linear coherent spaces which were shown to outperform single coherent space projection.

2.3 Summary

This chapter described a comprehensive review of state-of-the-art methods for gender classification and kinship verification using facial images. Major findings and motivations for this research work are summarized as follows:

- Among various local feature extraction methods, LBP has gained a lot of attention. Gabor-based methods have also been popular, but LBP gets an edge over Gabor features due to high computational efficiency. Hence, this work mainly focuses on LBP features. The goal of the thesis is to enhance the performance of LBP features for applications in soft biometric classification of facial images.
- Gender classification approaches have mainly used facial photographs. However, it is also interesting to infer gender information based on facial sketches which have received little attention. Further, facial sketches have been widely used in forensic applications and heterogeneous face recognition which matches facial photographs with sketches to determine the identity of suspects. Hence, this work addresses gender classification of facial sketches and evaluates the performance of several existing features for this task.
- With reference to kinship verification, most of the existing metric learning approaches aim to learn global transformation matrix. With the aim of improving discrimination
power, a framework is proposed in this work to learn multiple local distance metrics by adopting basic NRML approach.

- Existing methods have mainly used facial images to infer kinship information. However, a psychological study in [144] revealed that the ocular region carries more information regarding kinship than the other face parts. Inspired by their work, this thesis addresses periocular region-based kinship verification approach.
CHAPTER 3: MULTI-QUANTIZED LOCAL BINARY PATTERNS FOR GENDER CLASSIFICATION

This chapter presents a new feature extraction method namely, multi-quantized local binary patterns. For encoding the gray level difference (GLD) between a reference pixel and its neighbors, local binary pattern employs a binary quantization which retains the sign of GLD but discards the magnitude information. To improve the discrimination capability, the proposed method utilizes both the sign and magnitude components by performing multi-level vector quantization of GLD. The proposed method is evaluated on four publicly available datasets (FERET, PAL, CASIA, and FEI) through extensive experiments.

3.1 Introduction

For successful gender classification of facial images, it is very important to design efficient facial feature representation as it heavily affects the performance of a classifier. Further, such feature description should be robust against various types of image variations mentioned in Chapter 1. As the state-of-the-art facial feature extraction methods progressed, local appearance-based features have received increasing attention because of their robustness to illumination variation and high discrimination power. In particular, the introduction of local binary patterns (LBP) for face description has inspired many researchers to use them in various applications including face recognition, facial expression recognition, age estimation and gender classification. Inspired by the success of LBP, this thesis focuses on enhancing its performance for the soft biometric classification task. Towards this goal, this chapter presents multi-quantized local binary patterns (MQLBP) for facial gender classification task.

Being a local descriptor, the LBP considers a local neighborhood window and then computes gray level difference (GLD) between the neighboring pixels and a center pixel. For computational simplicity, this GLD is then quantized into two levels thereby retaining only the sign information of the GLD. However, the use of binary quantization scheme makes LBP very sensitive to random noise [166]. To enhance robustness against noise, Tan and Triggs [74] proposed a local ternary pattern (LTP) which employed a three-level quantization scheme. The other major drawback of LBP is that the information about the magnitude of GLD is completely lost. This loss of information reduces the discrimination capability of the LBP operator. To handle this issue, Ylioinas et al. [141] proposed to combine contrast information with LBP and showed improved gender classification performance. The other two relevant works namely extended local binary pattern (ELBP) [167] and completed local binary patterns (CLBP) [168] followed a different approach to utilize GLD magnitude information. Both the above methods explicitly encode the sign and magnitude of GLD. To retain the magnitude information, the ELBP operator slices the GLD into three different
layers, in addition to the fourth layer corresponding to sign information. Thus it generates four different patterns. In contrast, CLBP encodes magnitude information by comparing the magnitude of local GLD with the average magnitude of GLD of the whole image. Hence, it generates two different patterns namely CLBP_S and CLBP_M corresponding to sign and magnitude components respectively. In [168], the authors also proposed binary coding of center pixel (CLBP_C) which was fused with sign and magnitude components for further improvement. However, for a fair comparison, we have considered only two CLBP patterns corresponding to sign and magnitude information in this work.

This chapter also addresses the issue of enhancing the discrimination capability of LBP by utilizing the magnitude information of GLD. However, unlike CLBP and ELBP, it is proposed to generalize the concept of LBP by quantizing the GLD into multiple levels (Fig. 3.1). These multi-quantized local binary patterns (MQLBP) implicitly encode both the sign and magnitude information of GLD. Further, this method is conceptually different from LTP. Whereas LTP quantizes GLD into three fixed levels, the MQLBP follows a more natural philosophy to quantize GLD into the desired number of levels. To demonstrate enhanced discrimination capability of the MQLBP, it is evaluated on a facial gender classification task. A comprehensive experimental setup was designed using four different datasets. The results clearly indicated that the MQLBP improves discrimination power and noise tolerance as compared to LBP, CLBP, ELBP, and LTP.

![Fig. 3.1 Encoding schemes of (a) LBP, (b) LTP, (c) 2-level MQLBP and (d) 3-level MQLBP](image)

3.2 Existing local descriptors

3.2.1 Local binary pattern (LBP)

The basic LBP operator considers eight neighbors of a center pixel in a 3×3 window. The gray level difference is then computed between the neighboring pixels and a center pixel.
The GLD is further quantized into two levels to obtain an 8-bit binary number. This binary number is then represented in decimal form. Mathematically, the LBP code of a pixel centered at coordinates \((x_c, y_c)\) is given as follows:

\[
LBP(x_c, y_c) = \sum_{p=0}^{7} f_1(g_p - g_c)2^p
\]  

(3.1)

where \(g_c\) and \(g_p\) indicate a pixel intensity of a center pixel and its \(p^{th}\) neighbor within the window. The quantization function \(f_1(x)\) is defined as:

\[
f_1(x) = \begin{cases} 
1, & \text{if } x \geq 0; \\
0, & \text{otherwise}. 
\end{cases}
\]  

(3.2)

Thus, a total of 256 (\(2^8\)) different labels is possible depending on the relation between intensity values of the center pixel and its eight neighbors. An example of LBP calculation is shown Fig. 3.2.

![Diagram of LBP calculation](image)

**Fig. 3.2 Example of LBP calculation**

To consider different neighborhood sizes, the original LBP operator has been generalized by defining a local neighborhood as a set of \(P\) evenly spaced sampling points on a circle with radius \(R\). This extended LBP operator is denoted as \(LBP_{(P, R)}\) which consists of \(2^P\) bits for each pixel. However, a minor increase in the value of \(P\) results in a rapid explosion (due to geometric progression) of feature dimensionality which lowers the computational efficiency. Therefore, Ojala et al. [61] proposed to use only a subset of patterns. They
considered binary patterns as circular codes and proposed to use separate output labels for all those binary patterns which show at most two bitwise transitions from 0 to 1 and 1 to 0. Such patterns are known as uniform patterns. For example, 0001111 (2 transitions) and 11000011 (2 transitions) are uniform patterns while 1001101 (4 transitions) and 10110110 (6 transitions) are non-uniform patterns. All non-uniform patterns are accumulated into a single bin. Thus, the resulting LBP histogram contains less than $2^P$ bins. The uniform LBP patterns are denoted as $LBP_{PR}^{u}$. 

### 3.2.2 Local ternary pattern (LTP)

To reduce noise sensitivity of LBP, Tan and Triggs [74] proposed a three-level quantization scheme as shown in Fig. 3.1(b) to derive a method namely local ternary pattern. Using an additional thresholding parameter, they quantized GLD into three levels using a function as shown below:

$$f_2(x,t) = \begin{cases} 
  -1, & x \leq -t \\
  0, & -t < x < t \\
  1, & x \geq t 
\end{cases}$$  \hspace{1cm} (3.3)

To reduce feature dimensionality, the ternary code was split into two parts to generate two binary patterns by encoding positive and negative quantization levels. Hence, if a ternary code is 10-1-1 010-1 then, the respective upper and lower LTP codes would be 1000 0100 and 0011 0001. More details can be found in [74].

### 3.2.3 Completed local binary pattern (CLBP)

To enhance the discrimination capability, Guo et al. [168] proposed completed local binary patterns by utilizing sign and magnitude patterns from GLD. The component namely CLBP_S is the same as original LBP which retains the sign of GLD. Further, the magnitude information was utilized using an additional pattern (CLBP_M) as follows:

$$CLBP_{M_{PR}}(x_c, y_c) = \sum_{p=0}^{P-1} f_3\left[\left(g_p - g_c\right)\tau\right]2^p$$  \hspace{1cm} (3.4)

where the symbol $|\cdot|$ denotes an absolute value. The function $f_3(x, \tau)$ is defined as

$$f_3(x, \tau) = \begin{cases} 
  1, & x \geq \tau \\
  0, & x < \tau 
\end{cases}$$  \hspace{1cm} (3.5)

where the variable $\tau$ denotes the thresholding parameter which is set to mean value of the GLD magnitude of the whole image. More details can be found in [168].
3.2.4 Extended local binary pattern (ELBP)

The work in [167] followed a different approach to utilize the magnitude information of GLD. They used four different binary patterns corresponding to four layers obtained from GLD. The first layer retains the sign of GLD and hence returns the same code as LBP. The remaining layers encode the magnitude of GLD which can take any value in the range of (0, 255) assuming 8-bit images. They assumed that the neighboring pixels have a smaller difference in their intensities. Hence, they encoded GLD magnitude into 8 levels corresponding to range (0, 7). The first seven codes (0 to 6) represent the actual GLD magnitude, while all the magnitudes higher than six were assigned a common code of 7. These octal codes from all the neighbors were then converted to three-bit binary numbers which were used separately to build three LBP-like codes using respective bit position. More details and a numerical example can be found in [167].

3.3 Multi-quantized local binary pattern (MQLBP)

It is clear from equation (3.1) that the computation of LBP requires three major steps:

1. Compute the GLD between a center pixel and its neighbors.
2. Encode the GLD using a binary quantization function which retains the sign information only.
3. Get the LBP code by converting the binary pattern into a decimal form.

Resulting GLD after the first step can take up any value in the range of (-255, 255) assuming 8-bit gray scale images. The second step encodes the GLD using a binary quantization function. In other words, the output range of difference operator is quantized into two levels as shown in Fig. 3.1(a). Such two-valued encoding function leads to computational simplicity. However, the information about the magnitude of the difference is completely lost in the process which limits the discrimination capability of LBP. To overcome this limitation, Ylioinas et al. [141] proposed to combine LBP with contrast information to improve gender classification performance. The contrast information was computed using the local variance measure (VAR) as follows:

$$\text{VAR}_{P,R}(x_c, y_c) = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \quad \text{where} \quad \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p \quad (3.6)$$

As variance measure provides a continuous valued output, they proposed to quantize the feature space for obtaining histogram descriptor. However, there are two major issues with
this approach: (1) To determine the cut values for histogram bins, additional training stage is required. (2) The choice of the number of bins is also crucial for better classification performance. If the less number of bins are chosen, it reduces the discrimination ability while a large number of bins increase the feature size which may result in unstable histograms.

To avoid aforementioned issues, this work proposes a generalized form of LBP which implicitly captures both the magnitude and sign information from the gray level difference. The generalized form of LBP builds upon the basic idea of quantizing the output range of GLD operator. However, instead of limiting up to a binary quantization, it is proposed to quantize the output range into multiple levels and hence the resulting patterns will be known as multi-quantized local binary patterns (MQLBP). Multi-level quantization of GLD is carried out using a thresholding parameter \( t \) as shown in Fig. 3.1(c)-(d). It is important to note that MQLBP is conceptually different from LTP. As shown in Fig. 3.1(b) LTP quantizes GLD into three fixed levels, whereas the MQLBP extends the idea to quantize GLD into the desired number of levels. Moreover, LTP ignores the sign information when GLD is quantized into the middle level (level 0) which limits its discrimination power. As shown in Fig. 3.1(c)-(d), the proposed method overcomes this limitation by symmetrically quantizing the GLD with respect to zero. The proposed quantization function \( f_L(x,t) \) for level \( L \) \((L > 0)\) is formally defined as follows:

\[
f_L(x,t) = \begin{cases} 
0, & x < -(L-1)t \\
1, & -(L-1)t \leq x < -(L-2)t \\
\vdots \\
L-1, & -t \leq x < 0 \\
L, & 0 \leq x < t \\
L+1, & t \leq x < 2t \\
\vdots \\
2L-2, & (L-2)t \leq x < (L-1)t \\
2L-1, & x \geq (L-1)t 
\end{cases}
\]  

Thus 1-level quantization, as used in original LBP, segments the whole output range in two different parts. At each higher level of quantization, each sub-part is further quantized into two different segments. Hence, it results in a total of \( 2L \) segments and MQLBP code corresponding to \( i^{th} \) segment is computed as follows:

\[
\text{MQLBP}_{p,r,L}(x_c, y_c) = \sum_{p=0}^{p=1} \delta \left( f_L(g_p - g_c, t) - i \right) 2^p, \quad \begin{cases} 
i = 1, & L = 1 \\
0 \leq i \leq 2L-1, & L > 1 
\end{cases}
\]  

The binary function \( \delta(x) \) is defined as below:
\[
\delta(x) = \begin{cases} 
1, & x = 0 \\
0, & \text{otherwise}
\end{cases}
\]  

(3.9)

Fig. 3.3 shows the numerical example, which also highlights computational differences between LBP, LTP, and MQLBP. It is important to note that MQLBP is capable of retaining both the sign and magnitude information about the GLD computed in the first step outlined above. Hence, we hypothesize that MQLBP should enhance the discrimination ability of LBP. An example showing improved discrimination ability of 2-level MQLBP over LBP and CLBP is shown in Fig. 3.4. The figure shows sample face images of each gender from PAL database. As shown in the figure, the enlarged rectangular patches of both the images represent the same sign vector (CLBP_S). Hence, the resulting LBP and CLBP_S codes are the same as shown in the figure. However, it is difficult to state that both the patches share similar local structure because they belong to two different genders.

Fig. 3.3 Calculation of LBP, LTP and MQLBP operators. The LTP splits the ternary pattern into two binary patterns by retaining information about +1 and -1 in the corresponding patterns. The 2-level MQLBP generates four different binary patterns corresponding to each quantization level.
Though CLBP can improve the discrimination ability by the use of magnitude component (CLBP_M), their CLBP_S codes are still the same. Hence, out of two CLBP codes, only one of them is different in contrast to four different codes generated by MQLBP. Thus, the less correlation between the MQLBP codes of two different patches may provide better discrimination as compared to CLBP. The required codes were obtained by assuming the threshold \( t = 5 \), without loss of generality.

Before presenting experimental results, we would like to highlight the differences between the proposed work and the work in [141]. To make comparison easier, the differences are tabulated in Table 3.1.

Further, it is possible to have the same contrast measure for two different texture patterns. To understand this point, consider the illustrative example shown in Fig. 3.5 which shows two different blocks and resulting codes for LBP, variance (VAR) and MQLBP. As shown in the figure, the LBP and VAR result in the same codes for both the blocks. However, it is difficult to conclude that they have the same texture. Thus the combination of LBP and VAR fails to distinguish the difference between the patterns represented by these blocks. However, the proposed method provides different codes for each block and thus capable to discriminate between them. In this example, the VAR values are considered to be exactly the same. However, even if their continuous values are not exactly the same, their discrete values
may still fall into the same histogram bin after quantization which can limit the discrimination capability.

Table 3.1 Comparison of the proposed approach with the work in [141]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ylioinas et al. [141]</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoding of contrast information</td>
<td>Implicit</td>
<td></td>
</tr>
<tr>
<td>Tuning Parameters</td>
<td>Two:</td>
<td>One:</td>
</tr>
<tr>
<td></td>
<td>1. Number of histogram bins</td>
<td>1. Threshold</td>
</tr>
<tr>
<td></td>
<td>2. Bin thresholds</td>
<td></td>
</tr>
<tr>
<td>Nature of output</td>
<td>Continuous-valued which demands additional quantization step and extra training to determine the quantization thresholds</td>
<td>Discrete-valued output which eliminates the extra training cost</td>
</tr>
<tr>
<td>Computation</td>
<td>Requires to compute local variance which is a second-order operation</td>
<td>Needs to compute local difference which is the first-order operation</td>
</tr>
<tr>
<td>Role of center pixel within the neighborhood window</td>
<td>No active participation of the center pixel in computing the variance</td>
<td>Center pixel plays an important role as neighboring pixels are compared with it</td>
</tr>
</tbody>
</table>

Fig. 3.5 Comparison of discrimination ability of VAR and MQLBP
3.4 Face description using MQLBP

To describe each face image, a histogram is computed from each MQLBP image. However, a simple computation of the histogram of the whole image ignores the location information. The components of a face image follow the specific spatial order, i.e. forehead at the top, followed by eyes, nose, lips and jaw. Thus, efficient face representation requires spatial information which is retained by segmenting the face image into small regions as suggested in [63]. Hence, each MQLBP image is divided first into non-overlapping rectangular blocks. The local histograms are then computed from various blocks of the image. The local histograms are then concatenated to build a feature vector. These feature vectors of all MQLBP images are further concatenated to build a final face descriptor. The flowchart

Fig. 3.6 Flowchart of the proposed MQLBP algorithm
summarizing the steps involved in the proposed algorithm is illustrated in Fig. 3.6. It is clear that feature dimension is directly proportional to a number of levels used in computing MQLBP. To demonstrate more discrimination power of MQLBP while keeping lower feature dimensionality, we consider only two-level MQLBP in this work.

3.5 Experimental setup

To evaluate the performance of the proposed method, comprehensive experiments were carried out on four different datasets. The performance of the proposed method was also compared with several popular LBP variants on these datasets. Moreover, experiments were also performed on noisy data to investigate the noise tolerance capability of the proposed method.

3.5.1 Datasets

The proposed approach is evaluated on four different publicly available datasets: FERET [9], PAL [11], CASIA [169] and FEI [170].

