Human Facial Expression Recognition Using Computational Intelligence Techniques

Ph.D. (Computer Science)

By

Asim Munir

144-FBAS/PHDCS/F16

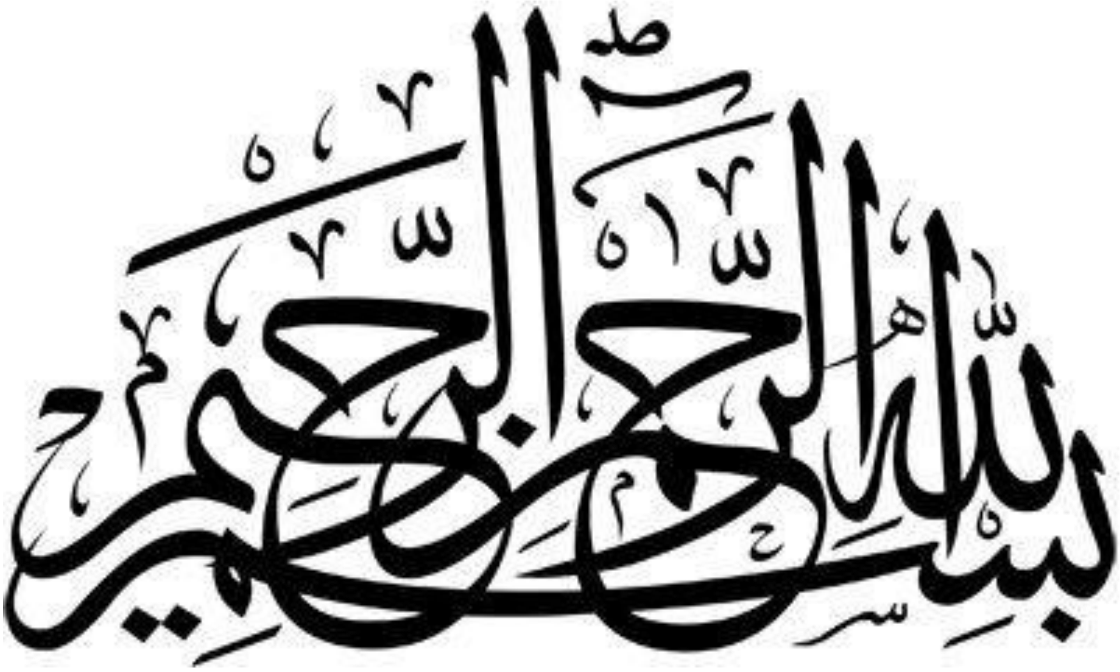
Supervisor

Dr. Ayyaz Hussain

Associate Professor

Department of Computer Science and Software Engineering
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad

(2020)



شُروع اَللہ کے پاک نام سے جو بڑا مہر بان نہایت رحم والا ہے ۔

In the name of ALLAH, The Most Gracious, The Most Merciful.

INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD
FACULTY OF BASIC & APPLIED SCIENCES
DEPARTMENT OF COMPUTER SCIENCE & SOFTWARE ENGINEERING

Date: __/__/2020

FINAL APPROVAL

It is certified that we have read this thesis, entitled ***Human Facial Expression Recognition Using Computational Intelligence Techniques*** submitted by **Mr. Asim Munir**, Registration No. 144-FBAS/PHDCS/F16. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University Islamabad for the award of the degree of **Doctor of Philosophy in Computer Science**.

COMMITTEE

Acknowledgment

I would love to bow my head before Allah Almighty, the Most Gracious, and the Most Merciful, whose countless blessings bestowed upon me throughout my life and especially during the whole course of my Ph.D.

I would like to express my sincere gratitude to my supervisor Dr. Ayyaz Hussain for his constant support, guidance and motivation during my Ph.D. It would never have been possible for me to take this work to completion without his incredible support and encouragement.

Next, I owe my genuine gratefulness to my parents, my sister, my brothers and other members of the family. Without their sincere prayers, I would not have reached this point.

I am also thankful to all my colleagues, friends, students and other persons who have directly or indirectly helped me in the completion of my research work.

Asim Munir

144-FBAS/PHDCS/F16

Declaration

I, Asim Munir, hereby state that my Ph.D. thesis titled ***Human Facial Expression Recognition Using Computational Intelligence Techniques*** is my own work except where due reference is made in the text and it has not been previously submitted by me for taking partial or full credit for the award of any degree at this university or anywhere else in the world. If my statement is found to be incorrect, at any time even after my graduation, the university has the right to revoke my Ph.D. degree.

Asim Munir

144-FBAS/PHDCS/F16

Dedication

To My Parents, Sister, Brothers,

Friends, Teachers

and

Students

Abstract

This research focuses on affective computing i.e. understanding human emotions through facial expressions that are valuable for mankind. Mainstream facial expression recognition is done in ultimate conditions and with obligatory actions associated with expressions. The endeavours are made in identifying and perceiving expressions in sufficiently less constrained. To tackle the imbalance of local illumination, the pre-processing phase increases facial expression recognition accuracy. This fact is demonstrated through experimentation on the static facial expression in the wild (SFEW) dataset which is considered as a challenging dataset and depicts the real-world illumination.

A novel algorithm merged binary pattern code (MBPC) is generated for every pixel. Two bits per neighbourhood are produced to form a 16-bit code per pixel. The effectiveness of the facial expression recognition mechanism is considerably improved by merging these local features. MBPC descriptor captures changes along fine edges and prominent patterns around eyes, eyebrows, mouth, bulges, and wrinkles of the face. This algorithm along with state of the art techniques is tested on the SFEW dataset. The illumination issues in the dataset are corrected using CLAHE in the pre-processing phase to upsurge the facial expression recognition accuracy. Results show that this algorithm works better for real-world images having the challenging light conditions and naturally exhibited expressions. The local illumination variations, as well as global illumination issues, are also managed using robust gradient patterns.

Another technique, that uses the proposed novel feature extraction mechanism named Binary Gradient Structure Descriptor (BGSD) to extract features from the image, is found robust to local illumination changes. The SFEW dataset is pre-processed using a normalization procedure to gain higher facial expression recognition accuracy. This scheme achieves the highest facial expression recognition accuracy of 99.6% using a variant of the support vector machine classifier. The Quadratic SVM classifier behaves consistently and gains higher facial expression recognition accuracy in all experiments.

The results of facial expression recognition are always promising when illumination imbalance and shading are corrected before the extraction of textural features.

Table of Contents

List of Abbreviations

Abbreviation	Explanation
AAM	Active Appearance Model
ACM	Active Contour Model
AFEW	Acted facial expression in wild
BPSO	Binary Particle swarm optimization
CK	Cohn Kanade Dataset
CLBP	Compound Local Binary Patterns
CNN	Convolutional Neural Network
DCT	Discrete Cosine Transforms
DWT	Discrete Wavelet Transforms
EBGM	Elastic Bunch Graph Matching
EMD	Empirical Mode Decomposition
EmtioW	Emotions in Wild
FERET	Facial Recognition Technology
FLD	Fisher's Linear Discriminant
GLCM	Gabor filter and grey level occurrences
HDR	High Dynamic Range
HMM	Hidden Markov Model
HOG	Histogram of oriented Gradients
HWIN	Homomorphic Wavelet-based Illumination Normalization
ICA	Independent Component Analysis
JAFFE	Japanese Female Facial Expression
KNN	K Nearest Neighbors
KPCA	Kernel Principal Component Analysis
KSOM	Kohonen Self-Organizing Map
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
LGBP	Local Gabor Binary Patterns
MBPC	Merged Binary Pattern Codes
MSBP	Multi-stage Binary Patterns

MSF	Masked Correlation filter
NDCT	Normalized Discrete Cosine Transform
PCA	Principal Component Analysis
PIF	Pose Invariant Flipping
PSO	Particle Swam Optimization
QMI	Quadratic Mutual Information
RCPR	Robust Cascaded Pose Regression
RGP	Robust Gradient Patterns
SFEW	Static facial expression in wild
SMO	Sequential Minimal Optimization
SURF	Speed up robust features
SVM	Support Vector Machines

List of Figures

List of Tables

Research Contributions

The following research papers related to this thesis are published/submitted in international journals during the Ph.D. research.

Published

1. Asim Munir, Ayyaz Hussain, Sajid Ali Khan, Muhammad Nadeem, Sadia Arshid. "Illumination invariant facial expression recognition using selected merged binary patterns for real world images", Optik, 2018 [Impact Factor: 1.914]

2. Sadia Arshid, Ayyaz Hussain, Asim Munir, Anum Nawaz & Sanneya Aziz "Multi-stage binary patterns for facial expression recognition in the real world." Cluster Computing 21, pages323–331(2018). [Impact Factor: 1.851]

Submitted

3. Asim Munir, Ayyaz Hussain, Muhammad Nadeem, Muhammad Rizwan Mughal, Sajid Ali Khan. "Local Gradient-Based Technique for Facial Expression Recognition from Images with Poor Illumination", Multimedia Tools and Applications, 2019.

Chapter 1

Introduction

1. Introduction

In recent decades, considerable research work is aimed at intelligent human-computer interaction. The main importance is given to build machines that would complement human life with the help of artificial intelligence. The research efforts are being carried out to devise innovative methods that are needed for advanced human-computer interaction and can be used in robotics or by a psychologist to determine the mental state of humans. This aim can be accomplished by introducing intelligence to computers in such a way that they start interacting with humans in the same manner as humans do with each other. Thus the non-verbal communication method is the future need that is anticipatory and human-centered.

This area needs to encompass a variety of challenges and one of those is to imitate human vision. Human vision involves the process of understanding human functionality and mimics that process intending to complement human life with intelligent machines. The goal is to enable a computer to see people, recognize them and interpret their emotions through gestures and expressions to predict human behavior.

A rapid knowledge about human behavior can be effectively captured by facial expression. It is a non-verbal means of communication among individuals and exhibits human emotions. Facial Expressions contain emotional signals and inner feelings of people about a subject that sometimes become complex to convey through the use of words and the way these words are vocalized. This states the significance of human facial expression in real-world situations where people express their thoughts and feelings when communicating with each other. Therefore, the challenging task of Facial Expression Recognition has attained academic value and has become the focus of research during recent years in the field of computer vision. Human emotions understanding by means of a computer with a certain level of acceptable accuracy is a difficult task because these are produced by highly flexible deformations of the face elements. There is diversity in the demonstration of the same expression by different people in real-world scenarios even belonging to the same region.

In the controlled environment, the task of facial expression recognition is biased because it gives better results when compared to that of an uncontrolled environment.

One of the most important factors is the variation in light conditions in an uncontrolled environment which can introduce unpredictable illumination effects on the face image.

Understanding expressions require the extraction of facial features. The flexibility of these facial features restricts the accurate recognition of expression that is introduced even by a slight disparity in light or the pose that may result in the variation of the local arrangement of texture. The mental state of a human can be identified through recognition of expression in the real-world [1]. Thus psychological study has made researchers consider the challenges in the fields of biometrics, security and human behavior identification. Furthermore, the analysis of facial expression is fruitful in robot vision, facial animation, and virtual reality.

Facial Expression Recognition is composed of features extraction and classification process applied to the image or video to classify facial expression into angry, disgust, fear, happy, neutral, sad or surprised emotion by the use of a computer. A typical facial expression recognition is composed of face detection and localization, image preprocessing, human face segmentation, feature extraction, and classification of expression.

1.1 Facial Expression

Face conveys signals that carry meaningful information that is best interpreted only by humans. The signals contain features that are interesting for social interaction such as gender, age, expression and more. This ability enables us to react differently with a person based on the information extracted visually from the face. To mimic such ability up to a considerable level, computer-based facial analysis is becoming widespread, covering applications such as identity recognition, gender determination, facial expression detection, etc. Features carried by face image are unique to every person and are characterized by highly deformable texture. The expression perception of computers is based on features that can be appearance-based or geometric-based. Appearance-based methods use texture information e.g. bulges, forehead lines; whereas geometric-based expression recognition extracts features from eyes, lips to analyze control points and then identify and recognize expression.

1.2 Facial Expression for Emotion Recognition

Communication among humans can be made through verbal and non-verbal signals. Humans possess the ability to recognize these signals with the least amount of effort and delay but as these signals are highly dependent on individuals so inculcating this capability to computers is the most challenging issue.

About over a hundred years ago, scientist Charles Darwin explained that one of the major means of conveying messages and emotions to another person is the exhibition of facial expression. Over time the human mind is capable of introducing this ability to recognize and comprehend facial expressions to the computers. For the communication among humans through facial expression psychologist, Mehrabian has given a formula "emotional expression language = 7% of the language + 38% of the sound + 55% of the facial expressions" [2].

Social setup is greatly influenced by facial expressions and it has an impact on the emotional states of the community. These are interactive indicators used as a source to convey some predefined meanings to persons in society. In some cases, these are considered more meaningful than words as these are observed visually and directly. The vision systems have offered sufficient improvement to analyze these natural signals. Researches have focused to develop automated tools for emotional and interactive research, affective computing, and responsive human-computer interfaces. Facial expression and analysis have unwrapped a wide range of dynamic applications and that has become a promising area of interest.

The researchers have focused on the feature extraction and recognition of facial expression to empower uneven interaction between computer machine and the users. Following this path, it will become very easy for humans to communicate to identify the mood of humans just like humans do with each other. Computer systems with this capability have a wide range of applications in basic and applied research areas, including man-machine communication, security, law enforcement, psychiatry, education, and telecommunications.

The traditional systems of interaction between humans and computers ignore most of the information communicated through these emotional states and only attend to the intentional contribution of the user. As mentioned, the paradigm is moving towards designs centered on the human being, so that the analysis of the affective states of the user becomes unavoidable. In the near future, humans will not interact with machines only through planned inputs, but also through their behavior, that is, emotional states [3] [4]. Therefore, the computer vision research community has shown great interest in analyzing and automatically recognizing facial expressions. Many application areas can benefit from a system that can recognize facial expressions that is human-computer interaction, entertainment, social robots, cheating detection, interactive video, behavioral monitoring and medical applications [5].

Emotion computing involves all human communicative signals. some studies agreed that the relative contribution of all signals i.e. audio, visual, body gesture are important and depend on the affected state of a human and the environment, while another group emphasis that facial expression is the most important signal in recognition emotional state of communicating person; and it correlates well with body gestures and verbal communication.

Existing approaches to detect the affected state of human are:

- In single modal methods, an image of a human face or speech is used to identify the state.
- In the other techniques, a sequence of posed expressions is given as input to the system.
- A limited number of elementary expressions that convey the emotions serve as a template of emotion. This includes happiness, sadness, disgust, anger, surprise, and fear.

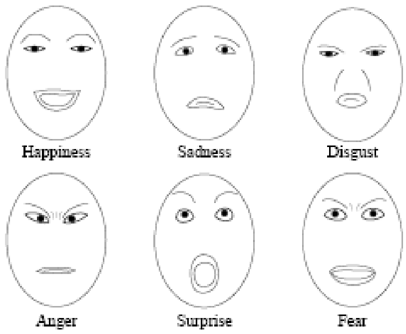


Figure 1.1: Basic Template of Expressions [6]

A neutral face is the one without any expression. This is specified by as face having no visible deformations and describe the state in which a person remains the most in a certain duration of the time. While other expressions of the face involve the changes in the shape of the muscles in any part of the face. Different types of deformations concerning the neutral face characterize the emotions and hence the emotional state of the person. These universal expressions are happiness, sadness, disgust, anger, surprise and fear and these expressions do not change too much from culture to culture.

1.3 Facial Expression Analysis

Facial expressions are classified into seven universal expressions: happiness, sadness, disgust, anger, surprise, fear and neutrality [7]. The variation in muscle movement helps to differentiate between facial expressions and expressions represented by the facial feature. Therefore, the extraction of the feature is a critical task to classify the emotional state/facial expressions.

The recognition of facial expression is a process of extracting features and classification of human facial images through the use of the computer, which allows the computer to conclude the emotional aspects of human facial expression, then achieves the goal of artificial intelligence of the human-computer interaction (HCI) system. In computer science, there are two main methodologies to analyze facial expression.

1. Vision-based method
2. Audio-based method

Some researchers have found that the combination of audio and visual signals gives better results [8]. However, because expressions can be emitted by a face without any sound, vision-based methods remain the hottest research area in human affection. Therefore, my research work focuses solely on methods based on vision. Vision-based methods for the analysis of expressions take an image or their sequence into account to start the work. In general, the system of recognition of facial expression based on vision consists of three steps in Figure 1.2.

1. Face detection
2. Local feature extraction
3. Expression recognition using classification

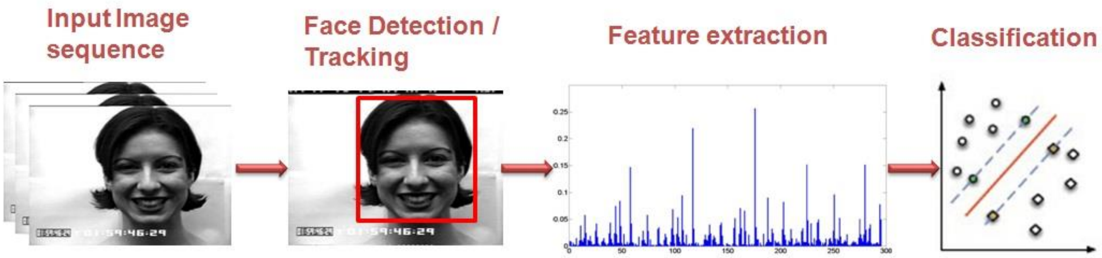


Figure 1.2: The Generic pipeline for FER algorithms [76]

In the first step, the recognition of facial expression consists of detecting the face from the given sequence of images. Tracking the face inside an image in the term as face detection. The second step in facial expression recognition is the feature extraction that differentiates the emotional state of the human. The ideal features should provide a good power to draw a clear boundary between class variations and provide strong bonds within-class.

Generally, facial features extraction methods are divided into two categories:

- Geometry-based
- Appearance-based

The last stage of expression recognition is classification taking features as input. Normally, for classification two approaches are followed. One directly performs the classification considering the features while others identify the specific action units (AU) for the depiction of the facial action coding system (FACS) [9].

It includes both measurements of facial features and recognition of expression. The general approach to Automatic Facial Expression Analysis (AFEA) systems consists of the following steps:

1.3.1 Image Preprocessing

In this procedure, images are made suitable for further processing. Preprocessing steps usually contain image normalization, smoothing, sharpening, contrast stretching or any procedure that is needed to obtain the best visual appearance of the image. The accuracy of expression recognition is highly manipulated by the shading and light variations.

1.3.2 Face Detection and Localization

The detection of a human face from an image is regarded as the foremost thing in facial expression recognition and later its exact boundary is identified. The face structure is identified through its structure. The variation in pose makes this task difficult but handled in a better way using feature invariant methods. The algorithms use features like textures, skin color and combined versions of skin tone, the scale of face, and shape [10]. Viola-Jones is one of the most useful algorithms to accurately detect the face from the image. [63]. Many standard templates of the human face can be applied on a piecewise or whole image to detect the face. Further, the model learning approach may be employed for the same task which may be based on Eigenface, neural networks, support vector machines, distribution-based, Markov model or Naïve-based classifiers [11].

1.3.3 Feature Extraction and Analysis

The facial features are obtained from the expressions generated by the deformation of the face area. After the detection and localization of the face, this mutational evidence makes a basis for the generation of such features. The deformation based essential features are obtained and these are noticeable around the position of eyes, eyebrows, nose, and shape of the lips. The changes or deformations define the texture that is unique as far as the basic expressions are concerned. Researchers focus on these changes to be captured using integral projections, calculating local difference or landmark formations [12]. The descriptors are formed by assembling these features based on some local and/or global criteria. Then the analysis of these features is mandatory to decide the emotional state of humans by using a pertinent classifier.

