Bayesian Edge-Texture Feature Descriptor for Human Face Recognition

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

Master of Engineering
in
Computer Science and Engineering

Submitted by

Abhijit Sarkar
Regn. No. - 163555 of 2022 - 2023
Exam Roll No. - M4CSE24002
Class Roll No. - 002210502003
Session - 2022 - 2024

Under the supervision of

Prof. Debotosh Bhattacharjee

Department of Computer Science and Engineering
Jadavpur University
188, Raja S.C. Mallick Rd,
Kolkata - 700032, West Bengal, India

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF RECOMMENDATION

This is to certify that the work embodied in this thesis entitled **"Bayesian Edge-Texture Feature Descriptor for Human Face Recognition"** has been satisfactorily completed by **Abhijit Sarkar** (Registration Number 163555 of 2022 - 2023; Class Roll No. 002210502003; Examination Roll No. M4CSE24002). It is a bonafide piece of work carried out under my supervision and guidance at Jadavpur University, Kolkata, for partial fulfillment of the requirements for the awarding of the **Master of Engineering** in **Computer Science and Engineering** degree of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, during the academic year 2022 - 2024.

Prof. Debotosh Bhattacharjee,
Department of Computer Science and Engineering,
Jadavpur University.
(Supervisor)

Forwarded By:

Prof. Nandini Mukherjee,
HOD, Department of Computer Science and Engineering,
Jadavpur University.

Prof. Dipak Laha,
DEAN, Faculty of Engineering and Technology,
Jadavpur University.

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled **"Bayesian Edge-Texture Feature Descriptor for Human Face Recognition"** is a bonafide record of work carried out by **Abhijit Sarkar** (Registration Number 163555 of 2022 - 2023; Class Roll No. 002210502003; Examination Roll No. M4CSE24002) in partial fulfillment of the requirements for the award of the degree of **Master of Engineering** in **Computer Science and Engineering** in the **Department of Computer Science and Engineering, Jadavpur University**, during the period of September 2022 to May 2024. It is understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the thesis only for the purpose of which it has been submitted.

Signature of Examiner 1
Date:

Signature of Examiner 2
Date:

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that the thesis entitled **"Bayesian Edge-Texture Feature Descriptor for Human Face Recognition"** contains literature survey and original research work by the undersigned candidate, as a part of his degree of **Master of Engineering** in **Computer Science and Engineering, Jadavpur University**. All the information have been obtained and presented in accordance with academic rules and ethical conduct.
I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Abhijit Sarkar

Examination Roll No.: M4CSE24002

Registration No.: 163555 of 2022 - 2023

Thesis Title: Bayesian Edge-Texture Feature Descriptor for Human Face Recognition

Signature of the Candidate:

Acknowledgements

The success and outcome of this thesis required a lot of guidance and assistance from many people, and I am extremely lucky to come across them during the thesis. I would like to take this opportunity to thank them for their guidance and support in everything I have accomplished. I owe my deep gratitude to my supervisor and thesis guide **Prof. Debotosh Bhattacharjee**, who was always there during my whole thesis work and guided me all along till the completion of my thesis work by providing all the necessary information that would be beneficial for my research and experiments. I am extremely thankful to him for his assistance and dedicated involvement in every step of the process. His constant suggestions have been an everlasting source of motivation and inspiration to me. It has been an honour to work under a person with wisdom yet a kind, simplistic attitude.

I would also like to thank **Prof. Nandini Mukherjee**, Head of the department, Department of Computer Science and Engineering at Jadavpur University, for extending her help to complete this thesis.

I had the pleasure of working in the **CMATER** lab of Jadavpur University. I acknowledge the generous collaboration and assistance of every member of the CMATER lab. It has been a wonderful experience working with every member of this lab. I learnt many things regarding research works and was inspired to do better in my thesis work from them.

I especially like to thank my classmate and one of the members of CMATER lab, **Akash Halder** who helped me in many steps of my thesis work.

At the end, I like to acknowledge both of my parents for their constant support during this thesis work.

Abhijit Sarkar

Examination Roll No.: M4CSE24002

Registration No.: 163555 of 2022 - 2023

Department of Computer Science and Engineering,

Jadavpur University.

Abstract

This thesis work describes a novel feature descriptor named Conditional Bayes Conjunctional Edge-Texture and Texture-Edge (CBCETTE) feature extractor for human face recognition. It is based on Euclidean inter-distances between the local centre of mass (LCOM) and pixels, which is proficient in creating an optimal feature set by combining edge and texture inclusive of invariant to illumination, rotation, and noise. Initially, the image is divided into blocks of size 3×3 and for each block, a local centre of mass (LCOM) is derived initially from the original image. Using this local centre of mass, from each block image an edge image and a texture image is extracted using a novel approach on the basis of the direction of light. Subsequently from the edge image and texture image, a probabilistic edge image and a probabilistic texture image are derived further to enunciate the combined edge texture feature set. Finally, Bayes' conditional probabilistic approach is applied on both the probabilistic edge and the probabilistic texture image to obtain the probabilistic edge-texture and vice-versa feature set. The optimum feature set is the conjunction of the two probabilistic feature sets. This final optimum conjunction probabilistic feature set is finally used for classification purposes. It has been successfully tested on various challenging databases like Yale B face database, Extended Yale B face database, Yale face database, Faces95 database, and ORL database. It provides the accuracy of 99.38%, 98.77%, 97.58%, 97.92%, and 93.75% respectively on Yale B face database, Extended Yale B face database, Yale face database, Faces95 database, and ORL database that exhibits supremacy of the proposed strategy for human face recognition under various conditions.

Contents

Abstract i

Table of Contents ii

List of fig iv

List of Tables v

List of Abbreviations vi

1 Introduction 1
- 1.1 Human Face Images . . . 1
- 1.2 Definition of Feature and Feature Extraction . . . 2
  - 1.2.1 Local Feature Extraction . . . 3
  - 1.2.2 Holistic Feature Extraction . . . 4
- 1.3 Human Face Recognition Employing Local Feature Descriptor . . . 4

2 Literature Survey 6
- 2.1 Challenges for Face Recognition . . . 6
  - 2.1.1 Illumination Variations . . . 6
  - 2.1.2 Pose Variations . . . 7
  - 2.1.3 Expression Variations . . . 7
  - 2.1.4 Rotation Variations . . . 8
  - 2.1.5 Accumulation of Noise . . . 8
  - 2.1.6 Occlusion . . . 8
  - 2.1.7 Ageing . . . 9
  - 2.1.8 Scale Variations . . . 9
- 2.2 Standard Databases for Face Recognition . . . 9
- 2.3 Trends in Feature Extraction Methodologies . . . 12
  - 2.3.1 Overview on Local Feature Extraction Methodologies . . . 12
  - 2.3.2 Overview on Holistic Feature Extraction Methodologies . . . 24
- 2.4 Conventional Performance Metrics . . . 26

3 Proposed Methodology 29
- 3.1 Concept of Pixel's distance from Local Centre of Mass . . . 29
  - 3.1.1 Bayesian Conditional Probabilistic Feature Extraction . . . 30
  - 3.1.2 Intuition Behind 3×3 Window . . . 38
  - 3.1.3 Intuition Behind Stride 1 . . . 39

3.1.4 Size of Feature Map 39
3.1.5 Properties of Proposed Method CBCETTE 40

4 Experiment and Result **41**
4.1 Experimental Setup 41
4.2 Experimental Details With Results 42
4.2.1 Yale B Face Database 42
4.2.2 Extended Yale B Face Database 44
4.2.3 Faces95 Face Database 45
4.2.4 Yale Face Database 47
4.2.5 ORL Face Database 48

5 Conclusion and Future Work **56**
5.1 Conclusion of the Present Work 56
5.2 Scope for Future Work 56

Bibliography **58**

List of Figures

2.1 Confusion Matrix 27

3.1 Distance Calculation From COM 31
3.2 3×3 neighbourhood of a pixel 32
3.3 Describes the process of calculating Conjunctional Edge-Texture Image 34
3.4 Shows the two-body system 36

4.1 Example of feature map images of Yale B 43
4.2 Example of feature map images of Extended Yale B 45
4.3 Example of feature map images of Faces95 47
4.4 Example of feature map images of Yale 48
4.5 Example of feature map images of ORL 50
4.6 Confusion Matrix of Yale B with 5-Fold Cross Validation 51
4.7 Confusion Matrix of Extended Yale B with 5-Fold Cross Validation 52
4.8 Confusion Matrix of Faces95 with 5-Fold Cross Validation 53
4.9 Confusion Matrix of Yale with 5-Fold Cross Validation 54
4.10 Confusion Matrix of ORL with 5-Fold Cross Validation 55

List of Tables

2.1 Widely Used Face Databases 10

4.1 Accuracy Comparison for Yale B 44
4.2 Accuracy Comparison for Extended Yale B 46
4.3 Accuracy Comparison for Faces95 47
4.4 Accuracy Comparison for Yale 49
4.5 Accuracy Comparison for ORL 50

List of Abbreviations

C

CBCETTE	**C**onditional **B**ayes **C**onjunctional **E**dge-**T**exture and **T**exture-**E**dge

F

FE	**F**eature **E**xtraction

H

HFR	**H**uman **F**ace **R**ecognition

L

LCOM	**L**ocal **C**entre **O**f **M**ass

P

PCA	**P**rincipal **C**omponent **A**nalysis

S

SVM	**S**upport **V**ector **M**achine

CHAPTER 1

Introduction

"The face is the mirror of the mind", the thought process of each mind is different than the other which is reflected in their face synopsis which conjointly creates overall face architecture that's why the appearance of each face is unique. There are more than 7 billion faces in the world, and more or less every face comprises the same constituents such as eyes, nose, and lips et al, but the emergence of each the face is different that's why the face is one of the most important biometric candidate keys of the human body [1]. That's why human face recognition (HFR) has emerged as the most demanding task in computer vision, pattern recognition, network security, content indexing and retrieval, and robotics for the past few years. Faces are the most easily accessible, either from digital devices or from the evidence provided by eyewitnesses, and easily recognizable among the biometric features. All over again, if the situation needs a real-time or quick resolution of a crime, then there is no other choice than face biometrics. Though progress made in face recognition in terms of computational time and performance accuracy with time, the identification process is not guaranteed all the time depending upon the environmental situation [2]. That makes human face recognition a challenging and interesting field of research work. The whole work could be divided into two subsets – namely feature extraction (FE) and classification. The phase called feature extraction helps to extract robust features from images of faces coming from different environmental conditions. Then using those features, classification models help to classify or recognise those images. Here all those things are thoroughly discussed in the chapter.

1.1 Human Face Images

Human face images are visual representations capturing the front-facing aspect of an individual's head, including facial features, expressions, and nuances.

These images serve as crucial input for face recognition systems, which aim to identify and verify individuals based on their unique facial characteristics. Face images are captured using a variety of imaging devices and techniques, ranging from traditional cameras to specialized sensors designed for biometric applications. Conventional cameras utilize visible light to capture color or grayscale images, while infrared cameras detect heat radiation to capture infrared images. Facial images can be obtained through different methods, including still photography, video surveillance, 3D scanning, and depth sensing technologies like structured light or time-of-flight cameras. A face image can be captured in different lighting conditions, different facial expressions, pose variations, and other unconstrained environmental factors. Different image-taking machines provide different quality of pictures and the distance between the face and the image-taking machine can change the scale of the subject and may introduce noise in images.

Human face images come in various types, including color, grayscale, and infrared, each offering different advantages and applications. Color images provide rich detail and realism, making them suitable for everyday applications such as personal device authentication. Grayscale images, which represent intensity levels of light without color information, are commonly used in biometric systems due to their simplicity and compatibility with image processing algorithms. Infrared images capture heat radiation emitted by the face, offering enhanced visibility in low-light conditions and enabling robust performance in environments where visible light may be limited or obscured.

1.2 Definition of Feature and Feature Extraction

In the context of human face recognition, 'feature' refers to specific characteristics or attributes of a person's face that can be quantitatively measured, analyzed, and compared with other faces. These features are essential for accurately distinguishing one individual from another. Some examples of facial features include the distance between the eyes, the width of the nose, the shape of the lips, the curvature of the eyebrows, and the texture of the skin. Different textural attributes of human faces can be prominent features. Edges delineate the boundaries and contours of facial features like the eyes, nose, and mouth, offering crucial structural information that defines the unique geometry of each face. Algorithms dedicated to edge detection are pivotal in extracting these features, enabling the creation of distinct representations for each individual. Texture refers to the intricate patterns and surface characteristics present on the skin, including wrinkles, pores, and variations in tone. Analyzing these textural details provides additional cues for distinguishing between faces, particularly in cases where facial structures may

be similar.

One of the fundamental steps in face recognition is feature extraction, which plays a vital role in capturing and representing the distinguishing characteristics or features of a person's face. This comes under image processing. For HFR, the feature extraction process becomes even more critical to accommodate the wide range of appearance variations encountered in unconstrained conditions. Feature extraction involves transforming raw image data into a compact and meaningful representation that can be used for subsequent analysis and comparison. The challenge lies in extracting meaningful facial features invariant to pose, illumination, expression, occlusion variations, etc. Due to these challenges, this becomes a more challenging field. A better human face recognition system could be built to solve these challenges. The human face shows intricate details, including the spatial relationships between texture patterns, facial landmarks, and the overall shape of facial components. The whole realm of feature extraction can be divided into two subcategories - local feature extraction and holistic or global feature extraction.

1.2.1 Local Feature Extraction

Local facial feature extraction is crucial in achieving optimal performance for facial recognition systems. Rather than considering the entire face, local feature extraction involves dividing the face into smaller regions and extracting features from each region independently. This approach allows to capture fine-grained details crucial for discriminating between individuals. Local feature descriptors have gained considerable attention due to their resilience to multiple variations by enforcing spatial locality in both pixel and patch levels. Local feature extraction methods often focus on extracting texture descriptors, shape information, or both from specific regions of the face, such as the eyes, nose, and mouth. Local features such as edge and texture play a pivotal role in effectively fulfilling the purpose of facial recognition. Edges represent boundaries between facial components, like the eyes and the surrounding skin. Texture patterns, on the other hand, encode detailed information about the surface appearance of facial regions. Both types of features are critical for capturing the unique characteristics of individuals, making them highly relevant for face recognition tasks. Local feature extraction techniques offer several advantages in face recognition. They can provide robustness against pose variations by focusing on relatively stable regions across different poses. Furthermore, by considering local regions independently, these methods can handle occlusions and variations in illumination more effectively. Since they analyse individual facial regions independently, they can still identify a person even if parts of their face are obscured or under different lighting conditions. This is particularly

important in surveillance and security applications where face images may suffer from partial occlusions or varying lighting conditions. Additionally, local feature extraction allows for efficient representation and comparison of facial attributes, as only relevant regions need to be considered for matching purposes. Computational resources can be saved by considering only relevant regions for matching purposes. This efficiency is crucial, especially in real-time applications or systems dealing with large databases of faces.

1.2.2 Holistic Feature Extraction

Holistic feature extraction for human face recognition involves capturing and analyzing the entire face as a single entity rather than focusing on specific facial components. This approach considers the spatial relationships and overall structure of facial features, including shape, texture, and color, to create a comprehensive representation of the face. By incorporating global information, such as the relative positions of eyes, nose, and mouth, holistic methods aim to capture the unique characteristics and variations in facial appearance. Holistic feature extraction techniques for human face recognition also leverage advanced algorithms for dimensionality reduction and feature selection to enhance computational efficiency and reduce the complexity of the recognition process. Additionally, they often integrate pre-processing steps like normalization and alignment to improve the consistency and alignment of facial images, ensuring reliable feature extraction across different poses, illuminations, and facial expressions. Furthermore, holistic approaches can facilitate the development of more robust and invariant representations of faces, enabling effective recognition performance even in challenging conditions such as variations in lighting, occlusions, and partial face obstructions. Techniques like principal component analysis (PCA), independent component analysis (ICA), and deep learning-based approaches are commonly employed for holistic feature extraction. Holistic representations hence sensitive to local deformation, rotation, and spatial scale, thus prone to facial variations such as expressions and poses.

1.3 Human Face Recognition Employing Local Feature Descriptor

In the field of digital image processing, classification is a huge task to perform. When it comes to classifying human faces, we call it human face recognition. If we go by extracting local features that would eventually make up a feature map for the given facial image, finding a local feature descriptor is a necessary task. Em-

ploying that local feature descriptor feature map images are generated and then machine learning algorithms are used to perform classification. Often as the feature map images are used for face recognition while using machine learning models rather than raw images, it provides better recognition or accuracy percentage. It happens because the feature map images provide more distinctive if possibly noise-free image sets to perform classification tasks. So, first is preparing local feature descriptors to extract local features to build a feature map. And then we proceed to classify images using those feature maps using machine learning algorithms.

CHAPTER 2

Literature Survey

Human face recognition is an actively researched topic because of its importance in security identification, smartphone, biometric-security, security monitoring, access control, etc. Francis Dalton formally proposed [3] the first face identification system in a paper entitled "Personal Identification and Description"[4]. Face recognition algorithms have seen significant application and improvement with the progress of technology. Different techniques have been applied with varying degrees of success. Many pieces of literature pave away the field of research of face recognition. They have contributed from understanding new feature extraction methodologies to having clear ideas about several challenges that one has to face for face recognition.

2.1 Challenges for Face Recognition

Human face recognition (HFR) is regarded as one of the most challenging and complex tasks in the area of pattern recognition because of its constraints caused by changes in the appearance of facial images, as the face acquisition process can undergo a wide range of variations [5]. The challenges that human face recognition poses, are illumination variations, pose variations, variations in expression, rotation variations, accumulation of noise, occlusion, ageing, scale variations, and rotation [5]. This section further describes these challenges in a way to better understand the impact of these challenges and the databases that could give those challenges to handle. We can clearly define those challenges by understanding several pieces of literature regarding the same.

2.1.1 Illumination Variations

Illumination is the term used for lighting variations that can make a face image look different. Illumination remains a challenge to robust human face recognition

systems where illumination-invariant features have to be extracted [6]. Though humans with the naked eye can recognize a face image regardless of illumination variations, a human face recognition system often fails to recognize those face images and misclassifies them due to a poor understanding of the features that could be beneficial in such a situation. Thus having a better understanding of features and extracting those illumination-invariant features would make a human face recognition system robust to illumination changes.