The FERET dataset is one of the most widely used databases in facial gender classification studies. It comprises good quality gray scale images of 1199 subjects. In our work, we employ the same subset as used in [12] which was also followed in [37, 127]. On visual verification, we have found one duplicate image which was discarded, making a total of 410 images.

The PAL dataset contains frontal images of 580 individuals. This dataset is very challenging as it contains individuals with large age variations (18 to 93 years) and five different ethnicities. To keep a balance between male/female ratio, this work has used a subset containing 221 images of each gender.

The CASIA dataset contains 2500 images from 500 young individuals from Asian origin. The images contain variations in pose and illumination. For our experiments, we have used a subset containing frontal face images of 454 subjects.

The FEI database consists of 400 images of 200 individuals. All images are frontal face images with two expressions: Neutral and smiling face. This dataset is balanced in terms of gender and contains 100 individuals from each gender. This work has used 200 images corresponding to neutral expression only.

The examples of cropped facial images from different datasets are depicted in Fig. 3.7. Table 3.2 presents a summary of images used in experiments.
Fig. 3.7 Examples of cropped facial images from different datasets: (a) FERET, (b) PAL, (c) CASIA, and (d) FEI

### 3.5.2 Evaluation protocol

The main concern in classification problem is how the classifier performs on unseen test data. To estimate the performance, test data is often omitted during training and model building. This reduces the size of training data. Moreover, the available dataset is only one possible realization from point of view of stochastic sampling process. Hence, the adequacy and diversity of error measure are decreased by fixed partitioning the dataset into training and test sets [171].
Table 3.2 Summary of images used in experiments

<table>
<thead>
<tr>
<th>Database</th>
<th>Male images</th>
<th>Female images</th>
<th>Total images</th>
</tr>
</thead>
<tbody>
<tr>
<td>FERET</td>
<td>209</td>
<td>201</td>
<td>410</td>
</tr>
<tr>
<td>PAL</td>
<td>221</td>
<td>221</td>
<td>442</td>
</tr>
<tr>
<td>CASIA</td>
<td>228</td>
<td>226</td>
<td>454</td>
</tr>
<tr>
<td>FEI</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>

To deal with such issues, five-fold cross validation scheme is a commonly used evaluation protocol in gender recognition literature [27, 28, 30, 44, 127, 138, 172, 173]. The five-fold cross validation scheme randomly divides the data into five non-overlapping subsets and at a time, only one subset (20% data) is used as a test set while the other four subsets (80% data) are used to train a classifier. This procedure is repeated five times so that each subset is used exactly once as a test set. Finally, the average of five different classification rates is computed. Thus, this protocol obtains five independent estimates of the error measure and every data samples are used for training and testing both. As these five error measures are averaged to obtain an estimate of overall error measure, typically it will be more robust than single measures [171].

For the reasons mentioned above, this work also employed five-fold cross validation scheme to evaluate the performance of gender classifier. Moreover, stratified partitions were used to ensure similar gender distribution in each fold as present in the whole dataset. Further, as suggested in [2, 42, 138] it was ensured to use only a single image per each subject in every dataset to avoid a classifier from learning identity-related information.

To compare the performance of different features, overall classification rate (OCR) has been used as defined below:

$$\text{OCR} = \frac{\text{Number of correctly classified images}}{\text{Total number of images}} \times 100\% \quad (3.10)$$

Further, the classification rates for each type of gender were also measured using two parameters: male classification rate (MCR) and female classification rate (FCR).

$$\text{MCR} = \frac{\text{Number of correctly classified male images}}{\text{Total number of male images}} \times 100\% \quad (3.11)$$

$$\text{FCR} = \frac{\text{Number of correctly classified female images}}{\text{Total number of female images}} \times 100\% \quad (3.12)$$
3.5.3 Parameter setting

To extract the features the original images were cropped after detecting the faces using a cascaded face detection algorithm [35]. Further to maintain uniformity, the cropped images were resized to 48×48 pixels as suggested in [12, 28]. The works in [12, 138] suggested to avoid automatic face alignment step as it does not apparently improve the classification performance. Hence, face alignment was not performed in this work. Further, for a fair comparison and to investigate the actual capability of local features, no other form of pre-processing was applied.

The proposed MQLBP method is compared with four different LBP variants: LBP, CLBP, ELBP and LTP. All methods used uniform patterns extracted with $P = 8$ and $R = 1$. The thresholding parameter $t$ required in LTP and MQLBP was varied from 1 to 15 for all experiments. To compute spatially enhanced histograms, the images were divided into non-overlapping blocks of size $8 \times 8$ resulting in a total of 36 blocks. To focus on the strength of the features rather than the classifier, we have used a linear support vector machine (SVM) for classification and fixed the soft margin parameter $C = 100$ as done in [37].

3.6 Experimental results

Different datasets contain variations in pose, expression, ethnicity and capturing device used for collecting images. Such variations cause different classification rates for each dataset as shown in Table 3.3 to Table 3.6. The second and third columns in the tables show the classification rates for the male and female category. The last column shows the mean overall classification rates after five-fold cross validation. The number in the brackets represents the standard deviation of classification rates of five folds. From these tables, it is clear that the proposed MQLBP method achieved the highest classification rates across all the datasets. Hence, the results demonstrate the higher discrimination capability of the proposed method. Further inferences from these tables are as follows:

1. Multi-quantization schemes (as used in LTP and MQLBP) perform significantly better than binary quantization scheme of LBP.

2. Although better than LBP, the results of LTP could not compete with the MQLBP. The reason is one extra quantization level used in MQLBP. The LTP used three quantization levels against four used in the MQLBP.
3. The MQLBP performed better than CLBP and ELBP. Both the latter methods use explicit encoding for sign and magnitude components of the GLD. In contrast, the MQLBP employs more natural and implicit encoding which turns out to perform better.

4. The maximum improvement in performance of MQLBP over other feature types is found on PAL dataset. This is a significant result because PAL dataset is relatively more challenging as the images contain large variations in age (18 to 93 years) and they cover five different ethnic groups: African, American, Asian, Caucasian and Hispanic.

5. Various methods performed differently for male and female categories across four datasets. However, the proposed method provides more balanced performance than other methods. To understand this point, mean absolute difference (MAD) was computed between the MCR and FCR across all the datasets. The values of MAD for different features are as follows: 2.60% for LBP, 1.24% for CLBP, 2.37% for ELBP, 1.77% for LTP and 1.21% for MQLBP. Hence, the proposed method resulted in the most balanced classification performance for the male and female category, although the CLBP also performed reasonably well. The LBP demonstrated the largest performance gap between the male and female categories across all the datasets.

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>93.76</td>
<td>90.55</td>
<td>92.20</td>
</tr>
<tr>
<td></td>
<td>(2.79)</td>
<td>(6.47)</td>
<td>(4.10)</td>
</tr>
<tr>
<td>CLBP</td>
<td>93.75</td>
<td>93.52</td>
<td>93.66</td>
</tr>
<tr>
<td></td>
<td>(3.73)</td>
<td>(3.80)</td>
<td>(2.18)</td>
</tr>
<tr>
<td>ELBP</td>
<td>93.74</td>
<td>90.54</td>
<td>92.20</td>
</tr>
<tr>
<td></td>
<td>(5.08)</td>
<td>(2.14)</td>
<td>(2.22)</td>
</tr>
<tr>
<td>LTP</td>
<td>94.69</td>
<td>93.04</td>
<td>93.90</td>
</tr>
<tr>
<td>(t = 1)</td>
<td>(5.31)</td>
<td>(4.47)</td>
<td>(3.34)</td>
</tr>
<tr>
<td>MQLBP</td>
<td>95.19</td>
<td>95.02</td>
<td>95.12</td>
</tr>
<tr>
<td>(t = 5)</td>
<td>(3.45)</td>
<td>(1.77)</td>
<td>(1.93)</td>
</tr>
</tbody>
</table>
Table 3.4 Classification rates (%) on PAL dataset

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>91.40</td>
<td>90.49</td>
<td>90.96</td>
</tr>
<tr>
<td></td>
<td>(1.89)</td>
<td>(4.93)</td>
<td>(2.84)</td>
</tr>
<tr>
<td>CLBP</td>
<td>92.78</td>
<td>92.78</td>
<td>92.76</td>
</tr>
<tr>
<td></td>
<td>(4.88)</td>
<td>(3.30)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>ELBP</td>
<td>91.88</td>
<td>90.48</td>
<td>91.18</td>
</tr>
<tr>
<td></td>
<td>(5.15)</td>
<td>(5.90)</td>
<td>(5.06)</td>
</tr>
<tr>
<td>LTP (t = 15)</td>
<td>93.70</td>
<td>93.68</td>
<td>93.67</td>
</tr>
<tr>
<td></td>
<td>(4.25)</td>
<td>(2.44)</td>
<td>(2.04)</td>
</tr>
<tr>
<td>MQLBP (t = 15)</td>
<td>95.96</td>
<td>95.03</td>
<td>95.48</td>
</tr>
<tr>
<td></td>
<td>(4.27)</td>
<td>(2.94)</td>
<td>(2.64)</td>
</tr>
</tbody>
</table>

Table 3.5 Classification rates (%) on CASIA dataset

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>91.67</td>
<td>89.39</td>
<td>90.53</td>
</tr>
<tr>
<td></td>
<td>(5.01)</td>
<td>(4.23)</td>
<td>(2.40)</td>
</tr>
<tr>
<td>CLBP</td>
<td>93.87</td>
<td>91.15</td>
<td>92.51</td>
</tr>
<tr>
<td></td>
<td>(4.17)</td>
<td>(5.67)</td>
<td>(1.67)</td>
</tr>
<tr>
<td>ELBP</td>
<td>91.25</td>
<td>89.36</td>
<td>90.31</td>
</tr>
<tr>
<td></td>
<td>(4.04)</td>
<td>(4.59)</td>
<td>(1.46)</td>
</tr>
<tr>
<td>LTP (t = 2)</td>
<td>93.00</td>
<td>91.59</td>
<td>92.29</td>
</tr>
<tr>
<td></td>
<td>(3.87)</td>
<td>(4.81)</td>
<td>(1.75)</td>
</tr>
<tr>
<td>MQLBP (t = 14)</td>
<td>94.30</td>
<td>91.58</td>
<td>92.95</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(4.29)</td>
<td>(1.71)</td>
</tr>
<tr>
<td>Feature</td>
<td>MCR</td>
<td>FCR</td>
<td>OCR</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>LBP</td>
<td>94.00</td>
<td>98.00</td>
<td>96.00</td>
</tr>
<tr>
<td></td>
<td>(6.52)</td>
<td>(2.74)</td>
<td>(3.35)</td>
</tr>
<tr>
<td>CLBP</td>
<td>95.00</td>
<td>97.00</td>
<td>96.00</td>
</tr>
<tr>
<td></td>
<td>(7.07)</td>
<td>(4.47)</td>
<td>(3.79)</td>
</tr>
<tr>
<td>ELBP</td>
<td>94.00</td>
<td>97.00</td>
<td>95.50</td>
</tr>
<tr>
<td></td>
<td>(6.52)</td>
<td>(4.47)</td>
<td>(4.11)</td>
</tr>
<tr>
<td>LTP</td>
<td>95.00</td>
<td>99.00</td>
<td>97.00</td>
</tr>
<tr>
<td>(t = 7)</td>
<td>(6.12)</td>
<td>(2.24)</td>
<td>(2.74)</td>
</tr>
<tr>
<td>MQLBP</td>
<td>98.00</td>
<td>99.00</td>
<td>98.50</td>
</tr>
<tr>
<td>(t = 14)</td>
<td>(2.74)</td>
<td>(2.24)</td>
<td>(1.37)</td>
</tr>
</tbody>
</table>

It is clear from the tables that the proposed method performed superior to other methods across all the datasets. The gain in classification rates of the proposed method over other feature types is graphically depicted in Fig. 3.8.

![Fig. 3.8 Performance gain of MQLBP over other feature types](image-url)
Examples of misclassified images from different datasets are shown in Fig. 3.9. The misclassifications occur when intra-class variations exceed inter-class variations. Some of the major sources of misclassifications are variations due to accessories, head pose, make-up and facial expressions. The top panel of the figure shows facial images which are really challenging and even difficult for humans to clearly determine their gender. The images in the second row are misclassified due to accessories like eye glasses, earrings, and scarf. Misclassifications due to pose variations are shown in the third row of the figure. The last row of the figure shows misclassification due to make-up or variations in facial expressions.

3.7 Effect of number of levels

Previous section presented experimental results for two-level MQLBP. It is interesting to study the effect of a number of levels on the classification performance. Hence, the threshold parameter was fixed to ten without loss of generality and the number of levels was varied from 2 to 4. All other parameters were kept as per the setting described in Section 3.5.3. The comparison of the performance for different number of MQLBP levels is depicted in Fig. 3.10.

It is clear from the figure that across all the datasets, the performance does not improve by increasing the number of levels of MQLBP. As we increase the number of levels, the length of the features also increases proportionately and such a large feature dimension might have made the job of classifier difficult. As two-level MQLBP provides better classification
performance with lower feature dimension, the experiments presented in the rest of the chapter were conducted using two-level MQLBP only.

3.8 Effect of age variations

Among various datasets, the PAL dataset provides age information of various subjects participated in the creation of the dataset. The images in the dataset are categorized into four categories based on age information: 18-29 years, 30-49 years, 50-69 years and 70-93 years. Hence, the dataset contains large age variations and it is interesting to analyze how various descriptors perform on different age groups.
Fig. 3.11 shows classification rates of various features for different age groups. The LBP descriptor performed well on the first two age groups (18-29 and 30-49 years). However, its performance dropped severely for subjects having age over 50 years. Similarly, the performance of CLBP also degraded significantly on 70-93 years age group. On the other hand, the LTP performed well on the last two age groups; however, its performance dropped on the second age group (30-49 years). The proposed MQLBP descriptor outperformed other descriptors on all but one age group (50-69 years) where LTP recorded the best classification rate. However, in comparison with other descriptors, the MQLBP performed more consistently with over 93.00% classification rates for all the age groups.

3.9 Experiments on noisy data

This experiment was designed to evaluate the robustness of the MQLBP against noise degradation. For this experiment, each dataset was split into two fixed parts with 60% data for the training set and remaining 40% data for the test set. The training set contained clean images without any type of noise. All the images in the test set were degraded by adding Gaussian noise with zero mean and different levels of variance corresponding to 5, 10, 15 and 20 dB signal-to-noise ratio (SNR). For each level of SNR, the experiment was repeated 20 times to reduce variability in classification rates. The examples of noisy images are illustrated in Fig. 3.12. The original sample image was taken from the PAL database.

![Examples of noisy images with different SNR: (a) 5 dB (b) 10 dB (c) 15 dB (d) 20 dB](image)

Fig. 3.12 Examples of noisy images with different SNR: (a) 5 dB (b) 10 dB (c) 15 dB (d) 20 dB

The plots of classification rate vs. threshold \( t \) at four different noise levels for each dataset are shown in Fig. 3.13 to Fig. 3.16. It is important to remember that the LBP, CLBP, and ELBP are independent of threshold \( t \) and hence they do not show any variations. It is clear from these plots that the performance of these three descriptors is degraded significantly in the presence of noise across all the datasets. In contrast, the LTP and MQLBP descriptors provided promising classification performance even in the presence of noise. Further, under the optimal setting of threshold, the MQLBP performs better than LTP at all noise levels for all the databases. Especially, it is known that detection of facial texture under a high level of
noise (5 dB SNR) is an extremely challenging task. Despite this fact, the MQLBP achieved classification rates above 60.00% in all datasets. Hence, these results clearly demonstrate the robustness of the proposed method against noise degradation.

Fig. 3.13 Classification performance in the presence of noise on FERET database

Fig. 3.14 Classification performance in the presence of noise on PAL database
Fig. 3.15 Classification performance in the presence of noise on CASIA database

Fig. 3.16 Classification performance in the presence of noise on FEI database
3.10 Comparison of computation time

The experimental results indicated that the MQLBP not only provides higher discrimination power but it is also robust against age variations and noise. These benefits have come at the cost of increased feature vector length which requires more computation time.

Table 3.7 shows the comparison of feature vector length and computation time for various local features. To calculate the computation time, 300 FERET images were randomly chosen. The features were extracted with parameter settings described in Section 3.5.3. For each image, the time required to obtain the specified feature vector was measured. Finally, the mean of computation time is reported for each of the feature vectors. The experiments were carried out using MATLAB 2013a on Intel quad-core CPU @ 3.40 GHz with 8 GB RAM. The proposed method requires 10.0 ms time with a feature size of 8496 which is 4 times of LBP and 2 times of CLBP and LTP. However, it is important to note that we have used unoptimized MATLAB code to measure computation time. Further, such computational cost is usually under control in view of high-performance computing machines being easily available nowadays.

Table 3.7 Comparison of computation time of different features

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature length</th>
<th>Computation time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>$59 \times 36 = 2124$</td>
<td>3.0</td>
</tr>
<tr>
<td>CLBP</td>
<td>$118 \times 36 = 4248$</td>
<td>5.3</td>
</tr>
<tr>
<td>LTP</td>
<td>$118 \times 36 = 4248$</td>
<td>5.2</td>
</tr>
<tr>
<td>ELBP</td>
<td>$236 \times 36 = 8496$</td>
<td>10.3</td>
</tr>
<tr>
<td>MQLBP</td>
<td>$236 \times 36 = 8496$</td>
<td>10.0</td>
</tr>
</tbody>
</table>

3.11 Summary

Local binary patterns (LBP) employ the binary quantization on a gray level difference between a center pixel and a neighboring pixel. This simple yet effective method, however, discards the magnitude of gray level difference. To address this issue, we have extended the vector quantization concept which is not restricted to binary quantization. The proposed MQLBP approach implicitly encodes both the sign and magnitude information of gray level
difference which improves the discrimination capability. In contrast to LTP, CLBP, and ELBP which use fixed number of quantization levels, the proposed method is more generic as the desired number of quantization levels can be obtained. To demonstrate the effectiveness of the proposed approach, comprehensive experiments were carried out using two-level MQLBP on four different datasets such as FERET, PAL, CASIA, and FEI. The results clearly indicated that the proposed method has several advantages including higher discrimination power, robustness against age variations and better noise tolerance capability.