1.3.4 Expression Classification

The classifier is invoked on the features that are obtained in the previous stage. The classifier builds a neural or statistical model to perform analysis that generally results in the appropriate expression. General techniques for expression recognition are neural nets, SVM's [13], statistical methods, Bayesians Belief nets [14], etc.

1.4 Challenges

Although, for recognition of facial expression different methods achieved good results, but there are still difficulties that must be addressed by the research community.

1. In real-time applications, we require a dimensionally rich feature vector for the recognition of facial expressions. Although it produces good results, but the recognition of facial expressions in real-time is computationally expensive, which makes it almost impossible to use for the recognition in real-time.
2. The real-world applications such as smart meeting, video conferencing and video surveillance required a facial expression system that works properly in low-resolution images. There are a few methods that recognize the emotional state in a very low-resolution image.
3. Images obtained from movies that mimic the real-world situation with complicated backgrounds and varying light conditions are more natural and make the expressions recognition task more challenging and complex. The data set compiled under the controlled environment is mostly used by the techniques. Such a synthetic dataset does not contain poor light and high pose variation issues. Besides this, the expression posed by different persons is artificial and projected using templates that draw a clear boundary between two distinct expressions. [15]. In real life images, the presence of such issues is mandatory that makes the emotion recognition task more challenging.

Other problems include estimation of the intensity of expression, recognition of spontaneous expression, recognition of micro-expressions, incorrect alignment problem, lighting and variation of a pose.

1.5 Research Motivation

The emotional state of a human can be identified by the facial expressions that may require immediate attention in response. This intelligence can be incorporated into computer machines with high reliability that understands the emotional condition to respond accordingly within an adequate time constraint. Moreover, machines would be able to work with humans more naturally if they can understand multiple people or a group of people in front of it. The impact of such a study will be more user-friendly machines that facilitate humans on the same grounds as humans can.

1.6 Problem Statement

A variety of techniques have focused on expressions that are exaggerated versions of the basic emotions and are deliberate. There is a need to develop advanced algorithms that can instantly do processing for the understanding of naturally transpiring human sentiments. In this regard, variation in light, shading, pose and orientation, local face deformations, gender, and ethnicity adds intra-class variability that makes the task of resolution of emotions challenging. The dimension of the features obtained from face images may be high that makes the task of classification slow or it may introduce a problem of overfitting. The classifiers may fail to provide good results due to excessive variability that is present in the expression recognition setup.

1.7 Data Set

A more challenging data set is more useful in obtaining results that are anticipated in the real world. Most of the datasets are fabricated under favorable conditions where the variations in light and pose are not significant so dealing with such datasets is easy.

The dataset used in this research is Static Facial Expressions in the Wild (SFEW) that portrays the real-world environment [52]. SFEW includes static and extreme conditions as far as light, pose and age are concerned, covering seven sorts of facial expressions. Its scope covers a wide age range, which is 1 to 70 years, which makes it one of its kind. In order to mimic the real world conditions Acted Facial Expressions in the Wild (AFEW) is extracted from movies that also contain lively and temporal facial sequence data. The search of expressions and related subject matter with temporal information is made successful with the help of Subtitles for Deaf and Hearing-impaired (SDH) and Closed Caption (CC). Apart from varying light conditions and expression types, SFEW encompasses individuals ranging from 1 to 70 years. The subject matter of the movie clips is annotated with attributes like Name, Age of Actor, Age of Character, Pose, Gender, Expression of Person and the overall Clip Expression and stored information in an extensible XML schema.

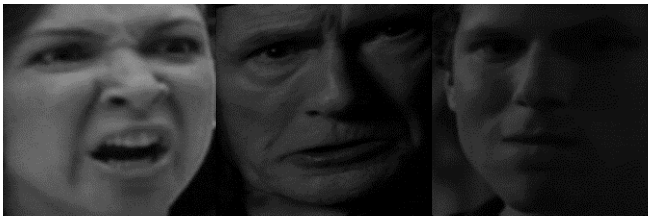
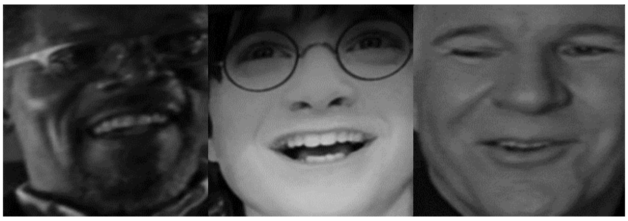
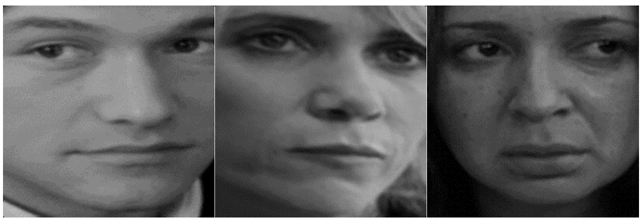
Emotion	Instances	Examples
Angry	164	
Disgust	48	
Fear	76	
Happy	173	
Sad	150	
Surprise	62	
Neutral	148	

Figure 1.3: Examples of Basic Seven Expressions from SFEW Dataset

The assorted idea of illumination conditions in the dataset makes it valuable for the examination of facial expressions as well as for face recognition, face alignment, age investigation and recognition of activities.

The dataset is divided into distinct parts using the 10-fold method. The classification model is built by training on 90% of the images while the other distinct 10% of the images are used for the testing.

1.8 Performance Measures

Mathematical and Computationally Intelligent techniques require the validation of the model as far as quantitative research is concerned. The models are trained using a specific dataset and later the verification is done based on the results. Using these results, Accuracy, Precision, Recall and F-measure are quantified to determine the performance of the proposed system.

In the wake of doing typical Feature Extraction and Classification, the following parameters are employed to discover how successful the model becomes for test datasets. Distinctive performance measurements are utilized to evaluate different Machine Learning Algorithms. For now, we will be focusing on the ones used for classification problems. We can use classification performance metrics such as Accuracy and confusion matrix.

1.8.1 Confusion Matrix

The performance of a classification model is depicted by the Confusion Matrix. It is assumed to be a straightforward and most useful way of confusion metrics utilized for finding the relevance and correctness of the classifier. For a specific dataset involving two or more classes, it can exhibit the behavior of the classifier against each class.

For a binary classification problem, usually, two values are used and likewise, sets of "classes” have two amounts. The actual classes are signified along with columns and the prediction outcomes are represented in rows as shown in Figure 1.4.

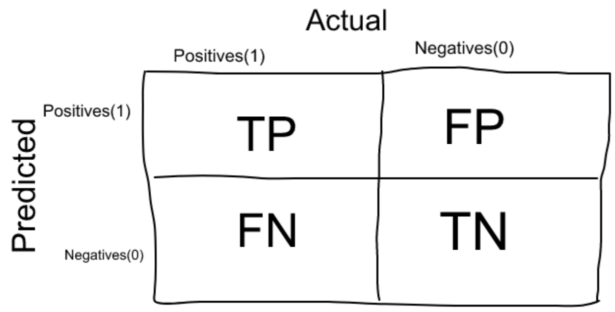


Figure 1.4: Example of Confusion Matrix

Main performance measures use the parameters presented in the Confusion matrix so it provides a base to evaluate the performance of a classifier.

The terms presented by a Confusion Matrix include [49]:

1. True Positives (TP): When the predicted class is True and that matches with the actual case that is normally represented by 1.
2. True Negatives (TN): When the classifier identifies a test as False that equals the real situation and is represented by 0.
3. False Positives (FP): False positives characterize the state when the predicted test case is identified as True but the real case is False. This indicates an error on part of the classifier.
4. False Negatives (FN): False negatives characterize the state when the predicted sample is False but in reality, it was labeled as True. This identifies that the classifier is not yet accurate.

Accuracy in arrangement issues is the number of right forecasts made by the model over various types of forecasts made.

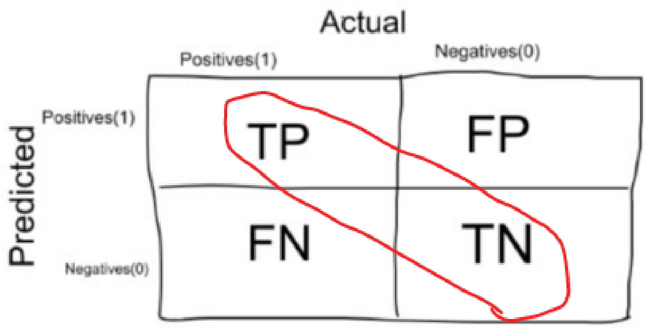


Figure 1.5: Example of Accuracy

The diagonal at 135 degrees shows the predictions called True Positives (TP) and True Negatives (TN) while other cells show the errors made by the classifier called False Positives (FP) and False Negatives (FN). For a given collection of training data, a set of test cases generate to compute the following measures.

1.8.2 Accuracy

An overall exactness of a classifier is dependent on the class that is corrected identified as true to find accuracy as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made} \tag{1.1}$$

Accuracy is a decent measure that gives equal weightage to both false positive and false negative cases of an objective variable.

1.8.3 Precision

For a given prediction class, this measure determines the accuracy and remains unaffected for alike outcomes.

$$Precision = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ true\ predictions + Total\ number\ of\ false\ positives} \tag{1.2}$$

1.8.4 Recall

Recall tells the fraction of occurrences of a particular class that is found relevant in the response of the system. It also termed as sensitivity and computed as:

$$\text{Recall} = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ true\ predictions + Total\ number\ of\ false\ negatives} \tag{1.3}$$

1.8.5 F-measure

A single measure that combines both precision and recall is the F-measure that is obtained using a weighted harmonic mean of these measures.

$$\text{F measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1.4}$$

F-measure obtains high value when both precision and recall are high.

1.9 Thesis Organization

The organization of the thesis is composed of five chapters.

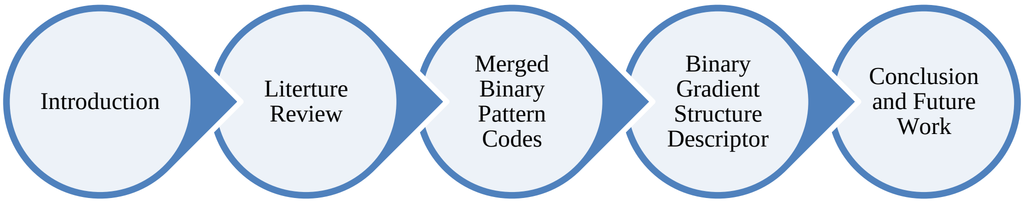


Figure 1.6 Thesis Organization

Chapter 1 includes a concise introduction to the topic of research. A literature review of the related work is presented in Chapter 2 to identify the research gap that lays the foundation to carry out this research. Chapter 3 includes the model that is employed to deal with the varying illumination while retaining the performance of the model. Some additional issues related to emotion recognition supported by the illumination normalization are discussed in Chapter 4. Chapter 5 contains the conclusion and highlights the future aspects of the research.

Chapter 2

Literature Review

2. Literature Review

This chapter describes the major techniques and concepts related to facial expression recognition. In computer vision, the recognition of facial expression is the process of identifying human facial expression, researchers are active in this era since the 1970s and they had done tremendous effort.

2.1 Expression Recognition

Facial Expression Recognition is the non-verbal communication judgment methodology related to the seven basic emotions: anger, fear, disgust, surprise, happiness, sadness and neutral. Each of these expressions possesses different characteristics and unique facial expressions. Figure 2.1 gives an illustration of these expressions.

Figure 2.1: Illustration of Expressions: anger, fear, disgust, surprise, happiness, sadness and neutral.

Based on existing literature, the design of the framework for recognition of facial expression is the three-step process: Face detection, Feature extraction, and expression classification. Dimension reduction was also added to the FER framework for increasing the accuracy rate and time efficiency [18].

2.1.1 Face Detection

In the first step, the recognition of facial expression consists of detecting the face from the given sequence of images. Tracking the face inside an image is termed as face detection. To detect the face from the image templates can be applied. Usually, the techniques for face detection work inside a rectangular region to identify Haar-like features as it is done in the Viola-Jones method [63].

2.1.2 Feature Extraction

The second step in the facial expression recognition framework is most important in terms of differentiation of the emotional state of the human. The ideal features can achieve the high accuracy of recognition if the extracted feature is inadequate the recognition accuracy will be compromised even if we chose the best classifier for facial expression recognition. The feature is extracted through two major approaches texture-based and geometric-based. In geometric based approaches, deals with the location and shape of facial components and texture-based methods, utilize the appearance features to focus on changing the local texture, which is shown to be more reliable [19]. Gabor wavelets [20] are very popular but the feature dimension is huge which causes computational complexity. An appearance-based approach, local binary pattern (LBP) [21], which is adopted by many researchers due to its strong classification capability and better calculations efficiency. In the literature, a large number of texture-based techniques are used for the recognition of expressions, for example, LBP, LGC, HOG, PCA and ICA, etc.

2.1.3 Expression Classification

The classification of expressions refers to labeling the new data by building a model set by training data that contain observations whose category is identified. The most common classification techniques are mentioned below:

- Support Vector Machines
- Decision Trees
- K-Nearest Neighborhood
- Ensemble Classifiers
- Instance-Based Learning
- Naive Bayes Classifiers

2.2 Previous Work

Jyoti et al. [22] provide a survey in zone-based areas of different appearance based feature extraction techniques with different image sizes and block sizes. The K-NN classifier is used to analyze the accuracy and compare the Local Binary Patterns with Local Gradient Coding, LGC-HD, LGC-VD, and HOG. Results show that LGC-VD gains higher accuracy in 16 x 16 block size and 256 x 256 image size.

Yanpeng et al. [23] use the active facial patches for gathering the key information of feature by using the fusion of PCA and LBP feature, they gain better grades for all expressions. All experiments are conducted on the CK+ database and the Softmax classifier used for expression classification. They are planning to implement the proposed technique to recognize the facial expressions from the video sequence.

Yan et al. [24] are interested in overcoming the occlusion problem in expression recognition, they proposed the novel algorithm by fusing the existed techniques to such as HOG and LBP, as HOG mainly focused on contour-based shape feature and LBP extract the features on grey-level images. They calculated the result of HOG with SRC and LBP with SRC than combining the resulted vector for a final decision. The CK+ dataset was used for experiments and expression classification is done by using SRC. There are still some limitations, such as the difficulty in determining the optimal size of local patches, the typical time consumption per image is greater and the classifier used by them is not very efficient for classifying facial expression. The performance of the proposed technique can be improved by using the neural network and the problem of huge feature dimensions can be addressed by using some feature reduction methods to reduce the computational cost.

Wei-Lun et al. [25] proposed a new framework that improves feature extraction from the very noisy images, includes the de-noising mechanism to reduce the effect of the unrelated feature and improve expression recognition. Propose es-LBP an extension of LBP to extract the features. Experiments were conducted on the JAFFE database, SVM and K-NN used for classification of expressions. The result can be improved while verifying that the reduction of the dimension of the unrelated features for the recognition of facial expression is important in the framework of recognition of facial expression.

Ying et al. [26] mainly focused on reducing the computational complexity and remove the useless information they optimize the LCG operator, the experiment conducted on the JAFFE database and the results show that the proposed LGC-HD is more efficient in extracting the features. But there is some limitation that restricts to obtain better grades. If the block number is smaller than each sub-block, the minutiae information cannot be extracted with precision and if the block number is greater than each sub-block the LGC-HD features for each sub-block are redundant and will influence the classification effect. If the LGC-HD operator is optimized using multi-scaling and used in multi-directions for feature extraction, then recognition accuracy will be increased.

Anusha et al. [27] present an efficient algorithm, in which features are extracted by using the robust local binary patterns technique followed by the Kullback Leibler (KL) Divergence classifier. To eliminate the local shading, highlighting and imbalance illumination effects, and image dataset is pre-processed with gamma correction technique. The pre-processing phased increase facial expression recognition accuracy. The JAFFE database is used for experiments and the method achieves the highest accuracy among the distance-based classification methods. But the proposed technique gets confused with classifying the sad and fear class of facial expressions.

Pierluigi et al. [28] provide a comprehensive study on Histogram of Oriented Gradients, experiments conducted on images with different conditions (e.g. image resolution, lighting conditions, etc.) selected from different databases (e.g. CK+, RBD) and for facial expression classification SVM classifier used.

Ling et al. [29] proposed a FER method based on two steps in the first step to extract the histogram function in LBP, ULBP, and LGC. In the second step, the feature vectors are optimized with an appropriate weight. These features are classified with SVM. The experiments were carried out in the CK + database which shows that the proposed method obtains a higher accuracy in recognition of facial expressions and also maintains the time efficiency.

Y Liu et al. [30] proposed the technique, in which most effective facial features are extracted from the salient areas of the face by using the LBP and HOG, the fusion of LBP and HOG causes the huge dimension, so for the reduction of dimension use the PCA technique. Facial expressions are classified by using many classifiers. The experiments are carried out on CK+ and JAFFE dataset, these datasets are pre-processed with gamma correction to achieve high recognition accuracy.

Jadisha et al. [31] proposed a methodology to overcome the occlusions, initially, RPCA used to reconstruct the occluded facial regions, Census Transform Histogram (CENTRIST) feature are extracted from the facial images and also patterns based on LBP, LGC and LGC-HD are extracted then feature vector reduced through PCA and LDA. Experiments were conducted on CK+ and JAFFE databases and expressions classification done through SVM and K-NN classifiers. The proposed method achieves high accuracy of recognition in occluded and non-occluded images without requiring high computational resources.

Faisal et al. [32] proposed an extension of LBP, the extra P bits are combining with the original LBP and named Combined Local Binary Pattern (CLBP). CLBP uses both sign and magnitude information, it is a coding scheme of 16 bits which causes the high dimension. To resolve the issue of high dimensions CLBP is divided into two sub-CLBP patterns of an eight-bit coding scheme named CLBP1 and CLBP2. The experiments performed on CK + and JAFFE datasets and classification of expressions done with the SVM classifier.

Xiaoyang et al. [33] contribute to making facial recognition more efficient and reliable under uncontrolled lighting conditions. The efficient illumination normalization was introduced to eliminate the uncontrolled lighting effects in the pre-processing phase. The advanced Local texture LTP descriptor was introduced which generalizes the LBP fragmentation where noise is low in uniform regions. After illumination normalization the fusion of two techniques i.e. Gabor wavelet and LBP for feature extraction and discriminative feature extraction technique. The experiment was carried out in the large-scale dataset FRGC-204 that contains a wide range of different lighting conditions.

Siddiqi et al. [34] propose a solution to the problem regarding varying light conditions used both in training and testing image subsets. The HL_FER system also provides the solution to the problems such as Similarity of expressions and accuracy. It includes the preprocessing steps to eliminate the light effect and a new automatic face detection scheme is presented. The HL_FER handled n cross-validation of three different datasets

(Cohn-Kanade dataset, the JAFFE dataset, and the AT&T dataset). It proposed, validate the accuracy that previous systems were unable to provide as those systems first worked doesn't involve the preprocessing step. In preprocessing using GHE it allows eliminating the light effect. It utilizes the techniques to work on the gray level and skin tone simultaneously to detect the face and uses the PCA and ICA techniques for global and local feature extraction. HL_FER also focuses on expression similarities, in this work the expressions are divided into three categories based on different parts of the face. It uses the LDA technique and HMM for recognizing the expression.

Huang and Yin [35] proposed a framework for human identification through robust facial features. The robustness is based on the gradients of the image called binary gradient patterns that are considered powerful representatives of local features. It combines the characteristics of both LBP and IGP domain to explore the relationship of the neighboring pixels. Local features are represented as binary strings. to enhance the discriminative power of the recognition system. Local features extracted by BGP show stronger orientation power than LBP and Gabor description. It increases discriminative power by simplifying computational capacity.

GBP is implemented in 3 steps:

- Compute the image gradient from multiple directions
- Encoding them into binary strings
- Division of structural and non-structural patterns

The experiments were conducted on the ARFace dataset and classification done through Fisher's Linear Discriminant Analysis (FDA) classifier use for recognition.