Several publicly available datasets of human face images provide researchers the face images with illumination variations to work on. The most famous datasets related to this are Yale Face Database, Yale B Face Database, Extend Yale B Face Database, AR face dataset, CMU PIE dataset, CAS-PEAL dataset et al.

2.1.2 Pose Variations

pose variation is the term used to describe the variations in head orientation and position. In unconstrained environments, face images can appear in various angles due to rotation or up-down movement of faces. This could be an easy task for humans to recognize a face image regardless of the pose variations, but certainly, it is a challenging task for human face recognition systems [6]. Thus extracting pose-invariant features from face images would make a human face recognition system robust to any kind of pose changes .

Several publicly available datasets such as ORL dataset, CMU Multi-PIE dataset, CMU PIE Dataset, Extended Yale B face database, FERET face dataset, Faces95 dataset et al. provide researchers human face images with pose variation to work on.

2.1.3 Expression Variations

Showing different expressions is a basic human nature. Human faces exhibit several expressions. These expressions depict the emotions and mental state of a particular person. Expressions change facial appearance and geometry in many ways of a particular person and that poses a significant challenge to recognize facial images with different expressions [6]. For humans, it may be an easy task but human facial recognition systems often misclassify facial images with variations in expression. Extracting good features of a facial image that would be expression-invariant is a challenging task. Finding out better features regarding this, would make the face recognition system much more robust to any kind of expression changes.

Several publicly available datasets of human face images provide researchers the face images with expression variations to work on. The most famous datasets re-

lated to this are Yale Face Database, AR face dataset, CMU PIE dataset, FERET dataset, Faces94 dataset, Faces95 dataset et al.

2.1.4 Rotation Variations

Rotation variations concerning the human face signify how tilted or turned the face is. It happens when a face is rotated around an axis. In a face image, the face could be rotated around any axis. Moreover, it may cause, only one specific side of the face is only visible. Even for the naked human eye, this challenge becomes sometimes too hard to classify. Well-distinguishable features need to be extracted to tackle this challenge for a human face recognition machine.

FEI dataset and UMist dataset give plenty of good face images of rotation variations to work on. Researchers may use this dataset to gain confidence in their research work regarding solving this challenge.

2.1.5 Accumulation of Noise

Noise in a facial image signifies hazy-burred images. Due to different settings of the image-capturing device or different environmental conditions or movement of the particular person while capturing using the image-capturing device can accumulate noise in facial image. To recognize a person from hazy-blurred images is not even an easy task for humans. Human face recognition systems often misclassify those facial images. In this kind of situation feature engineering plays a vital role and extracting good features of a facial image could make better classification in this kind of situation.

Publicly available Faces95 and Faces94 datasets provide some hazy-blurred images of the subject to do classification tasks. These two datasets can be challenging ones to work on.

2.1.6 Occlusion

Occlusion can be described as the obstruction or blockage of a specific part or segment of an image or entity. Occlusion of a face image can be specified as when a part of the face is blocked or covered by some entity either intentionally or unintentionally [6]. Nowadays masks and sunglasses are the most prominent examples of occlusion of face images. In real-world situations, it is also very crucial to work on these kinds of face images and to recognize persons from these face images with occlusion. In several situations, even humans fail to recognize a known person due to occlusion. Human face recognition systems many a time fail to classify images correctly due to occlusion. To solve this challenge, it is crucial

to find more distinguishable features and to extract those from face images to classify them more accurately, thus making a robust face recognition system for the occlusion problem.

Publicly available ROF, COFW, and AR datasets provide very promising images with occlusion to work on classification tasks. These datasets come in much more handy to researchers to experiment.

2.1.7 Ageing

As time passes by, the human face changes its appearance and texture. To compare or recognize faces from two different time frames, and if the time frames are large enough to show aging, it becomes a challenging task for machines and even for humans [6].To tackle this challenge, well-defined features need to be extracted that would compare faces from two very different time frames.

Publicly available FERET, and CAS-PEAL datasets provide face images to tackle this Ageing challenge for classification or recognition purposes.

2.1.8 Scale Variations

Scale variations mean a face could appear smaller and larger in two different images. Depending upon the distance and other factors between the subject and image image-taking instrument, a face can appear in different sizes. It is sometimes even hard for humans to classify those images. For a machine to perform human face recognition, certainly this scale variation is a challenge.

Publicly available Faces95, and GT datasets provide very promising images with scale variations to work on classification tasks. These datasets come in much more handy to researchers to experiment

In real-world situations, it is not just one of these challenges to work for a face recognition system, but a state-of-the-art face recognition system has to solve all these challenges as far as possible and has to provide a high accurate recognition rate for face images in any unconstrained environments. All of these challenges, alone and rather together, make HFR a field of research much more challenging and promising.

2.2 Standard Databases for Face Recognition

To evaluate the performance of human face recognition methodologies or systems, several databases are widely used. These databases are publicly available to work on and evaluate the performance of human face recognition methodologies

[5]. Here we are going to discuss some widely used face databases available to work on. In Table 2.1, we provide a short description of some widely used databases for benchmarking face recognition tasks.

Table 2.1: Widely Used Face Databases

Database Name	About
ORL / AT&T	There are a total of 400 facial images (40 subjects and 10 images each) to work on . The images are in grayscale. The images have varying poses and expressions.
Yale	There are a total of 165 facial images (15 subjects and 11 images each) to work on. The images are in grayscale. The images have varying illuminations and expressions.
Yale B	There are a total of 640 facial images (10 subjects and 64 images each) to work on. The images are in grayscale. The images have varying illuminations.
Extended Yale B	There are a total of 2,432 facial images (38 subjects and 64 images each) to work on. The images are in grayscale. The images have varying illuminations.
CMU PIE	There are a total of 41,368 colour facial images of 68 subjects to work on. The images have varying poses, illuminations, and expressions.
CMU Multi-PIE	There are more than 750,000 colour facial images of 337 subjects to work on. The images have varying poses, illuminations, and expressions.
AR	There are more than 4,000 colour facial images of 126 subjects to work on. The images have varying illuminations, expressions, and occlusion.
LFW	There are 13233 colour facial images of 5749 subjects to work on. And 1680 subjects have two or more images. The images have varying illuminations, expressions, poses, occlusion and scales.
LFW-a	It is an aligned version of LFW database.
JAFFE	There are a total of 213 facial images of 10 different Japanese female subjects to work on. The images are in grayscale. The images have varying expressions.

Database Name	About
FERET	The dataset consists of five distinct subsets, which are comprised of a gallery set denoted as Fa and four probe sets referred to as Fb, Fc, DupI, and DupII. The gallery set encompasses a solitary frontal image for each of the 1196 subjects. Furthermore, the DupI and DupII sets encompass 722 and 234 face images respectively. Due to the challenges of Ageing, it becomes a hard challenging dataset.
Faces94	There are a total of 3060 facial images (153 subjects and 20 images each) to work on. The images have varying expressions.
Faces95	There are a total of 1440 facial images (72 subjects and 20 images each) to work on. The images have varying scales, expressions, and illuminations.
Faces96	There are a total of 3040 facial images (152 subjects and 20 images each) to work on. The images have varying scales, expressions, and illuminations.
KDEF	There are a total of 4900 facial images 0f 70 subjects to work on. The images have varying expressions.
MIT-CBCL	There are a total of 2000 facial images (10 subjects and 200 images each) to work on. The images have varying illuminations and poses.
FEI	There are a total of 2800 facial images (200 subjects and 14 images each) to work on. The images have varying rotations, poses, and expressions.
SCface	There are a total of 4160 facial images (in visible and infrared spectrum) of 130 subjects. The images have varying illuminations, poses, noise and scales.
UMist	There are a total of 564 facial images of 20 subjects to work on. The images are in greyscale. The images have varying poses and rotations.
FRGC	It consists of 50,000 recordings divided into training and validation partitions. The images have varying illumination and expressions.
O2FN mobile	There are a total of 2000 facial images of 50 subjects to work on.The images have varying illumination and rotations.
CASIA-Face-5.0	There are a total of 2500 colour facial images (500 subjects and 5 images each) to work on. The images have varying illuminations, poses, expressions, and scales.

Database Name	About
CUFSF	It contains 1,194 different subjects' sketch-photo image pairs from the FERET database.
CAS-PEAL	There are 99,594 colour facial images of 1040 subjects to work on. The images have varying illuminations, expressions, poses, and accessories.
GT	There are a total of 750 colour facial images (50 subjects and 15 images each) to work on. The images have varying illuminations, expressions, poses, and scales.
Caltech 1999	There are a total of 447 facial images 0f 26 subjects to work on. The images have varying expressions, illuminations, and occlusions.

2.3 Trends in Feature Extraction Methodologies

Feature Engineering methodologies for human face recognition signify different feature extraction ways and finding out better feature maps of facial images that could provide robust human face recognition systems to any unconstrained environment. Feature extraction methodologies could be broadly classified into two categories - holistic feature extraction and local feature extraction.

2.3.1 Overview on Local Feature Extraction Methodologies

Local feature extraction methodologies extract features from small regions of a facial image independently. Thus the methodologies allow to capture of fine-grained details which are crucial for discriminating between individuals. It can be stated that Local Binary Pattern (LBP) [7] for extracting local features for human face recognition has been proven a crucial methodology that has influenced many more methodologies and given this segment of feature extraction a new horizon. Over the years many new methodologies have been come forward due to researchers. Here I provide an overview of several local feature extraction methodologies over the years that have come forward in several pieces of literature.

Ojala [7] shows that 'uniform' Local Binary Patterns (LBP) can be used as good discriminators for multiresolution gray-scale and rotation-invariant texture classification that eventually can be used for human face recognition. However, there are several limitations for LBP to unconstrained environments.

Liao et al.[8] propose multi-scale block local binary pattern (MB-LBP) which is a novel variation of LBP. In the MB-LBP method, the computation is performed by utilizing the mean values of block subregions as opposed to analyzing individ-

ual pixels. This approach offers increased resilience compared to the traditional LBP technique. Furthermore, MB-LBP encodes not only the microstructures but also the macrostructures of image patterns, thus yielding a more comprehensive representation of the image in comparison to the fundamental LBP operator. The efficient computation of MB-LBP can be achieved through the utilization of integral images.

Ahonen et al.[9] propose a new methodology for feature extraction for face recognition using the operator called Local Phase Quantization (LPQ) [10]. This methodology helps to extract those features that would be invariant to blurred noisy face images and illumination variation using LPQ. In face image analysis, histograms of LPQ labels calculated within local regions serve as a face descriptor, akin to the commonly utilized Local Binary Pattern (LBP) approach for describing face images. It takes into consideration the challenges of variations in noise (blurring), illumination, pose, and expression for face recognition tasks.

Zhang et al. [11] propose an innovative descriptor known as Local Derivative Pattern (LDP) for human face recognition, which captures high-order local patterns and shows it captures more discriminative information than LBP. The Local Derivative Pattern (LDP) serves as a comprehensive framework for encoding directional pattern features by utilizing localized variations in derivatives. To accurately capture the intricacies of local derivative direction variations, the nth-order LDP is introduced, which effectively encodes the (n-1)th order of such variations. The LDP templates excel at extracting high-order local information by effectively encoding distinct spatial relationships present in a specified local region. It takes into consideration the challenges of variations in illumination, pose, expression, ageing, and Occlusion for face recognition tasks.

Ahonen et al. [12] propose a novel rotation invariant image descriptor Local Binary Pattern Histogram Fourier (LBP-HF) features that can be used for facial images also. This descriptor is invariant to the rotational angle of the input texture image. It takes into consideration the challenges of variations in rotation, illumination, pose, and expression for face recognition tasks.

Vu et al. [13] propose a novel feature descriptor Patterns of Oriented Edge Magnitudes (POEM). The POEM descriptor is a spatial multi-resolution descriptor that is oriented and captures intricate details pertaining to the original image. The POEM descriptor is constructed based on multi-scale self-similarity, which in turn enhances its ability to withstand variations in the surroundings. Additionally, the POEM descriptor exhibits low complexity, rendering it suitable for real-time applications. Specifically, the POEM feature is constructed for each individual pixel by applying a self-similarity based structure to the oriented magnitudes. These magnitudes are determined by accumulating a local histogram of gradient

orientations across all pixels of the image cells, with the considered pixel serving as the center. It takes into consideration the challenges of variations in illumination, pose, occlusion, and expression for face recognition tasks.

Gong et al. [14] propose a new methodology for face recognition using Weber Local Descriptor (WLD) [15] feature. WLD is derived from Weber's Law, wherein it is postulated that the recognition of a stimulus change is contingent upon the magnitude of the ratio between said change and the original stimulus. In adherence to Weber's Law, it is proposed to utilize two ratios to delineate the textural characteristics of images, thereby facilitating the acquisition of differential excitation and orientation in WLD. Comprised of both a differential excitation component and an orientation component, WLD encompasses a plethora of localized texture information. In this approach, the authors initially partition face images into a series of sub-regions, subsequently extracting their respective WLD features. Additionally, the authors incorporate the Sobel descriptor to procure the orientation component. It takes into consideration the challenges of variations in illumination, pose, and expression for face recognition tasks.

Reza et al. [16] propose a face recognition method that exploits edge-based features of faces. Though there are many edge-based features, the author takes two edge-based features- Edge orientation features and Edge density features. Edges contain not only significant data regarding variations in faces but also present a straightforward approach for analysis. An edge indicates the boundary that separates two distinct classes of intensity. Using the edge orientation features, both global and local characteristics of a face are gathered. Edge density signifies edge focus on a special area. Edge density is calculated by dividing total edges by total area. The approach is also robust in challenging identification scenarios, including variations in facial orientation, eyewear, facial hair, hairstyle, and lighting conditions.

Hussain et al. [17] propose a new feature descriptor called Local Quantized Patterns (LQP) for human face recognition. LQP represents an expanded form of local patterns that uses wider local neighborhoods and/or improved quantization through domain-adaptive vector quantization. It additionally integrates a lookup table to enable local pattern features to incorporate a significantly larger number of pixels and/or quantization levels, all the while upholding simplicity and computational efficiency. The construction of these tables occurs in an offline manner, wherein all the codes are mapped to their nearest cluster centers, thus ensuring swift vector quantization without any additional overhead during runtime. LQP takes into consideration the challenges of illumination, expression variations, and ageing to solve.

Yuan et al.[18] propose a local feature extraction methodology using the combi-

nation of local binary pattern (LBP) and local phase quantization (LPQ) for face recognition. The images are divided into local blocks and from those bloks LBP histograms and LPQ histograms are created. To find out the final feature vector LBP and LPQ histograms are concatenated. It takes into consideration the challenges of variations in illumination, occlusion, and expression for face recognition tasks.

Nguyen et al. [19] propose Elliptical Local Binary Patterns (ELBP), a novel variant of LBP, for human face recognition. In the research, horizontal and vertical ellipse patterns are employed in ELBP to effectively capture the micro facial features presented in face images, both in the horizontal and vertical orientations. Moreover, ELBP is employed in the domain of face recognition, wherein a dimension reduction process is carried out using Whitened Principal Component Analysis (WPCA). It takes into consideration the challenges of variations in illumination, occlusion, and expression for face recognition tasks.

Ren et al. [20] propose Relaxed Local Ternary Pattern (RLTP) for human face recognition that solves the challenges faced by LBP to an extent. It can be easily observable that noise can overwhelm small pixel differences easily. To solve this issue, the authors introduce the concept of an 'uncertain' state to encode the small pixel differences. If the difference in pixel values falls within a threshold, it is transformed into an uncertain state. In this state, the exact sign and magnitude of the pixel difference cannot be determined precisely due to the potential influence of noise. A ternary code with three possible values is generated by representing large pixel differences with two distinct states and small pixel differences with a separate uncertain state. The pixel difference that belongs to the uncertain state is then evenly divided between the two distinct states in order to reflect the equal likelihood of the small pixel difference being positive or negative. In this manner, the code with three values is converted back into a code with two values. RLTP results in only one LBP histogram, whereas LTP results in two LBP histograms. It takes into consideration the challenges of variations in illumination, rotation, and expression for face recognition tasks.

Fan et al. [21] propose a new local feature descriptor called Local Vector Pattern (LVP) in high-order derivative space for face recognition. The vector of each pixel is used to construct a representation by calculating the values between the referenced pixel and the adjacent pixels with various distances from different directions. This representation generates a one-dimensional structure of micropatterns for the referenced pixel. By using pairwise direction vectors for each pixel, the LVP reduces the length of features through a comparative space transform. This transform encodes the various spatial relationships between the referenced pixel and its neighboring pixels. Moreover, the concatenation of LVPs gives more

distinctive features. The authors mention that recognition accuracy drops in the high-order derivative space due to influence of noise. It takes into consideration the challenges of variations in illumination, pose, expression, ageing, and occlusion for face recognition tasks.

Nguyen et al. [22] propose local patterns of gradients (LPOGs), a novel feature extraction methodology, for human face recognition. LPOG utilizes the technique of block-wise elliptical local binary patterns (BELBP) and local phase quantization (LPQ) operators on gradient images for the purpose of capturing local texture patterns in order to construct a feature vector for a face image. Two directional gradient images are computed from a single input image. The BELBP and LPQ operators are then applied separately to each gradient image, resulting in the generation of local pattern images. The histogram sequences of the non-overlapping subregions of these local pattern images are subsequently concatenated to form the LPOG vector for the given image. It takes into consideration the challenges of variations in illumination, pose, expression, noise, and occlusion for face recognition tasks.

Chen et al. [23] propose complete robust maximum margin criterion (CRMMC), a novel feature extraction methodology, for human face recognition. To mitigate potential correlations among features, we integrate an orthonormality restriction into CRMMC. In order to effectively harness the discriminatory information encompassed within the entirety of the feature space, we divide CRMMC into two distinct, but interrelated, orthogonal complementary subspaces, from which we extract the discriminant features. In this manner, CRMMC is capable of progressively extracting features through the resolution of two interconnected constrained optimization problems. It takes into consideration the challenges of variations in illumination, pose, expression, and rotation for face recognition tasks.