56
CHAPTER 4: COMPASS LOCAL BINARY PATTERNS FOR GENDER CLASSIFICATION

This chapter describes the design of new feature extraction method, compass local binary pattern (CoLBP) for facial gender recognition. To achieve robustness, the proposed method first computes directional edge responses using eight Kirsch compass masks. Then, the spatial relationships among the neighboring pixels in each edge response are exploited independently with the help of local binary pattern (LBP) to enhance the discrimination capability. Finally, spatial histograms computed from these LBP images are concatenated to build a face descriptor. Our proposed descriptor efficiently extracts discriminating information from four different levels, including gradient, regional, global and directional level. The proposed method was evaluated on three datasets (color FERET, LFW, and Adience) containing facial photographs. In spite of a wide range of challenges (low resolution, variations in pose, expression, and illumination) present in the datasets, the proposed method provided promising classification performance in comparison with several existing benchmark methods, thereby validating its robustness.

4.1 Introduction

Ever since the adoption of LBP for facial representation task [63], it has gained widespread interest among computer vision researchers. However, the direct operation on raw pixel intensities limits the discrimination capability and robustness of LBP. With the aim of enhancing the discrimination power and robustness, several researchers employed a different pipeline as shown in Fig. 4.1.

![Fig. 4.1 LBP-based facial representation pipeline](image-url)
The first stage in the pipeline generates the feature map from the input image. The feature map can be simply raw pixel intensities or it can include some form of filtering operations such as Gabor filtering. The local binary pattern coding can then be applied to the feature map. Finally, the spatial histogram is obtained by concatenating several regional histograms as described in Section 3.4.

Several researchers have employed convolution filtering to pre-process the image and generate the feature map. The LBP coding scheme is then applied on these filtered images. For example, Zhang et al. [174] proposed local Gabor binary pattern histogram sequence (LGBPShS) which first applied 40 multi-scale and multi-resolution Gabor filters on the original image. Then they applied LBP on these 40 Gabor magnitude images. The additional Gabor filtering step provides significant performance gain over original LBP algorithm. In contrast to using Gabor magnitude, Gabor phase information had been used in [175] to improve the performance. The method, namely histogram of Gabor phase pattern (HGPP) utilized 90 images from both the real and imaginary parts of Gabor phase to generate local and global Gabor phase patterns. Later Xie et al. [176] proposed local Gabor XOR patterns (LGXP) to encode Gabor phase information which was fused with Gabor magnitude to enhance the performance.

In contrast to Gabor filtering, Yang et al. [177] adopted Riesz transform based monogenic signal analysis for image decomposition. For encoding monogenic magnitude, they utilized LBP-like scheme while monogenic orientation was encoded using the quadrant bit coding method. The resulting pattern was named as a monogenic binary pattern (MBP). They evaluated the performance of MBP on face recognition task and demonstrated its superior performance over LBP.

The LGBPShS and MBP have been successfully applied to face recognition task and they are considered as the state-of-the-art methods in the literature. However, their performance on facial gender classification has not been evaluated. Hence, this chapter reports their performance on gender classification task. Further, the major drawback of these methods is their high feature dimension which increases the computational complexity. To reduce the computational complexity while retaining the discrimination capability, a new descriptor namely compass local binary pattern is proposed in this chapter.

The proposed method employs the Kirsch compass masks for image filtering to generate a feature map. The use of compass masks is inspired by several methods such as local directional patterns (LDiP) [29], enhanced LDiP (EnLDiP) [178], local directional...
number pattern (LDN) [179] and eight LDiP (ELDiP) [180]. These methods have also employed Kirsch compass masks for image filtering. However, they have adopted different pattern map coding schemes for designing the local descriptors.

Aforementioned methods (LDiP, EnLDiP, LDN, ELDiP) first convolve the original image with Kirsch compass masks to generate eight edge response images. Then they consider the relation among different edge response values of each pixel to extract useful structural information. However, for a given edge response image, the relationship among neighboring pixels representing gradient magnitudes is not considered. Different edge responses are obtained by filtering the original face image through eight different masks. Hence, they can be used independently to exploit the spatial relationship among neighboring pixels in each edge response image. Such spatial relationships among local directional gradients would provide more structural information than LDiP and its variants mentioned above. Hence, it is proposed to apply LBP on each of the eight edge responses obtained from Kirsch mask filtering. Thus, eight different LBP images will be generated from which the spatial histograms are computed as mentioned in [63]. These spatial histograms are then concatenated to build a single descriptor. As this descriptor is generated by applying LBP on compass mask filtered images, the proposed method is called as compass LBP (CoLBP).

The proposed method is similar to LGBPHS, HGPP, and MBP since it also adopts a serial combination of convolution filter and LBP for performance gain. Among these methods, HGPP relies on phase information while other methods utilize magnitude information. The other major difference between these methods lies in the choice of the convolution filter. The LGBPHS and HGPP methods employ Gabor filters which are specified by a continuous function that can be scaled and rotated to any arbitrary orientation. These methods demand more parameter specifications and computation efforts for constructing Gabor kernels which require modulation of Gaussian function by a complex plane wave. In contrast, the proposed method employs Kirsch compass masks which use fixed kernels oriented at 45° intervals. Hence, Kirsch masks can be implemented with relatively smaller computation efforts. Further, as mentioned earlier, Gabor-based methods require more convolution operations (40 and 90 for LGBPHS and HGPP respectively) that not only require more computation time but also increase the storage space due to severely high feature dimension. Similarly, as shown in Section 4.9, the MBP also requires relatively high feature dimension and computation time.
Classification rate and computational complexity are two important measures for comparing the performance of different feature extraction algorithms. The methods such as LBP and Kirsch mask based directional patterns (LDiP, LDN etc.) provide compact feature descriptors with very low computational efforts. However, they suffer from limited discrimination capability which can be improved by LGBPHS and MBP. On the other side, the LGBPHS and MBP demand high computational complexity. Hence, it is desirable to have a balance between classification performance and computational complexity. As an attempt to make this balance, compass LBP method is proposed in this chapter.

4.2 Existing local patterns

This section aims to provide a background to understand the proposed method. Here, some of the relevant local feature extraction methods based on Kirsch masks are briefly reviewed.

The LBP operates directly on raw pixel intensities which may limit its discrimination ability. In contrast, gradient images contain enhanced edge information. Further, gradient images are more stable than pixel intensities, making it more robust against noise and illumination variations [178, 180]. In view of these advantages, LDiP operator was derived in [29] using Kirsch compass masks \( \{ M^0, M^1, \ldots, M^7 \} \) shown in Fig. 4.2.

\[
\begin{bmatrix}
  -3 & -3 & 5 \\
  -3 & 0 & 5 \\
  -3 & -3 & 5 \\
  5 & -3 & -3 \\
  5 & 0 & -3 \\
  5 & -3 & -3 \\
  5 & -3 & -3 \\
  -3 & -3 & -3
\end{bmatrix}
\]

\[
\begin{bmatrix}
  -3 & 5 & 5 \\
  -3 & 0 & 5 \\
  -3 & -3 & -3 \\
  5 & -3 & -3 \\
  0 & -3 \\
  5 & -3 & -3 \\
  5 & 0 & 5 \\
  -3 & 5 & 5
\end{bmatrix}
\]

Fig. 4.2 Kirsch compass masks

The original face image is first convolved with Kirsch compass masks to obtain eight edge response images. Afterward, the bits corresponding to top three edge magnitudes are set to one while remaining bits are set to zero in an 8-bit LDiP code which is finally converted to a decimal form. A couple of following works considered two-digit octal numbers which were then converted to decimal code. Enhanced LDiP [178] used two most prominent edge directions. However, it discards the sign of edge responses, making it unable to differentiate intensity variations (dark to bright or vice versa). To overcome this limitation, Rivera et al.
[179] proposed local directional number pattern (LDN) which encodes the top positive and top negative directions. The aforementioned methods encode directional information using a subset of edge responses. The recent work in [180] urged that all the edge directions are important. Hence, they used all of them to build an 8-bit binary code by retaining their sign. The binary number was then converted to decimal to obtain a descriptor called eight local directional patterns (ELDiP). Fig. 4.3 illustrates the coding scheme of aforementioned local patterns.

![Fig. 4.3 Coding scheme of existing local patterns](image)

### 4.3 Proposed method

Unlike aforementioned methods, the proposed method treats each edge response separately to exploit the spatial relationships among neighbors within that edge response. Such treatment of encoding each edge response differently has potential to reveal more structural information which is so valuable for face representation. The steps of the proposed compass LBP method are outlined in Fig. 4.4.

Given the original image \( I \), its eight edge responses \( E_i \) are computed by convolving it with each of the Kirsch masks \( M_i \) as follows:
Fig. 4.4 The framework of the proposed method

\[ E^i = I \ast M^i, \ i = 0, 1, 2, ..., 7 \]  \hspace{1cm} (4.1)

The LBP is then applied on each of the edge responses to obtain compass LBP images \( C^i \):

\[ C^i = \text{LBP}_{p,R} \{ E^i \} = \sum_{p=0}^{P-1} f(\epsilon_p^i - \epsilon_c^i) 2^p \]  \hspace{1cm} (4.2)

where \( \epsilon_c^i \) and \( \epsilon_p^i \) indicate the edge response values of the center pixel and its \( p^\text{th} \) neighbor in \( i^\text{th} \) edge response \( E^i \). Fig. 4.5 shows the example of eight compass LBP images for the original face image shown in the center.

Finally, to describe each face, spatially enhanced histograms are computed by dividing each compass LBP image into rectangular blocks.

Let \( H_j^i \) denotes the histogram of \( j^\text{th} \) region \( R_j (j = 0, 1, 2, \ldots, N-1) \) in \( i^\text{th} \) compass LBP image \( C^i (i = 0, 1, 2, \ldots, 7) \). The histogram \( H_j^i \) is defined as:

\[ H_j^i(k) = \sum_{(x,y) \in R_j} \delta(C(x,y) - k), \ k \in [0, K] \]  \hspace{1cm} (4.3)
where \((x, y)\) denotes a pixel position in the region \(R_j\) and \(K\) represents the maximum possible value of compass LBP code. Without loss of generality, we have used uniform local binary patterns in this work. Hence, the value of \(K\) equals \(P(P-1)+2\) for \(P\) neighboring pixels.

Finally, the compass LBP histogram (CLH) is obtained by concatenating the histograms \(H_j^i\) for each region across all eight compass LBP images:

\[
\text{CLH} = \prod_{i=0}^{7} \prod_{j=0}^{N-1} H_j^i
\]  

(4.5)
where $\prod$ denotes the concatenation operation and $N$ represents the number of regions in the segmented face. The spatial histogram CLH represents the information at four levels:

1. Gradient level information is captured by different edge response images.
2. As the LBP codes are summed to create a histogram for each region, it produces regional level information.
3. Concatenation operation of all the regional histograms for each compass LBP image generates the global description.
4. Concatenation of histograms of different compass LBP images captures the directional information as well.

It is hypothesized that the combination of these four levels of information can enhance the facial characteristics.

To gain further insight, correlation coefficients were computed between different pairs of compass LBP images. The total number of pairs is given by $C_8^2$ resulting in 28 pairs. The correlation coefficient $\gamma$ between two images $C^i$ and $C^j$ each of size $m \times n$ is defined as:

$$
\gamma_{ij} = \frac{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (C^i(x,y) - \mu^i)(C^j(x,y) - \mu^j)}{\sqrt{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (C^i(x,y) - \mu^i)^2} \sqrt{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (C^j(x,y) - \mu^j)^2}}, \quad (i, j) \in [0, 7], \quad i \neq j \quad (4.6)
$$

where $\mu^i$ and $\mu^j$ are the mean of images $C^i$ and $C^j$ respectively.

If the correlation coefficient is +1, it means the two images have a positive linear relationship while -1 indicates a negative linear relationship. When the value of correlation coefficient is close to zero, it means that two images have a weak linear relationship between them. Fig. 4.6 shows the box plot of 28 correlation coefficients computed from a subset of the color FERET dataset containing 200 images. It is seen that different pairs of compass LBP images are not highly correlated. In fact, out of 5600 (28x200) pairs, 70.48% pairs had $|\gamma_{ij}| < 0.5$. Thus, different compass LBP images carry complementary information which is utilized effectively by the proposed method to enhance discrimination power.

Before presenting the experiments and results, a summary with comparison among different feature extraction methods is presented in Table 4.1.
4.4 Experimental setup

The datasets used in Chapter 3 mainly contained images captured under a controlled environment. In this chapter, we go further and evaluate the proposed CoLBP method on real-world face images captured in the wild. Since the images are captured in an uncontrolled environment, they pose several challenges like low resolution, occlusions, and variations in illumination, pose, and facial expressions. To deal with such challenges, the method proposed in this chapter has been evaluated on two recently proposed real-life face datasets: labeled faces in the wild (LFW) [72] and Adience dataset [172]. Further, since color FERET dataset [9] is one of the most widely used datasets, it is also considered in this work for benchmarking purpose. The original images of the datasets were converted into the grayscale format and then faces were detected using Viola-Jones face detector [26] with the exception of LFW dataset for which readily available cropped faces [181] were used. The cropped faces were resized to 48x48 pixels and no other kind of pre-processing was applied to have a fair comparison and to evaluate the actual capability of local features. Some examples of cropped faces from these datasets are shown in Fig. 4.7. The summary of the number of images used in each dataset is presented in Table 4.2.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Methodology</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>Threshold the gray level difference between the neighboring pixels and the center pixel</td>
<td>• Computationally efficient&lt;br&gt;• Invariant to monotonic illumination changes</td>
<td>• Operates on raw pixel intensity limiting its robustness and discrimination power.</td>
</tr>
<tr>
<td>LDiP</td>
<td>Threshold the directional edge responses based on their prominence</td>
<td>• Captures directional information from edge responses</td>
<td>• The choice of a number of directions has a significant impact on classification performance.&lt;br&gt;• Ignores the sign of edge magnitudes.</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>Encode the directions of top two positive edge responses</td>
<td>• Captures directional information from edge responses&lt;br&gt;• Compact encoding scheme</td>
<td>• Unable to differentiate intensity variations.&lt;br&gt;• Focuses on two prominent directions only and ignores the other edge responses.&lt;br&gt;• Discards spatial relation among neighbors of each edge responses.</td>
</tr>
<tr>
<td>LDN</td>
<td>Encode the directions of top positive and top negative edge responses</td>
<td>• Captures directional information from edge responses&lt;br&gt;• Capable of differentiating intensity variations&lt;br&gt;• Compact encoding scheme</td>
<td>• Focuses on two prominent directions only and ignores the other edge responses.&lt;br&gt;• Discards spatial relation among neighbors of each edge responses.</td>
</tr>
<tr>
<td>ELDiP</td>
<td>Threshold all the directional edge responses with reference to zero</td>
<td>• Captures directional information from edge responses&lt;br&gt;• Collects more information by using all the eight directional edge responses</td>
<td>• Discards spatial relation among neighbors of each edge responses.</td>
</tr>
<tr>
<td>CoLBP</td>
<td>Separately apply LBP on each of the edge responses</td>
<td>• Encodes complete structural and directional information from all eight edge responses overcoming the limitations of above methods</td>
<td>• Causes high dimensionality</td>
</tr>
</tbody>
</table>
Fig. 4.7 Examples of cropped face images from three datasets. The first two rows are from color FERET. The middle two rows are from LFW dataset and the last two rows are from Adience dataset.

Table 4.2 Summary of images used in experiments

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Color FERET</th>
<th>LFW</th>
<th>Adience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male images</td>
<td>587 (59.47%)</td>
<td>4259 (74.25%)</td>
<td>840 (47.81%)</td>
</tr>
<tr>
<td>Female images</td>
<td>400 (40.53%)</td>
<td>1477 (25.75%)</td>
<td>917 (52.19%)</td>
</tr>
<tr>
<td>Total images</td>
<td>987</td>
<td>5736</td>
<td>1757</td>
</tr>
</tbody>
</table>

4.4.1 Evaluation protocol and parameter settings

The experiments were carried out by following the same evaluation protocol as described in Section 3.5.2. The performance of various methods has been compared using classification rates (MCR, FCR, and OCR) defined in Section 3.5.2. As an additional measure of performance, receiver operating characteristic (ROC) curves have also been used to analyze the performance of various methods. The ROC curve is a plot of false positive rate (FPR) versus true positive rate (TPR). The TPR is defined as a ratio of the number of correctly classified positive samples to the total number of positive samples in the dataset. The FPR is given by the ratio of the number of misclassified negative samples to the total number of negative samples in the dataset.
The LBP features were extracted using the neighborhood size of $P = 8$ and radius $R = 1$. Further, uniform patterns with maximum two bitwise transitions per LBP code were utilized. The proposed method is also compared with patterns of oriented edge magnitudes (POEM) [182] and three-patch LBP (TPLBP) [183]. For a fair comparison with the proposed feature, the POEM feature was extracted using eight orientations. The LGBP$H$S descriptor was implemented using 40 Gabor filters at five different scales and eight different orientations. The other parameters were set as follows: The maximum frequency $\pi/2$, the standard deviation $2\pi$, and the frequency domain spacing factor between filters $\sqrt{2}$. The MBP descriptor was implemented using the codes provided by the authors [177] with following default filter parameters: The minimal wavelength 4, the multiplying factor of wavelength 0.64, the ratio factor 1.7 and the number of scales 3. For building face descriptors, spatial histograms were obtained (for all feature types) after dividing the local pattern images into disjoint rectangular blocks (sub-regions) of size $8 \times 8$ resulting in a total of 36 blocks per face image. The only exception is the TPLBP descriptor which required only 25 blocks since it considers neighborhood patches instead of pixels [183].