A thorough survey is presented by Shyla and Punithavalli [36] regarding the performance of numerous emotion recognition techniques. For the further improvement of the results, they supplemented the suggestions that may be incorporated. In order to effectively compare different techniques on the same scale meta-analysis of relevant data is also suggested. In addition to it, the color component of the image should also be explored. The techniques should handle the problems that are the property of real-life images including orientation variation and shading

Rasoulzadeh [37] proposes a facial expression recognition scheme based on fuzzy logic. As the first step in this technique, numerical data is extracted from specific areas of the face. A fuzzy model is given the data obtained from eyes, an area between both eyes including eyebrows, lips region and area between nose and lips. A triangular membership function is used for mapping between input and during the Fuzzification operation. This system achieves good accuracy apart from being simple at the same time for the JAFFEE database.

An approach that used LBP to extract features and then use of PCA to reduce the dimensionality is proposed by Yuan Luo et al. [38]. The reduced feature set is then classified by SVM to identify expression type. After experimentation for an intelligent wheelchair, they have claimed an accuracy of 89% and 83% for favorable and low illumination respectively.

Jun-Yong Zhu et al [39] explore the homogeneity criteria based on usage of either magnitude or direction along with the consideration of wavelength of different frequencies of light. Thus a new descriptor Logarithm Gradient Histogram is defined that covers all aspects of the illumination and is composed of two parts named logarithm gradient orientation (LGO) and logarithm gradient magnitude (LGM). The model is also capable of handling noise along with illumination imbalance.

Awais et al [40] present a new paradigm for the successful detection of face expression from face images in real-time. Contrary to other methods that take a long time dividing the picture into blocks or entire face images, the proposed method extracts the discriminatory feature from outgoing facial regions. In fact, by selecting the most biased features they reduce the data dimension. In case of occlusions, lighting and in the presence of noise, the proposed system can provide a high level of recognition accuracy. Three publicly available difficult data sets, such as MMI, CK+, and SFEW, are used to illustrate the robustness of the proposed system. Experimental findings demonstrate that the performance of the proposed framework is better than existing techniques, which show that geometric features with appearance-based features have considerable potential. The proposed WLD method works well on face images to extract the salient characteristics, but the variance of the local strength cannot be effectively interpreted using the standard WLD because various orientations of the neighborhood pixels are ignored. In the future, the experiments with ethnographical data sets are planned to take on this problem.

Yingying Wang et. al. [41] introduced a model to solve the problem of using the entire face image as input only. The program can generate unnecessary information and skip some important information during the extraction of the feature. It proposes a new CNN system that takes full advantage of three main regions based on the auxiliary sub-regional model in this paper and that modifies the learning results of the main task by giving different settings to increase the definite accuracy rate. The suggested approach has been tested based on four publicly available databases for facial expression: JAFFE, CK+, FER2013 and SFEW. The new method has shown better performance than the most advanced methods.

Sajjad et. al. [42] published an overview of the facial expression hybrid approach, incorporating local and global characteristics. Factual extraction is done with a uniform local ternary pattern (U-LTP) descriptor fusing the histogram of oriented gradients (HOG). Such features are derived not from various parts of faces such as eyes, nose and mouth, but the entire face picture. After the experimental analysis with the ULTP descriptor further improvement is explored for the dataset that contains noise and occlusion by choosing the most suitable HOG-parameters. The facial feeling is evaluated and categorized into seven global emotions: Positive, Anger, Anxiety, Disgust, Sad, Shock, and Neutral. A feature vector is obtained after fusing the features obtained through HOG and ULTP separately. The emotion classification is then performed on this feature vector. The performance of the model is evaluated on three publically available datasets that are JAFFE, MMI, and CK+. Emotion recognition accuracy achieved is 95.71, 98.20 and 99.68 for JAFFE, MMI and CK+ datasets respectively.

A. Barman et. al. [43] present this article considers (i) the standard distance signature obtained from the Active Appearance Model (AAM) grid, (ii) the standard texture signature derived from salient grid points, (iii) the stability indices resulting from these signatures and (iv) the related statistical measures, including a collection of training features for artificial neural modeling, such as the Multilayer perceptron, the Radio Perceptron, etc. The test results obtained after experiments support the effectiveness of the proposed technique on Cohn-Kanade (CK+), Japanese Female Facial Expression

(JAFFE), MMI and MUG databases. The combined distance-texture (D-T) signature is convincingly more effective than the distance signature. The efficiency of the technique proposed based on the combined D-T signature is demonstrated by its very positive performance compared with other arts.

I. M. Revina et. al. [44] provides an efficient FER approach by using Multi-Support Vector Neural Network (W-GOA-based MultiSVNN) as a proposed Whale-Grasshopper Optimization algorithm. The facial image characteristics are obtained with the Scale-Invariant Feature Transformation (SIFT) and the Scatter Local Directional Proponents (SLDP). In order to recognize facial expressions, the extracted features are categorized using the suggested classification. The proposed face recognition approach improves the accuracy of identification. The studies are performed using a database, such as the AU-Coded Expression Database of Cohn-Kanade and the Japanese Female Facial Expression (JAFFE) database. In terms of accuracy, TPR and FPR, the proposed algorithm is better than current methods and the values are respectively 0.96, 0.96 and 0.009.

Y. Yan et. al. [45] suggest the use of low-resolution, filter learning facial expression recognition. A novel IFSL method has been developed to provide efficient facial image representation. In this way, a new IFSL method is developed. The proposed IFSL approach mainly involves three steps: First, we integrate image filter learning into the Linear Discriminant Analysis (LDA) optimization process. A collection of discriminatory images (DIFs) filters corresponding to different facial expressions are learned by optimizing the cost function of LDA. The second element is to add the images filtered with the learned DIF to create the combined images. In the end, a regression learning method is used for subspace education, when the combined images generate an expression-conscious transformation matrix. The IFSL effectively removes nonrelevant information while retaining useful information in facial images, based on the transformational content. Experimental results are given on several common facial expression datasets to illustrate the efficiency of the proposed IFSL for the identification of LR facial expression. The proposed method shows superior results compared with several state-of-the-art approaches.

The literature reveals that the techniques have used the datasets that have fewer challenges because such data is obtained under controlled and laboratory environments where partial or all challenging factors can be controlled. Significant features play an important role in the recognition of facial expressions. In order to highlight the details that help in the extraction of such features, image enhancement becomes a necessary part of the pre-processing stage. Some techniques have done manual cropping of the data that compromises the usefulness of the system.

Table 2.1: Summary of literature review

Sr. No.	**Literature**	**Main Idea**	**Limitations / future work**	**Method / Technique / database**
01	"FACIAL EXPRESSION RECOGNITION ALGORITHM USING LGC BASED ON HORIZONTAL AND DIAGONAL PRIOR PRINCIPLE" [26] **(APRIL 2014) – ELSEVIER**	The LGC operates along the horizontal and diagonal direction. Then dimensionality is reduced to discard redundant features while preserving useful data that capture the texture from the face image. Real-time problems can be addressed using the LGC-HD operator.	**Limitations:** Local minutiae cannot be effectively located if the number of blocks is small as localization is not achieved. If the number of blocks is high then features inside a block may become redundant that will reduce the performance of the classifier. **Future Work:** The operator can be further optimized by	LGC based on horizontal and diagonal JAFFE database

			considering the features on multiple scales and different directions.	
02	"FACIAL EXPRESSION RECOGNITION USING LOCAL BINARY PATTERNS AND KULLBACK LEIBLER DIVERGENCE" [27] **(2015) – IEEE**	The pre-processing stage is comprised of five steps. The texture features are extracted using an LPB operator after the detection of eyes in the face. These features are represented as histograms. Using the basic expression a template is formed. The dissimilarity is measured as Kullback Leibler value to find the deviation with the test histogram.	**Limitations:** Confusion with classifying the sad and fear classes. **Future Work:** The video stream can be used as input to dynamically recognize the expressions.	Viola-Jones Algorithm in preprocessing LBP for features representation Kullback Leibler Divergence for emotion classification JAFFE Database
03	"FACIAL EXPRESSION RECOGNITION AND HISTOGRAMS OF ORIENTED GRADIENTS: A COMPREHENSIVE STUDY" [28]	An algorithmic pipeline pattern based on three stages is used algorithmic pipeline pattern is used Histogram of the gradient (HOG) applied on the registered face	**Limitations:** Non-frontal face issues Covers the orientation only in the range of [-30,30] degrees	Yale database B Chon-Kanade (CK+) Histogram of gradient(HOG) descriptor

	(2015) – Springer	SVM (support vector Machine) performs classification HOG perimeters are tuned		SVM (support vector Machine)
04	"FACIAL EXPRESSION RECOGNITION BASED ON TWO-STEP FEATURE HISTOGRAM OPTIMIZATION" [29] **2016 - IEEE**	The technique extracts the feature histogram based on the black-based LBP, LGC and ULBP. The feature vector contains information about the patterns based on pixels relationship. Local information inside a small region is exploited to produce patterns. Comprehensive texture information is obtained after the concatenation of histograms representing small regions. These feature vectors are optimized by a suitable weight.	**Limitation:** Weight values and block numbers are important to recognize quality. Use Statistic-based method **Future work:** Find a way to automatically obtain the optimal configuration so that the recognition rate can be further improved.	LBP, LGC and ULBP SVM CK+ database
05	"FACIAL EXPRESSION RECOGNITION WITH FUSION	Image normalization is done to obtain more useful features from main face areas.	The expression recognition rate is dropped if landmarks are	LBP, HOG for feature extraction

	FEATURES EXTRACTED FROM SALIENT FACIAL AREAS" [30] **March 2017 – Sensors**	Face and face landmarks are detected and the face is divided into several patches and compares it with a neutral face. These areas are utilized to obtain LBP and HOG features. After post-processing using the Z-score normalization of the features, classification is done. The feature set dimensionality is reduced using Principal Component Analysis.	not correctly identified. The effectiveness of gamma correction is limited by the less amount of data.	PCA dimension reduction JAFFE & CK+ database Softmax is used for classification. The Z-score method is used to fuse the features. Gamma correction is employed for feature data correction. Face landmarks are represented by a mixture of trees with a shared pool.
06	"FACIAL EXPRESSION RECOGNITION USING EXPRESSION SPECIFIC LOCAL BINARY PATTERNS AND LAYER	The scheme chooses three types of image variations to extract es-LBP data: large-scale, small-scale and interesting feature regions. The last part is further split into two portions	**Limitations:** The elimination of irrelevant features is to be verified for the proposed framework.	expression-specific Local Binary Patterns (es-LBP) PCA Marginal Fisher Analysis (MFA)

	DENOISING MECHANISM” [18] **(2013) – IEEE**	such that features can be extracted independently to approach the statistical composition. The size of the feature set is reduced by Marginal Fisher Analysis.		JAFFE database
07	“RECOGNITION OF OCCLUDED FACIAL EXPRESSIONS BASED ON CENTRIST FEATURES” [31] **(2016 – IEEE)**	The first stage of the technique initially restores the occlusion. The reconstruction uses a dual technique based on the RPCA principle. Then, are automatically detected. Census Transform Histogram produces a feature set from the of features is extracted from the facial fiducial points. The feature vector is converted into lower dimensional space for feature reduction. Then classification is performed using the reduced feature set. Four variations of PCA, LDA, KNN and SVM is	**Future Work:** The real occlusion is hard to handle that includes caps, facial hair, beard, sunglasses and other accessories. The CENTRIST scheme is to be improved for such challenges.	LBP, CENTRIST, LGC and LGC-HD CK+ & JAFFE K-NN & SVM

		used for extensive experimentations.		
08	"ACCURATE AND ROBUST FACIAL EXPRESSIONS RECOGNITION BY FUSING MULTIPLE SPARSE REPRESENTATION BASED CLASSIFIERS" [40] **AUGUST 2014 – ELSEVIER**	This paper focuses on overcoming the occlusion problem in FER. LBP is mainly used to extract texture features whereas HOG finds the contour features from the image. HOG and LBP are joined with SRC to obtain the characteristics of both texture and contour for facial expression recognition from image patches. The emotion type is determined by the classifier using these combined features. Experiments are conducted on eye region and mouth region occlusions.	**Limitations:** -The size of the patch cannot be estimated easily. -The classifier is not efficient -The time complexity of the method is high. **Future Work:** -Improve the performance by using the Neural Network - The use of a hybrid nature feature requires the use of dimensionality reduction mechanism.	HOG & LBP CK+ database SRC
9	"FACIAL EXPRESSION RECOGNITION	It selects three parts of the face including the	**Future Work:** Recognize the basic	LBP, ULBP & PCA

	WITH PCA AND LBP FEATURES EXTRACTING FROM ACTIVE FACIAL PATCHES" [23] **DECEMBER 2016 – IEEE**	forehead, cheeks and mouth, for classification. To obtain the local features, uniform pattern LBP is used. A regression classifier yields good classification accuracy on the features obtained after applying PCA. The pixel values are normalized before fusing. Then the concatenated LBP features are used by PCA to reduce the dimensions.	expressions from the videos using the proposed technique.	CK+ database Softmax regression classifier
10	"ROBUST FACE RECOGNITION WITH STRUCTURAL BINARY GRADIENT PATTERNS" [35] **MARCH 2017 - ELSEVIER**	A gradient-based robust technique to represent face texture known as binary gradient patterns (BGP). The relationship between neighboring pixels is identified using the gradient technique to represent LBP and IGO domain binary features structure. The orientation of the gradient reveals a stronger relationship to	**Future Work:** Enlarging the neighborhood and likewise increasing the number of directions the performance can be improved.	BGP and LPB ARface dataset Fisher's Linear Discriminant Analysis (FDA)

		better discriminate the texture. It is composed of three steps: Compute the image gradient from multiple directions Encoding them into binary strings Division of structural and non-structural patterns Fisher's Linear Discriminant Analysis (FDA) classifier is used for classification.		

2.3 Problem Statement

It is obvious that the presence of shading and the effects of local illumination in the images has remained a challenge for the research community to handle. This restricts the techniques to obtain higher accuracy and precision in facial expression recognition.

Apart from that, the current facial expression recognition system works well in a relatively controlled environment where lighting conditions are good and variations in the pose of an individual are low but tend to suffer when these variations are high. Further, the results have deteriorated if the expressions are more natural rather than deliberate. Images obtained in the uncontrolled environment where the illumination is close to the real world, high variations in posture and incorrect alignment problem requires the development of robust methods.

2.4 Research Questions

A thorough study has revealed that handling the effects of reflections, shading and imbalanced illumination effects are challenging for the researcher to deal with and gain better results. Such challenges raise the need to find responses to the following questions:

1. To what extent the Facial Expression Recognition system improves the results the local illumination effects are handled?
2. How to improve the existing methods or how to design new classification strategies to gain higher classification accuracy and better computational efficiency?
3. What will be the effect of dimensionality reduction on the computational performance of the whole scheme?
4. Which classifiers offer better performance on the local features extracted from face images and why?

2.5 State of the Art Techniques for Facial Expression Recognition

Features extraction techniques for facial expression recognition is divided into two main categories: geometric feature-based methods and appearance-based methods [46].

Geometric-based methods extract the features based on shape including the distances between points identified on the face, curves, blobs or other geometric parameters. The performance is degraded in the presence of noise or when illumination is poor while using geometric features.

On the contrary, the appearance-based features capture the local texture of the face. The spatial values of the intensity of the central pixel and the neighboring pixels inside a certain radius help to characterize the texture. The sensitivity of these features is low in the presence of shading and noise and thus is widely adopted to obtain high accuracy of the model.

Some of the appearance-based methods are discussed in the following section.

2.5.1 Local Binary Pattern (LBP)

Local Binary Pattern is featured as grayscale and invariant surface crude that depicts the spatial organization of the neighborhood surface of an image. A nearby neighborhood around every pixel of an image is selected by the LBP operator to capture the features binomially. The working of the LBP operator is precisely described as:

$$LBP_{P,R}\ (x_c, y_c) = \sum_{P=0}^{P-1} s\left(i_p, i_c\right) 2^P \tag{2.1}$$

$$s(x) = \begin{Bmatrix} 1, & if\ x\ \geq 0 \\ 0, & if\ x\ < 0 \end{Bmatrix} \tag{2.2}$$

Here, i_c is the gray estimation of the middle pixel (x_c, y_c), its neighbors gray estimation is i_p, P is the quantity of neighbors and neighborhood radius is represented by R. The fundamental LBP encoding process is represented in Figure 2.2.

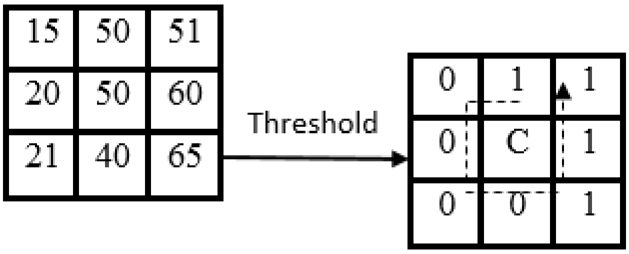


Figure 2.2: The basic LBP operator encoding process.

When exercised, the LBP operator finds the variations in the gray values around a pixel C to obtain a P-bit code. 8-neighbours are involved in the generation of individual bits considering a central pixel which are then arranged to provide a binary sequence of P-bits. On the off chance, a neighbor does not precisely fall into a pixel location, at that point the estimation of that neighbor is evaluated utilizing a bilinear interjection. The LBP descriptor of a block is represented by the histogram of that block obtained after applying the LBP operator.

The regularity adjustment in the pattern can be achieved by modifying the basic LBP operator to uniform LBP. The critical region of the image tends to provide more regularity. Such a uniform code is 00111111 where there is one transition from 0 to 1. Ojala et al. [47] observed that the uniform patterns of LBP are the central properties of the surface, which give a vast greater part of all back agony designs show in any surface picture. In this way, uniform patterns can portray critical nearby surface data, for example, a brilliant spot, level territory or dull spot, and edges of the positive and negative variable arch.

2.5.2 Combined Local Binary Pattern (CLBP)

The intrinsic LBP operator is expanded to generate 2×P binary CLBP codes for the focal pixel in light of the gray level estimations around a neighborhood. This feature extraction technique utilizes two bits to generate codes. One bit captures the sign and the other one holds the magnitude of the contrast between the middle and the neighboring gray values. In this scheme, the primary piece indicates the distinction between the middle and the neighboring pixel's gray value. The other item is utilized to encode the extent of the distinction regarding edge esteem, which is the mean difference M_{avg} of the contrast between the middle and the neighboring pixel values. The CLBP operator turns to "1" if the extent of the contrast between the central and the comparing neighbor is more prominent than the M_{avg} limit. Else, it is set to 0. In this manner, the pointer $s(x)$ is supplanted by the accompanying capacity:

$$s(i_p, i_c) = \begin{cases} 00 & i_p - i_c < 0,\ |i_p - i_c| \le M_{avg} \\ 01 & i_p - i_c < 0,\ |i_p - i_c| > M_{avg} \\ 00 & i_p - i_c \ge 0,\ |i_p - i_c| \le M_{avg} \\ 11 & otherwise \end{cases} \tag{2.3}$$

Where i_c is the gray estimation of the middle pixel, its neighbors P gray estimation is i_p and M_{avg} is the average magnitude difference amongst i_p and i_c in the nearby neighborhood. The encoding process of the CLBP operator is shown in Figure 2.3.

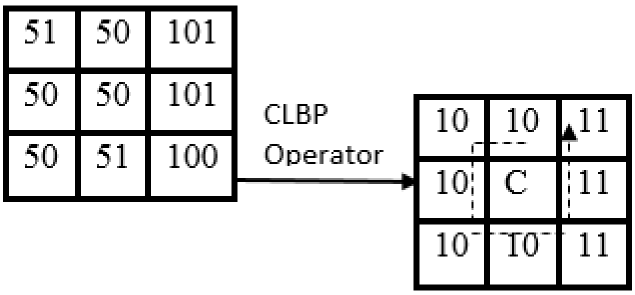


Figure 2.3: The basic CLBP operator encoding process.