In the paper [24] by Roy et al. concept of physics has been infused to get a novel local feature descriptor called local-gravity-face (LG-face) which is illumiantion-invatiant. LG-face incorporates the concept called the local gravitational force angle (LGFA). The LGFA represents the gravitational force direction exerted by the central pixel on neighboring pixels within a local neighborhood. Theoretical examination indicates that LGFA is an illumination-invariant feature by solely considering the reflectance component of the local texture impact from nearby pixels. Also, there is no requirement to estimate the illumination components (L) as they are automatically disregarded. Additionally, it maintains the integrity of edge information. It takes into consideration the challenges of variations in illumination, pose, expression, noise, and occlusion for face recognition tasks.

Yang et al. [25] introduce adaptive Local Ternary Pattern (ALTP) local feature descriptor enabling automatic selection of the threshold for LTP, which could not

be done easily for LBP and LTP. It is inspired by Weber's law. Further in the same paper, the authors introduce a center-symmetric adaptive local ternary pattern (CS-ALTP), a local feature descriptor, for human face recognition. CS-ALTP effectively divides CS-LBP into two distinct channels, namely positive and negative channels, to encompass a greater amount of discriminative information. It takes into consideration the challenges of variations in illumination, pose, expression, and noise for face recognition tasks.

Liu et al. [26] propose a local feature extraction methodology using LBP-like local descriptors and those descriptors are based on Angular Differences and Radial Differences. The authors extend the concept of LBP and introduce six kinds of local binary patterns- LBP_S, LBP_M, ADLBP_S, ADLBP_M, RDLBP_S, and RDLBP_M. Using the LBP concept, computing pixel difference including magnitude and sign between a central pixel and its neighbors, LBP_S denotes the operator derived from the sign component and LBP_M denotes the operator derived from the magnitude component. ADLBP_S denotes the sign component of the Angular Difference-based Local Binary Pattern (ADLBP) and ADLBP_M denotes the magnitude component of ADLBP. RDLBP_S denotes the sign component of the Radial Difference-based Local Binary Pattern (RDLBP) and RDLBP_M denotes the magnitude component of RDLBP. For these six feature extractors, histograms obtained from face images are concatenated to pursue face detection. This extended LBP-based feature extractor shows robustness to lighting variations, pose variations, expression variations, and ageing.

Wang et al. [27] propose local contourlet combined patterns (LCCP), a novel local feature descriptor, for face recognition. It is based on Non-Subsampled Contourlet Transform (NSCT). NSCT serves as a powerful analytical instrument capable of extracting image data at various levels of detail, diverse angles of orientation, and distinct frequency ranges. Thanks to its ability to analyze images at multiple scales and in multiple directions, as well as its anisotropy and shift-invariance, NSCT offers a comprehensive approach to image analysis. This technique not only enables multiresolution analysis, but also effectively captures the geometric structure and directional information present in images. These fundamental characteristics of NSCT contribute to its capacity for efficiently representing facial images. It takes into consideration the challenges of variations in illumination and expression for face recognition tasks.

Huang et al. [28] propose a feature descriptor based on image gradients. It is termed by the authors as Binary Gradient Patterns (BGP). To reveal local patterns in the gradient space, image gradient vectors are calculated across various orientations and translated into a collection of binary codewords. They are able to detect micro-oriented edges and retain strong local orientation, structures,

and textures. The authors also furthermore propose in the same paper a refined descriptor created by constructing BGP patterns based on a set of orientations derived from the magnitudes of image gradients, referred to as BGPM. This approach serves to strengthen the emphasis on both locality and orientation in the representation. It takes into consideration the challenges of variations in illumination, pose, expression, ageing, occlusion, and noise for face recognition tasks.

Chakraborty et al. [29] propose a local directional gradient pattern (LDGP) which is a local feature descriptor in high-order derivative space. The computation of the LDGP involves a 1D micropattern that encodes the connections between the higher-order derivatives of the reference pixel in four specific directions. The descriptor determines the correlation between the high-order derivatives of the referenced pixel in four distinct directions to calculate the micropattern that corresponds to the local feature. This proposed descriptor significantly decreases the length of the micropattern, resulting in reduced extraction and matching times, while still maintaining the recognition rate. In this paper, authors emphasize better recognition with less time complexity. It takes into consideration the challenges of variations in illumination, pose, expression, and noise for face recognition tasks.

Kar et al.[30] propose a novel method called local centre of mass face (LCMF) to extract local features for face recognition that would be invariant to illumination. In this approach, the luminance component (L) is automatically excluded, eliminating the necessity to calculate illumination. The fundamental basis of this approach relies on determining the angle of gradient between the centre of mass and the centre pixel within the local neighbourhood. Since the center of mass is a fundamental characteristic of mass distribution, its slope with the center pixel in the local neighbourhood endows the ability to preserve local edges. The angle of the slope, derived from the Center of Mass with the central pixel of the neighborhood, serves as a feature vector. This feature vector is aligned from the darkest part of the local area to the brightest part of the local area, as the center of mass is consistently positioned towards the brighter side of a mass dispersion and therefore encodes the orientation of the edges. It takes into consideration the challenges of variations in illumination, pose, expression, and occlusion for face recognition tasks.

Tran et al. [31] propose local intensity area descriptor (LIAD), a local texture descriptor, for face recognition. Each facial image is partitioned into smaller segments, from which LIAD histograms are derived and combined into a singular feature vector that signifies the facial image. In this case, each individual pixel within an image is represented by a numerical value obtained through the utilization of a trapezoidal numerical integration formula, which takes into account the intensity values of its surrounding eight pixels. When the outcome of the basic

LIAD is rounded down to the closest whole number that is lower than the initial value, it is referred to as LIADdown. Conversely, LIADup is employed when the opposite scenario occurs. It takes into consideration the challenges of variations in illumination, pose, expression, noise, scale and rotation for face recognition tasks.

Ramalingam et al. [32] propose a novel local feature extraction technique called Enhanced Local Binary Pattern (EnLBP) for robust human face recognition. This technique follows the basic principle of the LBP. The Enhanced Local Binary Pattern (EnLBP) method involves dividing an image into 3 x 3 regions. Within each region, the neighbouring pixels are summed with the center pixel, and the mean value is calculated to capture local information. The resulting mean values are stored as a vector for each region. To reduce dimensionality of the image, the center pixel is compared to its 8 neighbouring values in the resulting image. If the center pixel is smaller than its neighbours, the corresponding positions are set to 1; otherwise, they are set to 0. The resulting 8-bit binary values are arranged in a sequence, converted into a decimal value, and then substituted for the center position in each region. It takes into consideration the challenges of variations in illumination, pose, and expression for face recognition tasks.

Pillai et al. [33] propose local diagonal extrema number pattern (LDENP), a local feature descriptor, for human face recognition This methodology generates a concise representation of facial features that is acquired through the encoding of the face image's directional information. Unlike using masks, the proposed descriptor utilizes first-order derivatives for image encoding. The descriptor exclusively considers the values of local diagonal pixels rather than those of local neighboring pixels. As a result, it decreases the dimensionality of the feature descriptor. The descriptor in question computes micro-patterns by considering the number of extrema in the first-order local diagonal pattern. It takes into consideration the challenges of variations in illumination, pose, expression, ageing, occlusion, and noise for face recognition tasks.

Zhou et al. [34] propose a local feature extractor that infuses the information of the center pixel into the Center-Symmetric Local Binary Pattern (CS-LBP) to make a robust feature descriptor for human face recognition. It is called CS-LBP/Center. Furthermore, the authors introduce Weighted CS-LBP/Center or WCS-LBP/Center. The initial CS-LBP technique captures the texture without distinguishing each individual block. However, it is worth noting that certain blocks may possess more valuable discriminatory information compared to others. To leverage this aspect, when combining the histograms of CS-LBP/Center for each block, a weight is assigned to each block. Here comes WCS-LBP/Center into the picture. A high weight is assigned to blocks that exhibit rich texture and contain a greater amount of information, while a low weight is assigned to

blocks with poor texture. It takes into consideration the challenges of variations in illumination, pose, and expression for face recognition tasks.

Chakraborty et al. [35] propose a local feature descriptor called Centre Symmetric Quadruple Pattern (CSQP). It exhibits structural symmetry and captures facial asymmetry within a quadruple space encoding. To effectively handle the significant fluctuations in environmental conditions in the real world, it is necessary for the local neighborhood of these descriptors to expand its capacity to include a greater number of pixels. This expansion subsequently leads to an increase in the length of the descriptor's features. CSQP successfully addresses this limitation of the existing descriptors by encoding the pixels found within the local neighborhood using a quadruple-space approach. In doing so, CSQP is able to generate an eight bit pattern based on the analysis of 16 pixels within the local neighborhood. It increases recognition accuracy of faces in unconstrained environments. It takes into consideration the challenges of variations in illumination, pose, expression, scale, and rotqtion for face recognition tasks.

Hamdan et al. [36] propose to use moment-based Angular Radial Transform (ART) that extracts local features for human face recognition. The ART projection basis consists of the multiplication of a complex function with a cosine function. The resulting moments are also in complex form. These ART bases are orthogonal when considering the unit circle. It takes into consideration the challenges of variations in illumination, pose, and expression for face recognition tasks.

Bhattacharjee et al. [37] propose Pattern of Local Gravitational Force (PLGF), a novel local image descriptor, for human face recognition that is inspired from the concept of physics. PLGF represents a composite descriptor that amalgamates two distinct feature constituents: one being the Pattern of Local Gravitational Force Magnitude (PLGFM), and the other being Pattern of Local Gravitational Force Angle (PLGFA). The magnitude of the gravitational force in the local area is a quantification of the local irregular texture information, which encompasses edges, lines, and corners. In order to depict the intricate patterns of the microstructure present in the local irregular texture information of the gravitational force magnitude, a binary coding scheme resembling a local LBP is employed, denoted as PLGFM. The illumination-invariant local geometric structure of an image is represented by the direction of the local gravitational force. In terms of sensitivity, direction information surpasses magnitude information. PLGFA is responsible for encoding the local gravitational force angle exerted by the center pixel on all other pixels within a local neighborhood. At the end, to form the PLGF descriptor, the histograms of the two components are combined through concatenation. It takes into consideration the challenges of variations in illumination, pose, expression,

occlusion, and noise for face recognition tasks.

Bhattacharya et al. [38] propose a new local feature descriptor that is inspired by the concept of Van der Waal forces and it is called Local Force Pattern (LFP). LFP features efficiently capture the directional characteristics of facial textures, resulting in a more distinctive code. The authors employ the concept of Van der Waal's force between pixel pairs to compute the structure of each micro-pattern, extracting valuable directional information. LFP, a 2D descriptor, considers pixel-to-pixel relations in all directions, including diagonals, enhancing the ability to discriminate images with similar structural patterns but varying intensity transitions. The face image is partitioned into 'cells', small pixel boxes from which LFP features are extracted. These cells then form 'blocks', and the resulting forces are vector-summed to obtain the final feature vector. Subsequently, the authors concatenate these vectors to create a comprehensive face descriptor.

Kagawade et al. [39] propose a new method that exploits local directional gradient and the method is called multi-directional local gradient descriptor (MLGD). The method leverages the inherent similarity of facial images in small blocks. The computed weighted gradient attributes of facial depictions in distinct orientations and areas are established on the correlation amongst pixel components. These features, known as MLGD, effectively capture edge information by combining gradients in different directions. The accumulation of gradient changes in multiple directions enhances the extraction of edge and line details, resulting in improved discrimination within the face image class and among different classes of face images. It takes into consideration the challenges of variations in illumination, expression, and occlusion for face recognition tasks.

Yang et al. [40] propose a local feature descriptor called local multiple patterns (LMP) which is based on Weber's law. For the feature descriptor, Weber's ratio has been adapted to incorporate the direction of change. Consequently, the adjusted Weber's ratio is discretized into multiple intervals in order to produce numerous feature maps that capture various types of changes. Also, LMP combines the histograms of non-overlapping regions in feature maps to represent the image. To enhance the discriminative capability of LMP, the authors propose an enhanced version of LMP called MB-LMP. MB-LMP is designed to encompass both the micro and macro structures by employing a multi-scale block strategy. It takes into consideration the challenges of variations in illumination, pose, expression, occlusion, and noise for face recognition tasks.

Khanbebin et al. [41] propose arithmetic coded LBP (ACLBP), a new variant of LBP, for local feature extraction. It uses an arithmetic coding process during LBP calculation instead of applying original thresholds. The method proposed involves altering the LBP technique by employing distinct thresholds to compute

pixel differences. Through this approach, the authors implemented a genetic-based feature fusion technique, amalgamating LBP with Histogram of Oriented Gradients (HOG) and ACLBP. A unique aspect of the LBP calculation within each block involves the utilization of threshold values other than the central pixel. These alternative thresholds encompass neighboring pixels, the average, median, and maximum intensities of pixels within each block. It takes into consideration the challenges of variations in illumination, pose, expression, and scale for face recognition tasks.

Karanwal et al. [42] introduce neighborhood and center diference-based-LBP (NCDB-LBP) which is inspired by the foundational concepts of Local Binary Pattern (LBP). In the NCDB-LBP method, four labeled functions are introduced to capture robust features within a 3×3 pixel window. Two first-order derivatives are calculated at each position within the neighborhood. The first derivative is computed between the adjacent neighborhood and the current neighborhood, while the second derivative is computed between the center pixel and the current neighborhood. To account for all possible directions, the proposed NCDB-LBP method is implemented both in an anticlockwise (ac) and clockwise (c) manner within the 3×3 pixel window. These implementations are referred to as NCDB-LBPac and NCDB-LBPc descriptors, respectively. The resulting histograms from each subregion are subsequently merged to generate the complete feature size of the NCDB-LBPac and NCDB-LBPc descriptors. It takes into consideration the challenges of variations in illumination, pose, expression, and scale for face recognition tasks.

Khanbebin et al. [43] propose a new local feature called Distance Ratio Local Binary Pattern (DR-LBP). The authors consider three distances in the local neighbourhood. The first one would be the distance between the central pixel and the desired pixel. The remaining two distances correspond to the distances between the central pixel and the two nearest pixels to the specified pixel. The distance here signifies nothing but pixel value difference. From these three values give DR-LBP. It takes into consideration the challenges of variations in illumination, pose, and expression for face recognition tasks.

Karanwal et al. [44] propose orthogonal difference-Local binary pattern (OD-LBP), a local feature descriptor, for human face analysis. In OD-LBP, the process begins with computing three gray level differences for each orthogonal position. This is achieved by subtracting the values of the two closest neighborhoods and the center pixel from the corresponding orthogonal value. Subsequently, these gray level differences are compared with the outcome generated through a new comparison method. Finally, a row vector is formed by merging the binary patterns derived from both orthogonal groups. It takes into consideration the challenges of

variations in illumination, pose, expression, and scale for face recognition tasks.

Karanwal et al. [45] propose another local feature descriptor which is also a variant of LBP called Graph Based Structure Binary Pattern (GBSBP) for human face recognition. Here 3×5 and 5×3 pixel windows are used and they create four graphs. These four graphs give edge differences from 2 pixels that result in a histogram feature vector. It takes into consideration the challenges of variations in illumination, pose, expression, and scale for face recognition tasks.

Najmabadi et al. [46] propose a novel local feature extraction methodology for face recognition called Local Symmetric Directional Pattern (LSDP). LSDP algorithm allocates a maximum of four bits to represent the code of each individual image pixel. To extract the directional patterns from the facial textures, the authors leverage the directional information provided by the two most prominent edge responses. To capture the gradient information of the local neighborhood, LSDP employs the magnitude and direction of the edge responses. To extract the directional patterns from the facial textures, the authors leverage the directional information provided by the two most prominent edge responses. It takes into consideration the challenges of variations in illumination, pose, and expression for face recognition tasks.

Truong et al. [47] propose Weighted Statistical Binary Pattern (WSBP), new feature descriptors for local feature extraction for human face recognition. The descriptors leverage the linear topology and various orientations. Initially, the input image is partitioned into mean and variance moments. A unique variance moment, capturing distinctive facial attributes is created by extracting the root k-th. Subsequently, when constructing Sign and Magnitude components along four diverse directions using the mean moment, a weighting technique based on the new variance is employed for each component. The process involves concatenating the weighted histograms of Sign and Magnitude components to form a novel histogram of Complementary LBP across various directions. It takes into consideration the challenges of variations in illumination, pose, expression, occlusion ,and noise for face recognition tasks.

Karanwal [48] proposes a fusion-based local feature descriptor called Improved Local Descriptor (ILD). Four local feature descriptors are fused to make ILD. They are Local binary pattern (LBP) [7], Elliptical local binary pattern (ELBP) [19], Median binary pattern (MBP) [49], Local phase quantization (LPQ) [50]. After that all the global histograms are consolidated into a single histogram referred to as ILD. This integration process combines the information from various sources into a unified representation. ILD optimally leverages LBP, ELBP, MBP, and LPQ. LBP effectively gathers localized information within a singular scale. ELBP, on the other hand, captures resilient characteristics in both horizontal and

vertical orientations through the utilization of lower and higher scale pixels. By reducing image noise, MBP contributes to the enhancement of the overall quality. Additionally, LPQ quantizes the frequency components, further bolstering the effectiveness of ILD in the face recognition process. Consequently, ILD demonstrates its robustness in addressing the challenges for face recognition posed by unconstrained environments. It takes into consideration the challenges of variations in illumination, pose, expression, and scale for face recognition tasks.

Karanwal et al. [51] recently propose a novel local feature descriptor known as Triangle And Orthogonal LBP (TAO-LBP), which integrates two LBP variants, namely Triangle LBP (TLBP) and Orthogonal LBP (OLBP), in order to account for pose and expression variations. It takes into consideration the challenges of variations in illumination, pose, expression, and scale for face recognition tasks.

We can say by now that through the course of time many local feature extractors or descriptors or methodologies that incorporate some well-known pattern extractor have been created or proposed and in the future, many like these will come for better human face recognition to handle challenges regarding face recognition. The researchers always try to give better accuracy and recognition rates over several datasets. But time complexity is making sense in the modern era where a face recognition system has to provide better results within a limited time.