For classification purpose, a simple yet powerful model of linear support vector machine was used since this work focuses on the strength of the features rather than the classifier. The value of soft margin parameter ($C$) was fixed to $C = 100$ as done in [37].

### 4.5 Results on color FERET database

The color FERET is the most commonly used face database for benchmarking the performance of different algorithms for face and gender recognition. The dataset mainly contains color versions of the images used in gray FERET dataset. Further, numerous mistakes present in gray FERET dataset have been rectified in this version of the dataset. It consists of several images from 994 subjects taken under controlled conditions with variations in pose and facial expression. In this work, only single frontal image per subject have been used from $fa$ gallery. As mentioned in Table 4.2, the collection used in this work consists of 587 male images and 400 female images making a total of 987 images.

Table 4.3 shows the results of color FERET dataset. The second and third columns show the classification rates for the male and female category. The last column shows the mean overall classification rates after five-fold cross validation. The numbers in the brackets represent the standard deviation of classification rates of five folds. Following inferences can be drawn from the Table 4.3.
Table 4.3 Classification rates (%) of color FERET database

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>94.04</td>
<td>88.25</td>
<td>91.69</td>
</tr>
<tr>
<td>(2.40)</td>
<td>(4.01)</td>
<td>(1.22)</td>
<td></td>
</tr>
<tr>
<td>ELBP</td>
<td>96.60</td>
<td>87.75</td>
<td>93.01</td>
</tr>
<tr>
<td>(2.54)</td>
<td>(5.41)</td>
<td>(2.85)</td>
<td></td>
</tr>
<tr>
<td>TPLBP</td>
<td>91.48</td>
<td>80.25</td>
<td>86.93</td>
</tr>
<tr>
<td>(2.20)</td>
<td>(2.85)</td>
<td>(1.41)</td>
<td></td>
</tr>
<tr>
<td>LDiP</td>
<td>90.97</td>
<td>87.25</td>
<td>89.46</td>
</tr>
<tr>
<td>(2.54)</td>
<td>(5.11)</td>
<td>(2.67)</td>
<td></td>
</tr>
<tr>
<td>EnLDiP</td>
<td>91.99</td>
<td>87.00</td>
<td>89.97</td>
</tr>
<tr>
<td>(2.32)</td>
<td>(5.49)</td>
<td>(1.13)</td>
<td></td>
</tr>
<tr>
<td>ELDiP</td>
<td>90.63</td>
<td>80.00</td>
<td>86.32</td>
</tr>
<tr>
<td>(2.86)</td>
<td>(5.66)</td>
<td>(3.62)</td>
<td></td>
</tr>
<tr>
<td>LDN</td>
<td>93.36</td>
<td>86.50</td>
<td>90.57</td>
</tr>
<tr>
<td>(4.08)</td>
<td>(1.05)</td>
<td>(2.27)</td>
<td></td>
</tr>
<tr>
<td>POEM</td>
<td>94.38</td>
<td>86.00</td>
<td>90.98</td>
</tr>
<tr>
<td>(2.31)</td>
<td>(5.82)</td>
<td>(2.67)</td>
<td></td>
</tr>
<tr>
<td>LGBPHS</td>
<td>97.11</td>
<td>85.75</td>
<td>92.50</td>
</tr>
<tr>
<td>(0.96)</td>
<td>(4.01)</td>
<td>(1.57)</td>
<td></td>
</tr>
<tr>
<td>MBP</td>
<td>96.43</td>
<td>88.00</td>
<td>93.01</td>
</tr>
<tr>
<td>(2.19)</td>
<td>(2.88)</td>
<td>(2.08)</td>
<td></td>
</tr>
<tr>
<td>CoLBP</td>
<td>97.10</td>
<td>89.25</td>
<td>93.92</td>
</tr>
<tr>
<td>(1.65)</td>
<td>(3.81)</td>
<td>(2.04)</td>
<td></td>
</tr>
</tbody>
</table>

1. The proposed method performs the best achieving close to 94% overall classification rate, which is better as compared to state-of-the-art methods. The LGBPHS and MBP methods also provided competitive performance. Comparing the performance with various Kirsch mask based directional patterns such as LDiP, EnLDiP, ELDiP and LDN, the proposed method clearly performed significantly superior. The former methods ignore spatial relationships among the neighbors in each edge response. In contrast, the proposed method exploited such relationships to collect more structural and textural information resulting it to perform better.
2. The LBP showed the competitive classification rate of 91.69%. The LBP has also performed better on other applications like face recognition, face detection, texture classification and image retrieval. Thus, the results are in agreement with the literature showing good generalization ability of LBP for different applications. Further interesting observation is that the LBP outperformed LDiP. Even though LDiP captures the directional information, the encoding method of LDiP relies excessively on a number of most prominent edge directions which may generate unstable codes in uniform areas [75]. Hence, LDiP did not perform superior to LBP in our experiments. A similar observation was also reported in [75].

3. Male faces have been classified with higher rates for all the features. The performance gain achieved using the proposed method is largely due to improvement in male classification rate. From the table, it is clear that the proposed method achieves around 3% gain over the LBP in classifying male images. In contrast, such performance gain was only 1% for female faces. Thus, the results indicate that the female faces are more difficult to classify. A similar observation has also been reported in [28, 43, 44].

Fig. 4.8 ROC curves for color FERET
The performance of different features has also been evaluated using ROC curves as shown in Fig. 4.8. The vertical axis shows the true positive rate, which should be higher for better performance. At the same time, the false positive rate shown on the horizontal axis should be as low as possible. Hence, the closer the ROC curve is to the top left corner, the better is the performance. Further, a standard performance measure of ROC, area under the curve (AUC) is also reported in brackets with every feature type in the figure. As noted earlier, the ROC plots also clearly depict that the performance of MBP, LGBPHS, and the proposed method are comparable to each other and superior to rest of the methods. However, the advantage of the proposed method is lower feature dimension as compared to LGBPHS and MBP. This aspect is discussed in more detail in Section 4.9.

4.6 Results on LFW database

As mentioned earlier, the major challenges in facial gender recognition are low image resolution, occlusions, and variations in illumination, pose, and facial expressions. The color FERET dataset was designed under a controlled environment where above mentioned challenges were largely absent. To investigate the robustness of the proposed method, this work employed two challenging datasets containing real-world facial images. This section describes the results on LFW dataset while the next section presents the results on the more recently built Adience dataset.

The LFW dataset consists of more than 13,000 color photographs of faces from 5749 individuals. The images were collected from the web which includes aforementioned variations. The detection of facial texture is extremely challenging under such variations making it difficult to recognize gender. The dataset contains multiple images which are numbered sequentially for each the subject. In this work, the first image for each subject is used and their gender was labeled manually. It was difficult to determine gender from 13 images, hence total 5736 face images have been used in this work. It is important to note that prior work [44] used 7443 images in their work on gender recognition. The use of multiple images for some subjects raises a risk of learning identity-related information which may bias the gender recognition results.

The classification rates of various methods are given in Table 4.4. The MBP recorded the highest classification rate of 89.82%. However, the performance of the proposed method (89.70%) is very close to MBP and LGBPHS (89.40%) also. Moreover, the proposed method achieves the significant performance gain (5.74%) over LBP for LFW dataset in contrast to
2.23% gain achieved on color FERET dataset. Similarly, the performance gap between the proposed method and other Kirsch mask based directional patterns is much more prominent in this database. The same fact is also reflected in the ROC curves shown in Fig. 4.9. Understanding the challenging real-world conditions reflected in the LFW dataset, it is not surprising to observe the degradation in the performance as compared to color FERET dataset. However, the proposed method along with MBP and LGBPHS sustained those challenges much better as compared to other existing methods. This promising performance actually demonstrates the robustness of the proposed method.

Table 4.4 Classification rates of LFW database

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>88.61</td>
<td>70.55</td>
<td>83.96</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(2.29)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>ELBP</td>
<td>93.52</td>
<td>70.07</td>
<td>87.48</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(2.70)</td>
<td>(0.67)</td>
</tr>
<tr>
<td>TPLBP</td>
<td>88.66</td>
<td>61.34</td>
<td>81.62</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(2.35)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>LDiP</td>
<td>84.86</td>
<td>63.91</td>
<td>79.46</td>
</tr>
<tr>
<td></td>
<td>(0.85)</td>
<td>(1.62)</td>
<td>(0.68)</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>84.13</td>
<td>61.95</td>
<td>78.42</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(2.51)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>ELDiP</td>
<td>90.30</td>
<td>60.93</td>
<td>82.74</td>
</tr>
<tr>
<td></td>
<td>(1.06)</td>
<td>(1.29)</td>
<td>(0.85)</td>
</tr>
<tr>
<td>LDN</td>
<td>87.02</td>
<td>68.18</td>
<td>82.17</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(2.52)</td>
<td>(1.03)</td>
</tr>
<tr>
<td>POEM</td>
<td>94.25</td>
<td>68.31</td>
<td>87.57</td>
</tr>
<tr>
<td></td>
<td>(0.76)</td>
<td>(1.48)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>LGBPHS</td>
<td>97.07</td>
<td>67.30</td>
<td>89.40</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
<td>(2.92)</td>
<td>(0.75)</td>
</tr>
<tr>
<td>MBP</td>
<td>96.71</td>
<td>69.94</td>
<td>89.82</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(3.83)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>CoLBP</td>
<td>95.37</td>
<td>73.32</td>
<td>89.70</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(1.90)</td>
<td>(0.97)</td>
</tr>
</tbody>
</table>
Fig. 4.9 ROC curves for LFW dataset

4.7 Results on Adience database

The color FERET and LFW datasets were mainly designed to study face recognition. In contrast, the main purpose of the recently designed Adience dataset was to study age and gender recognition under challenging real-world conditions. Further, the dataset contains a wider range of challenging conditions as compared to LFW [172]. The dataset is partitioned into two categories: The frontal and the complete sets. In this work, the subset containing frontal images was used. Further, we discarded those images on which automatic face detection algorithm [26] failed. This resulted in a total 1757 images, including 840 males and 917 females. Hence, the dataset is the most balanced in terms of gender ratio. The classification performance of various descriptors on Adience dataset is reported in Table 4.5.

On comparing the aforementioned classification performance with that of LFW dataset, the reduction in classification rates of all the features is clearly evident. Hence, the results shown in Table 4.5 reflect the fact that the Adience dataset is more challenging than LFW.
dataset. Nevertheless, the proposed method clearly outperforms the other approaches by a significant margin. For example, the performance gain by the proposed method over the LBP, LGBPHS, and MBP is 7.80%, 3.07% and 4.26% respectively.

Table 4.5 Classification rates of Adience database

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>75.60</td>
<td>76.55</td>
<td>76.09</td>
</tr>
<tr>
<td></td>
<td>(3.74)</td>
<td>(2.51)</td>
<td>(2.15)</td>
</tr>
<tr>
<td>ELBP</td>
<td>76.61</td>
<td>78.63</td>
<td>77.69</td>
</tr>
<tr>
<td></td>
<td>(5.11)</td>
<td>(2.76)</td>
<td>(2.38)</td>
</tr>
<tr>
<td>TPLBP</td>
<td>71.07</td>
<td>77.10</td>
<td>74.22</td>
</tr>
<tr>
<td></td>
<td>(4.39)</td>
<td>(3.28)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>LDiP</td>
<td>69.40</td>
<td>72.08</td>
<td>70.80</td>
</tr>
<tr>
<td></td>
<td>(5.78)</td>
<td>(2.37)</td>
<td>(3.46)</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>69.17</td>
<td>72.63</td>
<td>70.97</td>
</tr>
<tr>
<td></td>
<td>(3.22)</td>
<td>(2.78)</td>
<td>(1.07)</td>
</tr>
<tr>
<td>ELDiP</td>
<td>68.57</td>
<td>75.57</td>
<td>72.23</td>
</tr>
<tr>
<td></td>
<td>(2.28)</td>
<td>(2.15)</td>
<td>(1.50)</td>
</tr>
<tr>
<td>LDN</td>
<td>73.57</td>
<td>73.06</td>
<td>73.31</td>
</tr>
<tr>
<td></td>
<td>(2.93)</td>
<td>(2.15)</td>
<td>(1.91)</td>
</tr>
<tr>
<td>POEM</td>
<td>76.31</td>
<td>79.17</td>
<td>77.80</td>
</tr>
<tr>
<td></td>
<td>(4.46)</td>
<td>(3.44)</td>
<td>(1.44)</td>
</tr>
<tr>
<td>LGBPHS</td>
<td>81.67</td>
<td>80.04</td>
<td>80.82</td>
</tr>
<tr>
<td></td>
<td>(3.40)</td>
<td>(2.06)</td>
<td>(2.14)</td>
</tr>
<tr>
<td>MBP</td>
<td>79.17</td>
<td>80.05</td>
<td>79.63</td>
</tr>
<tr>
<td></td>
<td>(4.23)</td>
<td>(3.11)</td>
<td>(1.94)</td>
</tr>
<tr>
<td>CoLBP</td>
<td>84.88</td>
<td>82.99</td>
<td>83.89</td>
</tr>
<tr>
<td></td>
<td>(3.17)</td>
<td>(2.44)</td>
<td>(1.92)</td>
</tr>
</tbody>
</table>

By observing the category wise classification rates, contrasting results can be seen for this dataset. While male faces have been consistently classified at a higher rate for previous two datasets, the opposite is true for most of the features for Adience dataset. Unlike, color FERET and LFW datasets, the difference between the performances of each category is
narrow for the majority of the features. This can be attributed to more balanced gender ratio present in this dataset.

The ROC curves shown in Fig. 4.10 also indicate superior performance of the proposed method which recorded the AUC of 0.9089. The second best AUC of 0.8892 was recorded by LGBPHS. Based on the results of three datasets, it can be concluded that the proposed method not only provides more discrimination capability but it is also robust against different real-world challenges. To graphically summarize the results of the three datasets, the performance gain by the CoLBP over other feature types is depicted in Fig. 4.11.

![ROC curves for Adience database](image)

Fig. 4.10 ROC curves for Adience database

To get insight about the prediction errors, misclassified images were visually analyzed for all the datasets. Representative examples of misclassifications are shown in Fig. 4.12. The LFW and Adience datasets contain images captured in an unconstrained environment. Hence, these images contain wide range of variations. During visual analysis it was found that the classifier failed very often for the images of children and senior subjects as shown in the top panel of the figure. A similar observation has also been reported in [184]. Misclassifications due to variations in ethnicity are shown in the second row of the figure. As can be seen,
classification of black females was found to be very challenging. Other major causes of misclassifications (shown from the third row of the figure) are large variations in pose, image quality, eye openness, occlusions and facial expressions.

Fig. 4.11 Performance gain by CoLBP over other feature types

Fig. 4.12 Examples of misclassifications from all the datasets
4.8 Comparison with other state-of-the-art results

Comparison of published results is a difficult task since standard evaluation protocols are not available in gender classification literature. Hence, there exists diversity in the choice of dataset, image size, pre-processing method, the subsets used for evaluation, the type of classifier and its tuning parameters. However, for completeness and to get an overall idea, the proposed method is compared with recently published approaches in Table 4.6.

Table 4.6 Performance comparison of the proposed method with other representative approaches

<table>
<thead>
<tr>
<th>Reference</th>
<th>Approach</th>
<th>Dataset</th>
<th>Number of images</th>
<th>CR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Makinen and Raisamo [2]</td>
<td>LBP</td>
<td>FERET</td>
<td>900</td>
<td>86.28</td>
</tr>
<tr>
<td>(2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jabid et al. [29]</td>
<td>LDiP</td>
<td>FERET</td>
<td>2000</td>
<td>95.05</td>
</tr>
<tr>
<td>(2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logoglu et al. [185]</td>
<td>Gradientfaces</td>
<td>FERET</td>
<td>2468</td>
<td>92.40</td>
</tr>
<tr>
<td>(2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ahmed and Kabir [75]</td>
<td>Directional ternary pattern (DTP)</td>
<td>FERET</td>
<td>1800</td>
<td>93.11</td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li et al. [4]</td>
<td>LBP</td>
<td>FERET</td>
<td>782</td>
<td>95.80</td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chu et al. [42]</td>
<td>Subspace SVM</td>
<td>FERET</td>
<td>800</td>
<td>91.13</td>
</tr>
<tr>
<td>(2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berber [186]</td>
<td>Gray-level co-occurrence matrix (GLCM)</td>
<td>FERET</td>
<td>644</td>
<td>93.11</td>
</tr>
<tr>
<td>(2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rai and Khanna [102]</td>
<td>Gabor + (2D)^2PCA</td>
<td>FERET</td>
<td>700</td>
<td>98.40</td>
</tr>
<tr>
<td>(2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shan [44]</td>
<td>Boosted LBP</td>
<td>LFW</td>
<td>13010</td>
<td>89.10</td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arigbabu [187]</td>
<td>Pyramid histogram of gradients (PHOG)</td>
<td>LFW</td>
<td>2759</td>
<td>89.30</td>
</tr>
<tr>
<td>(2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td>CoLBP</td>
<td>LFW</td>
<td>5736</td>
<td>89.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FERET</td>
<td>987</td>
<td>93.92</td>
</tr>
</tbody>
</table>
It is seen that the work in [102] has reported the highest classification rate on the FERET dataset. However, they used only 700 images from the FERET dataset in contrast to 987 images used in this work. Further, they have used two-fold cross validation scheme whereas this work employed five-fold cross validation which is more common in the literature. On the LFW dataset, the boosted LBP approach [44] obtained the best results. However, they have used multiple images per subject in their evaluation. This may bias the results since the classifier may learn identity-related information [27]. Based on preceding discussion and keeping the classification rates of the proposed method in mind, it is clear that the proposed method provides a competitive performance in comparison with state-of-the-art approaches.