In a CLBP technique encodes an image that computes a 16-bit code around a 3×3 neighborhood of a pixel. Thus it generates 2^{16} binary patterns after the utilization of 16-bit codes that help in labeling the pixels. The feature vector representing binary patterns can be decreased by examining some neighborhood data. This technique uses an approach where all binary CLBP patterns are isolated in two sub-CLBP models. To generate a sub-CLBP pattern, the bits are compared to half of the neighbors, where the total number of neighbors are represented by P in a 3x3 window. Typically, in a nearby neighborhood, the two sub-CLBP motifs are formed by linking the comparative assumptions of the binary grouping $(1, 2, 5, 6 \ldots 2P-3, 2P-2)$ and $(3, 4, 8 \ldots 2P-1, 2P)$, separately from the first CLBP 2P-bit [48].

As such, a 16-bit CLBP pattern is separated into two sub-CLBP 8-bit patterns, where the principal caption CLBP1 is acquired by linking the bit values relating to the neighbors in the four directions including north, east, south and west, separately. The second sub-CLBP pattern sub-CLBP2 is acquired by connecting the bit esteems comparing the neighbors in the north. The two sub-CLBP patterns are dealt with as discrete binary and consolidated codes amid the feature vector generation as indicated in Figure 2.4.

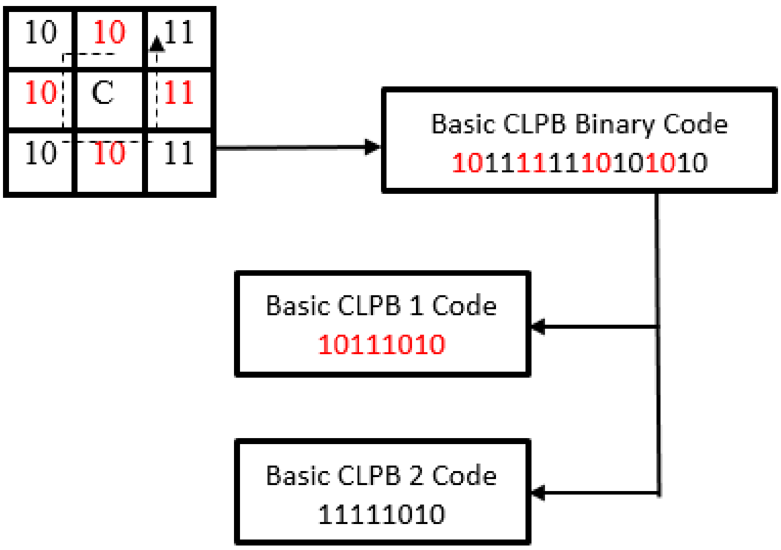


Figure 2.4: Division of 16-bit CLBP code into two sub-CLBP 8-bit portions.

2.5.3 Local Gradient Coding Horizontal and Diagonal (LGC-HD)

If we have a large set of facial data, we use the traditional LBP operator for feature extraction. The extracted feature vector's dimensions are high which reduced the recognition rate of the classifier. Assume that m is total sub-regions of the image, then differential feature dimensions obtained are $2^p x\, m$. Increasing the number of sub-regions cases the huge dimension and the recognition speed is reduced. The local gradient coding algorithm tends to improve the results since it reduces the dimension and increases the recognition speed.

LGC uses a 3x3 neighborhood pattern as shown in Figure 2.5. It is mathematically defined as:

$$LGC_P^R = s(g_1 - g_3)2^7 + s(g_4 - g_5)2^6 + s(g_6 - g_8)2^5 + s(g_1 - g_6)2^4 + s(g_2 - g_7)2^3 + s(g_3 - g_8)2^2 + s(g_1 - g_8)2^1 + s(g_3 - g_6)2^0 \quad (2.4)$$

G1	G2	G3
G4	GC	G5
G6	G7	G8

Figure 2.5: A 3 x 3 LGC operator

The additional optimization in the LGC operator is considered as LGC-HD that is obtained by the gradient computed along with the horizontal and diagonal directions. This additional optimization reduced the feature dimension and a change in the texture of the expression has produced such information. As the features are reduced, recognition accuracy is increased and computational cost is decreased [26].

$$LGC_P^R = s(g_1 - g_3)2^4 + s(g_4 - g_5)2^3 + s(g_6 - g_8)2^2 + s(g_1 - g_8)2^1 + s(g_3 - g_6)2^0 \quad (2.5)$$

The Local gradient difference value is extracted from the sub-blocks by the LGC-HD technique and hits the statistical histogram, eventually linking the entire LGC-HD histogram to the sub-blocks and generating the identification features of the complete image Figure 2.6.

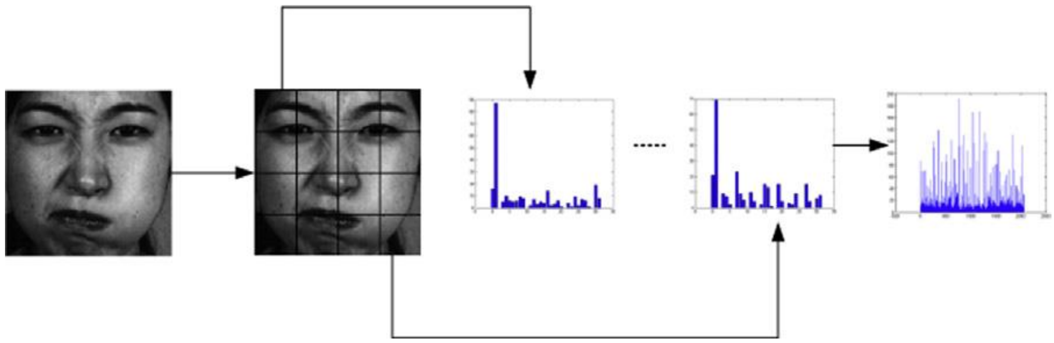


Figure 2.6: LGC-HD feature extraction procedure

2.6 Summary

The literature review focusing on the poor illumination conditions, shading, variation in pose, and extraction of local features is presented. Most of the techniques used a dataset that is taken in a contained environment like JAFFE and CK. The results offered by these techniques are good but tend to decrease when there is variation in illumination and pose. Moreover, some of the methods have not considered the higher dimension of the data representing the features that may affect the computational performance of the system.

The next chapter discusses the model of the first proposed solution in which Merged Binary Pattern codes are generated to extract the texture of the face after preprocessing. The novel feature structure is reduced in dimension and classification is performed to obtain the results.

Chapter 3

Technique I

Merged Binary Pattern Codes

3. Merged Binary Pattern Codes

This chapter discusses the model that is followed for facial expression recognition. It is composed of contrast enhancement, feature extraction using Merger Binary Pattern Codes, dimensionality reduction using PCA and then classification is enacted.

Comprehending expressions require the extraction of facial features. The flexibility of these facial features restricts the accurate recognition of expression that is introduced even by a slight variation in light or pose resulting in local intensity variation. This paper attempts to identify the mental state through recognition of expressions in the real world considering the challenges in the fields of biometrics, security and human behavior identification through psychological study. Furthermore, robot vision, facial animation and virtual reality also require analysis of facial expression.

Facial Expression Recognition is an extraction and classification process applied to the image or video to classify facial expression into angry, disgust, fear, happy, neutral, sad or surprised emotion by the use of the computer. A typical Facial expression recognition system consists of expression image preprocessing, face detection and face region segmentation, expression feature extraction and expression classification [56].

There exist a variety of approaches for automatic facial expression recognition and most of these use synthetic dataset where the emotions are intentionally expression under controlled environments JAFFE[50], CK[51]. Ekman [15] defines that such a dataset contains a discretionary expression. In a real-world environment, facial expressions are unconstrained with a varying pose, different age, occlusions, non-uniform illumination and low-resolution images thus making the task of emotion recognition real challenging as presented in the SFEW database[52]. Such data is collected from sources like the World Wide Web and TV broadcasts to represent the facial expression that has more importance in the real world.

In order to conserve local texture variations along with gradient difference, a novel method is proposed. Once the face image is extracted after it is detected, this method first improves the contrast by converting the input image into the frequency domain using Discrete Fourier transform (DFT) [53][54]. Then enhancement of the image is achieved using contrast limited adaptive histogram equalization (CLAHE) [65] in the frequency domain. Prior application of this method makes face images suitable for feature extraction by compensating complex background and poor illumination that mainly occur due to the use of accessories and/or uneven light reflections. MBPC extracts 16-bit code for a pixel using a 2-bit code per neighbourhood. The least bit of MBPC holds local texture difference whereas a higher bit specifies magnitude change of the gradient around the neighborhood of a pixel using a difference matrix. This practice helps in local and global illumination compensation.

The abundance of features obtained from both approaches demands a careful selection of influential features. A powerful tool PCA is used to reduce the number of dimensions without much loss of features obtained in the previous stage. The trained classifier is employed to test the sample facial expressions.

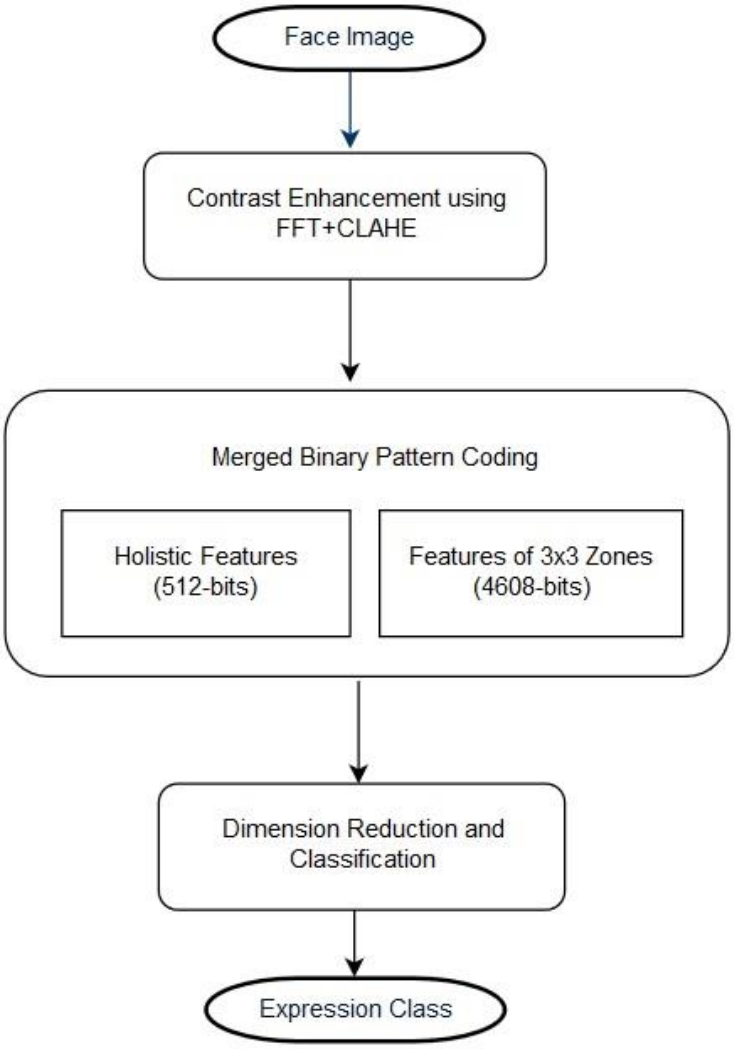


Figure 3.1: The Proposed Scheme I Model [57]

The proposed technique is composed of three major modules: Preprocessing to improve the contrast to handle poor illumination practices FFT+CLAHE method, robust features are obtained through MBPC, Principal Component Analysis is employed to reduce dimensionality and then classifier is used to identify the expression.

3.1 Contrast Enhancement using FFT+CLAHE

The face images obtained from the real-world may not be suitable for feature extraction before some preprocessing due to poor illumination or the presence of the shadow. The suggested approach performs enhancement of the image using contrast limited adaptive histogram equalization (CLAHE) in the frequency domain.

Here the proposed method is discussed step by step. In the first step, the image is processed using DFT which converts the image into low and high-frequency components. Dealing with low and high-frequency components separately allows better enhancement of the image. Next, the Gaussian filter is used on a low-frequency image to filter the noise and other artifacts. The low-frequency components characterize the illumination that is composed by applying the Gaussian filter. After the noise removal, CLAHE is implemented to enhance the contrast.

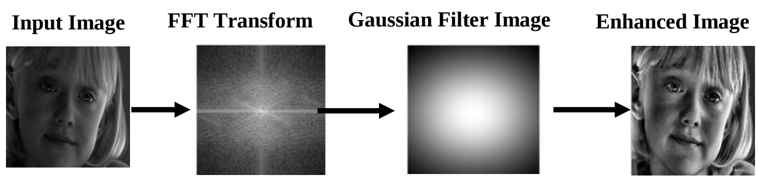


Figure 3.2: Contrast Enhancement

3.2 Features Extraction Using Merged Binary Pattern Coding (MBPC)

In order to observe and describe the effects of local and global deformations of the feature and its influence on the performance of facial expression recognition, this framework provides a platform to examine the texture-based methods for both full-face and division-based approaches.

This section of the study clarifies a 2-byte (16-bit) encoding mechanism that computes 2 bits involving the neighborhood around a pixel. These two bits are merged to capture the texture and fine details from the face images. A higher bit represents the difference of the sign as it is obtained in LBP coding. The lower bit is reserved for magnitude computed with the help of difference-matrix. This matrix holds a Gradient of respective neighboring pixels with a central pixel. The average of the difference matrix values is taken as a threshold to compare with the neighboring values of this matrix and thus binary code is obtained. If the value at the corresponding position is greater than the threshold, bit value 1 is captured otherwise 0. These two bits – sign and magnitude difference –are merged, holding higher and lower positions respectively, to obtain a 2-byte code. The 2-byte encoding produces a feature vector having length 2^{16}. The dimension reduction is achieved by decomposing 2-byte code into two subcodes: HV-code and D-code each of 8-bit in length. HV-code is composed of bits in the horizontal and vertical direction and D-code from the diagonal bits as explained in Figure 3.3. These codes are merged to manage feature vector having length 512.

Let δ_1 is the least significant bit and δ_2 is the higher bit of the code. Then δ_1 is computed as:

$$\delta_1(x_c, y_c) = \sum_{p=0}^{7} s(i_p - i_c)2^p \tag{3.1}$$

δ_2 initially requires the calculation of Gradient magnitude between the center pixel and its neighborhood obtained as:

$$g_n = |i_p - i_c| \tag{3.2}$$

The average value used as threshold can now be calculated as:

$$\bar{g} = \left({}^{1}\!/_{p}\right) \sum_{n=0}^{p-1} g_n \tag{3.3}$$

δ_2 for the neighboring values is obtained by comparing the corresponding value with Gradient average.

$$\delta_2(x_c, y_c) = \sum_{p=0}^{7} s(g_p - \bar{g})2^p \tag{3.4}$$

Where i_p is neighborhood pixel and i_c is a central pixel value in equation 3.2. g_p and $\bar{g}$ represent neighborhood gradients and average Gradient respectively, in equation 3.4.

s(x) obtains bit value from the sign for δ_1 and δ_2 in equation 3.1 and equation 3.4 as following.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.5}$$

δ_1 and δ_2 are merged to generate binary encoding where δ_1 is lower and δ_2 is a higher bit as shown in Figure 3.3. Further, the 16-bit code is decomposed into HV-code and D-code by selecting horizontal and vertical direction bits for HV-code and both directional bits for D-code. Arranging HV-code and D-code into two independent 8-bit codes result in feature reduction. The illustration is provided in Figure 3.3 given below:

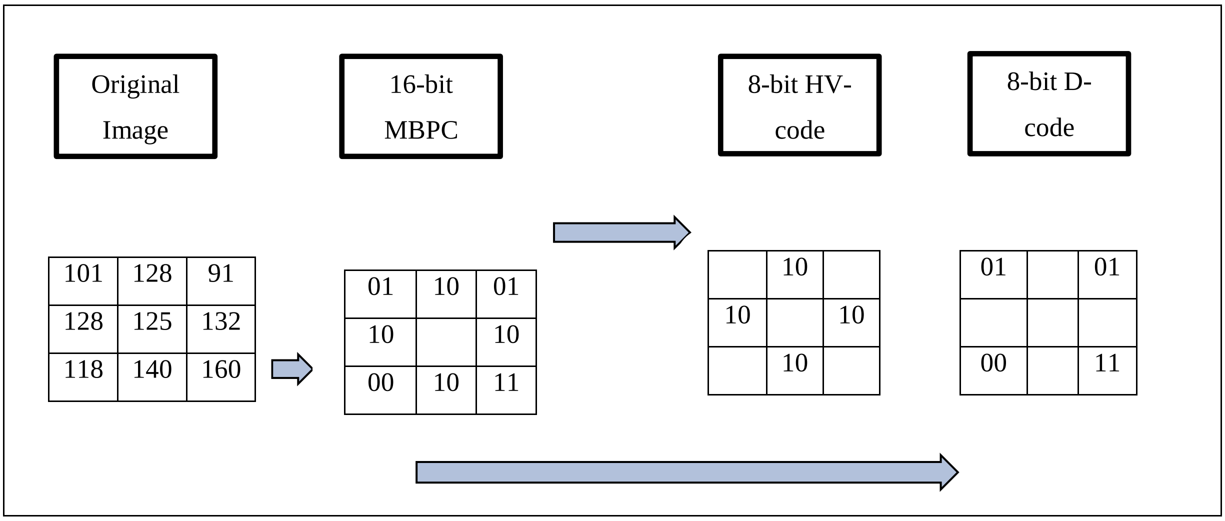


Figure 3.3: MBPC Descriptor

The formation of the MBPC descriptor for the holistic approach is illustrated in Figure 3.4.

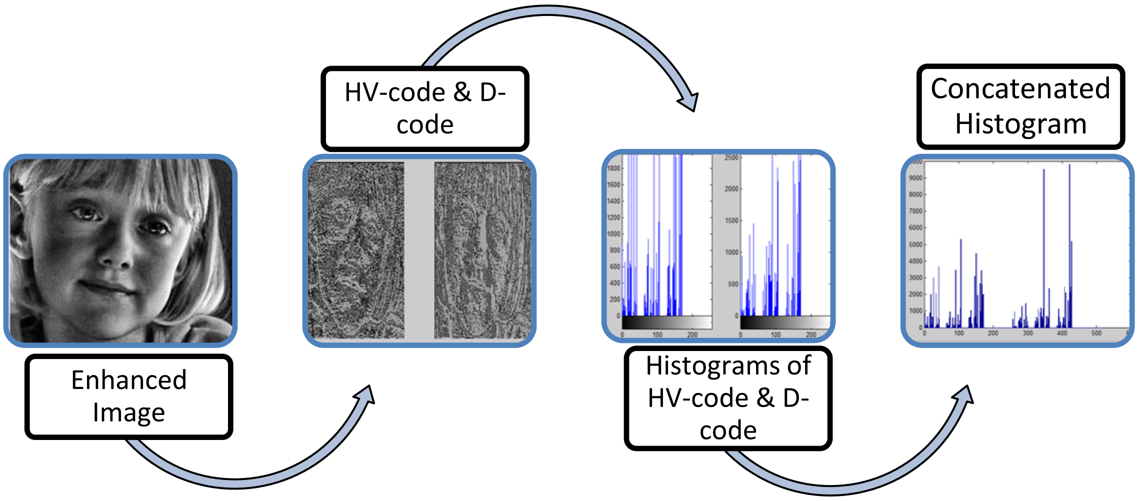


Figure 3.4: MBPC in Holistic Approach

The zone-based structure of the approach is shown below in Figure 3.5.