2.3.2 Overview on Holistic Feature Extraction Methodologies

Holistic feature extraction methodologies often take whole facial images as one single unit and try to extract features from there. Unlike local feature extraction processes where a whole image is divided into overlapping or non-overlapping small regions to gather features from local neighborhoods, holistic feature extraction processes try to gather features from a complete facial image. Holistic facial recognition employs a comprehensive approach by incorporating global facial information to facilitate the process of face recognition [52]. The global facial information is fundamentally depicted through a limited set of characteristics, which are directly derived from the pixel data of facial images. Holistic methodologies utilize the entire facial structure as input. Within these methodologies, each facial image is transformed into a high-dimensional vector, representing a signal, by amalgamating the values of all pixels forming the facial composition.

Fisher's Linear Discriminant, also known as Linear Discriminant Analysis, [52] identifies specific features that can differentiate between the faces of different or the same individuals [53]. Few characteristics are discovered through the opti-

mization of the Fisher Discriminant Criterion, as proposed by Fisher in 1936. This optimization involves maximizing the clustering of facial images belonging to the same individual, while simultaneously minimizing the clustering of facial images belonging to different individuals. Consequently, these characteristics can be utilized to ascertain the identity of individuals by grouping together the facial images of each individual.

Principal Component Analysis (PCA) is a method used to reduce the dimensions of images, commonly employed in image recognition and compression. By decreasing dimensions, irrelevant information is eliminated, and the facial structure is decomposed into uncorrelated components referred to as Eigen faces. The concept of utilizing eigenfaces was originally proposed by Kirby and Sirovich, who employed Karhunen-Loeve (KL) [54] transform to represent human faces. This particular approach proved highly successful in representing faces through the aforementioned analysis. In this approach, faces are expressed as a linear combination of eigenvectors with assigned weights, commonly referred to as eigenfaces. To facilitate face recognition, Principal Component Analysis (PCA) is applied by transforming the pixel values of an image into sets of eigenface feature vectors. These vectors are subsequently used for comparison, enabling the assessment of similarity between two facial images [54].

The application of the Discrete Wavelet Transform (DWT) is observed in the analysis of both images and signals [54]. This particular transform breaks down an image into a collection of fundamental functions known as wavelets, with this decomposition serving as the definition of the image's resolution. The Discrete Wavelet Transform (DWT) is achieved by subjecting the signal to a series of digital filters at varying scales. The act of scaling is accomplished by altering the signal's resolution through the process of subsampling. Within the framework of the DWT, the input sequence is decomposed into two sub-bands: one low-pass and one high-pass. Each sub-band contains half the number of samples as the original sequence [54].

Lai et al. [49] propose a holistic feature extraction methodology called spectroface. The wavelet transform and the Fourier transform are combined for making of spectroface. The wavelet transform is utilized on the face image to mitigate the impact of various facial expressions. Following the decomposition of the face image, the holistic Fourier invariant features (HFIF) are extracted from the low-frequency sub-band image. It takes into consideration the challenges of variations in illumination, pose, expression, rotation, occlusion and scale for face recognition tasks.

He et al [55] propose a holistic approach to feature extraction for human face recognition called Laplacianface. By utilizing the technique of Locality Preserving

Projections (LPP), the images of faces are transformed and assigned to a face subspace in order to facilitate analysis. In contrast to Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), which primarily focus on the Euclidean structure of the face space, LPP seeks to find an embedding that preserves local information. As a result, it is able to generate a face subspace that effectively captures the essential structure of the face manifold. The Laplacianfaces can be regarded as the most suitable linear approximations to the eigenfunctions of the Laplace Beltrami operator on the face manifold. While PCA and LDA aim to preserve the overall structure, it is often the case that the local structure holds greater importance in practical scenarios. In this section, the authors present a detailed explanation of Locality Preserving Projection (LPP), a novel algorithm designed to learn a subspace that both preserves the global structure and maintains the locality of the data. LPP strives to retain the intrinsic geometry of the data as well as the local structure. Due to its ability to preserve neighborhood relationships, Laplacianfaces demonstrate a greater capacity for capturing the underlying structure of the face manifold. Additionally, it is robust against the presence of outliers and noise, further enhancing its efficacy. It takes into consideration the challenges of variations in illumination, pose, expression, and noise for face recognition tasks.

Muqeet et al. [56] propose a methodology combining directional wavelet transform (DIWT) and local binary patterns (LBP), DIWTLBP. A proficient technique for evaluating direction, which relies on partitioning the image using a quadtree, has been executed to enable the adaptive selection of direction in specific regions of the face images. This facilitates the acquisition of DIWT sub-bands. The LBP histogram features are then derived from the chosen top-level DIWT sub-bands in order to obtain a local set of descriptive features. It takes into consideration the challenges of variations in illumination, pose, expression, rotation and scale for face recognition tasks.

It has been observed many holistic methods have proven their worth for face recognition purposes. Some of them are even being incorporated with local feature properties to give more robust techniques.

2.4 Conventional Performance Metrics

It is very important to evaluate the performance of human face recognition techniques on different datasets. Here different performance matrices come into the picture and all of these performance metrics can be expressed with the help of a confusion matrix as shown in Figure 2.1. A confusion matrix is nothing but a table for visual representation that summarizes the performance of a classification algorithm having predicted results and ground truth. Each of the attributes in

the matrix is defined as follows: True Positive (TP) signifies the cases where the algorithm or the technique accurately classifies the images as the positive class according to the ground truth, False Positive (FP) signifies the cases where the algorithm or the technique wrongly classifies the images as the positive class according to the ground truth, True Negative (TN) signifies the cases where the algorithm or the technique accurately classifies the images as the negative class according to the ground truth, False Negative (FN) signifies the cases where the algorithm or the technique wrongly classifies the images as the negative class according to the ground truth. From the confusion matrix, several key metrics such as accuracy, precision, recall/recognition rate, sensitivity, specificity, f1 score could be derived. The whole idea of the confusion matrix and derived key metrics from the confusion matrix can be easily extended for more than two classes

Ground Truth / Predicted Class	Positive Class	Negative Class
Positive Class	True Positive (TP)	False Positive (FP)
Negative Class	False Negative (FN)	True Negative (TN)

Figure 2.1: Confusion Matrix

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{2.3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2.5}$$

$$f1score = \frac{2 \times Precission \times Recall}{Precission + Recall} \tag{2.6}$$

With respect to human face recognition, recognition accuracy suggests how many images have been correctly classified into their respective classes over the whole testing database images. Error rate signifies (1-accuracy rate). Recall or Recognition Rate signifies how many images have been correctly classified over that particular class from the testing database. Precision signifies how many images

have been correctly classified over a particular class by the technique. Sensitivity follows the same principle of Recall. The F1 score is nothing but the harmonic mean of Precision and Recall and signifies the balance between these two. Also, Rank-k recognition accuracy or Rank-k recognition rate is used where k can be any positive integer. In rank(k), matching is corrected if one of the top k-classified images is correctly classified. ROC curve or Receiver Operating Characteristic curve is a graphical representation that could be derived from a confusion matrix and could give a fine insight into the performance of a classifier. It illustrates the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at various instances.

CHAPTER 3

Proposed Methodology

The proposed methodology that gives a feature map by extracting related local features from a facial image is described here. The Conditional Bayes Conjunctional Edge-Texture and Texture-Edge (CBCETTE) feature extractor would give edge-texture and texture-edge feature face. How CBCETTE would work is described in this chapter.

3.1 Concept of Pixel's distance from Local Centre of Mass

Parallel edge and texture images are produced using the local centre of mass (LCM) position by considering the two strategies - i) For edge image opus, the point nearest to the local centre of mass of the immediate neighbourhood (3×3) has the greatest illuminance as the mass is concentrated, i.e., minimum distance to LCM with maximum luminance is considered as 1, i.e., the edge, otherwise 0. ii) For texture image opus, the point(s) which is less than the mean distance of the pixels from the local centre of mass of the immediate neighbourhood (3×3) is considered as 1, which reflects the illumination variance inspired by the standard Local Binary Pattern (LBP) technique. Meanwhile, an edge picture is made by considering the point that all of the pixels make with the point of convergence of mass, i.e., the centre of mass of a particular square of pixels. The strategy of choosing edges takes significantly fewer events that appear differently with various methodologies like the Canny Edge [57] approach. From these edge and texture images, the probabilistic edge image, the probabilistic texture image, the Bayesian edge-texture image, the Bayesian texture-edge image, and finally the conditional Bayesian optimum conjunctional image of the original image are computed. In Figure 3.3, the way of getting the conjunctional edge-texture image from edge image and corresponding texture image is shown. The proposed methodology is

described in detail as:

Centre of Mass of n particles: The centre of mass is a concept in physics that refers to the point at which the mass of an object or a system of objects is considered to be concentrated. If a system consists of n particles of masses $m_1, m_2, m_3, \ldots, m_n$ having position vectors $r_1, r_2, r_3, \ldots, r_n$. The position vector of the centre of mass (CM) of the system -

$$r_{CM} = \frac{m_1 r_1 + m_2 r_2 + m_3 r_3 + \ldots + m_n r_n}{m_1 + m_2 + m_3 + \ldots + m_n} = \frac{\sum_{i=1}^{n} m_i r_i}{\sum_{i=1}^{n} m_i} \tag{3.1}$$

So in the 2D plane, the co-ordinates are -

$$X_{CM} = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3 + \ldots + m_n x_n}{m_1 + m_2 + m_3 + \ldots + m_n} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i} \tag{3.2}$$

$$Y_{CM} = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3 + \ldots + m_n y_n}{m_1 + m_2 + m_3 + \ldots + m_n} = \frac{\sum_{i=1}^{n} m_i y_i}{\sum_{i=1}^{n} m_i} \tag{3.3}$$

Calculation of position vector Local Centre of Mass: In the proposed system, the pixel intensity values are considered as the mass of the pixel, and the position of the local centre of mass (LCOM) is calculated for every 3×3 neighbourhood. So, the equation of LCOM becomes:

$$C(x', y') = \left(\frac{\sum_{i=1}^{n} I_i x_i}{\sum_{i=1}^{n} I_i}, \frac{\sum_{i=1}^{n} I_i y_i}{\sum_{i=1}^{n} I_i} \right) \tag{3.4}$$

where,

I = Intensity values of the pixels

x_i, y_i = x,y coordinate value of i^{th} pixel

n = number of pixels (for the above case it is 9 as 3×3 neighbourhood)

Calculation of minimal distance from the pixel to LCOM: The minimal distance between LCOM and each pixel in the 3×3 neighbourhood is calculated using Euclidean Distance.

Let the pixel coordinate for ith pixel in the neighbourhood is $P(x_i, y_i)$.

So, the minimal distance d(C,P) = $\sqrt{(x - x')^2 + (y - y')^2}$ (using the Euclidean Distance Formula)

3.1.1 Bayesian Conditional Probabilistic Feature Extraction

i. Binary Edge Image Extraction: To extract the binary edge feature matrix, we proposed a new approach by applying the new idea by following the direction

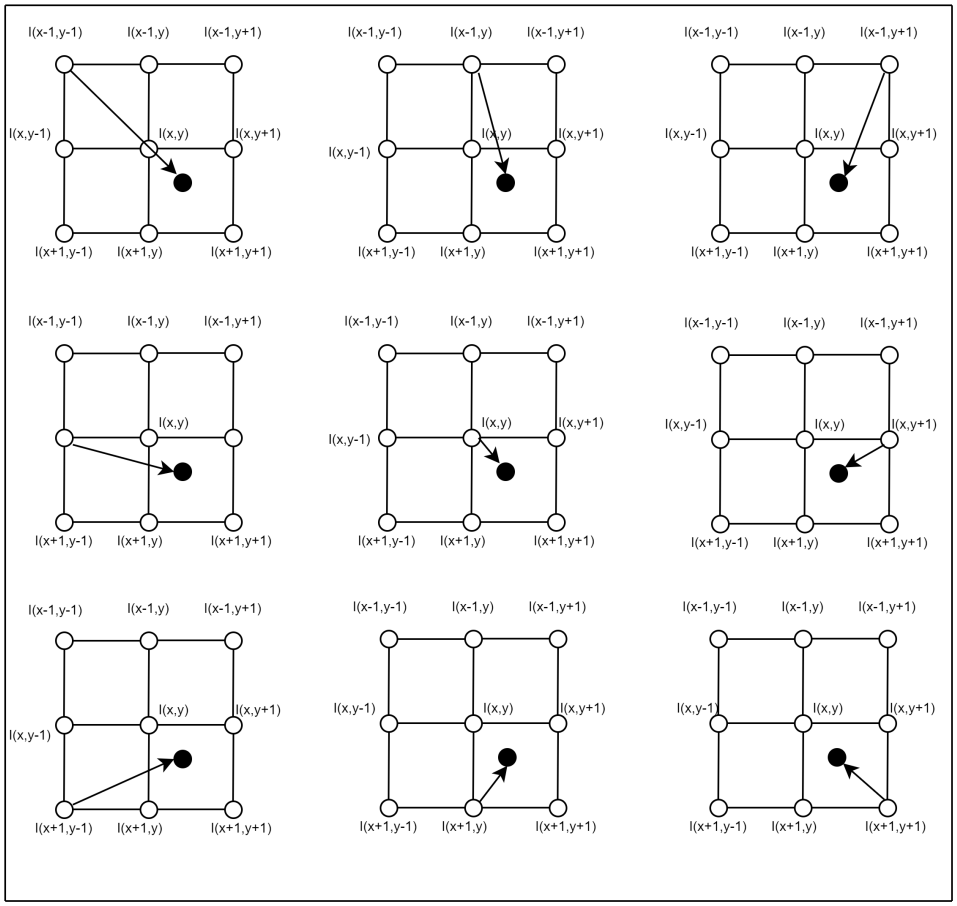


Figure 3.1: Distance Calculation From COM

of the illumination, where the maximum illumination concentration exerted from the centre of mass (COM) for each block I_B, considering the nearest pixel(s) (minimum distance from the COM) of that block as edge, i.e., one and the remaining pixel as 0, to obtain the binary edge matrix I_E finally.

ii. Binary Texture Image Extraction: To extract the texture feature matrix, we inherit the idea of the local binary pattern (LBP) by comparing all the distance neighbouring pixels' values from the COM. The pixels that are less than the mean distance are considered one and the remaining pixels are considered 0 of each block I_B. Finally, we obtain the binary texture matrix I_T. It is noted that I_E and I_T are invariant to illumination, rotation, and noise.

iii. Binary Intersectional Edge-Texture Image Extraction: Here, a conjunction of Binary Edge Image and Binary Texture Image is extracted with a concept of logical conjunction, as these are binary values, so normal arithmetic multiplication has been applied to obtain a binary intersectional edge-texture image extraction.

iv. Extraction of the Conditional Bayesian Probabilistic Edge-Texture and Texture-Edge Images: The Bayesian edge-texture and texture-edge image is inspired by the conditional probabilistic theory of Thomas Bayes. This gives a probabilistic measure of an incident if another incident has occurred. This Bayesian edge-texture and Bayesian texture-edge for pixels of an image is evaluated as: i) the probability of edge given a texture P(E/T) exists at that particular position and ii) the probability of texture given an edge P(T/E) exists at that particular position. Considering the above cases, the following equations are constructed:

$$P(E/T) = \frac{P(E \cap T)}{P(T)}, \qquad P(T/E) = \frac{P(T \cap E)}{P(E)} \tag{3.5}$$

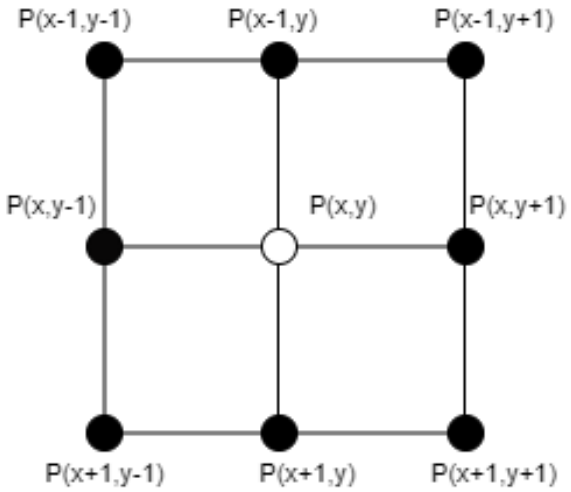


Figure 3.2: 3×3 neighbourhood of a pixel

To determine P(E/T) and P(T/E), the following parameters required are Probabilistic Edge (P(E)), Probabilistic Texture (P(T)), and Probabilistic Intersectional Edge Texture ($P(E \cap T) = P(T \cap E)$).
To calculate probabilistic edge:
P(E) = n(E)/n(S) Where, n(E) = number pixels containing edge, i.e., the number pixels with value 1 in 3×3 binary edge image block
n(S) = number of total pixels i.e., 9 as 3×3 binary edge image block.
We can write,
$[P(E)]_{(x,y)}$ = [p(x-1,y-1) + p(x-1,y) + p(x-1,y+1) + p(x,y-1) + p(x,y) + p(x,y+1) + p(x+1,y-1) + p(x+1,y) + p(x+1,y+1)] / 9
Where, p(x,y) signifies value at (x,y) position in 3×3 binary edge image block.
Calculate the probabilistic texture image and the probabilistic intersectional edge-texture image in a similar fashion.
To calculate probabilistic texture:

P(T) = n(T)/n(S)
Where, n(T) = number of pixels containing texture, i.e., the number of pixels with value 1 in 3×3 binary edge image block.
n(S) = number of total pixels i.e., 9 as 3×3 binary edge image block.
We can write,
$[P(T)]_{(x,y)}$ = [q(x-1,y-1) + q(x-1,y) + q(x-1,y+1) + q(x,y-1) + q(x,y) + q(x,y+1) + q(x+1,y-1) + q(x+1,y) + q(x+1,y+1)] / 9
Where, q(x,y) signifies value at (x,y) position in 3×3 binary texture image block.
For the case of $P(E \cap T)$ or $P(T \cap E)$ calculation, the Binary Intersectional Edge Texture image is used, and several pixels containing both edge and texture are used as events.
$P(E \cap T)_{(x,y)} = P(T \cap E)_{(x,y)}$ = [r(x-1,y-1) + r(x-1,y) + r(x-1,y+1) + r(x,y-1) + r(x,y) + r(x,y+1) + r(x+1,y-1) + r(x+1,y) + r(x+1,y+1)]/9
Where, r(x,y) signifies value at (x,y) position in 3×3 binary intersectional edge texture image block.