4.9 Computational complexity

This section compares the computational complexity of various methods using four measures: Feature dimension, feature extraction time, classifier training time and image classification time. These performance measures were obtained using the color FERET dataset with the parameter settings as mentioned earlier in Section 4.4.1. The experiments were carried out using MATLAB 2013a on Intel quad-core CPU @ 3.40 GHz with 8 GB RAM. To estimate the feature extraction time, 300 images were arbitrarily selected from the color FERET dataset. For every image, the time needed to extract the specified feature vector was recorded and their mean value is reported. In contrast, to estimate the classifier training time and image classification time, five-fold cross validation experiments were repeated 20 times and mean value of 100 (20 trials × 5 folds ) folds are reported. The color FERET dataset contains 987 images in total. Five-fold partition of these images for cross validation provided around 790 images in the training set and around 197 images in the test set for each fold. The classifier training time represents the average time required to train a linear SVM classifier on a training set of 790 feature vectors. The image classification time denotes the average time required to classify a single test image represented by a feature vector.

Table 4.7 provides an overall comparison of the above measures for various types of features. From the table, it is clear that the LBP not only requires less feature dimension, but it is also computationally the most efficient method among others. The methods such as EnLDiP and LDN, even though require slightly less feature dimension, demand more computation time due to eight convolutions required in obtaining edge responses. The feature size of the proposed method is 8 times more than LBP, but it is much smaller than the other two competitive state-of-the-art methods namely MBP and LGBPHS. The LGBPHS requires
40 convolutions in contrast to just 8 of the proposed method which significantly improves the computational speed as seen from the table.

Table 4.7 Comparison of computational complexity

<table>
<thead>
<tr>
<th>Feature</th>
<th>Dimension (block)</th>
<th>Dimension (Image)</th>
<th>Feature extraction time (ms)</th>
<th>Classifier training time (ms)</th>
<th>Classification time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>59</td>
<td>2124</td>
<td>3.0</td>
<td>219.4</td>
<td>0.9</td>
</tr>
<tr>
<td>ELBP</td>
<td>59</td>
<td>8496</td>
<td>10.3</td>
<td>375.5</td>
<td>5.0</td>
</tr>
<tr>
<td>TPLBP</td>
<td>256</td>
<td>6400</td>
<td>7.0</td>
<td>334.0</td>
<td>3.9</td>
</tr>
<tr>
<td>LDiP</td>
<td>256</td>
<td>9216</td>
<td>7.6</td>
<td>441.1</td>
<td>4.5</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>56</td>
<td>2016</td>
<td>5.1</td>
<td>240.0</td>
<td>0.9</td>
</tr>
<tr>
<td>ELDiP</td>
<td>256</td>
<td>9216</td>
<td>7.2</td>
<td>407.2</td>
<td>5.3</td>
</tr>
<tr>
<td>LDN</td>
<td>56</td>
<td>2016</td>
<td>5.1</td>
<td>229.3</td>
<td>0.9</td>
</tr>
<tr>
<td>POEM</td>
<td>59</td>
<td>16992</td>
<td>30.0</td>
<td>622.7</td>
<td>9.0</td>
</tr>
<tr>
<td>LGBPPhS</td>
<td>59</td>
<td>84960</td>
<td>226.2</td>
<td>2743.2</td>
<td>50.4</td>
</tr>
<tr>
<td>MBP</td>
<td>256</td>
<td>27648</td>
<td>116.8</td>
<td>952.9</td>
<td>15.9</td>
</tr>
<tr>
<td>CoLBP</td>
<td>59</td>
<td>16992</td>
<td>28.2</td>
<td>610.0</td>
<td>8.7</td>
</tr>
</tbody>
</table>

It is also important to note that, when comparing the computational complexity of various methods, their classification performance should also be taken into account. It is desirable to have a balance of classification performance and computational complexity. Thus, in comparison with LBP, even though the proposed method requires 8 times more feature dimension, it also provided remarkable gain in classification rates (2.23% on color FERET, 5.74% on LFW and 7.80% on Adience dataset). Moreover, although MBP and LGBPPhS also provided competitive classification performance, the proposed method has a clear advantage over them in terms of lower computational complexity. Further, high-performance computing systems are easily available nowadays. Hence, the moderate computational complexity of the proposed method can be usually under control. Based on the foregoing discussion, it can be concluded that the proposed method provides a good balance between a classification performance and computational complexity. Hence, it is a good candidate for high-performance facial gender classification.
4.10 Comparison among components of CoLBP

The proposed method combines information from different compass LBP images obtained using eight Kirsch compass masks. It is interesting to evaluate the gender classification performance of each of the eight components separately. The comparison of classification rates of different components is presented in Table 4.8. The numbers in the brackets represent the standard deviation of classification rates of five folds. As expected, the performance of individual components is inferior to their combination. Every individual component also performed lower than LBP on all the three datasets. Different edge responses obtained from Kirsch compass masks capture edge information from only one direction which could not provide sufficient discriminatory information. However, the proposed descriptor combines information from different components to achieve robustness and higher discrimination power.

<table>
<thead>
<tr>
<th>Component</th>
<th>Color FERET</th>
<th>LFW</th>
<th>Adience</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87.74 (1.11)</td>
<td>77.98 (1.29)</td>
<td>72.40 (3.19)</td>
</tr>
<tr>
<td>2</td>
<td>88.85 (3.16)</td>
<td>81.14 (1.37)</td>
<td>75.75 (1.47)</td>
</tr>
<tr>
<td>3</td>
<td>87.94 (2.00)</td>
<td>79.83 (0.91)</td>
<td>73.59 (2.14)</td>
</tr>
<tr>
<td>4</td>
<td>88.25 (2.07)</td>
<td>81.42 (1.07)</td>
<td>75.58 (1.35)</td>
</tr>
<tr>
<td>5</td>
<td>87.74 (2.02)</td>
<td>78.87 (0.70)</td>
<td>73.36 (1.58)</td>
</tr>
<tr>
<td>6</td>
<td>88.35 (1.68)</td>
<td>80.86 (1.12)</td>
<td>74.79 (3.17)</td>
</tr>
<tr>
<td>7</td>
<td>87.33 (1.42)</td>
<td>81.61 (0.59)</td>
<td>74.62 (1.27)</td>
</tr>
<tr>
<td>8</td>
<td>90.78 (0.98)</td>
<td>81.59 (0.95)</td>
<td>72.79 (1.55)</td>
</tr>
<tr>
<td>All</td>
<td>93.92 (2.04)</td>
<td>89.70 (0.97)</td>
<td>83.89 (1.92)</td>
</tr>
</tbody>
</table>
The performance gain by CoLBP over different components is plotted in Fig. 4.13. It is clear from the figure that the performance gain by CoLBP is much larger on the LFW and Adience datasets in comparison with the color FERET dataset. The average performance gain over different components by the CoLBP is 5.55% on color FERET, 9.29% on LFW and 9.78% on Adience dataset. This reflects the fact that the performance of individual components degraded severely on the LFW and Adience datasets due to the challenging real-world scenario. However, collectively they have demonstrated a promising performance in spite of challenging real-world conditions. This performance gain of CoLBP can be attributed to complementary information captured by different components.

### 4.11 Comparison of CoLBP and MQLBP

Finally, it is interesting to compare the performance of CoLBP with the MQLBP feature proposed in Chapter 3. Classification performance of both the methods is shown in Fig. 4.14. The differences in the performance of both the methods are marginal for color FERET and LFW datasets, whereas there is significant performance gap on the Adience dataset. The CoLBP performed superior to MQLBP on the Adience dataset, which is the most challenging dataset among the three.

![Chart showing performance gain of CoLBP over individual components](image)

Fig. 4.13 Performance gain of CoLBP over individual components
Local binary pattern (LBP) operates on raw pixel intensity which limits its discrimination capability and robustness. To enhance them, the properties of Kirsch masks and LBP have been utilized effectively in this chapter. The proposed compass LBP (CoLBP) method has been evaluated to classify facial gender on three publicly available datasets including color FERET, LFW, and Adience. The last two datasets are really challenging as they contain poor resolution images and variations in pose, expression, and illumination. The proposed method provided promising classification rates as compared to several state-of-the-art methods. Further, as compared to LGBPHS and MBP methods, the proposed method demands lower computational complexity.
CHAPTER 5: GENDER CLASSIFICATION OF FACIAL SKETCHES

This chapter describes the problem of gender classification of facial sketches which has remained unaddressed in the literature. This work evaluates the performance of various LBP variants on gender classification of facial sketches. Experiments were carried out on two publicly available sketch datasets. The experimental results clearly revealed the applicability of various LBP variants which have shown promising classification performance on the two datasets.

5.1 Introduction

Prior chapters described novel feature extraction algorithms for gender classification of facial photographs. However, apart from visible light sources, facial images are also captured from various other sources such as infrared devices, thermal imaging devices, and sketch artists. It is an interesting idea to evaluate the performance of various LBP algorithms on different modalities. To this end, this work focuses on gender classification of facial sketches. The choice of sketch modality is made since sketch images have been used for image retrieval and object detection tasks [188, 189] but it has not received attention in gender classification literature. In contrast, gender classification of thermal and near-infrared (NIR) images has been conducted in [190, 191]. Further, gender classification of facial sketches has useful application in forensic analysis. The use of sketches in criminal investigations is well known. For instance, when the photo of a suspect is not available, a forensic sketch of the suspect is prepared by an artist based on the description provided by the eye witness. Under this scenario, the sketch image can then be used to determine the identity of a suspect by matching it with the mug shot images available in the dataset. Such type of system which matches facial images of two different modalities (sketch versus photo) is known as heterogeneous face recognition (HFR) system. Gender recognition of facial sketches can be used for indexing purpose to reduce the search space of a large-scale HFR system. Klare et al. [192] demonstrated the importance of gender and other demographic information for the heterogeneous face recognition task. However, they manually labeled the sketches with gender information. Thus, it would be of great value to design an algorithm which can automatically identify the gender of sketch images. With these motivations, the present work evaluates the performance of various LBP variants on gender classification of facial sketches.

5.2 Datasets

The present work employed two sketch image datasets namely CUHK face sketch FERET (CUFSF) and CUHK face sketch (CUFS) database. Both the databases were released
by Chinese university of Hong Kong to study face photo-sketch synthesis and recognition. The CUFS dataset contains 606 facial sketches of subjects from three different datasets: CUHK student database, AR database [115] and XM2VTS database [130]. The CUFSF was released later, which contains 1195 sketch images of subjects from FERET dataset. However, some subjects were duplicated with multiple identities. After fixing this error, 987 sketches of distinct subjects were used in this work. Some example images from both the datasets are shown in Fig. 5.1. The summary of the number of images used in each dataset is presented in Table 5.1.

![Fig. 5.1 Example images from sketch datasets: (a) CUFSF, (b) CUFS. For both the datasets, the top two rows contain sketch images of male subjects and the bottom two rows contain images from female subjects.](image-url)
Table 5.1 Summary of images used in experiments

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CUFSF</th>
<th>CUFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male images</td>
<td>587 (59.47%)</td>
<td>369 (60.89%)</td>
</tr>
<tr>
<td>Female images</td>
<td>400 (40.53%)</td>
<td>237 (39.11%)</td>
</tr>
<tr>
<td>Total images</td>
<td>987</td>
<td>606</td>
</tr>
</tbody>
</table>

Table 5.2 Classification rates of CUFSF database

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>95.06</td>
<td>88.75</td>
<td>92.50</td>
</tr>
<tr>
<td></td>
<td>(1.86)</td>
<td>(4.42)</td>
<td>(1.53)</td>
</tr>
<tr>
<td>ELBP</td>
<td>95.06</td>
<td>87.50</td>
<td>92.00</td>
</tr>
<tr>
<td></td>
<td>(1.38)</td>
<td>(4.24)</td>
<td>(1.82)</td>
</tr>
<tr>
<td>TPLBP</td>
<td>92.16</td>
<td>84.25</td>
<td>88.96</td>
</tr>
<tr>
<td></td>
<td>(1.93)</td>
<td>(3.60)</td>
<td>(1.27)</td>
</tr>
<tr>
<td>LDiP</td>
<td>92.33</td>
<td>83.75</td>
<td>88.86</td>
</tr>
<tr>
<td></td>
<td>(2.91)</td>
<td>(3.85)</td>
<td>(1.65)</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>94.21</td>
<td>88.75</td>
<td>92.00</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(4.24)</td>
<td>(1.35)</td>
</tr>
<tr>
<td>ELDiP</td>
<td>96.42</td>
<td>81.25</td>
<td>90.27</td>
</tr>
<tr>
<td></td>
<td>(1.39)</td>
<td>(3.95)</td>
<td>(1.35)</td>
</tr>
<tr>
<td>LDN</td>
<td>93.02</td>
<td>88.50</td>
<td>91.19</td>
</tr>
<tr>
<td></td>
<td>(1.82)</td>
<td>(4.18)</td>
<td>(1.80)</td>
</tr>
<tr>
<td>POEM</td>
<td>95.57</td>
<td>89.00</td>
<td>92.91</td>
</tr>
<tr>
<td></td>
<td>(2.83)</td>
<td>(4.63)</td>
<td>(2.11)</td>
</tr>
<tr>
<td>LGBPHS</td>
<td>96.60</td>
<td>89.00</td>
<td>93.52</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td>(2.85)</td>
<td>(1.44)</td>
</tr>
<tr>
<td>MBP</td>
<td>96.94</td>
<td>90.00</td>
<td>94.12</td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td>(2.65)</td>
<td>(1.50)</td>
</tr>
<tr>
<td>MQLBP</td>
<td>96.93</td>
<td>89.50</td>
<td>93.92</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
<td>(1.90)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>CoLBp</td>
<td>96.25</td>
<td>92.00</td>
<td>94.53</td>
</tr>
<tr>
<td></td>
<td>(1.54)</td>
<td>(3.01)</td>
<td>(1.35)</td>
</tr>
</tbody>
</table>
5.3 Experimental results

All the experiments were conducted by following the evaluation protocol and parameter settings as described in Section 4.4.1.

5.3.1 Results on CUFSF dataset

The CUFSF dataset contains the same number of images as used in color FERET dataset described in Chapter 4. It contains 587 sketch images from male subjects and 400 images from female subjects. The classification rates of various features are presented in Table 5.2. Among various features, the lowest classification rate of 88.86% was recorded by local directional pattern (LDiP) feature. The compass LBP (CoLBP) feature recorded the highest classification rate of 94.53%. However, the performance of MQLBP, LGBPHS, and MBP was very close to that of CoLBP. The similar performance can also be seen in ROC plots of different features shown in Fig. 5.2.

![ROC curves for CUFSF dataset](image-url)

Fig. 5.2 ROC curves for CUFSF dataset
Table 5.3 Classification rates of CUFS database

<table>
<thead>
<tr>
<th>Feature</th>
<th>MCR</th>
<th>FCR</th>
<th>OCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>91.06</td>
<td>82.28</td>
<td>87.62</td>
</tr>
<tr>
<td></td>
<td>(4.99)</td>
<td>(4.54)</td>
<td>(2.88)</td>
</tr>
<tr>
<td>ELBP</td>
<td>92.67</td>
<td>84.80</td>
<td>89.60</td>
</tr>
<tr>
<td></td>
<td>(3.95)</td>
<td>(2.85)</td>
<td>(2.53)</td>
</tr>
<tr>
<td>TPLBP</td>
<td>91.60</td>
<td>81.82</td>
<td>87.79</td>
</tr>
<tr>
<td></td>
<td>(4.00)</td>
<td>(6.19)</td>
<td>(3.76)</td>
</tr>
<tr>
<td>LDiP</td>
<td>91.60</td>
<td>83.97</td>
<td>88.61</td>
</tr>
<tr>
<td></td>
<td>(4.27)</td>
<td>(5.69)</td>
<td>(3.99)</td>
</tr>
<tr>
<td>EnLDiP</td>
<td>90.24</td>
<td>82.70</td>
<td>87.29</td>
</tr>
<tr>
<td></td>
<td>(3.82)</td>
<td>(2.92)</td>
<td>(2.48)</td>
</tr>
<tr>
<td>ELDiP</td>
<td>91.06</td>
<td>76.79</td>
<td>85.47</td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(2.77)</td>
<td>(3.46)</td>
</tr>
<tr>
<td>LDN</td>
<td>90.79</td>
<td>80.59</td>
<td>86.80</td>
</tr>
<tr>
<td></td>
<td>(4.61)</td>
<td>(3.51)</td>
<td>(2.69)</td>
</tr>
<tr>
<td>POEM</td>
<td>92.95</td>
<td>85.65</td>
<td>90.10</td>
</tr>
<tr>
<td></td>
<td>(2.80)</td>
<td>(3.87)</td>
<td>(2.12)</td>
</tr>
<tr>
<td>LGBPHS</td>
<td>94.03</td>
<td>83.08</td>
<td>89.76</td>
</tr>
<tr>
<td></td>
<td>(3.29)</td>
<td>(5.12)</td>
<td>(2.40)</td>
</tr>
<tr>
<td>MBP</td>
<td>94.84</td>
<td>85.20</td>
<td>91.09</td>
</tr>
<tr>
<td></td>
<td>(2.98)</td>
<td>(4.35)</td>
<td>(1.61)</td>
</tr>
<tr>
<td>MQLBP</td>
<td>94.29</td>
<td>85.63</td>
<td>90.92</td>
</tr>
<tr>
<td></td>
<td>(4.15)</td>
<td>(4.15)</td>
<td>(2.62)</td>
</tr>
<tr>
<td>CoLBP</td>
<td>95.39</td>
<td>84.81</td>
<td>91.25</td>
</tr>
<tr>
<td></td>
<td>(2.27)</td>
<td>(2.42)</td>
<td>(1.26)</td>
</tr>
</tbody>
</table>

5.3.2 Results on CUFS dataset

The CUFS dataset contains 369 images from male subjects and 237 images from female subjects, making a total of 606 images. The classification performance of different features is shown in Table 5.3. The lowest classification rate of 85.47% was recorded by the enhanced LDiP (ELDiP) feature. Further, the CoLBP recorded the highest classification rate of 91.25% on this dataset also. The ROC plots of different features are depicted in Fig. 5.3. It is clear
from the figure that the performance of MQLBP, LGBPHS, MBP and CoLBP is very close to each other.