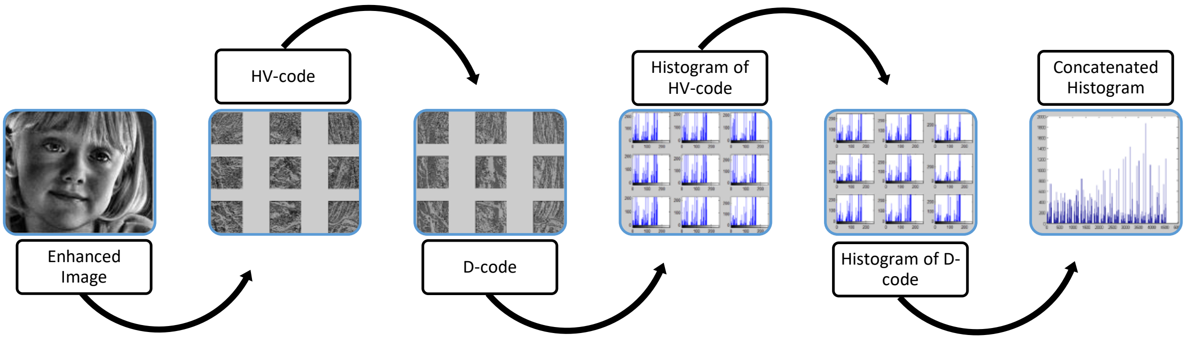


Figure 3.5: MBPC in zone-based approach

3.3 Dimension Reduction and Classification

The data dimensions become high after extraction of local texture features from the face image that contains fine details necessary to differentiate among dissimilar expressions. For the elimination of the least useful features, a feature selection step is needed. Several approaches are available for dimensionality reduction and this includes Principal Components Analysis (PCA) and Fisher's Linear Discriminant (FLD) that is generalized into Linear Discriminant Analysis (LDA). PCA tends to find the projection to explore the orthogonal features whereas FLD draws a line to provide maximum separation between classes. FLD is effective if the data representing the features are normally distributed [75]. But when the distribution is multi-modal then good separation cannot be achieved. We have used PCA to reduce the data dimensions while preserving important characteristics.

In the last stage of the proposed solution, the resultant prominent features are provided to classifiers to build the model to perform the classification task. Sequential minimal optimization (SMO), Simple Logistic, K-nearest neighbors (KNN*)*, multilayer perceptron *(*MLP) and decision tree-based J48 classifiers are utilized to classify the expressions of the face. J48 has provided promising results and are discussed in section 3.5.

3.4 Experimental Set-up

Static Facial Expressions in the Wild (SFEW) database [52] contains seven different expressions (anger, disgust, fear, happy, neutral, sad and surprise). This data has been selected to perform experimentation because it has been formed by selecting video frames from the AFEW data-set and gives real-world reflection. The database of 700 images includes facial expressions with different face poses, a wide range of age, occlusions, varying focus, different image resolution and improper illumination.

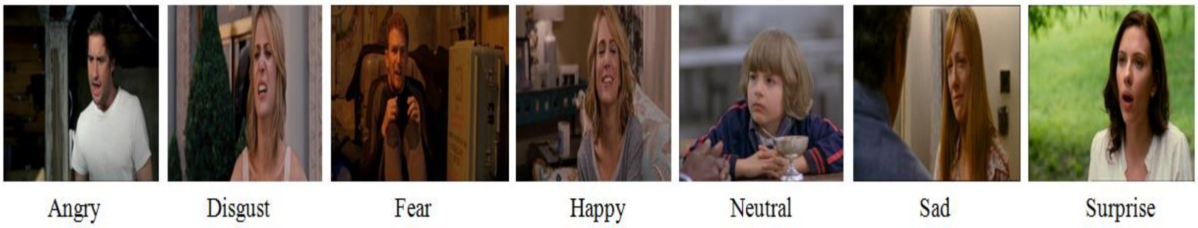


Figure 3.6: Seven Basic Expressions from SFEW dataset

For the detection of a face from an image, Viola-Jones' algorithm is used and then the output face image is normalized and cropped as shown in Figure 3.7.

Figure 3.7: The cropped image after applying the Viola-Jones algorithm

The pre-processing step (presented in section 3.1) is performed on all the input images before feature extraction. A comparison of the visual quality of some of the enhanced images using FFT+CLAHE and histogram-based methods is shown in Figure 3.8.

Figure 3.8: The first row shows the original low contrast images. Images processed using Histogram based method are shown in the second row. The third row displays the images processed using FFT+CLAHE approach

After performing an experiment on different images with low illumination, it has been observed that the histogram equalization just enhances the global contrast of an image but it does not help in preserving important information. However, through the proposed technique we have enhanced the contrast of an image in such a way that it preserves texture information.

After preprocessing, features are extracted by using the MBPC technique. A feature vector is summarized using PCA that is provided to the classifiers for the facial expression classification.

3.5 Results

The impact of local information extracted using the MBPC descriptor is observed for both the full-face image and zone-based approach. The sample dataset is partitioned unevenly into ten equal-sized sub-samples. A ten-fold cross-validation structure is chosen for the classification.

In the holistic approach, features are extracted in the form of histograms from the entire image containing the face. Face image is divided into nine equal parts and features are extracted from all parts. A combined histogram is formed from the individual histograms representing each local region. PCA is applied to both approaches for feature selection and a reduced feature set is used for classification.

Features with high variance are selected using the PCA algorithm. A number of feature sets (FS) are generated: FS-40, FS-60, FS-80 and FS-100. It is observed that as we increase the number of features after feature selection using PCA, the accuracy rate tends to decrease. It happens because redundant features are normally misclassified.

Table 3.1: Results of LBP, CLBP, LGC-HD, LGC-HVD and MBPC techniques using holistic approach using SFEW database

Techniques →	**LBP**			**CLBP**			**LGC-HD**			**LGC-HVD**			**MBPC**		
Classifiers ↓	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**
SMO	36.2	38.2	**38.2**	23.5	30	**30**	85.9	85.9	**85.9**	90.8	89.4	**89.4**	93.7	93.5	**93.5**
Simple Logistic	78	78	**78.23**	85.3	85	**85.3**	92	92	**92.35**	93	92	**92.94**	94.7	94.1	**94.1**
KNN	71	70	**70**	84.1	84	**84.1**	93	92	**92.35**	94	93	**93.52**	88.5	87.1	**87.1**
MLP	59.5	55.9	**55.9**	48.5	48.2	**48.2**	56.8	56.5	**56.5**	56.8	56.5	**56.5**	90.2	89.4	**89.4**
J48	74.4	73.5	**73.5**	80.5	79.4	**79.4**	86.5	86.5	**86.5**	88.3	88.2	**88.2**	96.5	96.5	**96.5**

The maximum average accuracy rate of 96.5% has been obtained by using a simple j48 classifier for FS-40 in a holistic approach as shown in table 3.1. We have also noticed that the proposed framework gives better accuracy for all five classifiers which indicates that the extracted and selected features are robust to variation in illumination and other artifacts. As the framework utilized a minimum number of features for classification which also indicates that the proposed framework is computationally less expensive.

Figure 3.9 shows that the accuracy of SMO, Simple Logistic, KNN, MLP and J48 well-known classifiers is higher when features are extracted using Merged Binary Pattern Codes after image pre-processing. The Proposed MBPC technique was able to provide the best results using a holistic approach.

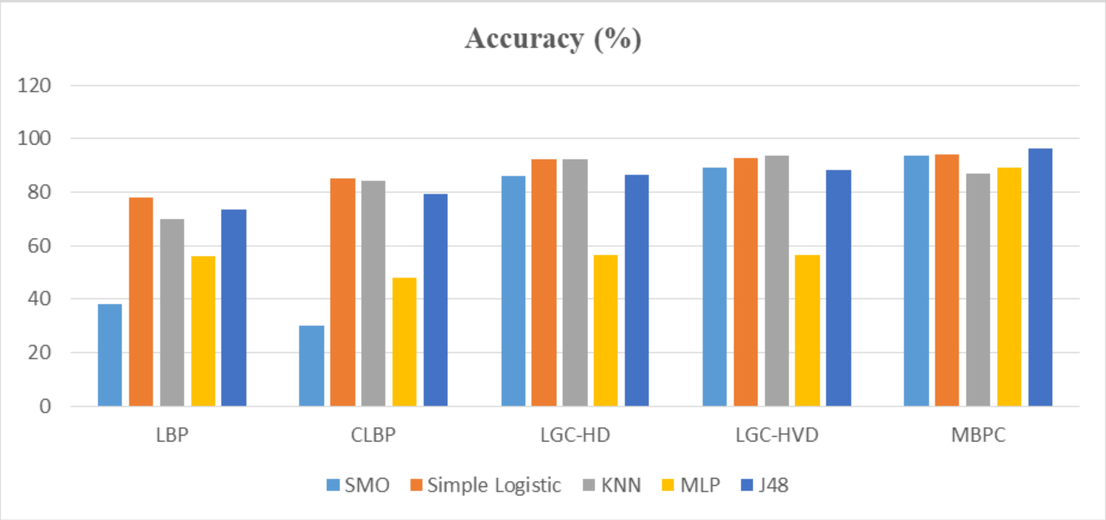


Figure 3.9 Comparison of Performance of State of the Art Feature Extraction Techniques with the Proposed Technique Using Holistic Approach

There is a considerable reduction in recognition rates in the division based approach when features are extracted using MBPC for the SFEW dataset. The useful correlation among different features is wasted during the division of face image into zones. An achieved accuracy of 67.2% still outperforms the state of the art techniques when a zone-based approach is adopted.

The total number of extracted features for zone-based face images using the proposed MBPC framework is 4608. PCA is also applied in that case in order to select the most discriminative features out of 4608 features. Feature sets (FS) of FS-60, FS-80, FS-100 and FS-120 are generated in this experiment. The best average accuracy has been obtained using FS-60 for a simple KNN classifier.

Table 3.2: Results of LBP, CLBP, LGC-HD, LGC-HVD and MBPC techniques using 3x3 zone-based approach using SFEW database

Techniques → / Classifiers ↓	LBP			CLBP			LGC-HD			LGC-HVD			MBPC		
	Precision (%)	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Accuracy (%)**
SMO	36.3	37.6	**37.6**	48.4	42.4	**42.4**	54.3	53.5	**53.5**	47.1	48.8	**48.8**	65.2	64.4	**64.4**
Simple Logistic	36.1	40	**40**	52.2	50.6	**50.6**	43.6	41.8	**41.8**	44.5	42.4	**42.4**	56.7	56.7	**56.7**
KNN	45.4	44.1	**44.1**	46.6	46.5	**46.5**	55.5	51.2	**51.2**	56.7	51.2	**51.2**	68.4	67.2	**67.2**
MLP	45.7	44.3	**44.3**	40.6	39.8	**39.8**	47.5	46.7	**46.7**	45.3	44.6	**44.6**	58.5	57.4	**57.4**
J48	31.2	31.2	**31.2**	36.6	37.1	**37.1**	42.5	42.9	**42.9**	43.5	42.9	**42.9**	66	65.5	**65.5**

Although there is a decline in performance in a zone-based approach but still the use of the MBPC features offers better expression resolution ability for different classifiers including SMO, Simple Logistics, KNN, MLP and J48. The results show that MBPC is offering a good resolution of expressions when compared with other state-of-the-art techniques that include LBP, CLBP, LGC-HD, LGC-HVD as shown in Figure 3.10.

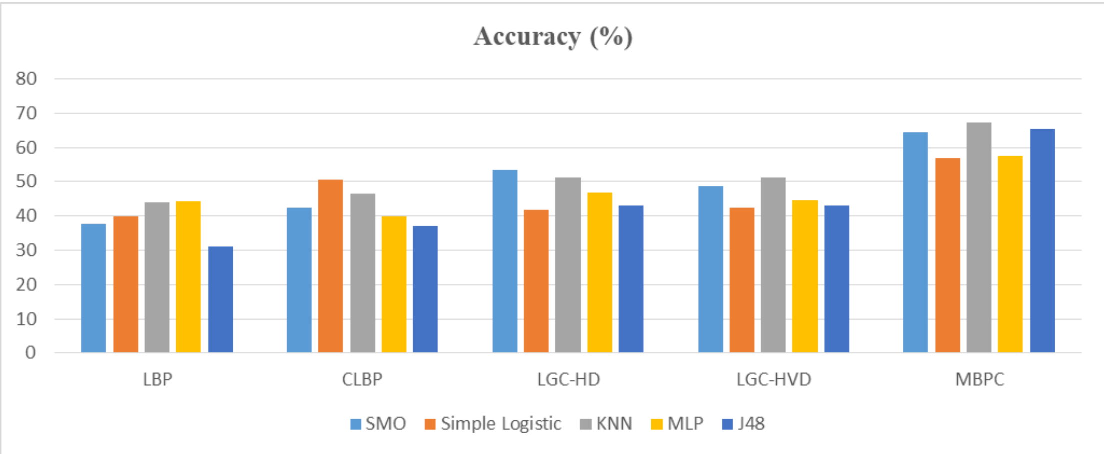


Figure 3.10 Comparison of Performance of State of the Art Feature Extraction Techniques with the Proposed Technique Using Zone-Based Approach

For feature extraction computation of gradient and the sign difference involves the processing of each pixel once so the asymptotic complexity of the method is O(MN) where M and N are the rows and columns respectively.

3.6 Summary

This section discussed a novel and potent technique MBPC for facial expression recognition in poor light conditions while dealing with the pose adaptation at the same time. In the preprocessing stage of the model, enhancement is made successful with the use of CLAHE to improve the visual appearance and making an image suitable to extract the features. The dimensionality of the features thus obtained is reduced using Principal Component Analysis so that the non-redundant can be eliminated. The comparison of the proposed holistic approach is made with the state of the art techniques LBP, LGC-HD, LGC-HVD and CLBP using dataset close to the real world. As a result of this experimentation, the performance of the proposed scheme is the best among all techniques shown in table 3.2.

The next chapter discusses the novel feature descriptor called BGSD that extracts low-level features from the image obtained after pre-processing. The results are obtained using a variant of the SVM classifier to affirm the effectiveness of the model.

Chapter 4

Technique II

Binary Gradient Structure Decsriptor

4. Binary Gradient Structure Descriptor (BGSD)

The traditional systems of interaction between humans and computers ignore most of the information communicated through these emotional states and only attend to the intentional contribution of the user. As mentioned, the paradigm is moving toward designs centered on the human being, so that the analysis of the affective states of the user becomes unavoidable. The conventional methods of interaction with computers will soon be replaced by showing the behavior, that is, emotional state by the humans in the near future [3] [4]. Therefore, great interest is exhibited by the computer vision researcher society, for the analysis and automatic facial expression recognition. A system capable of recognizing the emotions through facial expressions can benefit many application areas including human-computer interaction, human behavior monitoring and medical responses, entertainment, social robots, cheating detection and interactive video [5].

Initially, facial expressions are the area of interest for philosophers and thinkers and then it becomes an experimental study. The computer vision researchers express that the facial expression recognition system is the one that correctly categorizes the facial features representing the universal emotions: anger, fear, disgust, happiness, sadness, surprise and neutrality, as presented by Ekman in 1971 [6]. The humans can do this task in a very limited time, in our routine, without any additional effort, but computer systems still face the challenge of recognizing facial expression in variations in lighting, posture variations. These factors are effectively handled in this research study using image normalization that provides an image that is most suitable to extract the features using a novel method Binary Gradient Structure Descriptor. BGSD extracts the local structural features of the face image that were given to a variant of the SVM classifier. The classification results exhibit that the proposed scheme has outperformed the state of the art techniques used for emotion recognition.

The key contributions of this research are mentioned below:

- The experiments are conducted on the SFEW database that contains images with poor illumination and shading. Based on the comparison done with existing techniques, the local illumination normalization has made considerable improvements in results.

- This research aims to introduce the features that effectively perform facial expression recognition in real-world scenarios. The novel feature extraction method Binary Gradient Structure Descriptor obtains high accuracy for classifying and recognizing the emotional state of humans.

Static Facial Expressions in the Wild (SFEW) database is selected for experimentation including images with static and tough conditions covering seven kinds of facial expressions. It covers the broad age images, that is, (1-70 years), which makes it unique in terms of age. The database contains images that mimic real-world illumination conditions.

Technique II

Although the research community contributes a lot in the field of facial expression recognition, for each module there is a wide range of algorithms, the challenging task is how to combine them and solve the problem of local illumination. The proposed system architecture to resolve the local illumination problem can be described in Figure 4.1.

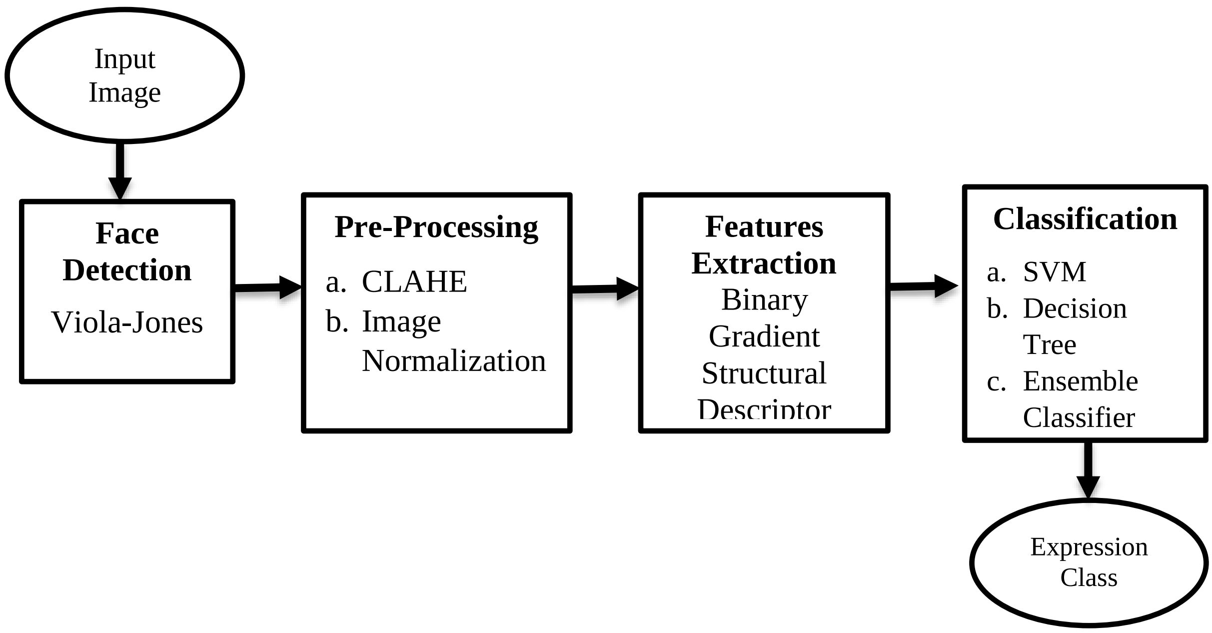


Figure 4.1: Proposed Model II

The Proposed framework is completed in 4 phases respectively: Face detection, the pre-processing phase, feature extraction phase and a classification phase. The major contribution of this framework is to carry out experiments to compare the existing algorithms in terms of facial expression recognition accuracy [77].

4.1 Face Detection

The first stage of recognition of facial expression is the detection of the face from the given sequence of images. The most popular face detection algorithm for face detection called Viola-Jones object detection is applied. Viola-Jones method for face detection finds Haar-like features from a rectangular region inside the image that makes it a very accurate and faster algorithm for face detection [63]. The face detection procedure includes some steps that are performed consecutively on the input image. First, the classifier trained for face detection looks for a face in the image.

Figure 4.2: Face Detection with Voila Johns Algorithm

After the face is found in an input image, two classifiers are invoked on the upper part of the face for the detection of eyes. The left and right eyes are detected separately - in the regions from the upper left and upper part as shown in Figure 4.3.