Then, all the values are consigned to the Bayes theorem:

$$P(E/T) = \frac{P(E \cap T)}{P(T)} \quad if P(T) > 0 \tag{3.6}$$

$$= 0 \quad if P(T) = 0 \tag{3.7}$$

$$P(T/E) = \frac{P(T \cap E)}{P(E)} \quad if P(E) > 0 \tag{3.8}$$

$$= 0 \quad if P(E) = 0 \tag{3.9}$$

v. Extraction of Conditional Bayes Conjunctional Optimum Edge-Texture: The main goal is to have an optimum set of features that should combine both edge and texture features. The combination of P(E/T) and P(T/E), i.e., conditional Bayes conjunctional edge-texture and texture-edge feature face ($I_{CBCETTE}$). This is achieved by: $I_{CBCETTE}$ = P(E/T) + P(T/E) – max (P(E/T), P(T/E)) = min (P(E/T), P(T/E))

We will prove that this extracted set $I_{CBCETTE}$ = X (say) is optimum.

Theorem 1 *It is to prove that the extracted ($I_{CBCETTE}$) feature set X is optimal*

Let X be a random variable over the space Ω. As discussed our case is a two-category decision problem:

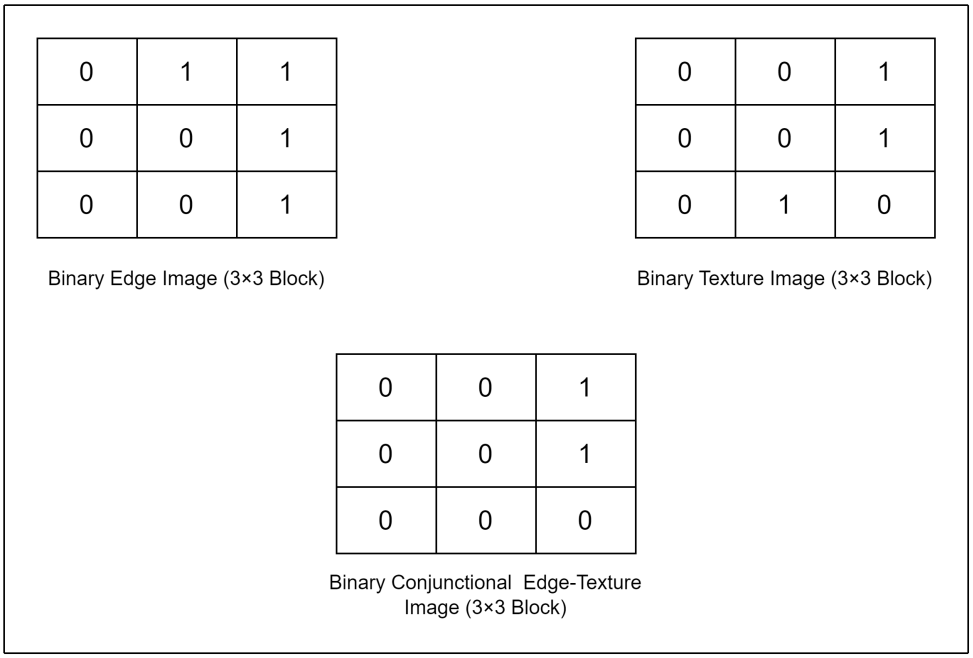


Figure 3.3: Describes the process of calculating Conjunctional Edge-Texture Image

H_1 : X ϵ ω_1
H_2 : X ϵ ω_2
Where H1 and H2 are the hypotheses.
Now, the optimal decision is -
Choose H_1 when $P(X|\omega_1)P(\omega_1) \geq P(X|\omega_2)P(\omega_2)$
Choose H_2 when $P(X|\omega_1)P(\omega_1) < P(X|\omega_2)P(\omega_2)$
A basic decision Rule: partition the space Ω into two disjoint regions R_1 and R_2.
Choose H_1: ϵ R_1
Choose H_2: ϵ R_2 H1 and H2 are the hypotheses. Our aim is to prove that this decision rule is not optimal.
So, partition the space Ω into two disjoint regions Ω_1 and Ω_2.
Choose H_1: ϵ Ω_1
Choose H_2: ϵ Ω_2
Where, $\Omega_1 = \{X\epsilon\Omega : P(X|\omega_1)P(\omega_1) \geq P(X|\omega_2)P(\omega_2)\}$, $\Omega_2 = \{X\epsilon\Omega : P(X|\omega_1)P(\omega_1) < P(X|\omega_2)P(\omega_2)\}$.
Now, there is an error in this decision rule:
$P(error) = P(X\epsilon R_1, \omega_2) + P(X\epsilon R_2, \omega_1) = P(X\epsilon R1|\omega_2)P(\omega_2) + P(X\epsilon R_2|\omega_1)P(\omega_1)$
$\implies P(error) = \int_{R_1} P(X|\omega_2)P(\omega_2)\,dx + \int_{R_2} P(X|\omega_1)P(\omega_1)\,dx$
Bayes error = P(Bayes_error) = $\int_{\Omega_1} P(X|\omega_2)P(\omega_2)\,dx + \int_{\Omega_2} P(X|\omega_1)P(\omega_1)\,dx$
Let, $\Delta(error) = P(error) - P(Bayes_error)$
As discussed earlier $R_1 = (R_1 \cap \Omega_1) \cup (R_1 \cap \Omega_2), R_2 = (R_2 \cap \Omega_1) \cup (R_2 \cap \Omega_2), \Omega_1 = (\Omega_1 \cap R_1) \cup (\Omega_1 \cap R_2), \Omega_2 = (\Omega_2 \cap R_1) \cup (\Omega_2 \cap R_2)$

Now $\Delta(error) = \int_{R_1} P(X|\omega_2)P(\omega_2)\,dx + \int_{R_2} P(X|\omega_1)P(\omega_1)\,dx - \int_{\Omega_1} P(X|\omega_2)P(\omega_2)\,dx - \int_{\Omega_2} P(X|\omega_1)P(\omega_1)\,dx$
or $\Delta(error) = P(\omega_2)\left[\int_{R_1} P(X|\omega_2)\,dx - \int_{\Omega_1} P(X|\omega_2)\,dx\right] + P(\omega_1)\left[\int_{R_2} P(X|\omega_1)\,dx - \int_{\Omega_2} P(X|\omega_1)\,dx\right]$
or, $\Delta(error) = P(\omega_2)\left[\int_{R_1\cap\Omega_2} P(X|\omega_2)\,dx - \int_{\Omega_1\cap R_2} P(X|\omega_2)\,dx\right] + P(\omega_1)\left[\int_{R_2\cap\Omega_1} P(X|\omega_1)\,dx - \int_{\Omega_2\cap R_1} P(X|\omega_1)\,dx\right]$
or, $\Delta(error) = \int_{R_1\cap\Omega_2} [P(X|\omega_2)P(\omega_2) - P(X|\omega_1)P(\omega_1)]\,dx + \int_{R_2\cap\Omega_1} [P(X|\omega_1)P(\omega_1) - P(X|\omega_2)P(\omega_2)]\,dx$

Recall, $\Omega_1 = \{X\epsilon\Omega : P(X|\omega_1)P(\omega_1) \geq P(X|\omega_2)P(\omega_2)\}$,
$\implies \int_{R_2\cap\Omega_1} [P(X|\omega_1)P(\omega_1) - P(X|\omega_2)P(\omega_2)]\,dx \geq 0$
$\Omega_2 = \{X\epsilon\Omega : P(X|\omega_1)P(\omega_1) < P(X|\omega_2)P(\omega_2)\}$,
$\implies \int_{R_1\cap\Omega_2} [P(X|\omega_2)P(\omega_2) - P(X|\omega_1)P(\omega_1)]\,dx \geq 0$

Thus, when $\Delta(error) \geq 0$, $\Delta(error) = P(error) - P(Bayes_error)$, $\implies P(error) \geq P(Bayes_error)$.
Hence it is proved that the Bayes Decision Rule is optimal. Hence the $I_{CBCETTE}$ feature set X is optimal.

Theorem 2 *It is to prove that a conjunctional feature set of optimal infused features is also optimal.*

We have already seen that $P(error) \geq P(Bayes_error)$. Now note that both P(E/T) and P(T/E) are optimal in their own respects since both correspond to Bayes decision rules. However, P(Bayes_error), in their case, is proportional to the probability value of the decision rule. So, min(P(E/T), P(T/E)) ensures that the P(error) is minimized for each pixel. This makes our feature optimum.

Theorem 3 *It is to prove that for a neighbourhood w of size (3×3) the pixel distances from local COM position is illumination invariant.*

Considering the Illumination Reflectance Model(IRM) proposed by Horn [58], an image I(x,y) can be expressed as I(x,y) = R(x,y)*L(x,y), where L(x,y) is the luminance component and R(x,y) is the reflectance component. Generally the reflectance component R consists information about the characteristics feature of the image, whereas L the luminance component, exhibits slow spatial variation with no discontinuities. It is assumed that L varies very slowly and by contrast, R varies abruptly, Thus if one can eliminate the L component, then the R component is still able to represent the key facial features, which are the most vital information for a face recognition algorithm. Abolishing the L component will make the R(x,y)

as an illumination-invariant and hence makes the image I as illumination invariant and will increase the sharpness of I. Though unscrambling L(x,y) and R(x,y) is an ill-posed problem. Hence, for the elimination of the L component, the division-based method has been considered.

Now considering $P(x_i, y_j)$ is the position vector of a pixel where $C(x', y')$ is the LCOM of that (3×3) neighbourhood.

Assuming that $L(x, y) = L(x \pm 1, y) = L(x, y \pm 1) = L(x \pm 1, y \pm 1) =$ L (constant, say) for the smallest neighbourhood of the centre pixel P(x,y).

$$
\begin{aligned}
d_{(c,p)} = \sqrt{(x'-x)^2 + (y'-y)^2} &= \sqrt{\left(\frac{\sum_{i=1}^{3} I(x_i)x_i}{\sum_{i=1}^{3} I(x_i)} - x_i\right)^2 + \left(\frac{\sum_{j=1}^{3} I(y_j)y_j}{\sum_{j=1}^{3} I(y_j)} - y_j\right)^2} \\
&= \sqrt{\left(\frac{\sum_{i=1}^{3} L * R(x_i)x_i}{\sum_{i=1}^{3} L * R(x_i)} - x_i\right)^2 + \left(\frac{\sum_{j=1}^{3} L * R(y_j)y_j}{\sum_{j=1}^{3} L * R(y_j)} - y_j\right)^2} \\
&= \sqrt{\left(\frac{\sum_{i=1}^{3} R(x_i)x_i}{\sum_{i=1}^{3} R(x_i)} - x_i\right)^2 + \left(\frac{\sum_{j=1}^{3} R(y_j)y_j}{\sum_{j=1}^{3} R(y_j)} - y_j\right)^2} \qquad (3.10)
\end{aligned}
$$

In equation(3.10) $d_{(c,p)}$ indicates the distance from LCOM to any point in the corresponding block won't contain L value.So, the value $d_{(c,p)}$ is illumination invariant. Hence it is proved.

Theorem 4 *It is to prove that for a neighbourhood w of size (3×3), the pixel distances from the local COM position are rotation invariant.*

According to physics, for a rigid body position of each particle is constant hence if it is rotated at a certain angle with respect to any axis the inter-particle distance will be constant. So, the unit particle mass and the inter-particle distance is constant for rigid body hence the local distance from the COM is always constant. From Figure 3.4 suppose for a two-body system mass of two particles are m_1 and

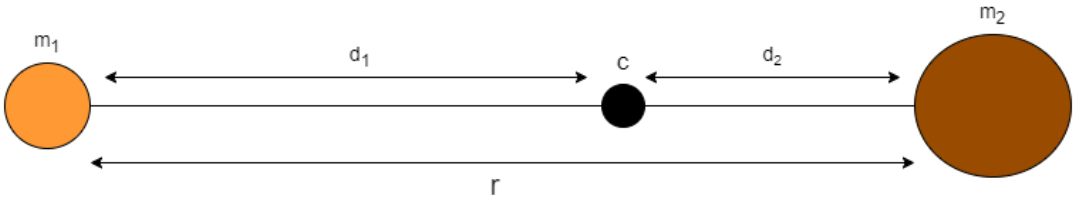


Figure 3.4: Shows the two-body system

m_2 and their inter-particle distance is r and centre of mass of the system is C.

From m_1 and m_2 mass the distance from COM is d_1 and d_2 respectively. According to the formula of COM.

$$d_1 = \frac{m_2 \times r}{m_1 + m_2} \tag{3.11}$$

$$d_2 = \frac{m_1 \times r}{m_1 + m_2} \tag{3.12}$$

From equations (3.11) and (3.12) states that the distance from centre of mass d_1 and d_2 only depends on the values of the mass of the particle (m_1, m_2) and inter-particle distance (r), both the properties of the rigid system is constant hence distance from the COM (d_1, d_2) is constant. Hence the theorem is proved.

Theorem 5 *It is to prove that for a neighbourhood w of size (3×3), the pixel distance from local COM position is robust to the Gaussian noise Z with mean zero and standard deviation σ at 99% confidence interval.*

If Z $N(0, \sigma)$ then $P(-2\sigma \leq Z \leq 2\sigma) > 99\%$.
We have assumed that Gaussian noise,ϵ_i, where i ranges over the local neighbourhood, have maximum standard deviation $\sigma = 0.04$.
$-0.08 \leq \epsilon_i \leq 0.08$ with probability >99% where $\epsilon_i \sim N(0.004)$
Let, L be the lower bound of M and U be the upper bound of M, where M represents the pixel distance from the local centre of mass position after the addition of noise. Thus,

$$\left(L = \frac{\sum_{i=1}^{N} m_i d_i - .08 \sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} m_i + .08N}\right) \leq \left(M = \frac{\sum_{i=1}^{N} m_i d_i + \sum_{i=1}^{N} \epsilon_i d_i}{\sum_{i=1}^{N} m_i + \sum_{i=1}^{N} \epsilon_i}\right) \leq \left(U = \frac{\sum_{i=1}^{N} m_i d_i + .08 \sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} m_i - .08N}\right) \tag{3.13}$$

Then we need to prove that $|M - L| \leq \frac{1}{100}M$, and $|U - M| \leq \frac{1}{100}M \Leftrightarrow |U - L| \leq \frac{M}{100} + \frac{M}{100} = \frac{M}{50}$
Now,

$$|U - L| = .16 \left[\frac{(\sum_{i=1}^{N} m_i)(\sum_{i=1}^{N} d_i) + (\sum_{i=1}^{N} m_i d_i)N}{(\sum_{i=1}^{N} m_i)^2 - (.08N)^2}\right] \tag{3.14}$$

This to be noted that $(\sum_{i=1}^{N} m_i)(\sum_{i=1}^{N} d_i) \leq (\sum_{i=1}^{N} m_i d_i)N$
Also,

$$(\sum_{i=1}^{N} m_i)^2 - (.08N)^2 = (\sum_{i=1}^{N} m_i + .08N)(\sum_{i=1}^{N} m_i - .08N) \tag{3.15}$$

From assumptions,

$$(\sum_{i=1}^{N} m_i + .08N) \geq (\sum_{i=1}^{N} m_i + \epsilon_i) \tag{3.16}$$

$$(\sum_{i=1}^{N} m_i - .08N) \geq 4N \tag{3.17}$$

- with very high probability since$(m_i \epsilon \{0, 1, \ldots, 256\} uniformly)$
From equations (3.15) (3.16) and (3.17), it can be written -

$$(\sum_{i=1}^{N} m_i)^2 - (.08N)^2 \geq 4N(\sum_{i=1}^{N} m_i + \epsilon_i) \tag{3.18}$$

Also,

$$\left(\frac{(\sum_{i=1}^{N} m_i)(\sum_{i=1}^{N} d_i) + (\sum_{i=1}^{N} m_i d_i)N}{(\sum_{i=1}^{N} m_i)^2 - (.08N)^2}\right) \leq \left(2N\frac{(\sum_{i=1}^{N} m_i d_i)}{4N(\sum_{i=1}^{N} m_i + \epsilon_i)} = \frac{1}{2}\frac{(\sum_{i=1}^{N} m_i d_i)}{(\sum_{i=1}^{N} m_i + \epsilon_i)}\right) \leq \left(\frac{5}{8}\frac{\sum_{i=1}^{N} m_i d_i - .08(\sum_{i=1}^{N} d_i)}{\sum_{i=1}^{N} m_i + \epsilon_i}\right)$$
$$[As, \epsilon_i \geq -0.08, m_i \geq 1] \tag{3.19}$$

$$\implies 0.16\left[\frac{(\sum_{i=1}^{N} m_i)(\sum_{i=1}^{N} d_i^2) + (\sum_{i=1}^{N} m_i d_i^2)N}{(\sum_{i=1}^{N} m_i)^2 - (.08N)^2}\right] \leq \frac{2}{100}\frac{(\sum_{i=1}^{N}(m_i + \epsilon_i)d_i^2)}{(\sum_{i=1}^{N}(m_i + \epsilon_i))} = \frac{1}{50}\frac{(\sum_{i=1}^{N}(m_i + \epsilon_i)d_i^2)}{(\sum_{i=1}^{N}(m_i + \epsilon_i))} \tag{3.20}$$

Similarly, it can be proved the same for $\sigma = .01, .02, .03$. Hence proved, invariant to Gaussian noise (Z) with mean zero and standard deviation.
Thus, the position of LCOM d is noise invariant at 99% confidence interval. Hence the distance from the Local Centre of Mass position after addition of noise is invariant at 99% confidence interval.

3.1.2 Intuition Behind 3×3 Window

Window size being odd, we could quickly compute the $I_{CBCETTE}$ for pixel (x,y). Also, the 3x3 block has a relatively small receptive field compared to larger filter sizes like 5x5 or 7x7. It means that each output element in the feature map depends on a smaller region of the input image. This allows the methodology

to capture fine-grained details and local patterns in the image. 3x3 block is also computationally efficient. The small size of the 3x3 filter provides some level of translation invariance, making the methodology more robust to local shifts and distortions in the input image. The smaller 3x3 blocks can help the model be more vital to noise, illumination, and rotation variations, as they focus on capturing local information rather than global patterns, which can be more sensitive to changes in the image. The features learned by these blocks tend to be more generalizable and can be fine-tuned for specific tasks, like recognizing faces from different racial or age groups.