![ROC curves for CUFS dataset](image)

**Fig. 5.3** ROC curves for CUFS dataset

### 5.4 Discussion

The experimental results presented in this work demonstrate the viability of performing gender classification of facial sketches. From the results on both the datasets, it is clear that the LBP and its variants provide promising gender classification performance on sketch modality also. While preparing sketch images, the artists use various pencil strokes and shades to highlight the facial texture information. The success of different LBP variants can be attributed to their capability in efficiently encoding the texture information present in facial sketches.

Among various types of descriptors, the CoLBP has performed the best on both the datasets. Fig. 5.4 summarizes the performance gain by CoLBP over other features on both the
datasets. Considering LBP as the benchmark method, the classification rates of CoLBP are 2.03% and 3.63% higher than LBP on CUFSF and CUFS datasets respectively. Examples of misclassified samples by CoLBP method are shown in Fig. 5.5. Top two panels show misclassification from CUFSF dataset and the bottom two panels show misclassifications from CUFS dataset. Different factors such as variations in age, hair style and facial expressions and accessories (e.g. eyeglasses, earrings, cap) might have confused the classifier causing such misclassification.

![Performance gain by CoLBP over other feature types](image)

**Fig. 5.4** Performance gain by CoLBP over other feature types

![Illustration of misclassified samples](image)

**Fig. 5.5** Illustration of misclassified samples
Further, it is clear from the results in Table 5.2 and Table 5.3, that the sketch images of male subjects have been consistently classified at a higher rate than female subjects for all the feature types. Similar observation was also reported earlier in Section 4.5. Moreover, the difference between the performance of male and female subjects is larger on CUFS dataset as compared to CUFSF dataset. This point is more clearly depicted using the bar chart shown in Fig. 5.6. The vertical axis shows the difference between the classification rates of male and female subjects, i.e. MCR – FCR.

![Chart showing performance difference between male and female subjects for various LBP features on CUFS and CUFSF datasets.](chart.png)

**Fig. 5.6 Difference in classification rates of male and female subjects**

5.5 Summary

This chapter addressed an interesting problem of the gender classification of facial sketches. The chapter clearly described the useful forensic application of the targeted problem. This work evaluated the applicability of various LBP features for the gender classification of facial sketches. The experiments were carried out on two publically available facial sketch datasets: CUFSF and CUFS. The experimental results clearly demonstrated the efficiency of the LBP and its variants in the classifying gender of facial sketches. Among various features, the CoLBP feature provided the highest classification rates of 94.53% and 91.25% on CUFSF and CUFS dataset respectively. However, the other features such as LGBPHS, MQLBp, POEM and MBP also performed competitive classification performance. These features also recorded classification rates higher than 90.00% on both the datasets.
The LBP and its variants have already shown their effectiveness in classifying gender of facial photographs. This work goes one step further and demonstrates their potential on sketch modality also. Based on the experimental results, it can be concluded that the LBP and variants can also be used for classifying gender of facial sketches.
This chapter describes a kinship verification framework using a block-based neighborhood repulsed metric learning (BNRML) method. The proposed method learns multiple local distance metrics from different blocks of the images represented by local patterns. Moreover, to contemplate diversity in discrimination power of different blocks, weighted score-level fusion scheme is used to obtain a similarity score of an image pair. The proposed framework is validated using full face images from two challenging datasets. Moreover, this chapter explores the effectiveness of periocular region in verifying kinship. Experimental results clearly demonstrated the potential of periocular features for kinship verification. Finally, the fusion of periocular and face traits was also performed, which provided highly competitive results as compared to state-of-the-art methods.

6.1 Introduction

Previous chapters addressed gender classification of human faces. Apart from gender, human face also provides cues for extracting various other attributes such as identity, age, and emotion. The high information contents encoded by human face have made it a popular object of study in computer vision. It has fostered various specialized research areas such as face recognition, age estimation, gender classification and facial expression analysis. Further, several other areas are also emerging such as affective computing and prediction of facial aesthetics. Kinship verification is one such emerging area which is receiving growing interest among researchers in recent years.

Humans are social animals. Their social structure is centered on the relationships between different groups of people. Kinship is the most basic of all human relationships. It is formally defined as a relationship between two family members who are biologically related with genetic overlaps [151]. Due to genetic overlapping, family members often share a facial resemblance. Hence, it is natural for humans to infer kinship cues from facial images. Further, several psychological studies [144, 193-195] also validated the importance of facial appearance for reliably verifying kin relations among humans. This finding has attracted the attention of computer vision researchers to design algorithms for extracting kinship information from facial images. Further motivation for studying the kinship verification comes from the possible potential applications in the following areas:

- Social media analysis
- Family album management
- Forensic science
- Historical and genealogical research
- Searching missing family members
The aim of automatic kinship verification is to determine whether the individuals present in a pair of face images are related or not. It is important to realize the challenges posed by automatic kinship verification since it requires matching features of two different persons. Further, the degree of kin relation (e.g. parent-child, siblings, and grandparent-child) is also unknown. Hence, it has to simultaneously deal with variations in age, gender, and ethnicity. The larger appearance gap between the two facial images makes it more challenging than a typical face verification task. Moreover, the use of unconstrained facial images raises additional difficulties such as low resolution, occlusions, and variations in expression, pose, and intensity.

Based on the aforementioned motivations, this work addresses kinship verification of facial images. Among various degrees of kinship, parent-child relation has gained more attention in the literature and the datasets of parent-child image pairs are also publicly available. Hence, similar to the other studies [10, 160, 162, 164, 196], this work also considers kinship verification of parent-child image pairs which can be categorized into four types: father-son (F-S), father-daughter (F-D), mother-son (M-S) and mother-daughter (M-D).

For representing facial appearance, local binary pattern (LBP) has been a popular choice in automatic kinship verification studies [159, 164, 196]. However, other variants of LBP have not been evaluated on kinship verification task. Hence, this work evaluates the performance of several LBP variants on two challenging datasets: KinFaceW-I and KinFaceW-II [10].

Metric learning methods have gained substantial attention in recent times for kin similarity measurement. One of the promising methods is a neighborhood repulsed metric learning (NRML) method which was also used as a baseline method in kinship verification in the wild evaluation [196]. Hence, this work employs NRML method to evaluate kinship verification. Further, to extract more discriminative information, a block-based NRML (BNRML) method is proposed in this work. The face images are first divided into multiple blocks and local features are extracted from each block. Then NRML is applied separately to respective blocks of a pair of images to learn multiple distance metrics. Hence, unlike original NRML method which learns a single global distance metric, BNRML method learns multiple local distance metrics at different spatial locations corresponding to different blocks of the image. The idea of BNRML is inspired from block-based Fisher linear discriminant (BFLD) [176] which showed promising performance on face recognition task. Further, the idea of multiple metric learning has also been successfully applied in person re-identification task [197, 198].
Systematic literature review on existing kinship verification methods was carried out in Section 2.2. Based on the review, it was observed that the existing methods have mainly focused on full face images to extract kin information. However, the human face is composed of different parts such as eyes, nose, mouth and jaw. It is commonly observed that children inherit certain facial part more prominently than the others from their parents. Hence, it is required to investigate the role of different face parts for automatic kinship verification. To this end, the present work focuses on the periocular region for verifying kin relations.

6.2 Periocular biometrics

The periocular region refers to the area of the face directly surrounding the eye. It has recently emerged as a novel biometric modality in computer vision literature. The human face consists of various parts such as forehead, eyes, nose, mouth, chin, and cheeks. In comparison with other facial parts, the periocular region carries more geometric information and textural details [199]. Further, several studies in psychology literature investigated the role of various facial parts for person identification. Such studies provided ample evidence that the upper part of the face and eyes, in particular, provide more useful information for identity recognition [200-204]. The lead opened by these findings attracted the attention of several computer vision researchers to focus on the periocular region for automatic person identification.

The work on periocular biometrics was pioneered by Park et al. [205]. They studied the feasibility of periocular region for automatic person identification and reported the encouraging performance of periocular region. Inspired by this work several researchers proposed different approaches to demonstrate the success of periocular biometrics [206-208].

The growing interest among researchers can be attributed to the several advantages offered by the periocular region. Periocular region can be useful in challenging scenarios where the whole face is not visible due to partial occlusion from masks or scarves. Further, the periocular region is considered to provide more invariance to variations in pose, facial expressions, and age. In comparison with the iris, one of the most popular biometric traits, periocular images can be captured at a long distance and without user cooperation. Apart from aforementioned advantages, other motivating factors are:

- Easy availability of periocular images from existing face and ocular datasets.
- No need of additional capturing device to acquire periocular images.

Based on the preceding motivations, researchers have used periocular images under following different scenarios:
• Primary biometric trait: As mentioned earlier, several works [205, 206, 209] focused on using periocular region independently for person identification task.

• Multimodal fusion: To enhance the performance of person identification, some group of researchers attempted the fusion of periocular information with other biometric modalities such as face [210, 211], iris [212, 213] and sclera [214].

• Soft biometrics: Inspired from the success of periocular region for person identification task, several researchers investigated the capability of periocular features for soft biometric classification. They have demonstrated the success of periocular features in gender classification [173, 215, 216] and ethnicity estimation [215] tasks.

The preceding discussion clearly suggests that the periocular region contains useful information for person identification, gender classification, and ethnicity estimation. However, the study on experimental evaluation of periocular features for kinship verification is lacking in the literature. With this motivation, this work investigates the effectiveness of periocular region for automatic kinship verification. Further, this idea is also supported by a psychological study [144] which found that the ocular region carries more information regarding kinship than the other face parts.

### 6.3 Background of NRML

For successful kinship verification, it is required to have a larger intra-class similarity (similarity between a true parent-child pair) than inter-class similarity (similarity between a non-kin pair). However as mentioned in Section 6.1, this task is very challenging due to variations in age, gender, and ethnicity between the image pairs. This often causes large appearance gap between intra-class samples such that the similarity between them gets smaller than inter-class similarity leading to eventual misclassification. The NRML [10] method aims to reduce such misclassifications. It learns a robust distance metric such that the non-kin samples, lying in the neighborhood of the samples with kin relation, are repulsed as far as possible. At the same time, the samples with kin relation are projected as close to each other as possible. More specifically, it computes following three types of distances for each positive parent-child image pair in the training dataset:

• Type-1: Distance between a parent image and \(k\) images of other children which are nearest to it. (\(k\) non-kin pairs)
- Type-2: Distance between a child image and $k$ images of other parents which are nearest to it. ($k$ non-kin pairs)

- Type-3: Distance between true parent-child images (1 kin pair)

The NRML then finds a projection matrix by solving an optimization problem to simultaneously maximize Type-1 and Type-2 distances and minimize Type-3 distances. In other words, the distances between non-kin samples are maximized whereas the distances between kin samples are minimized. Mathematical details on the formulation of NRML are given in Appendix A. More technical details can be found in [10].

![Processing flow of the proposed method during training stage](image-url)

Fig. 6.1 Processing flow of the proposed method during training stage
6.4 Proposed method

The aim of automatic kinship verification is to determine whether the persons represented by a pair of face images are related or not. Hence, it is a binary classification problem. This section describes the proposed kinship verification method in detail. For clarity, the steps required in proposed method are explained separately for training stage and testing stage.

6.4.1 Training stage

The framework of proposed method during training stage is depicted in Fig. 6.1. It involves three major steps: Feature extraction, metric learning, and block weight learning. The technical details of each step are described in the following subsections.

6.4.1.1 Feature extraction

Let \( I = \left\{ \left( I_p^i, I_c^i, l_i \right) \right\}_{i=1}^N \) be the training set of \( N \) labeled samples. Each training sample consists of a pair of parent-child images denoted by \( I_p^i \) and \( I_c^i \) respectively. The binary kinship information is denoted by class labels \( l_i \in \{1,0\} \), where \( l_i = 1 \) indicates positive kinship (persons in images are related to each other) and \( l_i = 0 \) indicates negative kinship (persons in images are not related to each other). Further, two sets \( T = \{i|l_i = 1\} \) and \( F = \{i|l_i = 0\} \) are defined to indicate indices of training samples with positive and negative kinship respectively.

The processing starts with the computation of local features (e.g. LBP) for each image \( I_i \) (parent or child). The number of output patterns varies depending on the choice of feature type. For example, the LBP returns only one pattern whereas the LTP return two patterns. These patterns are considered as separate channels as suggested in [74]. Let the number of patterns be represented by the variable \( c \). Then as shown in Fig. 6.1, each type of local pattern is divided into \( b \) blocks and histograms are computed for each block resulting in total \( B = c \times b \) histograms per image. Let \( D = \left\{ \left( h_p^{i,j}, h_c^{i,j} \right) | l_i = 1, 2,..., N, j = 1, 2,..., B \right\} \) be the set of histogram features extracted from each of the \( N \) image pairs in the training set \( I \). The feature vectors \( h_p^{i,j} \in \mathbb{R}^m \) and \( h_c^{i,j} \in \mathbb{R}^m \) denote the histogram of the \( j^{th} \) block from the \( i^{th} \) parent and child images respectively. This set of training feature vectors is then used for metric learning which is explained in the next section.
6.4.1.2 BNRML metric learning

The objective of metric learning is to learn a transformation matrix using a training set so as to improve kinship verification performance in the transformed subspace. For this task, the NRML method concatenates the histograms of all the blocks for building a single face descriptor per image. These descriptors are then used to learn a single global distance metric. With the aim of improving discrimination power, we extend the idea and use the histograms from each respective block of image pairs to independently learn multiple distance metrics. The resulting method, known as block-based NRML (BNRML) is described as follows:

From the training feature set $D$, obtain multiple subsets corresponding to each block of positive image pairs. Let the subset $D_j = \{(h^p_{i,j}, h^c_{i,j})|i \in T\} \in D$ represents the pairs of feature vectors for every $j^{th}$ block of image pairs with positive kinship. Each subset $D_j$ is used independently to learn NRML transformation matrix $W_j \in R^{m \times d}$, where $l \leq m$. The original feature vectors $h^p_{i,j}$ and $h^c_{i,j}$ are then transformed to $u^p_{i,j} \in R^d$ and $u^c_{i,j} \in R^l$ as follows:

$$u^p_{i,j} = (W_j)^T h^p_{i,j}, \quad u^c_{i,j} = (W_j)^T h^c_{i,j}, \quad i = 1, 2, ..., N, \ j = 1, 2, ..., B \quad (6.1)$$

6.4.1.3 Block weight learning

In order to assign a weight to every block, cosine similarity between each pair of projected feature vectors of the respective block is calculated as follows:

$$s_{i,j} = \text{sim}(u^p_{i,j}, u^c_{i,j}) = \frac{u^p_{i,j} \cdot u^c_{i,j}}{|u^p_{i,j}| \cdot |u^c_{i,j}|}, \quad i = 1, 2, ..., N, \ j = 1, 2, ..., B \quad (6.2)$$

where $s_{i,j}$ denotes the similarity score of the $j^{th}$ block of the $i^{th}$ image pair.

Let $S^\text{Pos}_j = \{s_{i,j} | i \in T\}$ denotes the set of similarity scores of positive image pairs for $j^{th}$ block. Similarly, $S^\text{Neg}_j = \{s_{i,j} | i \in F\}$ denotes the set of similarity scores of negative image pairs for $j^{th}$ block. The weight of each block $j$ is calculated using Fisher’s separation criterion [174] as follows:

$$\omega_j = \frac{(m^\text{Pos}_j - m^\text{Neg}_j)^2}{v^\text{Pos}_j + v^\text{Neg}_j}, \quad j = 1, 2, ..., B \quad (6.3)$$

where, $m^\text{Pos}_j$ and $v^\text{Pos}_j$ represent intra-class mean and variance of similarity scores within set $S^\text{Pos}_j$. Similarly, $m^\text{Neg}_j$ and $v^\text{Neg}_j$ denote inter-class mean and variance of similarity scores within
the set \( S_j^{\text{Neg}} \).

Since different histograms are extracted from different regions of an image, their discriminative capabilities are also different. The weight \( \omega_j \) can be considered as a measure of discrimination power of a block. Such block weighting scheme has been successfully applied to face recognition task [174] which motivated us to use it for the proposed kinship verification task.

### 6.4.2 Testing stage

Fig. 6.2 shows the processing steps required to perform during the testing phase. To verify the kin relation between a pair \((I^p_q, I^c_q)\) of parent and child images presented during the testing stage, spatial histograms are first computed from the blocks of local patterns. Then for every block, the extracted feature vectors are projected on NRML space using respective transformation matrices learned during the training stage. For each block \(j\), similarity scores \( s_{q,j} \) between a pair of local feature vectors in the learned NRML space are calculated using cosine distance as defined in Equation (6.3). Subsequently, the similarity scores of all the blocks are fused together using a weighted sum rule to obtain an overall similarity score \( S \) between a pair of images:

\[
S(I^p_q, I^c_q) = \sum_{j=1}^{B} \omega_j s_{q,j}
\]  

(6.4)

Finally, the overall similarity score between a pair of test images is compared with a predefined threshold to make a decision. If the similarity score is above the threshold the persons in an image pair are identified as kins otherwise, they are declared as non-kins.
6.5 Information fusion of left and right periocular regions

The proposed BNRML framework can be extended to fuse information from different regions. To demonstrate the idea of information fusion of different regions, this section describes the fusion of left and right periocular regions. This work employs such information fusion to improve the performance. The fusion scheme is depicted in Fig. 6.3. As shown in the figure, the feature extraction, and metric learning steps are performed independently for all the blocks of both the left and right periocular image pairs. Thus, assuming $B$ blocks for each of the left and right periocular regions, the total of $2B$ blocks are formed and their weights are learned using Equations (6.2) and (6.3) during the training stage. Similarly, during the testing stage, the overall similarity score is obtained by fusing the similarity scores from all the $2B$ blocks using the weighted sum rule given by Equation (6.4). Thus the proposed BRNML framework is generic in nature and can be easily applied to fuse information from different regions.