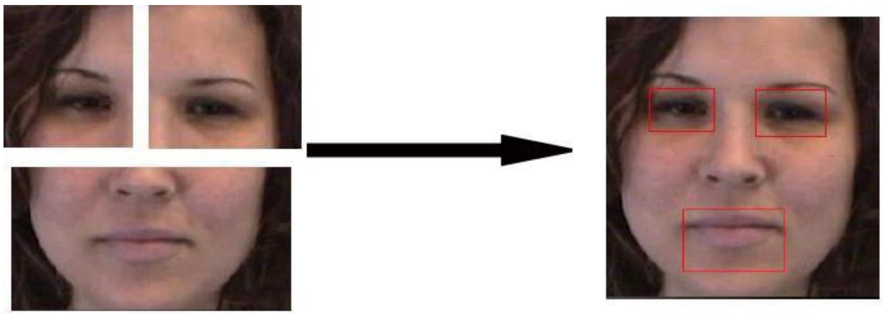


Figure 4.3: Eyes and moth detection with Voila Johns Algorithm

Finally, the mouth region is located with the fourth classifier looking at the lower part of the face. The detected area is reduced to improve the efficiency of the algorithm. Having face locations and facial landmarks, face representation can be formed. If there are more faces detected in the image, the algorithm is the largest one for further processing.

4.2 Pre Processing

The essential step to improve the recognition of facial expression accuracy is the pre-processing phase. The pre-processing pipeline helps to eliminate the imbalance and local illumination effect and increase recognition accuracy. Based on the proposed facial expression recognition framework different types of pre-processing techniques such as, image resizing, noise removal and illumination normalization are used.

The images are resized to reduce the mathematical complexity in the facial expression recognition framework. In literature, there are many techniques available for resizing the face images. We use the bicubic interpolation method to resize images. Later noise removal is performed using 3rd stage of the Gaussian filter.

4.2.1 Illumination Normalization

It is a pre-processing chain that is executed before the feature extraction techniques, the different stages are designed to eliminate the local illumination effects, shading and reflections while preserving the essential information of facial expressions shown in Figure 4.4.

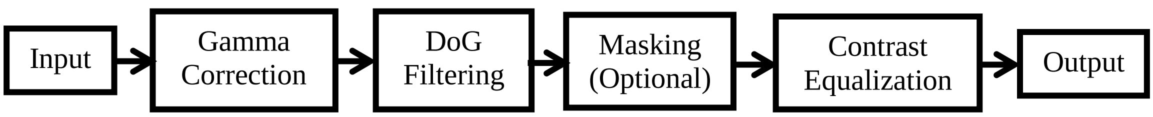


Figure 4.4: Illumination Normalization Pre-processing Chain

First, Gamma correction is applied to the input image to adjust the intensity values. To improve this lighter region, we chose this specific value that keeps the object's information intact, even with the lighting effect. Further, to tackle illumination inconsistencies and shading problems the bandpass technique is used which works on the high-frequency components to remove shading while preserving the essential information. Fine Gaussian filtering is applied with two standard deviation values, the difference of which coincides with the bandwidth. The selected values are $\sigma 1 = 1$ and $\sigma 2 = 2$ for the two Gaussian filters.

Masking is used only when we considered the hair, beard, etc. in facial regions that are irrelevant and should be masked at that point.

The last stage of the illumination normalization pipeline is contrast equalization which provides a global scale by rescaling the intensities of the image to standardize a solid measure of the general contrast or intensity variation. The two-stage process is defined as:

$$Im(x_m, y_m) = \frac{Im(x_m, y_m)}{\left(mean(|Im(x'_m, y'_m)|^{\beta})\right)^{\frac{1}{\beta}}} \tag{4.1}$$

$$Im(x_m, y_m) = \frac{Im(x_m, y_m)}{\left(mean\left(min(\tau, |Im(x'_m, y'_m)|)^{\beta}\right)\right)^{\frac{1}{\beta}}} \tag{4.2}$$

The effect of large values is reduced by one component of compression β and is a cut-off limit for large values. Some extreme values in the processed images are eliminated by a nonlinear tan hyperbolic function given in equation 4.3.

$$Im(x_m, y_m) \leftarrow \tau \tanh(Im(x_m, y_m)/\tau) \tag{4.3}$$

The extreme values limit is in the range of $(-\tau, \tau)$.

Figure 4.5: An example of illumination normalization stages: input image, Gamma correction, DoG filtering, and contrast normalization.

4.2.2 Contrast Limited Adaptive Histogram Equalization (CLAHE)

A histogram displays the shape and spread of continuous sample data having L levels [0, L-1] and is given by the discrete function:

$$h(r_k) = n_k \tag{4.4}$$

where r_k represent the intensity of k^{th} level and n_k gives the total number of pixels in the image with r_k intensity [8]. A normalized histogram is calculated through the following equation:

$$p_r(r_k) = \frac{n_k}{MN} \qquad k = 0,1,2,\ldots,L-1 \tag{4.5}$$

where $p_r(r_k)$ is the probability of the intensity level r_k that appears in an image and is equal to 1.

The histogram equalization (HE) is used to enhance the contrast of the image globally by transforming the histogram of the original image into a uniform histogram and is given by the following equation:

$$s_k = (L-1)\sum_{j=0}^{k} p_r(r_{kj}) \qquad k = 0,1,2,\ldots,L-1 \tag{4.6}$$

where s_k is the new histogram distribution.

This process is based on the fact that image quality is uniform across all areas, and a grayscale map provides similar growth for all image areas. However, this approach fails, if the grayscale distribution changes from one region to another. In this case, the local histogram equalization approach gives the best results [64]. Even in some cases, this method cannot solve the problem, when the distribution of gray levels is strongly localized, it may be undesirable to transform images with very low contrast using global histogram equalization. In these cases, the mapping curve can include segments with large gradients. This problem is solved by the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm proposed in [65]. First, the image is decomposed into several overlapping fixed-size regions. A histogram of each region is calculated and then based on the desired contrast-increasing threshold, a limited clip is selected for the histogram. Local histograms are redistributed in such a way that their height does not exceed the boundary of the clip. The clip boundary β is determined using the following equation:

$$\beta = \frac{MN}{L}\left(1 + \frac{\alpha}{100}\,(s_{max} - 1)\right) \tag{4.7}$$

Where α represent the factor for clipping, $\frac{MN}{L}$ is the exact limit for a clip.

Although histogram equalization improves the overall contrast of the image but it may make the background more prominent than the region of interest. In this case, CLAHE reduces this artifact by limiting the region sizes. The comparison is shown in Figure 4.6.

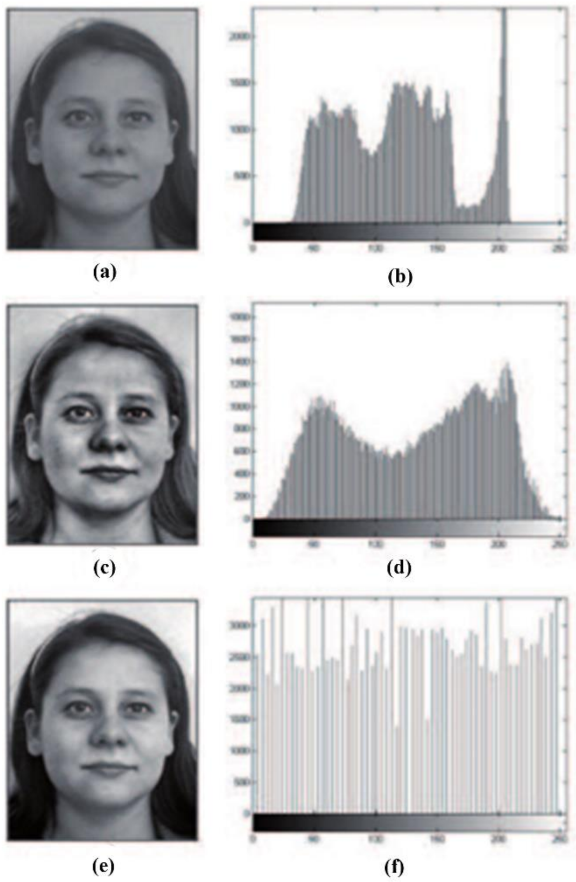


Figure 4.6: The difference between the histograms of Image Normalization and CLAHE methods. (b) is the histogram of (a), (d) is the histogram of (c) and (f) is the histogram of (e)

4.3 BGSD Features Extraction

There is another way of capturing the local properties of the facial image using image gradient orientation (IGO) illustration. The image is divided into blocks and then a histogram of the IGO domain is generated that is considered robust to minor deformations in face texture and local illumination variation. As far as illumination is concerned, the IGO domain is more suitable than LBP.

Assembling a descriptor is the main aim that can incorporate the features of both LBP and IGO while keeping this scheme computationally proficient. For this reason, a smaller representation, IGO histogram of four canisters in the interval $[0, 2\pi]$ is used for BGP. Thus, the customary IGO is figured along the horizontal and vertical gradients. The area encompassed by a pixel is taken into account which is comprised of four neighbors in two directions.

Keeping in view this fact, four-receptacle HIGO is stretched out to different bearings, producing another feature descriptor named Binary Gradient Patterns (BGP). In particular, the BGP establishes connections between symmetric neighbors of a candidate pixel in *m* numerous directions. The calculation of a basic and essential BGP operator of four directions is exhibited in Figure 4.7 [17].

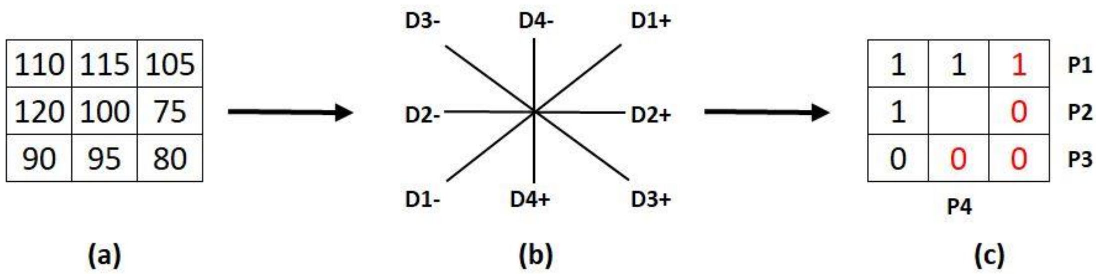


Figure 4.7: Basic description of BGP technique: (a) 8 neighbors around a central pixel (b the essential BGP technique for four directions (c) the resulting string of principal operation shown in red and bold and associated binary bits in black and plain.

An arrangement of nearby neighbors of a focal pixel is presented as shown in Figure 4.7 (a). At that point, four directions D1, D2, D3 and D4 are used to compute a couple of binary numbers, an essential (P_i^+) and related (P_i^-) by associating two symmetric neighbors along every path and are shown in Figure 4.7 (b) and (c).

$$P_i^+ = \begin{cases} 1 & if\ D_i^+ - D_i^- \geq 0 \\ 0 & if\ D_i^+ - D_i^- < 0 \end{cases} \tag{4.8}$$

$$P_i^- = 1 - P_i^+$$

$i = 1, 2, 3 \ldots m$

where P_i^+ and P_i^- present the force estimations of the pixels corresponding to areas as presented in Figure 4.7 (b).

In the end, the marking of the central pixel is determined by four binary numbers using the following equation.

$$S = \sum_{i=1}^{m} 2^{i-1} P_i^+ \tag{4.9}$$

The primary and complementary numbers in four directions generate 8 binary numbers. Only the fundamental binary bits are utilized to ascertain labels that depict every conceivable fluctuation of BGP patterns. The number of principle binary bits control the number of BGP pattern labels (L_s) and depends on the number of directions(m), $L_s = 2^m$. Subsequently, there are $S^m \in \{0,1,2,\ldots,2^{m-1}\}$ conceivable labels for BGP operator in m directions. This number (2^m) is considerably smaller than the number of LBP labels (2^{2m}).

The proposed descriptor is aimed to identify and isolate significant local detail that is a good representative of texture. BGSD technique extracts the sixteen unique labels using a BGP descriptor by increasing the number of directions and the size of the neighborhood. If N represents the number of neighbors around a pixel within a certain radius of the square then BGSD is computed as:

$$L_j^{sN} = \begin{cases} 2^{j-1} - 1 & if\ j \leq m \\ 2^m - L_{2m-j+1}^{sN} - 1 & if\ j > m \end{cases} \tag{4.10}$$

Where j ranges from 1 to N, m is the number of directions and equal to N/2 and L^S is the label of the structure.

The paired structures of these labels are shown in Figure 4.8 (a). It is obvious that each of these labels includes four 0s and 1s, derived by the pairs. Surprisingly, out of these 15, certain labels contain a stream of 4 consecutive bits having value 0 while the remaining stream of 4 bits has value 1. These 8 patterns describe the Binary Gradient

Structure. These are the main labels and are exhibited by bold outlines of the box while the other eight labels with intermittent patterns are represented as thin lines.

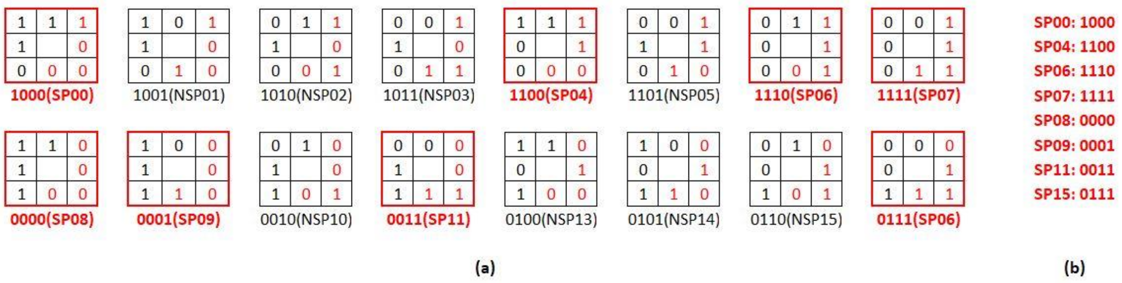


Figure 4.8: Binary Gradient Structure Representation using Bold Boxes

These persistent 1's of the structure describe the steadier nearby changes in surface that mainly illustrate the actual edge of the surface. The more concrete form of the facial texture is registered by these profoundly steady structures. The formation also includes the weak structures, represented by the secondary labels and shown by thin outlines. These structures carry subtle details that may be common across all expression types. So, the selection of 8 labels out of 16 decreases the dimensionality of the structural features. These labels with irregular 1s and 0s contain subjective changes in the nearby surface that do not represent the main texture and may characterize the noise.

4.4 Classification

In this research work, the effectiveness of the features is explained using a classification based on three well-known machine learning algorithms: Support Vector Machine [48], Decision Tree Learning Algorithm and Ensemble Tree Classifier.

4.4.1 Support Vector Machines (SVM)

The Support Vector Machine classifier is the most popular margin classifying in the area of the visual pattern [66]. The linear classifier was used in the earliest pattern recognition frameworks. If x is the given pattern to a class $y = \pm 1$, then

$$\hat{y}\,(x) = w^{\perp}\,\Phi\,(x) + b \tag{4.11}$$

is the sign of linear discriminant taking before the pattern transformed into a feature vector $\Phi(x)$.

The Hyperplane $\hat{y}\ (x) = 0$ define a decision limit in a feature space. The problem specifies a feature vector $\Phi\ (x)$ that is usually selected by hand. In a training set $(x_1, y_1) \cdots (x_n, y_n)$ learning procedure is executed to determine the w and b parameters. Choosing the hyperplane for effective separation ability that maximizes the margin is the one that leaves as much space as possible between the hyperplane and the nearest example [63]. The complex nonlinear classifiers are calculated using the linear mathematics of the optimal hyperplanes.

4.4.2 Decision Tree Learning Algorithm

The decision tree learning utilizes a prescient model to move from the perceptions represented in the branches to decisions about the target estimation of the objective. It is one of the prescient demonstrating approaches utilized as a part of insights, information mining and machine learning [67]. The basic algorithm used in decision trees is known as the ID3 algorithm [68]. The ID3 algorithm constructs decision trees using a top-down greedy approach. The information gain is a factual property that measures how well a given characteristic isolates the training cases as indicated by their objective classification. While an attribute with high information gain (left) divides the data into groups with an odd number of positives and negatives and, as a result, helps separate them from each other.

The impurity in the training samples is given as entropy and the effectiveness of attribute while the classification of training data is measured as the information gain. An attribute A, examples of the relative sample, information gain is represented as $Gain\ (S, A)$ showing in Figure 4.9 and defined as:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} rac|S_v||S| \cdot Entropy(S_v) \qquad (4.12)$$

$$InformationGain = Entropy(parentnode) - [AverageEntropy(children)] \qquad (4.13)$$

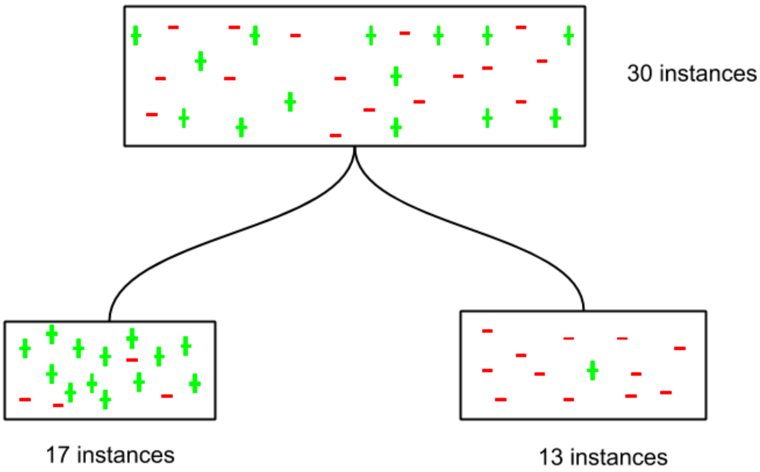


Figure 4.9: Decision tree sorting instances based on information gain.

4.4.3 Ensemble Learning Algorithm

The ensemble is the art of combining various learning algorithms to improvise on the stability and predictive power of the model. In machine learning, multiple learning algorithms are used in ensemble methods to gain better predictive performance than any of the other learning algorithms could be obtained alone.

A single hypothesis is represented by the training ensemble which is used for making the predictions that why it is a supervised learning algorithm. This hypothesis, be that as it may, isn't contained inside the speculation space of the models from which it is built. Along these lines, it can be demonstrated that the gathering has greater adaptability in the capacities they can speak to. This adaptability may, in principle, enable them to over-fit the training data in excess of an individual model would, yet practically speaking, some get together strategies (especially bagging) tend to diminish issues identified with over-change of the training.

Observationally, ensemble tends to create better outcomes when there is a critical decent variety between the models [38]. Numerous ensemble techniques, along these lines, look to advance decent variety among the models they consolidate [39]. Despite the fact that it may not be natural, more irregular calculations can be utilized to create a more grounded ensemble than exceptionally consider calculations [40] [41]. In any case, it has been demonstrated that utilizing algorithms of robust learning calculations is more viable than utilizing procedures that endeavor to dumb-down models to advance decent variety.

4.4 Experimental Setup

The experiments are conducted on seven classes of all-inclusive facial appearances: happiness, sadness, anger, fear, disgust, surprise and neutral. In this research, the major focus is to remove the local illumination effects to gain better recognition results. Two different techniques are applied to grayscale images of the given dataset SFEW to handle the local illumination.

We have employed different classifiers to evaluate the performance of the proposed algorithm using a k-fold cross-validation technique where k is 10. In 10-fold cross-validation, the dataset is equally divided into 10 subsets. The models are trained on the 9 subsets where a single set is used for testing. This whole procedure is iterated 10 times (k-folds) by choosing each of the k subsets once for the testing. The final value is obtained by the mean estimations from k-folds.

In the pre-processing chain that image normalization is performed before the extraction of features that joins a progression of stages intended to neutralize the impacts of illumination varieties, local shading, and reflections while protecting the fundamental components of the visual appearance as shown in the Figure 4.10.

The CLAHE strategy usually increases the global contrast of numerous images, by changing the original image histogram to a uniform histogram that is, endeavoring to make uniform the appropriation intensity pixels of the image as shown in Figure 4.10.