3.1.3 Intuition Behind Stride 1

Stride refers to the step size at which the block/filter moves over the input image during the feature extraction process. Setting the stride to 1 means the filter moves one pixel at a time, ensuring that every pixel is considered during the operation. With a stride of 1, the output feature map will have the exact spatial resolution as the input image. This preservation of spatial resolution is crucial for tasks that require precise localization, such as face recognition. A higher stride would downsample the feature map, which could lead to a loss of fine-grained spatial information. Smaller strides allow the model to extract more local features and finer details from the input image. When the stride is set to 1, each location in the output feature map is influenced by a relatively small region in the input image, enabling the model to capture delicate patterns and intricate structures. When the stride is set to 1, adjacent convolutional operations largely overlap, and the output features are combined more effectively. This overlapping nature ensures the model can detect features occurring at different locations and orientations, leading to better feature representations. Using a stride of 1 lets the model precisely localize objects or regions in the input image without losing spatial details. More considerable strides can make the model more robust to small translations in the input image. This translational invariance is desirable in some cases, such as image classification, where the position of an object in the image does not affect its class label.

3.1.4 Size of Feature Map

We can deduce the size of the feature map while using a 3×3 window block and horizontal and vertical stride 1. Let the input image's size m×n and the size of the feature map after applying the methodology $m' \times n'$.The formula of convolution gives us -

$$m' = \left[\frac{m + 2p - k}{s}\right] + 1, \qquad n' = \left[\frac{n + 2p - k}{s}\right] + 1 \tag{3.21}$$

Where,
p signifes padding size. Here we take p as 0.
k signifies window/kernel size. Here k is 3.
s signifies stride. Here s is 1.
So, we get,m' = (m-2) and n' = (n-2).
The size of the feature map we get (m-2)×(n-2) for m×n input image size.

3.1.5 Properties of Proposed Method CBCETTE

From the above theorems and mentioned ways, we can point out some properties of the proposed method.

i. The proposed method (CBCETTE) extracts a combined feature set of edge and texture using the Bayesian Conditional Probabilistic approach which is the optimum feature set that has been proved mathematically, making the method really very computationally efficient.

ii. The extracted optimum feature by the proposed method (CBCETTE) is invariant to illumination, rotation, and noise that has been proved mathematically.

iii. A novel distance from the local Centre of mass along the direction of light and inspired Local Binary Pattern (LBP) concept has been applied to extract the edge features and texture features guided by the illumination makes the edge and texture features a very good local characteristic descriptor.

iv. The proposed method CBCETTE neither requires any pre-processing or any training.

v. Finally, the proposed method has probabilistic optimized feature set with a minimum data loss rate.

CHAPTER 4

Experiment and Result

The whole experiment, from the experimental setup to taking different human face image datasets to do recognition or, in other words, classification tasks, has been discussed in this chapter. Also in this chapter, accuracy as the metric is used to discuss the results that have been achieved from the experiment. Also, tables of comparisons of different methodologies with our methodology are given in this chapter with respect to different human face image datasets.

4.1 Experimental Setup

For experimental purposes, Google Colab which is a cloud-based platform to perform machine learning-related works, is used. It provided 12.7 GB of RAM and 107.7 GB of disk space. Python as a programming language is used for experimental purposes.

Several Python packages or libraries and third-party packages or libraries are used. The 'numpy' library which provides numerical operations using Python is used to manipulate image vectors. The 'os' package which provides a portable way of using operating system-dependent functionality, is used to read files that contain the images. The 'OpenCV' or 'cv2' library which helps to read or write images, to apply several pre-built operations on images, to manipulate images like transforming colour images to grayscale images, is used for experimental purposes. The 'math' library is used to perform some mathematical operations. The 'matplotlib' library which provides several methods for plotting the data, showing images, and even reading images, is used in experiments. Third-party toolkit modules such as 'mtcnn' and 'dlib' are used to detect faces and crop faces automatically.

The 'Scikit-learn' or 'sklearn' package which is a free machine learning library for Python language, is used in the experiment. The 'Scikit-learn' not only provides several machine learning models but also provides several methodologies and mod-

ules to operate on data or manipulate data before fitting machine learning models and it also provides several evaluation matrices, and confusion matrices to evaluate over-predicted results.
The face image datasets that are used in this experiment are - Yale B face database, Extended Yale B face database, ORL face database, Faces95 face database, and Yale face database.

4.2 Experimental Details With Results

Several experimental steps including the pre-processing of images before applying our proposed methodology to find feature-map are the same for all the databases that are used in this experiment. For each of the databases, the images are extracted and each of the subjects has been transferred into a grayscale image and only the front face of each of the subjects has been cropped. The reasons behind to use of grayscale images are that it reduces complexity as each pixel of grayscale image has a value between 0 to 255 and also being one channel, it is less complex to work on and also maintains consistent features, and less prone to sensitivity to lighting conditions. For this, the image vectors can be easily operated on and the proposed methodology could be applied to find feature maps. For the experiment the images are reshaped according to the algorithm and after that feature map of each of the images is computed and stored for classification purposes. K-fold cross-validation is used for training and testing purposes. It helps to find best hyperperameters during training. It helps to get clear idea about accuracy over all the images during testing. Before each k-fold cross-validation, shuffling is done to avoid any kind of bias. Then for classification purposes machine learning model like a Support Vector Machine (SVM) which is called 'SVC' (Support Vector Classifier) module in 'sklearn', is trained on the train set and is tested on the test set to find out accuracy. The hyperparameters are - kernel, degree, gamma, c, coef0. The 'kernel' signifies the SVM's kernel and maps the dataset into higher dimensions according to it. The 'degree' signifies the kernel's degree. The 'gamma' signifies the kernel's coefficient. The 'c' signifies l2 regularization that helps to reduce overfitting. It must be strictly positive. The 'coef0' is only useful when the kernel is polynomial one. The 'GridSearchCV' method of sklearn is used to fine-tune those hyperparameters with k-cross validation.

4.2.1 Yale B Face Database

The Yale B Face Database consists of 640 images, that is 10 persons having 64 images each. The images were captured under different lighting conditions and

various facial expressions. There are 9 poses x 64 illumination conditions for each person (class). So, we can say there are 10 classes and each class has 64 images in other words. Each grayscale cropped image of the database size is (192,168). Each image is reshaped to (202,202). So, the feature map for each image would be (200,200).

The images are flattened to feed to the machine learning classifier. So each of the flattened images' shape becomes (40000,). The GridSearchCV method of sklearn is used to find out the best hyperparameters of SVC that control the accuracy of the model regarding classification. The SVC model takes different hyperparameters. To find out the best hyperparameters, several values of those hyperparameters are given to the GridSearchCV method and 5-fold cross-validation is used for training purposes. The model is then trained with all those values in combination to find out the best hyperparameters that produce the best accuracy for the model that is trained using feature maps.

The value of each hyperparameter that gives the best result is given below - kernel := poly, c := 0.1, degree := 4, gamma := scale, coef0 := 0.

5-fold cross-validation performed over the dataset using these hyperperameters to get mean training and testing accuracy. Mean training accuracy is 100% , and mean testing accuracy is 99.38%. The 5 fold accuracies of training are [1.0, 1.0, 1.0, 1.0, 1.0] and The 5 fold accuracies of testing are [0.9921875, 0.9921875, 0.9921875, 1.0, 0.9921875] respectively.

The result proves our methodology enables the classifier to do perfect classification or recognition on the Yale B face database. The confusion matrix is given in Figure 4.6 to have better insights. The comparisons of performances for the Yale B face database with other methods are given in Table 4.1. Table 4.1 shows the proposed methodology with svm gives better result with respect to accuracy than almost every feature descriptor or feature extraction methodology based classifi-



Figure 4.1: Example of feature map images of Yale B

Table 4.1: Accuracy Comparison for Yale B

Method	Accuracy(%)
PCA [59]	68.6
Wavelet [59]	28.8
LDA [59]	60
ESTIM [59]	89.5
Randon and Wavelet [59]	37.7
DOMF [59]	37.7
LPP [60]	90.5
OLPP [60]	92
DLPP [60]	93.4
CLPP [60]	94.1
FLBP-BLCMt [61]	94.54
ALBP+BCD [61]	86.45
Gabor Wavelet LBP [42]	92
NCDLBP + SVM [42]	100
Proposed Method + SVM	**99.38**

cation on the Yale B face database.

4.2.2 Extended Yale B Face Database

The Extend YaleB Face Database consists of 2432 images, that is 38 persons having 64 images each. The images were captured under different lighting conditions and various facial expressions. There are 9 poses x 64 illumination conditions for each person (class). So, we can say there are 10 classes and each class has 64 images in other words. Each grayscale cropped image of the database size is (192,168). Each image is reshaped to (202,202). So, the feature map for each image would be (200,200).

The images are flattened to feed to the machine-learning classifier which is SVM in this experiment. So each of the flattened images' shape becomes (40000,). The GridSearchCV method of sklearn is used to find out the best hyperparameters of SVC that control the accuracy of the model regarding classification. The SVC model takes different hyperparameters. To find out the best hyperparameters several values of those hyperparameters are given to the GridSearchCV method and 5 cross-validation is used for training purposes. The model is then trained with all those values in combination to find out the best hyperparameters that produce the best accuracy for the model that is trained using feature maps.

The value of each hyperparameter that gives the best result is given below - kernel := poly, c := 0.05, degree := 4, gamma := scale, coef0 := 0.

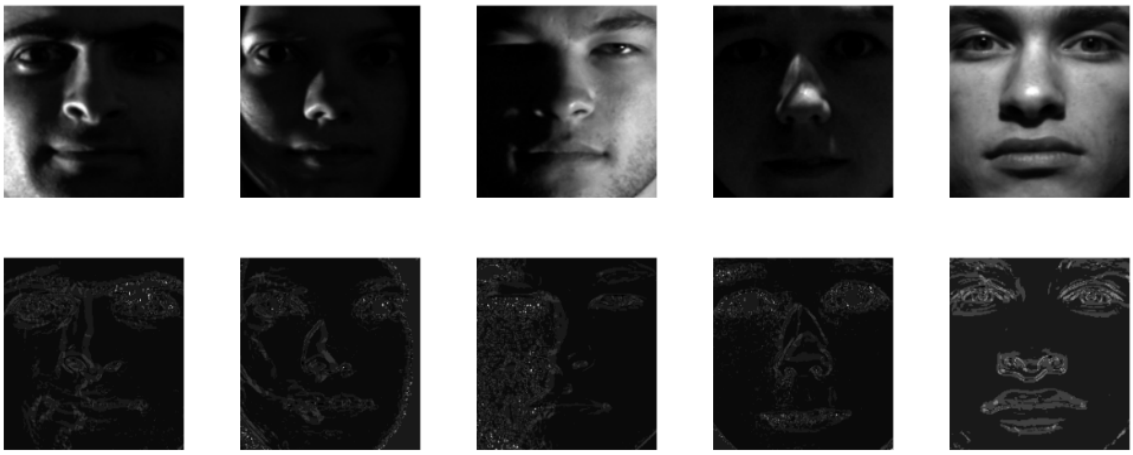


Figure 4.2: Example of feature map images of Extended Yale B

5-fold cross-validation performed over the dataset using these hyperperameters to get mean training and testing accuracy. Mean training accuracy is 100% , and mean testing accuracy is 98.77% . The 5 fold accuracies of training are [1.0, 1.0, 1.0, 1.0, 1.0] and The 5 fold accuracies of testing are [0.9897330595482546, 0.9876796714579056, 0.9876543209876543, 0.9814814814814815, 0.9917695473251029] respectively.

The result proves our methodology enables the classifier to do better classification than most other feature extraction techniques on the Extended Yale B face database. The confusion matrix is given in Figure 4.7 to have better insights. The comparisons of performances for the Extended Yale B face database with other methods are given in Table 4.2.

4.2.3 Faces95 Face Database

The Faces95 Face Database consists of 1440 colour facial images that is 72 subjects /persons having 20 images each. The images have varying scales, expressions, and illuminations. As the front face of each image needs to be cropped, the 'dlib' library is used. It works with the principals of HOG and CNN to detect faces. Also after getting the resultant face images, manual checking is incorporated cause in several instances it fails to provide better results or provides no result at all. After the whole thing, all the cropped face images are acquired and converted into grayscale images. All the images are of different sizes due to scale factors.

Before applying the proposed methodology to get the feature map for each image all images are resized into (60,60). So the size of each feature map image is (58,58).

The images are flattened to feed to the machine learning classifier. So each of the flattened images' shape becomes (3364,). Support Vector Classifier (SVC) is

Table 4.2: Accuracy Comparison for Extended Yale B

Method	Accuracy(%)
PCA [23]	52.23
LDA [23]	76.11
LC-LBP [62]	71.05
LNDP [62]	62.78
SSR [42]	92
LN+ICA [42]	98.1
HE+Gabor [42]	98.4
GWT [42]	86.7
GW + DKPCA [42]	98.8
MMC [23]	78.37
CRMMC [23]	78.9
CLBP [62]	78.9
TAO-LBP +SVM [51]	86.65
NCDLBP + SVM [42]	99.23
Proposed Method + SVM	**98.77**

used for classification purposes. The GridSearchCV method of sklearn is used to find out the best hyperparameters of SVC that control the accuracy of the model regarding classification. The SVC model takes different hyperparameters. To find out the best hyperparameters, several values of the hyperparameters are given to the GridSearchCV object and 5 cross-validation is used for training purposes. The model is then trained with all those values in combination to find the best hyperparameters that produce the best accuracy for the model trained using feature maps.

The value of each hyperparameter that gives the best result is given below - kernel := poly, c := 0.03, degree := 2, gamma := scale, coef0 := 0.1.

5-fold cross-validation performed over the dataset using these hyperperameters to get mean training and testing accuracy. Mean training accuracy is 100% , and mean testing accuracy is 97.92%. The 5 fold accuracies of training are [1.0, 1.0, 1.0, 1.0, 1.0] and The 5 fold accuracies of testing are [0.9618055555555556, 0.9930555555555556, 0.9826388888888888, 0.9756944444444444, 0.9826388888888888] respectively.

The confusion matrix is given in Figure 4.8 to have better insights. The comparisons of performances for the Faces95 face database with other methods are given in Table 4.3. The result proves our methodology enables the classifier to do certainly good classification on the Faces95 face database.



Figure 4.3: Example of feature map images of Faces95

Table 4.3: Accuracy Comparison for Faces95

Method	Accuracy(%)
LBPH [63]	98.73
PCA + SVM [63]	98.35
Image Enhancement+PCA+LDA [63]	99.37
PRC + OCC + ICC [64]	76
Gabor [64]	83
MOGP [64]	98.61
Proposed Method + SVM	**97.92**

4.2.4 Yale Face Database

The Yale Face Database consists of 165 images, that is 15 persons having 11 images each. Each person or class has different conditions like center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, wink. So, we can say there are 11 classes and each class has 15 images in other words. As the front face of each image needs to be cropped, 'mtcnn'(Multi-Task Cascaded Convolutional Neural Networks) library is used. Before applying the proposed methodology to get the feature map for each image all images are resized into (70,70). So the size of each feature map image is (68,68).
The images are flattened to feed to the machine learning classifier. So each of the flattened images' shape becomes (4624,). Support Vector Classifier (SVC) is used for classification purposes. The GridSearchCV method of sklearn is used to find out the best hyperparameters of SVC that control the accuracy of the model regarding classification. The SVC model takes different hyperparameters. To find out the best hyperparameter several values of the hyperparameters are given to the GridSearchCV object and 5 cross-validation is used for training purposes. The model is then trained with all those values in combination to find out the best hy-

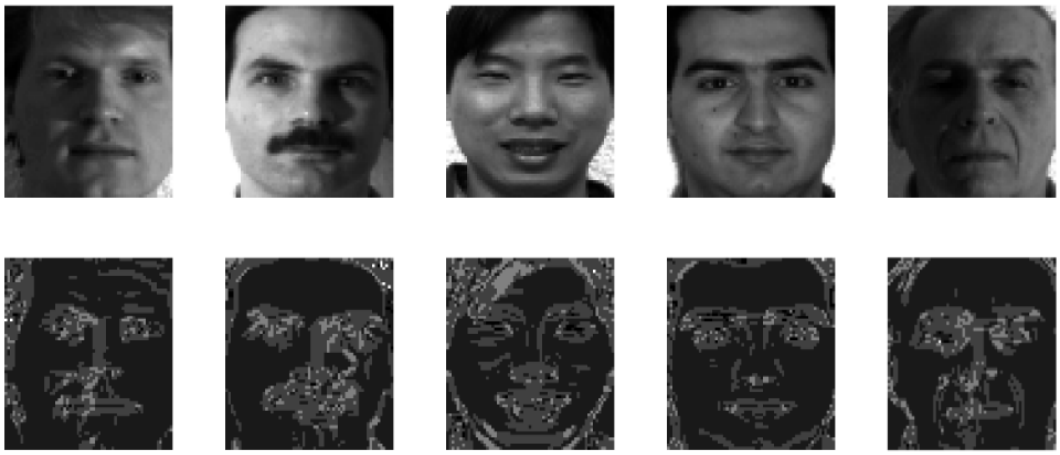


Figure 4.4: Example of feature map images of Yale

perparameters that produce the best accuracy for the model that is trained using feature map.
The value of each hyperparameter that gives the best result is given below -
kernel := poly, c := 0.05, degree := 2, gamma := scale, coef0 := 0.
5-fold cross-validation performed over the dataset using these hyperperameters to get mean training and testing accuracy. Mean training accuracy is 100% , and mean testing accuracy is 97.58% . The 5 fold accuracies of training are [1.0, 1.0, 1.0, 1.0, 1.0] and The 5 fold accuracies of testing [0.9393939393939394, 1.0, 0.9393939393939394, 1.0, 0.9393939393939394] are respectively.

The result proves our methodology enables the classifier to do certaily better classification than most other feature extraction techniques on Yale face database. The confusion matrix is given in Figure 4.9 to have better insights. The comparisons of performances with respect to Yale face database with other methods are given in Table 4.4.