Fig. 6.3 Information fusion of left and right periocular regions using BNRML framework

6.6 Experimental setup

This section describes the experimental methodology used to evaluate the performance of the proposed method. Comprehensive experiments were carried out on two different datasets which are briefly described in Section 6.6.1. Evaluation protocol and parameter settings are presented in Section 6.6.2.
Fig. 6.4 Sample images from the two datasets: (a) KinFaceW-I, (b) KinFaceW-II. The two adjacent images in each row represent a kin relation. The four rows from top to bottom show images for father-son (F-S), father-daughter (F-D), mother-son (M-S) and mother-daughter (M-D) respectively.

6.6.1 Datasets

The proposed kinship verification framework is evaluated on two publicly available kinship datasets: KinFaceW-I and KinfaceW-II [10]. The datasets contain real-world images collected from the internet. Since the images were captured in the wild, the faces in the images show large variations in illumination, pose, and expressions. Further, there were no restrictions in terms of age, gender, and ethnicity. Both the datasets contain aligned and cropped face images of size $64 \times 64$ pixels. Further, both the datasets include four types of kin relations: father-son (F-S), father-daughter (F-D), mother-son (M-S) and mother-daughter (M-D). There are two major differences between both the datasets: (1) the image pairs present in KinFaceW-I were obtained from different photographs whereas most of the image pairs in KinFaceW-II were collected from the same photographs. (2) The KinFaceW-II dataset is relatively larger in size than KinFaceW-I as shown in Table 6.1 which summarizes the number of images available in each dataset. Some example images from both the datasets are displayed in Fig. 6.4.

6.6.2 Evaluation protocol and parameter settings

To evaluate the performance of the proposed framework, pre-specified training/testing splits [196] have been used for performing 5-fold cross validation. The image pairs with kin
relation were considered as positive samples and those without kinship were labeled as negative samples. Further, NRML method uses identity information of training samples to learn distance metric. Hence, image-unrestricted setting as suggested in [196] was employed as evaluation protocol.

Table 6.1 Summary of images used in the experiments

<table>
<thead>
<tr>
<th></th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>KinFaceW-I</td>
<td>156</td>
<td>134</td>
<td>116</td>
<td>127</td>
<td>533</td>
</tr>
<tr>
<td>KinFaceW-II</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>1000</td>
</tr>
</tbody>
</table>

The NRML method was implemented using the Matlab codes provided by the authors of [10]. The codes were used with the default settings which considered the neighborhood size $k = 5$ and NRML feature dimension was set to 30. Further, principal component analysis (PCA) was applied first to reduce feature dimension [196]. However, the major advantage of the proposed block-based framework is that it avoids the small sample size (SSS) problem since each block is represented by 59-dimensional histogram. Hence, feature dimension was not reduced for BNRML method. Nevertheless, for a fair comparison between both the methods, the feature vectors of all the blocks were projected onto PCA space of corresponding block before computing BNRML.

6.7 Results on full face images

To benchmark the performance of the proposed BNRML method against NRML, the experiments were first performed on full face images. Further, MQLBP features proposed in this thesis were used for experiments. The MQLBP features were extracted using the neighborhood size of $P = 8$ and radius $R = 1$. The MQLBP threshold was empirically set to $t = 6$ for all the experiments. Moreover, uniform patterns with maximum two bitwise transitions per code were utilized. To compute spatial histograms, each of the four MQLBP images was divided into non-overlapping blocks of size $8 \times 8$ [196]. Hence it will result in a total of 64 blocks per MQLBP image. Hence, each whole face image was represented by 256 ($64 \times 4$) 59-dimensional histogram features which were used independently to learn local distance metrics. Whereas, the original NRML method concatenates these histograms to build a feature vector and learns a single global distance metric.
The performance of the proposed BRNML method is compared with NRML in Table 6.2 and Table 6.3 for KinFaceW-I and KinFaceW-II dataset respectively. It is clearly evident that the proposed method consistently outperformed the NRML method for each type of kin relation on both the datasets. Since face images usually reside on non-linear manifold [217], learning a single global distance metric (as done in NMRL) may not be adequate. Hence, the proposed method operated at block-level to learn multiple distance metrics from various patches of image pairs. Further, the information gained from various blocks was fused together using score-level fusion. These factors collectively contributed to achieve superior performance by the proposed method. Similar results can also be seen in Fig. 6.5 and Fig. 6.6 which present ROC curves of both the methods for the two datasets.

Table 6.2 Comparison of mean verification accuracy (%) of NRML and BNRML on KinFaceW-I dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRML</td>
<td>76.62</td>
<td>69.46</td>
<td>70.69</td>
<td>70.93</td>
<td>71.93</td>
</tr>
<tr>
<td>BNRML</td>
<td>83.98</td>
<td>77.64</td>
<td>74.51</td>
<td>78.79</td>
<td>78.73</td>
</tr>
</tbody>
</table>

Table 6.3 Comparison of mean verification accuracy (%) of NRML and BNRML on KinFaceW-II dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRML</td>
<td>72.80</td>
<td>69.20</td>
<td>69.80</td>
<td>73.60</td>
<td>71.35</td>
</tr>
<tr>
<td>BNRML</td>
<td>83.40</td>
<td>78.80</td>
<td>78.00</td>
<td>79.60</td>
<td>79.95</td>
</tr>
</tbody>
</table>

The pictorial view of the performance gain by the proposed method over NRML is depicted in Fig. 6.7. It can be seen that the performance gain is higher on KinFaceW-II dataset for different types of kin relations except for mother-daughter (M-D) relation. The overall performance gain by the proposed method over NRML is 6.81% on KinFaceW-I and 8.60% on KinFaceW-II datasets.

104
Fig. 6.5 ROC curves for NRML and BNRML method on KinFaceW-I dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D
Some illustrative examples of misclassifications from KinFaceW-II dataset are shown in Fig. 6.8. The four rows from top to bottom show images for father-son (F-S), father-daughter (F-D), mother-son (M-S) and mother-daughter (M-D) respectively. Each row in the figure shows four parent-child pairs out of which first two pairs are false negatives and the other two pairs are false positives. As seen from the figure, false negatives occurred due variations in age, pose, illumination and image quality of the image pairs that eventually caused large intra-class distance. For example, the second image pair in the second row contains illumination variation due to shadow. On the other hand, the first pair in the third row exhibits variation in the pose for the images of a mother and a son. As shown in Fig. 6.8, the visual analysis of different false positives revealed that the image pairs in such cases often shared common facial expressions which might have confused the classifier.
The proposed scheme employed Fisher’s separation criterion to assign weights to each block. The assumption is that the discrimination capability of different facial blocks may not be uniformly distributed. To understand the impact of block weighting scheme, the experiments were performed without employing the block weighting scheme. Hence, to obtain an overall similarity between a test pair, the similarity scores of different blocks will be fused using simply a sum rule. Equivalently, the weights $\omega_j$, $j = 1, 2, \ldots, B$ in Equation (6.4) will be set to one. The verification rates with and without block weighting scheme are summarized in Table 6.4 and Table 6.5. It is clear from the tables that weighting the blocks improve the
performance. Nevertheless, the performance of BNRML method without block weighting is still superior to basic NRML method.

Table 6.4 Effect of block weighting on mean verification accuracy (%) for KinFaceW-I dataset

<table>
<thead>
<tr>
<th>Block Weighting</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>80.46</td>
<td>72.05</td>
<td>73.24</td>
<td>75.64</td>
<td>75.35</td>
</tr>
<tr>
<td>Yes</td>
<td>83.98</td>
<td>77.64</td>
<td>74.51</td>
<td>78.79</td>
<td>78.73</td>
</tr>
</tbody>
</table>

Table 6.5 Effect of block weighting on mean verification accuracy (%) for KinFaceW-II dataset

<table>
<thead>
<tr>
<th>Block Weighting</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>79.40</td>
<td>75.40</td>
<td>74.20</td>
<td>76.00</td>
<td>76.25</td>
</tr>
<tr>
<td>Yes</td>
<td>83.40</td>
<td>78.80</td>
<td>78.00</td>
<td>79.60</td>
<td>79.95</td>
</tr>
</tbody>
</table>

6.7.2 Comparison of different LBP variants

The proposed BNRML framework can be used with other LBP variants. This work compares the performance of five popular LBP variants: (1) LBP, (2) MBP, (3) CoLBP, (4) LTP and (5) MQLBP. The LTP threshold was empirically set to $t = 6$ for all the experiments. The other parameters used for various features are summarized in Table 6.6.

The total number of blocks mentioned in the table is given by:

$$\text{Total number of blocks} = \frac{\text{Number of blocks}}{\text{Pattern}} \times \frac{\text{Number of patterns}}{\text{Image}}$$

Table 6.7 and Table 6.8 present the mean verification accuracy of different features for the two datasets. Among various methods, the MBP provided the lowest mean verification accuracy on both the datasets. The MBP feature has shown its success on face recognition and gender classification task, but it could not achieve similar performance on kinship verification task. The best verification rates were recorded by the MQLBP method proposed in this thesis.
Comparing the mean verification rates with the LBP feature, the MQLBP improved the performance by 1.56% and 2.40% on KinFaceW-I and KinFaceW-II dataset respectively.

Table 6.6 Parameters of different features for BNRML framework

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number of patterns</th>
<th>Total number of blocks</th>
<th>Feature dimension per block</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>1</td>
<td>64x1=64</td>
<td>59</td>
</tr>
<tr>
<td>MBP</td>
<td>3</td>
<td>64x3=192</td>
<td>256</td>
</tr>
<tr>
<td>CoLBP</td>
<td>8</td>
<td>64x8=512</td>
<td>59</td>
</tr>
<tr>
<td>LTP</td>
<td>2</td>
<td>64x2=128</td>
<td>59</td>
</tr>
<tr>
<td>MQLBP</td>
<td>4</td>
<td>64x4=256</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 6.7 Mean verification accuracy (%) of different features under BNRML framework on KinFaceW-I

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>83.68</td>
<td>76.11</td>
<td>74.09</td>
<td>74.79</td>
<td>77.17</td>
</tr>
<tr>
<td>MBP</td>
<td>80.80</td>
<td>73.15</td>
<td>75.83</td>
<td>77.21</td>
<td>76.75</td>
</tr>
<tr>
<td>CoLBP</td>
<td>85.25</td>
<td>75.01</td>
<td>74.09</td>
<td>78.76</td>
<td>78.28</td>
</tr>
<tr>
<td>LTP</td>
<td>81.76</td>
<td>76.51</td>
<td>74.93</td>
<td>78.39</td>
<td>77.90</td>
</tr>
<tr>
<td>MQLBP</td>
<td>83.98</td>
<td>77.64</td>
<td>74.51</td>
<td>78.79</td>
<td>78.73</td>
</tr>
</tbody>
</table>

Table 6.8 Mean verification accuracy (%) of different features under BNRML framework on KinFaceW-II

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>82.80</td>
<td>72.80</td>
<td>78.60</td>
<td>76.00</td>
<td>77.55</td>
</tr>
<tr>
<td>MBP</td>
<td>77.60</td>
<td>73.40</td>
<td>72.00</td>
<td>74.60</td>
<td>74.40</td>
</tr>
<tr>
<td>CoLBP</td>
<td>83.20</td>
<td>71.20</td>
<td>73.20</td>
<td>74.20</td>
<td>75.45</td>
</tr>
<tr>
<td>LTP</td>
<td>82.00</td>
<td>77.40</td>
<td>77.80</td>
<td>79.40</td>
<td>79.15</td>
</tr>
<tr>
<td>MQLBP</td>
<td>83.40</td>
<td>78.80</td>
<td>78.00</td>
<td>79.60</td>
<td>79.95</td>
</tr>
</tbody>
</table>
6.8 Performance evaluation on periocular images

One of the goals of this work is to evaluate the effectiveness of the periocular region for kinship verification in the wild. As suggested in [215], the periocular region is defined as the region surrounding eyes which may or may not include eyebrows. To extract the periocular regions from face images, Matlab implementation of cascaded object detection algorithm [26] was employed. All the automatically segmented periocular regions were visually verified and misclassifications were rectified by manually segmenting the periocular regions. Fig. 6.9 shows the examples of image pairs with kin relations using full face images and periocular regions.

![Fig. 6.9 Example image pairs with kinship relation for two datasets: (a) KinFaceW-I and (b) KinFaceW-II. Four kin relations from top to bottom are father-son (F-S), father-daughter (F-D), mother-son (M-S), mother-daughter (M-D) respectively. Each row shows full face images of parent and child at the corners. Periocular images of parent and child are shown at the top and bottom respectively in the middle column of each row.](image-url)

The extracted periocular images were resized to 16 × 24 pixels. The MQLBP features were extracted from periocular images with the parameter settings described in Section 6.7. To obtain spatial histograms, each local pattern image was partitioned into 12 non-overlapping blocks of size 4×8. For metric learning, the BNRML method was used since its
effectiveness has already been validated in Section 6.7. Further, the left (L) and the right (R) periocular regions were considered independently initially to evaluate the performance of periocular features. Later, information fusion of the both the regions was performed as described in Section 6.5.

The verification results for two datasets are shown in Table 6.9 and Table 6.10 respectively. Further, the results have also been depicted using scatter plots shown in Fig. 6.10 and Fig. 6.11 for the purpose of visual interpretation. Following inferences can be drawn from the results:

1. The verification results of left and right periocular regions by themselves are very encouraging. The mean verification accuracy of both the regions are 72.01% and 70.15% respectively on KinFaceW-I dataset. Corresponding accuracies on KinFaceW-II dataset are 74.40% and 74.15%. It is important to remember that the periocular region has already demonstrated its potential for face recognition, ethnicity classification, and gender recognition task. Our experiments extend this knowledge and demonstrate the suitability of the periocular region for automatic kinship verification.

<table>
<thead>
<tr>
<th>Region</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left (L)</td>
<td>73.71</td>
<td>68.32</td>
<td>71.52</td>
<td>74.50</td>
<td>72.01</td>
</tr>
<tr>
<td>Right (R)</td>
<td>72.76</td>
<td>69.42</td>
<td>71.09</td>
<td>67.33</td>
<td>70.15</td>
</tr>
<tr>
<td>Left + Right</td>
<td>76.26</td>
<td>70.91</td>
<td>74.93</td>
<td>74.47</td>
<td>74.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left (L)</td>
<td>76.20</td>
<td>75.20</td>
<td>73.60</td>
<td>72.60</td>
<td>74.40</td>
</tr>
<tr>
<td>Right (R)</td>
<td>76.60</td>
<td>75.60</td>
<td>72.80</td>
<td>71.60</td>
<td>74.15</td>
</tr>
<tr>
<td>Left + Right</td>
<td>80.20</td>
<td>79.60</td>
<td>76.60</td>
<td>76.20</td>
<td>78.15</td>
</tr>
</tbody>
</table>

Table 6.9 Mean verification accuracy (%) of periocular region on KinFaceW-I dataset

Table 6.10 Mean verification accuracy (%) of periocular region on KinFaceW-II dataset
2. The fusion of left and right periocular region improves the performance on both the datasets. The gain in mean verification accuracy of the fused region over left and right periocular region are 2.13% and 3.99% on KinFaceW-I dataset. Corresponding performance gains are 3.75% and 4.00% on KinFaceW-II dataset.

3. The mean verification accuracy of all the regions was consistently higher on KinFaceW-II dataset than KinFaceW-I dataset. The images of parent-child pairs in KinFaceW-II dataset were captured from the same photographs whereas they were captured from different photographs in case of KinFaceW-I dataset. Hence, the
similar acquisition conditions present in KinFaceW-II dataset may have produced better performance.

Similar results can also be verified by the ROC curves depicted in Fig. 6.12 and Fig. 6.13.

![ROC curves](image)

Fig. 6.12 ROC curves of periocular region on KinFaceW-I dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D

### 6.8.1 Comparison with different LBP variants

Previous section demonstrated encouraging performance of the MQLBP features extracted from the periocular region for kinship verification task. In this section, the performance of other LBP variants has been evaluated. Further, the performance of different variants was evaluated for fused periocular regions (L + R) since it showed the best performance during experiments described in the previous section. Various features have been
extracted using parameters settings described in Section 6.7.2. The parameters are summarized in Table 6.11. The total number of blocks mentioned in the table is given by:

\[
\text{Total number of blocks} = \frac{\text{Number of blocks}}{\text{Pattern}} \times \frac{\text{Number of patterns}}{\text{Periocular region}} \times \frac{\text{Number of periocular regions}}{\text{Image}}
\]  

(6.6)

Fig. 6.13 ROC curves of periocular region on KinFaceW-II dataset: (a) F-S, (b) F-D, (c) M-S, (d) M-D

Table 6.12 and Table 6.13 present the mean verification accuracy of different features for the two datasets. The results of all the features are quite promising since they all recorded the mean verification accuracy of 69.00% or more on both the datasets. Among various methods, the MBP recorded the lowest mean verification accuracy whereas the MQLBP performed the best on both the datasets. Further, all the features demonstrated higher
verification rates on KinFaceW-II dataset as compared to KinFaceW-I. Considering the performance of LBP as the baseline performance, the MQLBP achieved a performance gain of 2.06% and 4.50% on KinFaceW-I and KinFaceW-II datasets respectively.