The experiments show that the overall contrast is enhanced through both schemes. The images enhanced through CLAHE have better subjective appearance whereas image normalization has obtained results that improve the structural details of the images. Ironically, we have obtained the images that can contain better texture information of the image contents.

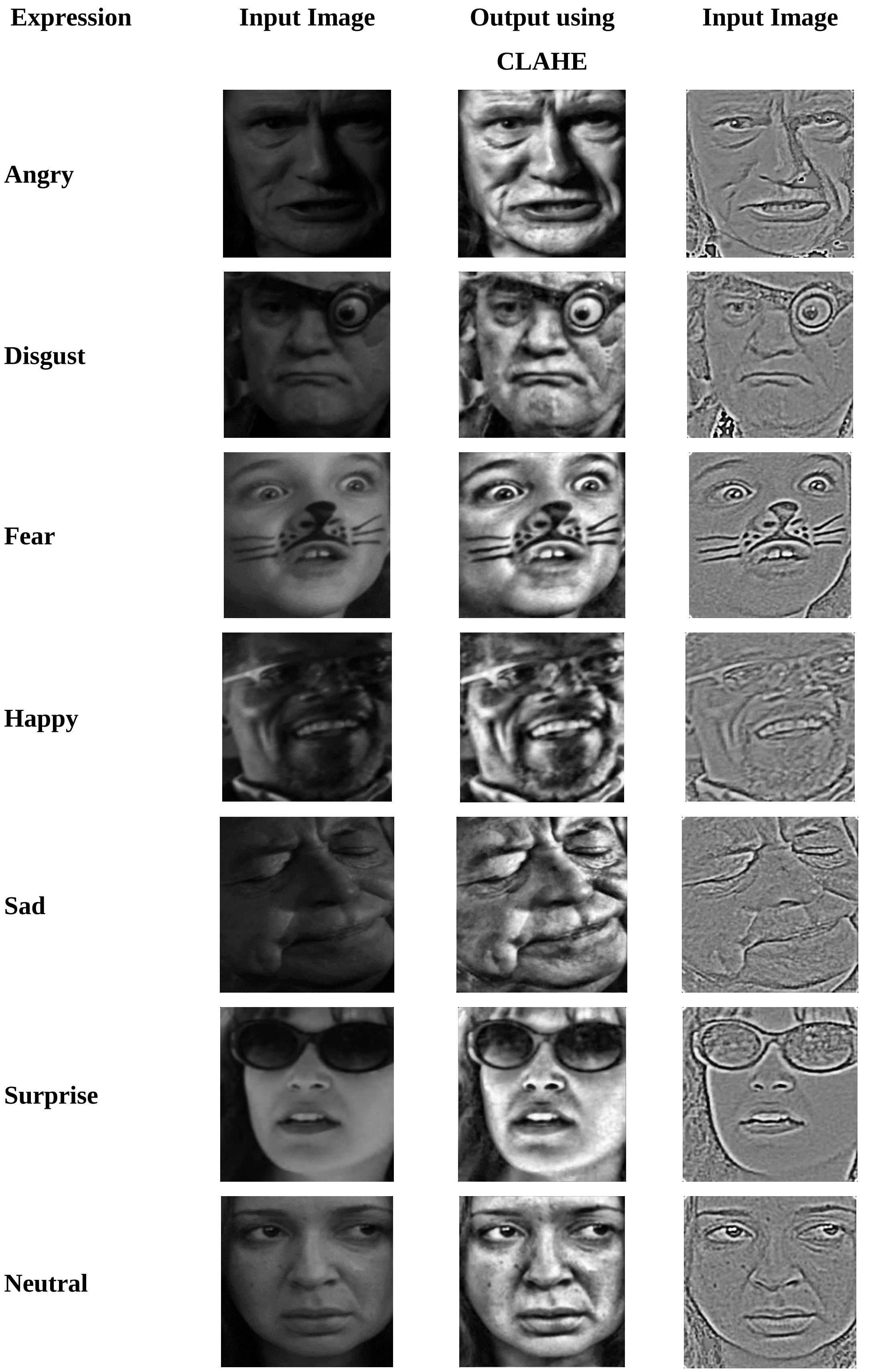


Figure 4.10: Enhancement of images using CLAHE and Image Normalization for each of the expression

After applying these pre-processing methods, Binary Gradient Structure Descriptor is used to extract the features from these images. The classification using SVM, Decision trees and ensemble classifier is performed on the features presented by the descriptor to identify human expressions.

4.5 Results

The classification results of the proposed technique are presented along with the baseline feature extraction techniques and the performance of classifiers based on their variations are generated and a thorough analysis is presented. The results are generated using three versions of the dataset. The first version of the data is not pre-processed while the other two variations are pre-processed by CLAHE and image normalization respectively.

4.5.1 LGC-HD Technique Results on Three Variations of Datasets

Following is the graphical comparison of Local Gradient Coding Horizontal and Diagonal (LGC-HD) features extraction technique using three variations of the dataset termed as normal, pre-processed with Contrast Limited Adaptive Histogram Equalization (CLAHE) and pre-processed with images normalization procedure composed of gamma correction, DoG filtering and contrast equalization.

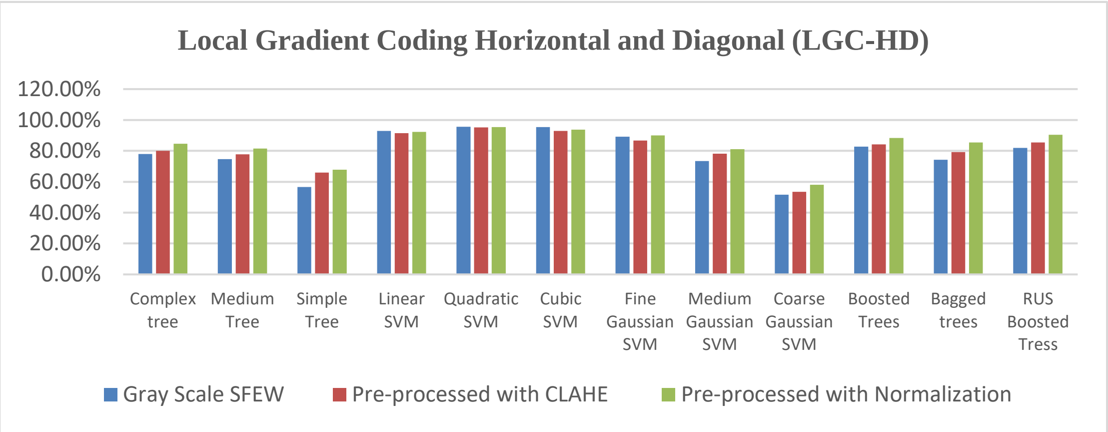


Figure 4.11 Results of LGC-HD Technique on Dataset Variations

From the results shown in Figure 4.11, it is obvious that for SFEW Dataset Images, pre-processing using image normalization, LGC-HD achieves the highest accuracy of

95.5% for support vector machine classifier variant that is Quadratic SVM. The confusion matrix presented in table 4.1 exhibits this fact.

Table 4.1: LGC-HD Quadratic SVM Confusion Matrix for SFEW dataset Pre-processed with Image Normalization

A	**B**	**C**	**D**	**E**	**F**	**G**	**← Classified As**
99%	8%						**A=Angry**
1%	88%	5%					**B= Disgust**
	4%	92%	2%				**C=Fear**
		3%	97%	2%			**D=Happy**
			1%	94%	6%		**E=Sad**
				4%	90%	1%	**F=Surprise**
					3%	99%	**G=Natural**
99%	**88%**	**92%**	**97%**	**94%**	**90%**	**99%**	**Positive Prediction**
1%	**12%**	**8%**	**3%**	**6%**	**10%**	**1%**	**False Discovery Rate**

4.5.2 CLBP Technique Results on Three Variations of Datasets

The superiority of image normalization as a pre-processing technique is also established in Figure 4.12. The three variations of the dataset were used to extract features using Combined Local Binary Pattern (CLBP) technique and the classification is performed on these sets. It is obvious that image normalization has performed better than Contrast Limited Adaptive Histogram Equalization (CLAHE) for the same set of classifiers.

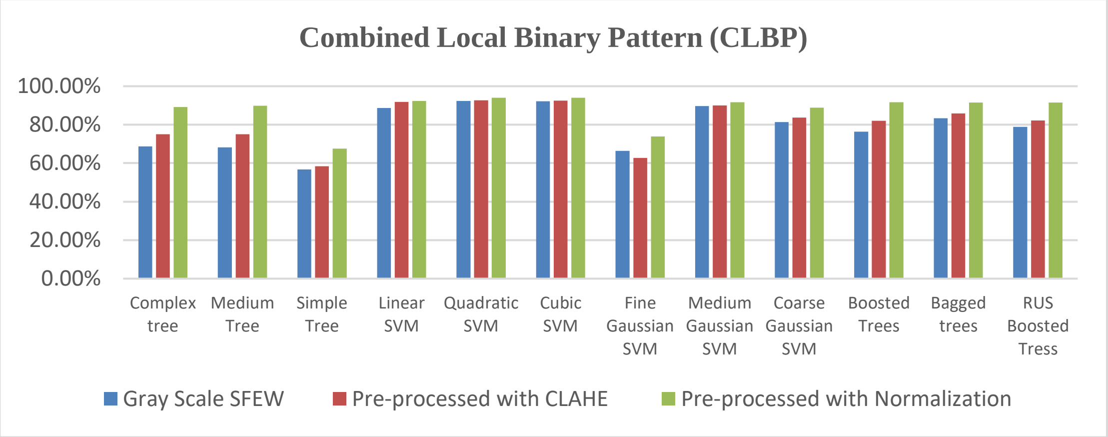


Figure 4.12 Results of CLBP Technique on Dataset Variations

The confusion matrix of the classifier is shown in table 4.2 when features are extracted using the CLBP technique with image normalization as a pre-processing stage.

Table 4.2: CLBP Quadratic SVM confusion matrix for SFEW dataset pre-processed with Image Normalization Steps

A	B	C	D	E	F	G	← Classified As
98%	5%						**A=Angry**
2%	95%	6%					**B= Disgust**
		85%	4%				**C=Fear**
		9%	94%	3%			**D=Happy**
			2%	96%	5%		**E=Sad**
				1%	89%	5%	**F=Surprise**
					5%	95%	**G=Natural**

4.5.3 Results of Proposed BGSD Technique for Three Variations of Datasets

The graphical comparison using Binary Gradient Structure Descriptor (BGSD) feature extraction technique for the same three datasets variants that is normal means without any preprocessing, pre-processed with Contrast Limited Adaptive Histogram Equalization (CLAHE) and images pre-processed with normalization procedure reveal that BGSD can represent the structure in a better way when image normalization is performed. Figure 4.13 verifies this fact.

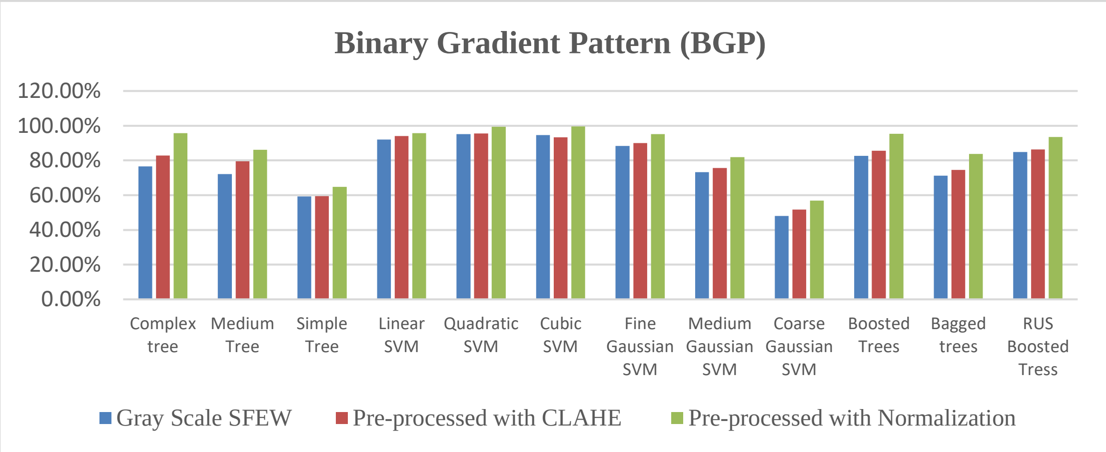


Figure 4.13 Results of Proposed Technique on Dataset Variations

Table 4.3 shows the confusion matrix of Cubic SVM using the features extracted using Binary Gradient Structure without using enhancement.

Table 4.3: Confusion Matrix of Cubic SVM using BGSD features for SFEW dataset

A	**B**	**C**	**D**	**E**	**F**	**G**	**← Classified As**
100%	8%						**A=Angry**
3%	88%	4%					**B= Disgust**
	4%	92%	2%				**C=Fear**
		4%	98%				**D=Happy**
				99%	6%		**E=Sad**
				1%	83%	5%	**F=Surprise**
					11%	95%	**G=Natural**
100%	**88%**	**92%**	**98%**	**99%**	**83%**	**95%**	**Positive Prediction**
	12%	**8%**	**2%**	**1%**	**17%**	**5%**	**False Discovery Rate**

Likewise, the confusion matrix using Cubic SVM for the features obtained from images enhanced exercising the CLAHE technique is presented in Table 4.4.

Table 4.4: Confusion Matrix of Cubic SVM using BGSD features for SFEW dataset pre-processed with CLAHE

A	**B**	**C**	**D**	**E**	**F**	**G**	**← Classified As**
100%	4%						**A=Angry**
	92%	5%					**B= Disgust**
	4%	91%	2%				**C=Fear**
		4%	97%	2%			**D=Happy**
			1%	95%	6%		**E=Sad**
				3%	89%	2%	**F=Surprise**
					4%	98%	**G=Natural**
100%	**92%**	**91%**	**97%**	**95%**	**89%**	**98%**	**Positive Prediction**
	8%	**9%**	**3%**	**5%**	**11%**	**2%**	**False Discovery Rate**

Using SFEW dataset normalized images, Cubic SVM employing structural features provide the best results exhibited in the following confusion matrix given in Table 4.5.

Table 4.5: Confusion Matrix of Cubic SVM using BGSD features for SFEW dataset pre-processed with Image Normalized Steps

A	**B**	**C**	**D**	**E**	**F**	**G**	**← Classified As**
100%							**A=Angry**
	98%	1%					**B= Disgust**
	2%	99%					**C=Fear**
			100%				**D=Happy**
				100%	1%		**E=Sad**
					99%		**F=Surprise**
						100%	**G=Natural**
100%	**98%**	**99%**	**100%**	**100%**	**99%**	**100%**	**Positive Prediction**
	2%	**1%**			**1%**		**False Discovery Rate**

The above tables show that the SFEW image dataset pre-processed and illumination normalized with gamma correction, DoG filtering, and contrast equalization improves the classification accuracy and gains better results.

Figure 4.14 provides the graphical comparison of three feature extraction techniques named Local Gradient Coding Horizontal and Diagonal (LGC-HD), Combined Local Binary Pattern (CLBP) and Binary Gradient Structural Descriptor (BGSD) used to extract the feature of grayscale images of SFEW dataset pre-processed with Contrast Limited Adaptive Histogram Equalization (CLAHE).

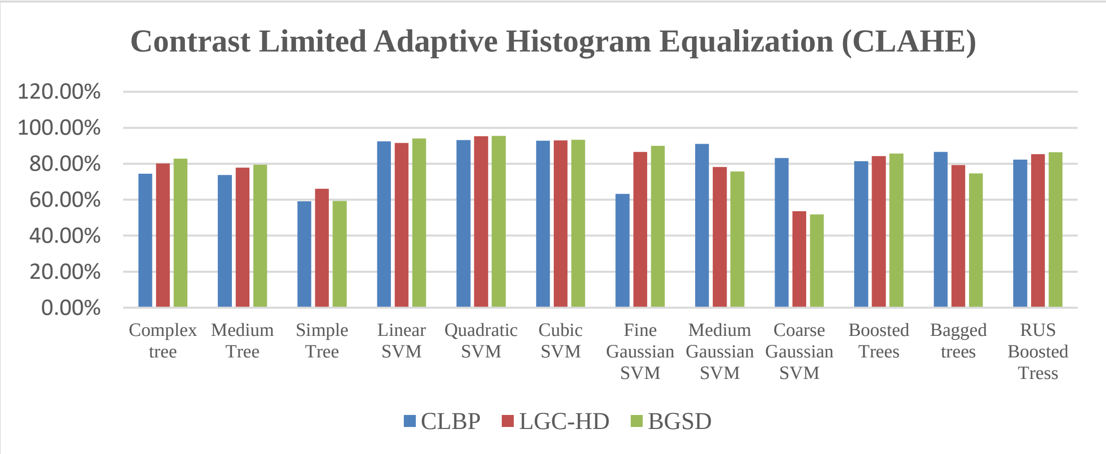


Figure 4.14: LGC-HD, CLBP and BGSD techniques result for SFEW dataset pre-processed with CLAHE

The following is the graphical comparison of three feature extraction techniques named LGC-HD, CLBP and Binary Gradient Structural Descriptor (BGSD) used to extract the feature of grayscale images of SFEW dataset pre-processed with Images normalization steps gamma correction, DoG filtering and contrast equalization.

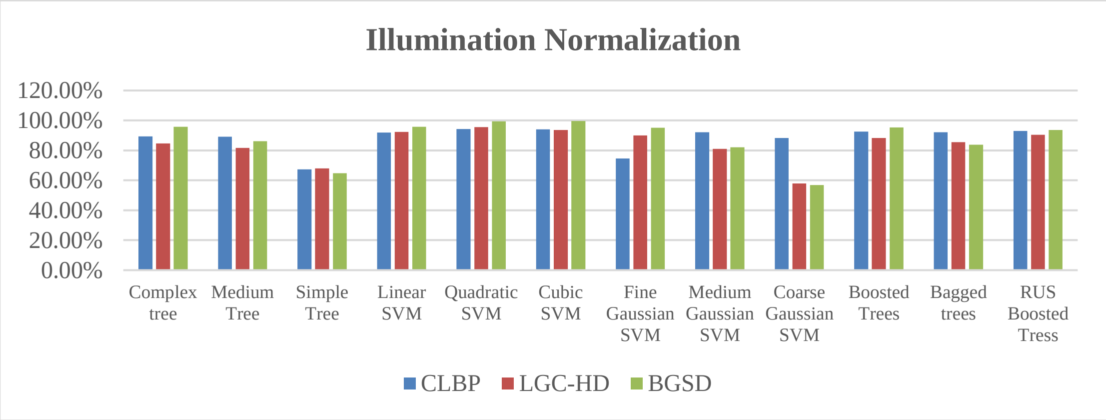


Figure 4.15: LGC-HD, CLBP and BGSD Techniques Results for SFEW Dataset Pre-processed with Image Normalization

The performance of emotion recognition is also compared with the framework proposed by Ying et al. [26] that used the JAFFE dataset for experimentation. The results show that the proposed LGC-HD is more efficient in extracting the features. This framework achieved an average recognition rate of 85.5% for the SVM. Faisal et al. [32] proposed a scheme in which the experiments are performed on CK + and JAFFE datasets and classification of expressions is done with an SVM classifier. This technique achieved an average recognition rate of 88.1%. Huang and Yin [35] proposed a framework that is powerful and robust. The robustness is based on the gradients of an image called binary gradient patterns that are considered powerful representatives of local feature and named as BGP. The experiments conducted on the ARFace dataset and classification done through Fisher's Linear Discriminant Analysis (FDA) that achieve the average recognition rate of 98.8%. The mechanism proposed in our study is proficient enough to obtain an average recognition rate of 95.3%, 95.7% and 99.6% for Boosted Trees, Complex Tree and Cubic SVM classifiers respectively. The results are listed in Table 4.6 given below.