4.2.5 ORL Face Database

ORL Database consists of 400 images, that is 10 persons having 40 images each. All images are frontal and slight tilt of the head. So, we can say there are 10 classes and each class has 40 images in other words. Each grayscale cropped image of the database size is (80,70). Before applying the proposed methodology to get the feature map for each image all images are resized into (70,70). So the size of each feature map image would be (68,68).
The images are flattened to feed to the machine learning classifier. So each of the flattened images' shape becomes (4624,). Support Vector Classifier (SVC) of sklearn is used for classification purposes. The GridSearchCV method of sklearn is used to find out the best hyperparameters of SVC that control the accuracy of the

Table 4.4: Accuracy Comparison for Yale

Method	Accuracy(%)
LPP [60]	70.7
OLPP [60]	81.3
DLPP [60]	72
CLPP [60]	77.3
MPLBP-LH [43]	80.28
LBP [43]	83.4
LBP + PCA [43]	89.2
EL_LBP(horizontal) [43]	93.65
CLTP_S/M/C [43]	83.13
LapCLTP_S/M/C [43]	85.13
DR_LBP [43]	94.09
NCDB-LBPac+FLDA+SVM (RBF) [42]	100
Proposed Method + SVM	**97.58**

model regarding classification. The SVC model takes different hyperparameters. To find out the best hyperparameters several values of the hyperparameters are given to GridSearchCV method and 5 cross-validation is used for training purposes. The model is then trained with all those values in combination to find the best hyperparameters that produce the best accuracy for the model trained using feature maps.
The value of each hyperparameter that gives the best result is given below -
kernel := poly, c := 0.1, degree := 4, gamma := 0.1, coef0 := 0.1.
5-fold cross-validation performed over the dataset using these hyperperameters to get mean training and testing accuracy. Mean training accuracy is 100% , and mean testing accuracy is 93.75% . The 5 fold accuracies of training are [1.0, 1.0, 1.0, 1.0, 1.0] and The 5 fold accuracies of testing are [0.9125, 0.925, 0.9125, 0.9625, 0.975] respectively.

The result proves our methodology enables the classifier to do quite well on the ORL face database. The confusion matrix is given in Figure 4.10 to have better insights. The comparisons of performances for the ORL face database with other methods are given in Table 4.5.

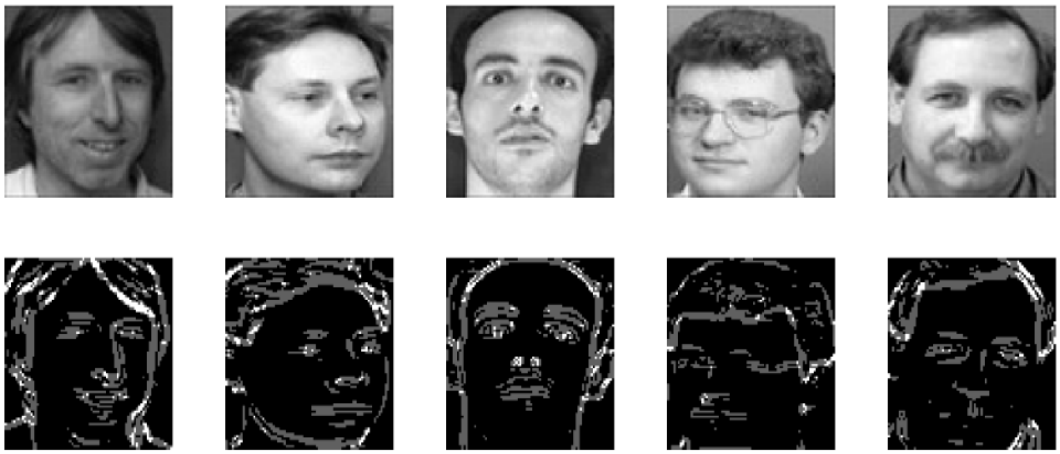


Figure 4.5: Example of feature map images of ORL

Table 4.5: Accuracy Comparison for ORL

Method	Accuracy(%)
PCA [65]	60.3
Wavelet [65]	83.1
LDA [65]	83.8
ESTIM [65]	78.3
Randon and Wavelet [65]	74.4
DOMF [65]	88.4
LPP [60]	92
OLPP [60]	95.5
DLPP [60]	96
CLPP [60]	97.5
LBP [60]	86.1
LBP + PC [60]A	91.4
EL_LBP(horizontal [43])	93.62
DR_LBP [43]	95.95
GABOR + PCA +SVM [66])	98.62
NSST [66])	99.32
Proposed Method + SVM	**93.75**