Table 6.11 Parameters of different features for BNRML framework

<table>
<thead>
<tr>
<th>Feature</th>
<th>Number of patterns</th>
<th>Total number of blocks</th>
<th>Feature dimension per block</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>1</td>
<td>12×1×2=24</td>
<td>59</td>
</tr>
<tr>
<td>MBP</td>
<td>3</td>
<td>12×3×2=72</td>
<td>256</td>
</tr>
<tr>
<td>CoLBP</td>
<td>8</td>
<td>12×8×2=192</td>
<td>59</td>
</tr>
<tr>
<td>LTP</td>
<td>2</td>
<td>12×2×2=48</td>
<td>59</td>
</tr>
<tr>
<td>MQLBP</td>
<td>4</td>
<td>12×4×2=96</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 6.12 Mean verification accuracy (%) of different periocular features on KinFaceW-I dataset

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>72.12</td>
<td>70.14</td>
<td>73.30</td>
<td>72.76</td>
<td>72.08</td>
</tr>
<tr>
<td>MBP</td>
<td>68.90</td>
<td>67.12</td>
<td>70.27</td>
<td>69.70</td>
<td>69.00</td>
</tr>
<tr>
<td>CoLBP</td>
<td>78.87</td>
<td>68.23</td>
<td>71.09</td>
<td>73.19</td>
<td>72.85</td>
</tr>
<tr>
<td>LTP</td>
<td>76.28</td>
<td>70.51</td>
<td>73.70</td>
<td>72.47</td>
<td>73.24</td>
</tr>
<tr>
<td>MQLBP</td>
<td>76.26</td>
<td>70.91</td>
<td>74.93</td>
<td>74.47</td>
<td>74.14</td>
</tr>
</tbody>
</table>

Table 6.13 Mean verification accuracy (%) of different periocular features on KinFaceW-II dataset

<table>
<thead>
<tr>
<th>Feature</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>74.60</td>
<td>72.40</td>
<td>75.00</td>
<td>72.60</td>
<td>73.65</td>
</tr>
<tr>
<td>MBP</td>
<td>70.00</td>
<td>72.00</td>
<td>70.40</td>
<td>67.40</td>
<td>69.95</td>
</tr>
<tr>
<td>CoLBP</td>
<td>78.00</td>
<td>71.60</td>
<td>73.60</td>
<td>72.40</td>
<td>73.90</td>
</tr>
<tr>
<td>LTP</td>
<td>79.40</td>
<td>79.00</td>
<td>77.00</td>
<td>72.80</td>
<td>77.05</td>
</tr>
<tr>
<td>MQLBP</td>
<td>80.20</td>
<td>79.60</td>
<td>76.60</td>
<td>76.20</td>
<td>78.15</td>
</tr>
</tbody>
</table>
6.8.2 Comparison with full face images

This section compares the performance of periocular region with full face images. Further, the information from both the periocular and face images was fused using BNRML framework to investigate the scope of gain in performance. The mean verification rates of three different facial regions on two datasets are depicted using bar charts in Fig. 6.14 and Fig. 6.15. As expected, the mean verification rates of periocular region are lower than the face images for both the datasets. Nevertheless, the periocular features achieved more than 74.00% of mean verification rates on both the datasets. Further, the verification performance of periocular region was within five percentage points of the facial images. Moreover, the periocular region when fused with face enhanced the verification performance. Overall, the results support the viability of periocular region for automatic kinship verification.
Table 6.14 Comparison of verification accuracy (%) with existing methods on KinFaceW-I dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature</th>
<th>Region</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRML [196]</td>
<td>LBP</td>
<td>Face</td>
<td>81.43</td>
<td>69.76</td>
<td>67.23</td>
<td>72.87</td>
<td>72.82</td>
</tr>
<tr>
<td></td>
<td>HOG</td>
<td></td>
<td>83.68</td>
<td>74.64</td>
<td>71.56</td>
<td>79.96</td>
<td>77.46</td>
</tr>
<tr>
<td>ASL [196]</td>
<td>LBP</td>
<td>Face</td>
<td>85.51</td>
<td>76.54</td>
<td>69.93</td>
<td>74.36</td>
<td>76.59</td>
</tr>
<tr>
<td></td>
<td>HOG</td>
<td></td>
<td>86.90</td>
<td>76.48</td>
<td>70.62</td>
<td>79.75</td>
<td>78.44</td>
</tr>
<tr>
<td>ESL [164]</td>
<td>LBP</td>
<td>Face</td>
<td>81.70</td>
<td>71.10</td>
<td>69.60</td>
<td>74.30</td>
<td>74.10</td>
</tr>
<tr>
<td></td>
<td>HOG</td>
<td></td>
<td>83.90</td>
<td>76.00</td>
<td>73.50</td>
<td>81.50</td>
<td>78.60</td>
</tr>
<tr>
<td>Proposed</td>
<td>MQLBP</td>
<td>Periocular (L+R)</td>
<td>76.26</td>
<td>70.91</td>
<td>74.93</td>
<td>74.47</td>
<td>74.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face (F)</td>
<td>83.98</td>
<td>77.64</td>
<td>74.51</td>
<td>78.79</td>
<td>78.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fusion (L+R+F)</td>
<td>84.95</td>
<td>76.13</td>
<td>76.21</td>
<td>79.16</td>
<td>79.11</td>
</tr>
</tbody>
</table>

Table 6.15 Comparison of verification accuracy (%) with existing methods on KinFaceW-II dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature</th>
<th>Region</th>
<th>F-S</th>
<th>F-D</th>
<th>M-S</th>
<th>M-D</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRML [196]</td>
<td>LBP</td>
<td>Face</td>
<td>79.20</td>
<td>71.60</td>
<td>72.20</td>
<td>68.40</td>
<td>72.85</td>
</tr>
<tr>
<td></td>
<td>HOG</td>
<td></td>
<td>80.80</td>
<td>72.80</td>
<td>74.80</td>
<td>70.40</td>
<td>74.70</td>
</tr>
<tr>
<td>ASL [196]</td>
<td>LBP</td>
<td>Face</td>
<td>84.24</td>
<td>79.45</td>
<td>75.98</td>
<td>77.04</td>
<td>79.18</td>
</tr>
<tr>
<td></td>
<td>HOG</td>
<td></td>
<td>87.51</td>
<td>80.82</td>
<td>79.78</td>
<td>75.63</td>
<td>80.94</td>
</tr>
<tr>
<td>ESL [164]</td>
<td>LBP</td>
<td>Face</td>
<td>80.50</td>
<td>72.20</td>
<td>72.80</td>
<td>71.60</td>
<td>74.30</td>
</tr>
<tr>
<td></td>
<td>HOG</td>
<td></td>
<td>81.20</td>
<td>73.00</td>
<td>75.60</td>
<td>73.00</td>
<td>75.70</td>
</tr>
<tr>
<td>Proposed</td>
<td>MQLBP</td>
<td>Periocular (L+R)</td>
<td>80.20</td>
<td>79.60</td>
<td>76.60</td>
<td>76.20</td>
<td>78.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Face (F)</td>
<td>83.40</td>
<td>78.80</td>
<td>78.00</td>
<td>79.60</td>
<td>79.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fusion (L+R+F)</td>
<td>83.60</td>
<td>79.40</td>
<td>79.80</td>
<td>80.80</td>
<td>80.90</td>
</tr>
</tbody>
</table>
6.8.3 Comparison with previous work

Table 6.14 and Table 6.15 compare the performance of the proposed method with existing state-of-the-art methods. The methods evaluated under image-unrestricted setting were used for fair comparison. As shown in the tables, the existing approaches have mainly used histogram of oriented gradients (HOG) [218] and LBP features with different metric learning methods such as NRML, ESL and asymmetric metric learning (ASL). Further, the results tabulated for comparison are directly cited from respective papers. The promising performance of periocular region is clearly evident from the comparison. Further, the fusion of periocular and full face region proposed in this work provided highly competitive performance in comparison with state-of-the-art methods.

6.9 Summary

In this chapter, a block-based neighborhood repulsed metric learning (BNRML) framework has been proposed to address kinship verification task. Experimental results under image-unrestricted protocol on two kinship datasets have clearly indicated superior performance of the proposed framework in comparison with NRML method. Further, on comparing the performance of several variants of local binary patterns under BNRML framework, the multi-quantized local binary pattern features were found to perform better than others.

Apart from experiments on full face images, this work also investigated the effectiveness of periocular region on kinship verification task. Such study has been carried out for the first time to the best of our knowledge. Experimental results have clearly demonstrated the viability of periocular region for kinship verification. Specifically, a fusion of both the left and right periocular region performed superior to the individual periocular region. Finally, the fusion face and periocular traits using BRNML framework demonstrated highly competitive performance as compared to the state-of-the-art methods.
7.1 Conclusion

Soft biometrics refers to a set of characteristics providing some information that is not distinctive enough to determine the identity of a person. Soft biometric traits can be extracted from distance without user’s cooperation and they can be used to improve the performance of classical biometric systems. Further, their human compliance nature makes them suitable for many other potential applications. Due to aforementioned advantages, soft biometrics has recently received a lot of attention among computer vision researchers.

This research work pursued the classification of two soft biometric attributes (gender and kinship) using facial images. The major concentration is on finding the discriminative facial representations which are robust against various challenging factors such as low image resolution, variations in pose, illumination, and expression. Hence, the methods proposed in this work have been evaluated on real-life datasets which contain images captured in an unconstrained environment.

This study first focused on gender classification of facial images. Existing gender classification algorithms were systematically reviewed and the focus was narrowed down to local binary patterns (LBP). Based on thoughtful analysis of LBP, the method was generalized by defining a multi-level thresholding function. The resultant multi-quantized LBP (MQLBP) implicitly captures both the sign and magnitude information from the gray level differences between neighboring pixels and a center pixel. Comprehensive experiments were carried out on four publically available datasets including FERET, PAL, CASIA, and FEI. Comparison of MQLBP with different LBP variants clearly demonstrated higher classification rates of the proposed method. Further, gender classification performance of several descriptors was also analyzed with reference to age variations. Among various descriptors, the MQLBP performed most consistently across different age groups. Finally, the performance of the proposed MQLBP descriptor was also evaluated on noisy input images and the results indicated better noise tolerance capability of the proposed method.

The basic LBP operates directly on raw pixel intensity. However, gradient images can also be analyzed to discover buried local patterns acting as a source of facial features. Based on this idea, this work developed a new feature descriptor called compass local binary patterns (CoLBP). This feature integrates the properties of Kirsch compass masks and LBP to
achieve higher discrimination power. The performance of this descriptor was evaluated on challenging real-life datasets such as LFW and Adience. In comparison with several existing benchmark methods, the proposed descriptor provided promising classification performance. Specifically, the CoLBP method demonstrated competitive classification rates in comparison with two state-of-the-art methods: Local Gabor binary pattern histogram sequence (LGBHS) and monogenic binary pattern (MBP). However, the analysis on computational complexity demonstrated that the proposed method has a clear advantage over them in terms of lower computational burden. Thus, the proposed CoLBP method offers a good balance between a classification performance and computational complexity.

In addition to facial photographs, this thesis also addressed the idea of gender recognition of facial sketches which has gained very little attention. Facial sketches are widely used in forensic investigations. Sketch-based gender classification can be a valuable tool for filtering a large dataset in a heterogeneous face recognition system (HFR). With this motivation, this work evaluated the performance of several LBP variants on two facial sketch datasets. The results clearly demonstrated the success of LBP variants on this modality as well. It is important to remember that the LBP and its variants have already shown their effectiveness in classifying gender of facial photographs. This work goes one step further and demonstrates their potential on sketch modality also. Moreover, the proposed CoLBP descriptor showed higher classification rates than other LBP variants on both the datasets.

Another major problem addressed in this work is kinship verification in an unconstrained setting. Since the feature extraction algorithms have already been discussed earlier, we concentrated on a metric learning task. Based on literature review, neighborhood repulsed metric learning (NRML) was used to evaluate the performance of proposed feature extraction algorithms. Further, an NRML method was extended to obtain more discriminative information. The NRML method learns a single global distance metric which may not capture adequate information as face images usually reside on the non-linear manifold. To improve the performance, this work proposed block-based NRML (BNRML) method. The performance of proposed method was evaluated on two datasets: KinFaceW-I and KinFaceW-II. The proposed method significantly improved the performance of NRML method across four kin relations present in both the datasets. Further, among various feature descriptors, the MQLBP feature proposed in this work achieved the highest classification rates on both the datasets.

Existing kinship verification approaches use whole face images to extract kinship cues. Inspired by the success of periocular biometrics, this work addresses an interesting problem of
kinship verification from periocular images. The word periocular refers to the region of a face surrounding eyes. To study the effectiveness of periocular features, this work segmented the periocular regions from face images and then extracted local features from it. Using BNRML framework, the performance of different features was evaluated on two kinship datasets. The results were very promising and it demonstrated the feasibility of periocular features for kinship verification task. Further, the information from the left and right periocular regions was fused to improve the classification performance. It is important to note that the periocular region has already demonstrated its potential for face recognition and other related tasks. This work broadens this knowledge and validates the suitability of the periocular region for kinship verification.

7.2 Future work

This dissertation presents several feature representation algorithms for classification of two soft biometric attributes using face images. However, the research in this field is ongoing and this work also opens up possible future directions which are listed below:

- An interesting future work is to evaluate the performance of proposed methods for classification of other soft biometric attributes like ethnicity and age.
- This work is based on existing features like LBP and Gabor wavelets. Such features have been originally proposed for texture classification problem. Later, they were introduced into face analysis tasks. Hence, it is appealing to go back from face analysis task and evaluate the performance of MQLBP and CoLBP features on texture classification task.
- In addition to feature extraction step, typical pattern recognition pipeline also includes feature selection step before classification. Therefore, it is imperative to explore various feature selection algorithms for reducing feature dimension in order to improve overall system performance.
- This work explored LBP-based texture features only. However, various shape and geometric features can also be discovered for soft biometric classification. Another important aspect is to perform information fusion of different types of features and obtain unified feature representation so as to improve the overall performance.
- For kinship verification task, this work proposed block-based NRML (BNRML) framework. This idea can be applied to other metric learning methods and comparative analysis of different methods can be performed.
This work proposed kinship verification from periocular features. It is interesting to evaluate the performance of other facial parts (e.g. nose, mouth) in kinship verification. Further, information fusion from different face parts can be performed to investigate any performance gain over the whole face.

This work employed four types of kin relations: Father-son, father-daughter, mother-son and mother-daughter. This can be extended over to other kin relations such as siblings and grandparent-child. Further, for different types of kin relations, it will be interesting to investigate the effects of age variations and degree of kin relation on the performance. For example, the age difference among siblings is usually much smaller than parent-child or grandparent-child. Hence, the impact of age variations on overall performance can be studied and efforts can be made to mitigate such influences.
PUBLICATIONS FROM THE WORK

Journals


BIBLIOGRAPHY


127


APPENDIX A

Technical details of neighborhood repulsed metric learning (NRML)

The NRML method was proposed by Lu et al. [10]. For clarity in understanding, the mathematical details of NRML are reproduced below.

Let $F = \{(f_i^p, f_i^c, l_i)\}_{i=1}^N$ be the training set of $N$ labeled samples. Each training sample consists of a pair of feature vectors from parent-child images denoted by $f_i^p$ and $f_i^c$ respectively. The binary kinship information is denoted by class labels $l_i \in \{1, 0\}$, where $l_i = 1$ indicates positive kinship and $l_i = 0$ indicates negative.

The distance metric $d$ is defined as follows:

$$d(x, y) = \sqrt{(x-y)^T A (x-y)} \quad (A.1)$$

where $A$ is $m \times m$ square matrix which is symmetric and positive semidefinite. Alternatively, we can seek a nonsquare matrix $W$ of size $m \times l$, where $m \leq l$, such that

$$A = WW^T \quad (A.2)$$

Then, (A.1) can be rewritten as

$$d(x, y) = \sqrt{(x-y)^T WW^T (x-y)} = \sqrt{(u-v)^T (u-v)} \quad (A.3)$$

where $u = W^T x$ and $v = W^T y$.

The NRML method computes three types of distances for each positive parent-child image pair in the training dataset and solves an optimization problem as follows:

$$\max_A J(A) = J_1(A) + J_2(A) - J_3(A)$$

$$= \frac{1}{Nk} \sum_{r=1}^{N} \sum_{s=1}^{k} d^2(f_r^p, f_{r,s}^c) + \frac{1}{Nk} \sum_{r=1}^{N} \sum_{t=1}^{k} d^2(f_r^c, f_{r,t}^p) - \frac{1}{N} \sum_{r=1}^{N} d^2(f_r^p, f_r^c) \quad (A.4)$$

where $f_{r,s}^c$ denotes the $s^{th}$ $k$-nearest neighbor of $f_r^c$ and $f_{r,t}^p$ denotes the $t^{th}$ $k$-nearest neighbor of $f_r^p$ respectively. The term $J_1$ ensures that the nearest neighbors $f_{r,s}^c$ of $f_r^c$ should be separated from $f_r^p$ as far as possible. Similarly, the term $J_2$ ensures that the nearest neighbors $f_{r,t}^p$ of $f_r^p$ should be separated from $f_r^c$ as far as possible. In contrast, the term $J_3$ ensures that $f_r^p$ and $f_r^c$ are drawn as close to each other as possible.

Combining (A.3) and (A.4), the term $J_1(A)$ can be simplified to the following form:
\[ J_1(A) = \frac{1}{Nk} \sum_{r,s=1}^{N} (f_r^p - f_r^c)^T W W^T (f_r^p - f_r^c) \]
\[ = tr(W^T H W) \] (A.5)

where
\[ H_1 = \frac{1}{Nk} \sum_{r,s=1}^{N} (f_r^p - f_r^c) (f_r^p - f_r^c)^T \]

Similarly, \( J_2(A) \) and \( J_3(A) \) can be simplified as
\[ J_2(A) = tr(W^T H_2 W) \] (A.6)
\[ J_3(A) = tr(W^T H_3 W) \] (A.7)

where
\[ H_2 = \frac{1}{Nk} \sum_{r,s=1}^{N} (f_r^p - f_r^c) (f_r^p - f_r^c)^T \] and
\[ H_3 = \frac{1}{N} \sum_{r=1}^{N} (f_r^p - f_r^c) (f_r^p - f_r^c)^T \].

The transformation matrix \( W \) can now be obtained by solving the following optimization problem:
\[ \max W J(W) = tr(W^T (H_1 + H_2 + H_3) W) \]
subject to \( W^T W = I \) (A.8)

The constraint \( W^T W = I \) is used to keep the optimization problem well posed with respect to the scale of \( W \). The following eigenvalue problem can then be used to obtain \( W \):
\[ (H_1 + H_2 + H_3) \omega = \lambda \omega \] (A.9)