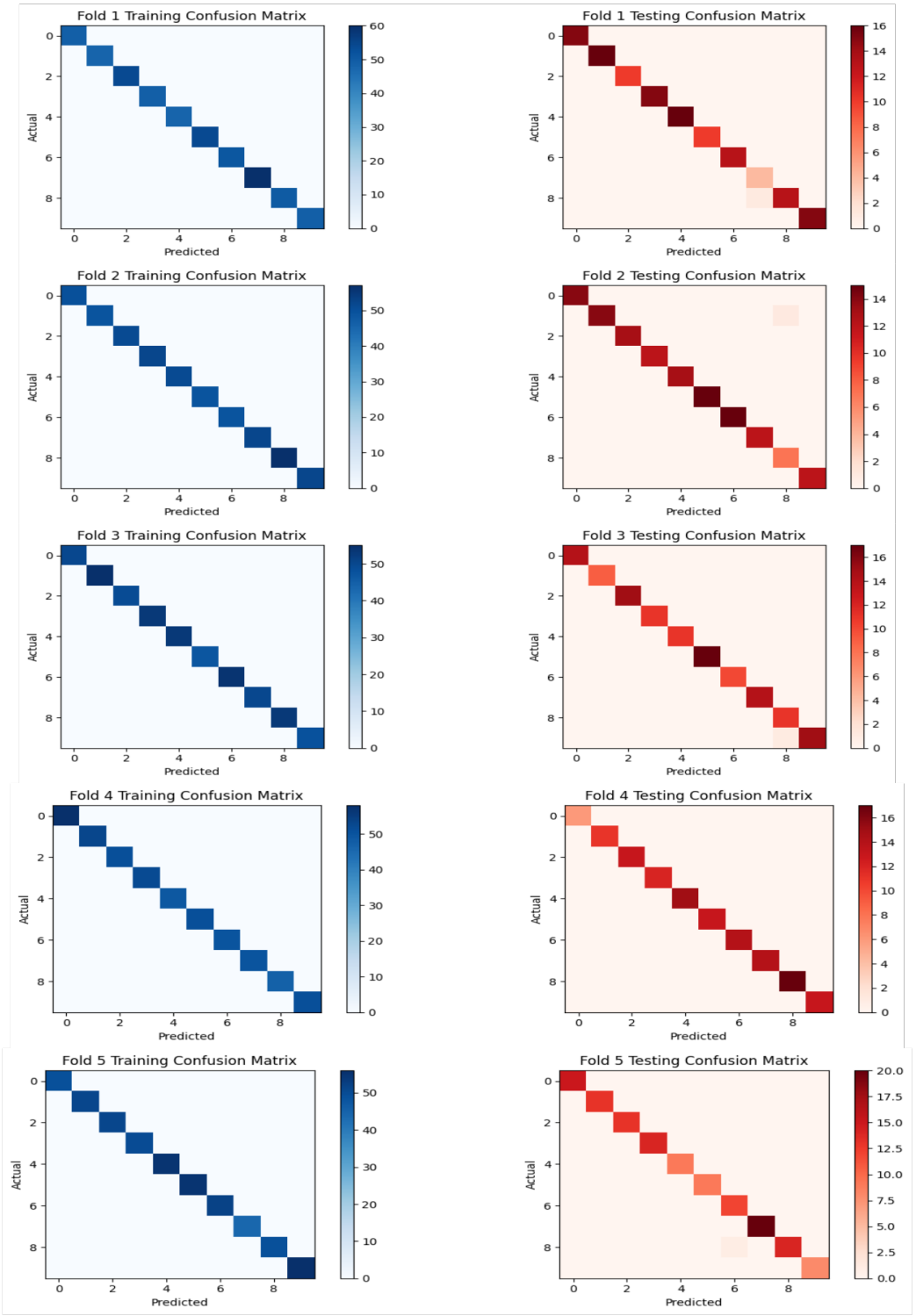


Figure 4.6: Confusion Matrix of Yale B with 5-Fold Cross Validation

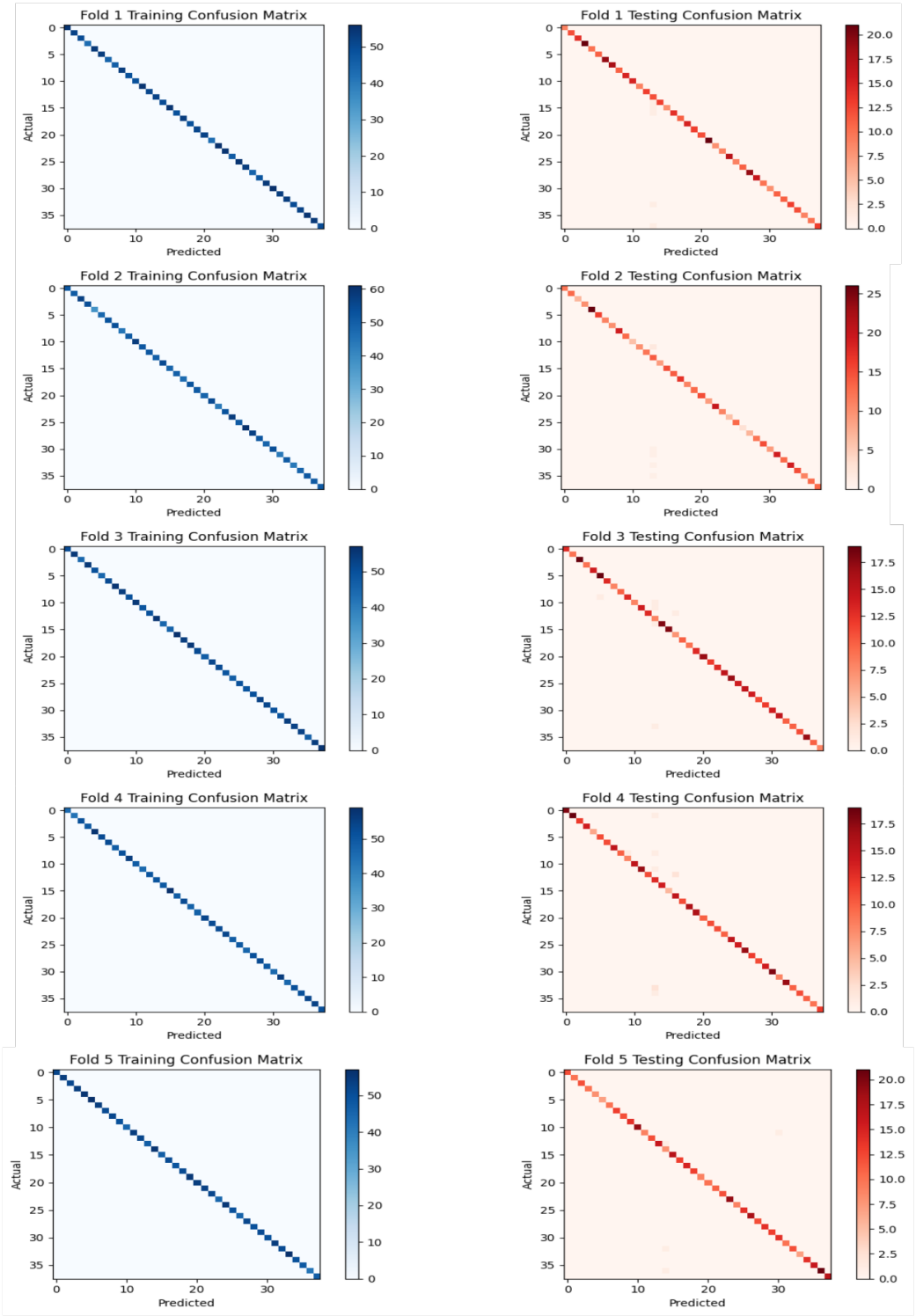


Figure 4.7: Confusion Matrix of Extended Yale B with 5-Fold Cross Validation

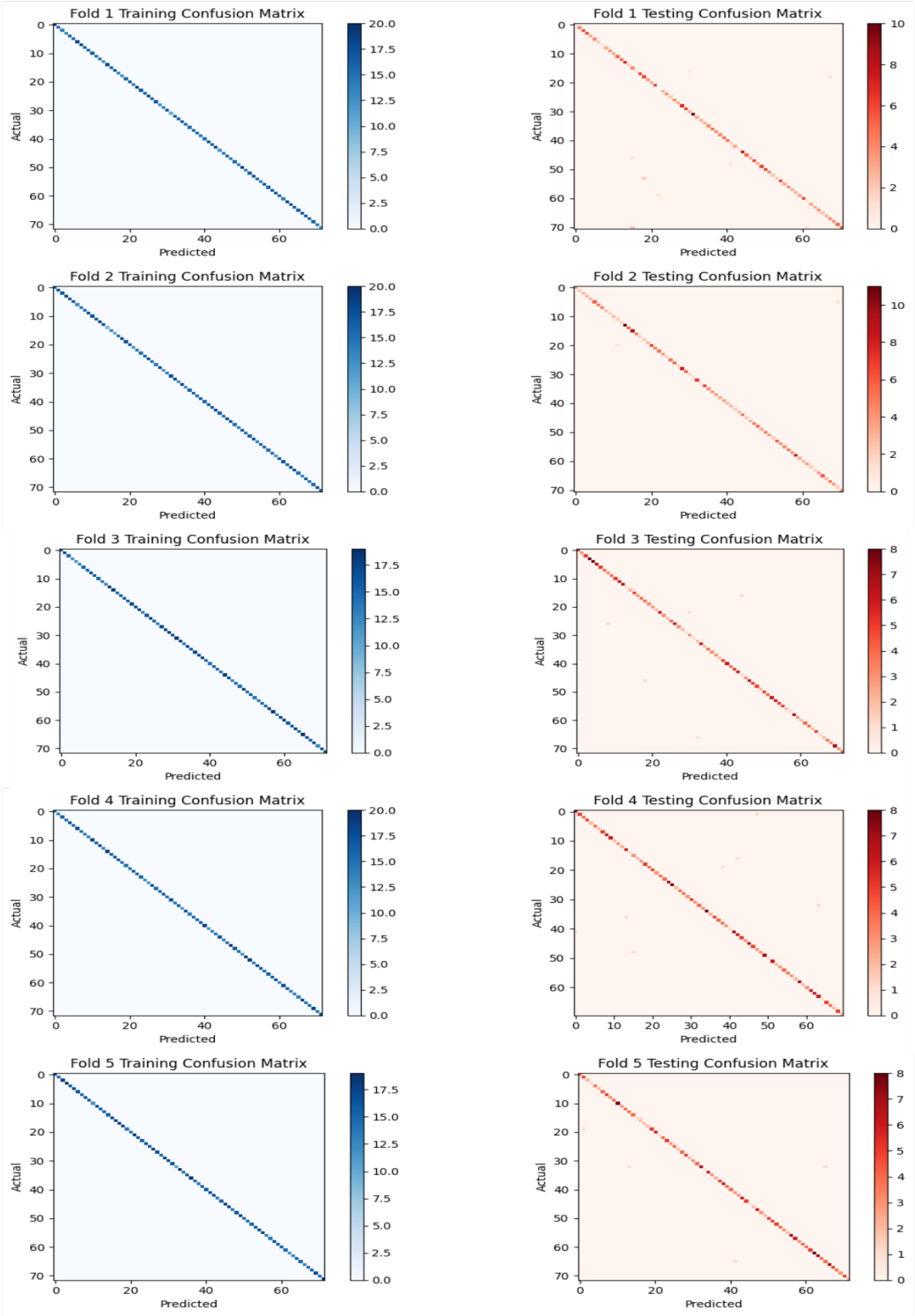


Figure 4.8: Confusion Matrix of Faces95 with 5-Fold Cross Validation

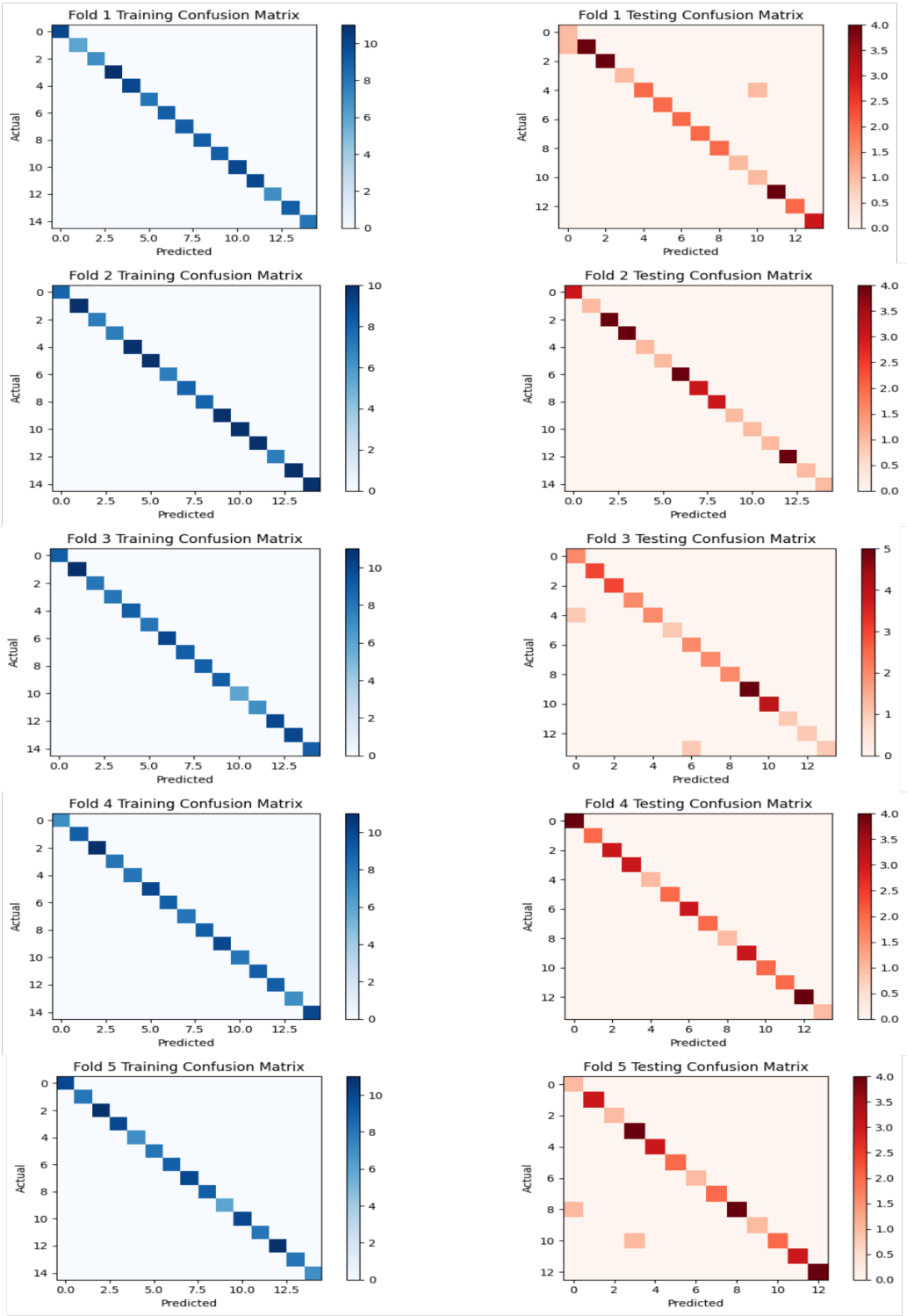


Figure 4.9: Confusion Matrix of Yale with 5-Fold Cross Validation

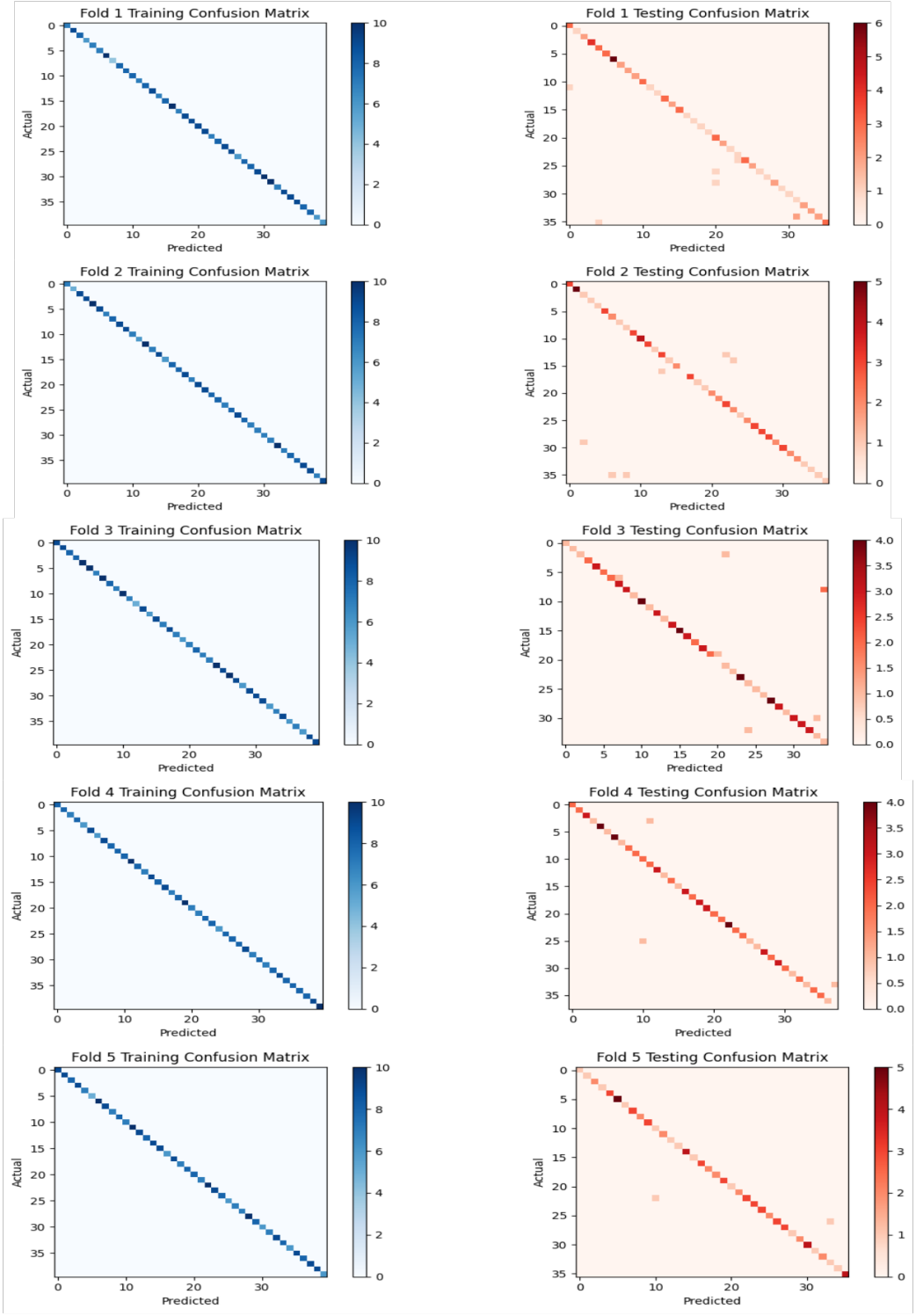


Figure 4.10: Confusion Matrix of ORL with 5-Fold Cross Validation

CHAPTER 5

Conclusion and Future Work

In this chapter, a brief description is given of the overall workflow, challenges, and problems faced during this thesis work. Future scope is also provided to show how this novel feature descriptor could be useful not only in the domain of human face recognition but also in other domains like medical imaging.

5.1 Conclusion of the Present Work

The primary goal of this thesis work is to find a novel feature descriptor or extractor for recognizing human faces in unconstrained environments with significant accuracy that would open a new horizon of research. Also in this thesis work, it is one of the main concerns regarding the feature extractor that it should be simple but would provide distinct features and computationally it would be viable. CBCETTE is being successful regarding these matters. It provides significant accuracies on standard datasets such as Yale, YaleB, Extended Yale B, Yale and Faces95. All these datasets contain human face images in unconstrained environments. It also provides significant results in small datasets such as the ORL face database without any image augmentation or any extra computation.

5.2 Scope for Future Work

The scope of this work is huge. CBCETTE can be used not only just for human face recognition but also for medical image recognition and texture recognition. It is a hand-made feature descriptor that comes with mathematical significance. However, the amalgamation of CBCETTE with the features generated from deep learning models could open a new horizon of pattern recognition or feature learning for any recognition task not only just human face recognition. As CBCETTE is based on properties of edge and texture and it is very much known that edge and

texture play a vital role in generating features of any kind of image, with some more research and subtle new addition with this work, this proposed method could be used to recognize other things also.

Bibliography

[1] L. I. Millett and J. N. Pato, "Biometric recognition: Challenges and opportunities," 2010.

[2] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine*, IEEE, 2004, pp. 184–193.

[3] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[4] F. Galton, "Personal identification and description," *Journal of anthropological institute of Great Britain and Ireland*, pp. 177–191, 1889.

[5] M. O. Oloyede, G. P. Hancke, and H. C. Myburgh, "A review on face recognition systems: Recent approaches and challenges," *Multimedia Tools and Applications*, vol. 79, pp. 27 891–27 922, 2020.

[6] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, "Classical and modern face recognition approaches: A complete review," *Multimedia tools and applications*, vol. 80, pp. 4825–4880, 2021.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[8] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, August 27-29, 2007. Proceedings*, Springer, 2007, pp. 828–837.

[9] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *2008 19th international conference on pattern recognition*, IEEE, 2008, pp. 1–4.

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[11] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE transactions on image processing*, vol. 19, no. 2, pp. 533–544, 2009.

[12] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram fourier features," in *Image Analysis: 16th Scandinavian Conference, SCIA 2009, Oslo, Norway, June 15-18, 2009. Proceedings 16*, Springer, 2009, pp. 61–70.

[13] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I 11*, Springer, 2010, pp. 313–326.

[14] D. Gong, S. Li, and Y. Xiang, "Face recognition using the weber local descriptor," in *The First Asian Conference on Pattern Recognition*, IEEE, 2011, pp. 589–592.

[15] J. Chen, S. Shan, C. He, *et al.*, "Wld: A robust local image descriptor," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1705–1720, 2009.

[16] M. R. Reza, A. Abdolrahaman, and E. A. Reza, "A robust face recognition method using edge-based features," in *2012 IEEE Symposium on Computers & Informatics (ISCI)*, IEEE, 2012, pp. 185–188.

[17] S. U. Hussain, T. Napoléon, and F. Jurie, "Face recognition using local quantized patterns," in *British machive vision conference*, 2012, 11–pages.

[18] B. Yuan, H. Cao, and J. Chu, "Combining local binary pattern and local phase quantization for face recognition," in *2012 International Symposium on Biometrics and Security Technologies*, IEEE, 2012, pp. 51–53.

[19] H.-T. Nguyen and A. Caplier, "Elliptical local binary patterns for face recognition," in *Asian conference on computer vision*, Springer, 2012, pp. 85–96.

[20] J. Ren, X. Jiang, and J. Yuan, "Relaxed local ternary pattern for face recognition," in *2013 IEEE international conference on image processing*, IEEE, 2013, pp. 3680–3684.

[21] K.-C. Fan and T.-Y. Hung, "A novel local pattern descriptor—local vector pattern in high-order derivative space for face recognition," *IEEE transactions on image processing*, vol. 23, no. 7, pp. 2877–2891, 2014.

[22] H.-T. Nguyen and A. Caplier, "Local patterns of gradients for face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 8, pp. 1739–1751, 2015.

[23] X. Chen, Y. Cai, L. Chen, and Z. Li, "Discriminant feature extraction for image recognition using complete robust maximum margin criterion," *Machine Vision and Applications*, vol. 26, pp. 857–870, 2015.

[24] H. Roy and D. Bhattacharjee, "Local-gravity-face (lg-face) for illumination-invariant and heterogeneous face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1412–1424, 2016.

[25] W. Yang, Z. Wang, and B. Zhang, "Face recognition using adaptive local ternary patterns method," *Neurocomputing*, vol. 213, pp. 183–190, 2016.

[26] L. Liu, P. Fieguth, G. Zhao, M. Pietikäinen, and D. Hu, "Extended local binary patterns for face recognition," *Information Sciences*, vol. 358, pp. 56–72, 2016.

[27] Y. Wang, S. Yu, W. Li, L. Wang, and Q. Liao, "Face recognition with local contourlet combined patterns," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2016, pp. 1273–1277.

[28] W. Huang and H. Yin, "Robust face recognition with structural binary gradient patterns," *Pattern Recognition*, vol. 68, pp. 126–140, 2017.

[29] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local directional gradient pattern: A local descriptor for face recognition," *Multimedia Tools and Applications*, vol. 76, pp. 1201–1216, 2017.

[30] A. Kar, S. Sarkar, and D. Bhattacharjee, "Local centre of mass face for face recognition under varying illumination," *Multimedia Tools and Applications*, vol. 76, pp. 19 211–19 240, 2017.

[31] C.-K. Tran, C.-D. Tseng, P.-J. Chao, *et al.*, "Local intensity area descriptor for facial recognition in ideal and noise conditions," *Journal of Electronic Imaging*, vol. 26, no. 2, pp. 023 011–023 011, 2017.

[32] S. P. Ramalingam, R. Nadesh, and N. SenthilKumar, "Robust face recognition using enhanced local binary pattern," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 1, pp. 96–101, 2018.

[33] A. Pillai, R. Soundrapandiyan, S. Satapathy, S. C. Satapathy, K.-H. Jung, and R. Krishnan, "Local diagonal extrema number pattern: A new feature descriptor for face recognition," *Future Generation Computer Systems*, vol. 81, pp. 297–306, 2018.

[34] N. Zhou, A. G. Constantinides, G. Huang, and S. Zhang, "Face recognition based on an improved center symmetric local binary pattern," *Neural Computing and Applications*, vol. 30, pp. 3791–3797, 2018.

[35] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Centre symmetric quadruple pattern: A novel descriptor for facial image recognition and retrieval," *Pattern Recognition Letters*, vol. 115, pp. 50–58, 2018.

[36] B. Hamdan and K. Mokhtar, "Face recognition using angular radial transform," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 2, pp. 141–151, 2018.

[37] D. Bhattacharjee and H. Roy, "Pattern of local gravitational force (plgf): A novel local image descriptor," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 2, pp. 595–607, 2019.

[38] S. Bhattacharya, G. S. Nainala, S. Rooj, and A. Routray, "Local force pattern (lfp): Descriptor for heterogeneous face recognition," *Pattern Recognition Letters*, vol. 125, pp. 63–70, 2019.

[39] V. C. Kagawade and S. A. Angadi, "Multi-directional local gradient descriptor: A new feature descriptor for face recognition," *Image and Vision Computing*, vol. 83, pp. 39–50, 2019.

[40] W. Yang, X. Zhang, and J. Li, "A local multiple patterns feature descriptor for face recognition," *Neurocomputing*, vol. 373, pp. 109–122, 2020.

[41] S. Najafi Khanbebin and V. Mehrdad, "Genetic-based feature fusion in face recognition using arithmetic coded local binary patterns," *IET Image Processing*, vol. 14, no. 15, pp. 3742–3750, 2020.

[42] S. Karanwal and M. Diwakar, "Neighborhood and center difference-based-lbp for face recognition," *Pattern Analysis and Applications*, vol. 24, pp. 741–761, 2021.

[43] S. Najafi Khanbebin and V. Mehrdad, "Local improvement approach and linear discriminant analysis-based local binary pattern for face recognition," *Neural Computing and Applications*, vol. 33, pp. 7691–7707, 2021.

[44] S. Karanwal and M. Diwakar, "Od-lbp: Orthogonal difference-local binary pattern for face recognition," *Digital Signal Processing*, vol. 110, p. 102 948, 2021.

[45] S. Karanwal, "Graph based structure binary pattern for face analysis," *Optik*, vol. 241, p. 166 965, 2021.

[46] M. Najmabadi and P. Moallem, "Local symmetric directional pattern: A novel descriptor for extracting compact and distinctive features in face recognition," *Optik*, vol. 251, p. 168 331, 2022.

[47] H. P. Truong, T. P. Nguyen, and Y.-G. Kim, "Weighted statistical binary patterns for facial feature representation," *Applied Intelligence*, vol. 52, no. 2, pp. 1893–1912, 2022.

[48] S. Karanwal, "Improved local descriptor (ild): A novel fusion method in face recognition," *International Journal of Information Technology*, pp. 1–10, 2023.

[49] J. H. Lai, P. C. Yuen, and G. C. Feng, "Face recognition using holistic fourier invariant features," *Pattern Recognition*, vol. 34, no. 1, pp. 95–109, 2001.

[50] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and Signal Processing: 3rd International Conference, ICISP 2008. Cherbourg-Octeville, France, July 1-3, 2008. Proceedings 3*, Springer, 2008, pp. 236–243.

[51] S. Karanwal and M. Diwakar, "Triangle and orthogonal local binary pattern for face recognition," *Multimedia Tools and Applications*, pp. 1–27, 2023.

[52] S. Karamizadeh, S. M. Abdullah, and M. Zamani, "An overview of holistic face recognition," *IJRCCT*, vol. 2, no. 9, pp. 738–741, 2013.

[53] D. S. Pankaj and M. Wilscy, "Comparison of pca, lda and gabor features for face recognition using fuzzy neural network," in *Advances in Computing and Information Technology: Proceedings of the Second International Conference on Advances in Computing and Information Technology (ACITY) July 13-15, 2012, Chennai, India-Volume 2*, Springer, 2013, pp. 413–422.

[54] J. Keche, M. Wanjari, and M. Dhore, "Face recognition using feature extraction technique,"

[55] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[56] M. A. Muqeet and R. S. Holambe, "Local binary patterns based on directional wavelet transform for expression and pose-invariant face recognition," *Applied Computing and Informatics*, vol. 15, no. 2, pp. 163–171, 2019.

[57] W. Rong, Z. Li, W. Zhang, and L. Sun, "An improved canny edge detection algorithm," in *2014 IEEE international conference on mechatronics and automation*, IEEE, 2014, pp. 577–582.

[58] B. K. Horn, "Determining lightness from an image," *Computer graphics and image processing*, vol. 3, no. 4, pp. 277–299, 1974.

[59] Y. Liu, Y. Tong, Z. Wang, X. Cao, and L. Meng, "Research on dimensionality reduction in unconstrained face image based on weighted block tensor sparse graph embedding," *Signal, Image and Video Processing*, vol. 17, no. 5, pp. 1873–1881, 2023.

[60] Y. Tan, Y. Zhao, and X. Ma, "Contourlet-based feature extraction with lpp for face recognition," in *2011 International Conference on Multimedia and Signal Processing*, IEEE, vol. 1, 2011, pp. 122–125.

[61] M. A. Talab, S. Awang, and M. D. Ansari, "A novel statistical feature analysis-based global and local method for face recognition," *International Journal of Optics*, vol. 2020, pp. 1–17, 2020.

[62] S. Karanwal, "A comparative study of 14 state of art descriptors for face recognition," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 12 195–12 234, 2021.

[63] R. Aggarwal, S. Bhardwaj, K. Sharma, *et al.*, "Face recognition system using image enhancement with pca and lda," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, 2022, pp. 1322–1327.

[64] T. Cevik, N. Cevik, and M. Zontul, "A local-holistic graph-based descriptor for facial recognition," *Multimedia Tools and Applications*, vol. 82, no. 13, pp. 19 275–19 298, 2023.

[65] Y. W. Wong, K. P. Seng, and L.-M. Ang, "Dual optimal multiband features for face recognition," *Expert systems with Applications*, vol. 37, no. 4, pp. 2957–2962, 2010.

[66] Y. Peng, L. Li, S. Liu, J. Li, and X. Wang, "Extended sparse representation-based classification method for face recognition," *Machine Vision and Applications*, vol. 29, pp. 991–1007, 2018.