Table 4.6: Comparison of the Proposed Technique with state of the art techniques

Sr. No.	**Techniques**	**Dataset Used**	**Classifiers**	**Accuracy**
1	CLBP	JAFFE	SVM	85.5%
		CK+		88.1%
2	LGC-HD	JAFFE	SVM	88.9%
3	BGP	ARFace dataset	CNN	98.8%
4	Proposed Technique (BGSD)	SFEW	Boosted Trees	95.3%
			Complex Tree	95.7%
			Cubic SVM	**99.6%**

The results are obtained using a 10-fold cross-validation method. The most fascinating part of our methodology is that it offers the best results for a basic Cubic SVM classifier. Overall the discriminating power of the BGSD feature is supplemented by the high value of accuracy achieved during expression recognition for other classifiers as well. This fact is shown in table 4.7.

Table 4.7: Summary of the Results

Classifier	**CLBP** Pre-processed with		**LGC-HD** Pre-processed with		**Proposed Technique (BGSD)** Pre-processed with	
	CLAHE	Illumination Normalization	CLAHE	Illumination Normalization	CLAHE	Illumination Normalization
Complex tree	75.1	89.2	80.1	84.7	82.8	**95.7**
Cubic SVM	92.5	94	92.9	93.7	93.4	**99.6**
Boosted Trees	82.1	91.6	84.3	88.3	85.6	**95.3**

The structural features of the image are computed using gradient operation for each pixel to obtain the optimized pattern thus the asymptotic complexity of the BGSD is O(MN) where M and N are the rows and columns respectively.

4.6 Summary

The effectiveness of facial expression recognition is highly dependent on the local intensity variations in the real world images. The issues related to shading and poor light are dealt with image normalization and even better than CLAHE. Further, the use of the novel technique BGSD has provided a feature set that is proved to be invariant to both global and local light imbalances. BGSD captures the local structure of the face region consistently and efficiently. Different classifiers are used to test these features and the best results are obtained by non-linear SVM. The performance of the proposed methodology is tested using the SFEW dataset and it surpasses the other state of the art techniques.

In the last chapter, conclusions based on the experimentation performed in Chapters 3 and 4 are presented. Also, future improvements are provided that need dealing to evolve the models.

Chapter 5

Conclusion and Future Enhancements

5. Conclusion and Future Enhancements

Local texture-based techniques for facial expression recognition are focused during this research by dealing with imbalanced illumination, shading and pose variation. These challenges are faced in real-life scenarios. These aspects are explored through experiments to recognize emotions in real scenarios.

The task of facial expression recognition is sufficiently easy and involves much less time and a simpler model to correctly identify an expression from images obtained under a controlled environment, by using computers. Also in such lab-controlled environments, the expressions are well-posed that reduces the ambiguity among a range of distinct expressions. But the complexity of facial expression recognition is increased if images are captured from events happening in the real world. In such images, light conditions may add shading or illumination variation that is sufficiently present on the face and there may exist large intra-class variation of a certain expression that makes the job of expression recognition very challenging. Also, the camera may not be correctly capturing the face from the front and there may be variation in the pose. This study has focused on the challenges that are linked to the images that are obtained in a real-world scenario.

5.1 Conclusion

In the first module of this research, a new framework Merged Binary Pattern Coding (MBPC) is successfully used for facial expression recognition in real-world situations. It encourages the use of contrast enhancement schemes to handle imbalance illumination after face localization. The extraction of local features using gradient and sign difference respectively makes it robust against pose variation, minor occlusion and slight local illumination problem left in the preprocessing stage. The number of features is reduced by retaining only important features using PCA to improve the performance of classifiers. The results describe that the proposed technique surpasses all the texture-based techniques that include LBP, LGC-HD, LGC-HVD and CLBP for both holistic and division-based approaches using a dataset containing real-world images. Moreover, the holistic approach offers significantly higher performance than a division based approach.

A new framework Binary Gradient Structure Descriptor is proposed in the second module of this study that extracts the textural structure from the images that mimic the real-life illumination for facial expression recognition. The results obtained using the SFEW dataset generally illustrate that if we handle the imbalance local illumination in the pre-processing phase, the facial expression recognition accuracy is increased after applying feature extraction techniques. Image normalization has successfully retained the structural information while dealing with poor illumination better than CLAHE.

Among all the discussed feature extraction techniques, Binary Gradient Structure Descriptor (BGSD) achieves the highest facial expression recognition accuracy of 99.6% when classified with a Cubic Support Vector Machine classifier. Features based on the gradient difference construct a descriptor that can handle pose variations and insignificant local illumination issues skipped in the pre-processing stage. This Support vector machine variant is very consistent in gaining higher facial expression recognition accuracy exhibited through experimentation.

Moreover, we have explored two feature extraction algorithms whose worst-case time complexity is equal to the size of the gray-scale image. The performance of these techniques is amplified by the pre-processing techniques to make images more suitable for feature extraction. An attempt is made to improve the results compared to the state of the art techniques used for the same purpose.

5.2 Future Enhancements

More reliable features to identify the behavior of humans may be obtained by effectively dealing with the issues present in real-life images. These issues include shading, poor illumination, pose variation and different occlusions. Preserving details in the preprocessing stage is the key element to capture accurate information for facial expression recognition from real-life images. It is more likely that these images may contain different types and levels of noise in addition to poor illumination. Our focus in the next stage of research will be to effectively handle the noise in the preprocessing stage to obtain even more promising results.

Apart from that, the higher resolution in the number of emotions in addition to the existing range i.e. happy, sad, anger and surprise, etc. will be considered. For advancing the human-computer interaction system more effort from the research community is required to introduce the more complex emotional states in addition to the six basic expressions such as pain, fatigue and mental state such as thinking, agreeing, disagreeing, laying and frustration as they have numerous application areas.

The accumulative emotion of a group and crowd present in videos may be identified that will help to provide security to different monitored zones. Furthermore, the missing features due to occlusion can be reconstructed that involves more experimentation after guessing the features missed due to the obstacles.

References

References

[1] Sidney D'Mello and Rafael Calvo. "Beyond the basic emotions: What should affective computing compute?" In Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems, pages 2287–2294, 2013.

[2] A. Mehrabian, "Communication without words", Psychol. Today 2 (9), 52–55, 1968.

[3] J. Zhang, Q. Zhang, and J. Hu. "RGB color centroids segmentation (CCS) for face detection" ICGST International Journal on Graphics, Vision and Image Processing, 9:1-9, 2009.

[4] K. Scherer, T. Banziger, and E. Roesch, editors "A Blueprint for Affective Computing" Oxford University Press, 2010.

[5] B. Golomb, T. Sejnowski, "Benefits of machine understanding of facial expression" in NSF Report- Facial Expression Understanding, Salt Lack City, UT, pp 55-71, 1997.

[6] L. Zhang, D. Tjondronegoro, and V. Chandran, "Facial expression recognition experiments with data from television broadcasts and the World Wide Web," Image Vis. Comput., vol. 32, no. 2, pp. 107–119, 2014.

[7] Izard, Carroll E, "Innate and universal facial expressions: Evidence from developmental and cross-cultural research", *Psychological Bulletin,* Vol 115, No. 2, pp. 288299, Mar. 1994.

[8] Z. Zeng, Y. Hu, Y. Fu, T. S. Huang, G. I. Roisman, and Z. Wen. "Audio-visual emotion recognition in adult attachment interview" *In Proceedings of the 8th international conference on Multimodal interfaces,* 2006.

[9] P. Ekman and W. Friesen "The facial action coding system: A technique for the measurement of facial movements" *Consulting Psychologist*, 1978.

[10] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.

[11] M.-H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods.," in Fgr, 2002, vol. 2, p. 215.

[12] Ghimire, Deepak & Lee, Joonwhoan. "Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines". Sensors. 13. 7714-7734 (2013).

[13] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, 2009.

[14] I. Kotsia and I. Pitas, “Facial expression recognition in image sequences using geometric deformation features and support vector machines,” *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, 2007.

[15] P. Ekman, “An argument for basic emotions,” Cogn. Emot., vol. 6, no. 3–4, pp. 169–200, 1992.

[16] D. Ghimire, J. Lee, Z.-N. Li, S. Jeong, S. H. Park, and H. S. Choi, “Recognition of facial expressions based on tracking and selection of discriminative geometric features,” *Int J Multimed. Ubiquitous Eng*, vol. 10, no. 3, pp. 35–44, 2015.

[17] X. Ben, P. Zhang, R. Yan, M. Yang, and G. Ge, “Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation,” *Neural Comput. Appl.*, pp. 1–18, 2015.

[18] Wei-Lun Chao, Jun-Zuo Liu,Jian-Jiun Ding, Po-Hung Wu "Facial Expression Recognition Using Expression Specific Local Binary Patterns and Layer Denoising" Graduate Institute of Communication Engineering National Taiwan University, IEEE, 2013.

[19] J. Whitehill, M. S. Bartlett, and J. Movellan, “Automatic facial expression recognition” Social Emotions in Nature and Artifact, London, U.K.: Oxford Univ. Press, 2013.

[20] Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S. “Comparison between geometry-based and gabor-wavelets based facial expression recognition using multilayer perceptron” In Automatic Face and Gesture Recognition, IEEE (1998).

[21] Ojala, T., Pietikäinen, M., Harwood, D. “A comparative study of texture measures with classification based on featured distributions” Pattern recognition, 29(1), 51-59 (1996).

[22] Kumari, J., Rajesh, R., Pooja & K.M. “Facial expression recognition: a survey” Procedia Computer Science 58, 486–491 (2015).

[23] Yanpeng Liu, Yuwen Cao, Yibin Li, Ming Liu, Rui Song" Facial Expression Recognition with PCA and LBP Features Extracting from Active Facial Patches"

Real-time Computing and Robotics (RCAR), IEEE International Conference 6-10 June 2016.

[24] Yan Ouyang, Nong Sang, Rui Huang "Accurate and robust facial expressions recognition by fusing multiple sparse representations based classifiers" Neurocomputing 149 (2015) 71–78 Elsevier 2014.

[25] Wei-Lun Chao, Jun-Zuo Liu, Jian-Jiun Ding, Po-Hung Wu "Facial expression recognition using expression-specific local binary patterns and layer denoising mechanism" Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on 10-13 Dec. 2013.

[26] Ying Tong, Rui Chen, Yong Cheng "Facial expression recognition algorithm using LGC based on horizontal and diagonal prior principle" *Optik - International Journal for Light and Electron Optics Volume 125, Issue 16, August 2014, Pages 4186–4189.*

[27] Anusha Vupputuri, Sukadev Meher "Facial Expression recognition using Local Binary Patterns and Kullback Leibler divergence" *Communications and Signal Processing (ICCSP), 2015 International Conference on 2-4 April 2015.*

[28] Pierluigi Carcagnì, Marco Del Coco, Marco Leo, Cosimo Distante "Facial expression recognition and histograms of oriented gradients: a comprehensive study" *Springerplus. 2015; 4: 645.*

[29] Ling Gan, Sisi Si "Facial expression recognition based on two-step feature histogram optimization" *3rd International Conference on Materials Engineering, Manufacturing Technology and Control (ICMEMTC 2016).*

[30] Liu, Y.; Li, Y.; Ma, X.; Song, R. "Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas". *Sensors* **2017**, *17*, 712.

[31] Jadisha Yarif Ramírez Cornejo , Helio Pedrini "Recognition of occluded facial expressions based on CENTRIST features" *Acoustics, Speech, and Signal Processing (ICASSP), 2016 IEEE International Conference on 20-25 March 2016.*

[32] Faisal Ahmed, Hossain Bari,Emam Hossain "Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP)" *The International Arab Journal of Information Technology, Vol. 11, No. 2, March 2014.*

[33] Xiaoyang Tan, William Triggs “Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions” *IEEE Transactions on Image*

Processing, Institute of Electrical and Electronics Engineers, 2010, 19 (6), pp.1635-1650.

[34] M. H. Siddiqi et al., "Human facial expression recognition using curvelet feature extraction and normalized mutual information feature selection," Multimed. Tools Appl., vol. 75, no. 2, pp. 935–959, 2016.

[35] W Huang, H Yin "Robust face recognition with structural binary gradient patterns" Pattern Recognition, March 2017 - Elsevier.

[36] E. Mary Shyla, M. Punithavalli "Performance analysis of facial expression recognition schemes". Internation journal of Computer Applications-NCATIT, (2013).

[37] Rasoulzadeh, M. "Facial Expression recognition using Fuzzy Inference System" International Journal of Engineering and Innovative Technology (IJEIT) Volume 1, Issue 4, April 2012.

[38] Luo, Yuan & Wu, Cai-ming & Zhang, Yi. "Facial expression recognition based on fusion feature of PCA and LBP with SVM". Optik. 124. 2767-2770, (2013).

[39] Jun-Yong Zhu, Wei-Shi Zheng, Feng Lu, Jian-Huang Lai, "Illumination invariant single face image recognition under heterogeneous lighting condition", Pattern Recognition, Volume 66, 2017, Pages 313-327, ISSN 0031-3203.

[40] Awais Mahmood, Shariq Hussain, Khalid Iqbal, and Wail S. Elkilani, "Recognition of Facial Expressions under Varying Conditions Using Dual-Feature Fusion," Mathematical Problems in Engineering, vol. 2019, Article ID 9185481, 12 pages, 2019. https://doi.org/10.1155/2019/9185481.

[41] Wang, Yingying & Li, Yibin & Song, Yong & Rong, Xuewen. "Facial Expression Recognition Based on Auxiliary Models. Algorithms". 12. 227. 10.3390/a12110227, (2019)

[42] Sajjad, M., Shah, A., Jan, Z. et al. "Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery". Cluster Comput 21, 549–567 (2018).

[43] Asit Barman, Paramartha Dutta,"Facial expression recognition using distance and texture signature relevant features" Applied Soft Computing, Volume 77, 2019, Pages 88-105, ISSN 1568-4946.

[44] I. Michael Revina, W.R. Sam Emmanuel, "Face Expression Recognition with the Optimization based Multi-SVNN Classifier and the Modified LDP Features"

Journal of Visual Communication and Image Representation, Volume 62, 2019, Pages 43-55, ISSN 1047-3203, https://doi.org/10.1016/j.jvcir.2019.04.013.

[45] Yan Yan, Zizhao Zhang, Si Chen, Hanzi Wang, "Low-resolution Facial Expression Recognition: A Filter Learning Perspective", Signal Processing (2019).

[46] J. Anil and L. P. Suresh, “Literature survey on face and face expression recognition,” in *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*, 2016, pp. 1–6.

[47] Ojala T., Pietikainen M., and Maenpaa T., "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 971-987, 2002.

[48] Faisal Ahmed, Hossain Bari, and Emam Hossain, "Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP)", The International Arab Journal of Information Technology, Vol. 11, No. 2, March 2014.

[49] Marina Sokolova, Guy Lapalme, "A systematic analysis of performance measures for classification tasks", Information Processing & Management, Volume 45, Issue 4, 2009, Pages 427-437, ISSN 0306-4573.

[50] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "*The Japanese female facial expression (JAFFE) database"*. 1998.

[51] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 2010, pp. 94–101.

[52] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, “Acted facial expressions in the wild database,” *Aust. Natl. Univ. Canberra Aust. Tech. Rep. TR-CS-11*, vol. 2, 2011.

[53] C.V. Loan, “Computational Frameworks for the Fast Fourier Transform,” *Society of Industrial and Applied Mathematics* (SIAM), 1992.

[54] C.R. Gonzalez, E. Richard, "Digital image processing", 4th ed. *New Jersey: Pearson Education;* 2008.

[55] X. Huang, S.-J. Wang, G. Zhao, and M. Piteikainen, “Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection,” in

Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 1–9.

[56] Sadia Arshid, Ayyaz Hussain, Asim Munir, Anum Nawaz, Sanneya Aziz. "Multi-stage binary patterns for facial expression recognition in real world", Cluster Computing, 2017.

[57] Asim Munir, Ayyaz Hussain, Sajid Ali Khan, Muhammad Nadeem, Sadia Arshid. "Illumination invariant facial expression recognition using selected merged binary patterns for real world images", Optik, 2018.

[58] P. S. Subramanyam and S. A. Fegade, "Face and Facial Expression Recognition-A Comparative Study," 2013.

[59] M. R. Faraji and X. Qi, "Face recognition under varying illuminations using logarithmic fractal dimension-based complete eight local directional patterns," *Neurocomputing*, vol. 199, pp. 16–30, 2016.

[60] A. Hussain, M. S. Khan, M. Nazir, and M. A. Iqbal, "Survey of various feature extraction and classification techniques for facial expression recognition," in *Proceedings of the 11th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications, and Proceedings of the 11th WSEAS International Conference on Signal Processing, Robotics and Automation, and Proceedings of the 4th WSEAS International Conference on Nanotechnology; World Scientific and Engineering Academy and Society (WSEAS)*, 2012, pp. 138–142.

[61] L. U. O. Yuan, C. Wu, and Y. Zhang, "Facial expression feature extraction using hybrid PCA and LBP," *J. China Univ. Posts Telecommun.*, vol. 20, no. 2, pp. 120–124, 2013.

[62] S. A. Khan, A. Hussain, and M. Usman, "Facial expression recognition on real world face images using intelligent techniques: A survey," *Opt.-Int. J. Light Electron Opt.*, vol. 127, no. 15, pp. 6195–6203, 2016.

[63] M. Kolsch and M. Turk. "Analysis of rotational robustness of hand detection with a viola-jones detector". In International Conference on Pattern Recognition, pages 107-110, 2004.

[64] Ali M. Reza. "Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement." Jurnal of VLSI Signal Processing, vol. 38, pp 35-44, 2004.

[65] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.M. ter Haar Romeny, J.B. Zimmerman, and K. Zuiderveld, "Adaptive Histogram Equalization and its Variations," Computer Vision, Graphics and Image Processing, vol. 39, 1987, pp. 355-368.

[66] Vladimir N. Vapnik. "Statistical Learning Theory". Wiley-Interscience, 1998.

[67] Quinlan, J. R. 1986. "Induction of Decision Trees". Mach. Learn. 1, 1 (Mar. 1986), 81–106.

[68] Mitchell, Tom M. "Machine Learning". McGraw-Hill, 1997. pp. 55–58.

[69] Rokach, L. (2010). "Ensemble-based classifiers". Artificial Intelligence Review. 33 (1-2): 1–39.

[70] Kuncheva, L. and Whitaker, C., "Measures of diversity in classifier ensembles", Machine Learning, 51, pp. 181-207, 2003.

[71] Sollich, P. and Krogh, A., "Learning with ensembles: How overfitting can be useful", *Advances in Neural Information Processing Systems*, volume 8, pp. 190-196, 1996.

[72] Brown, G. and Wyatt, J. and Harris, R. and Yao, X., "Diversity creation methods: a survey and categorization". *Information Fusion*, 6(1), pp.5-20, 2005.

[73] Lior Wolf, Tal Hassner, and Yaniv Taigman. "Descriptor Based Methods in the Wild". *Faces in Real-Life Images workshop at the European Conference on Computer Vision (ECCV)*, Marseille, 2008.

[74] S. A. M. Al-Sumaidaee, S. S. Dlay, W. L. Woo and J. A. Chambers, "Facial expression recognition using local Gabor gradient code-horizontal diagonal descriptor," *2nd IET International Conference on Intelligent Signal Processing 2015 (ISP)*, London, 2015, pp. 1-6.

[75] Richard O. Duda, Peter E. Hart, David G. Stork, " Pattern Classification", 2nd Edition, JOHN WILEY & SONS, INC., 2001.

[76] Rizwan Ahmed Khan. Detection of emotions from video in non-controlled environment. Image Processing [eess.IV]. Université Claude Bernard - Lyon I, 2013. English. ffNNT : 2013LYO10227f

[77] Asim Munir, Ayyaz Hussain, Muhammad Nadeem, Muhammad Rizwan Mughal, Sajid Ali Khan, "Local Gradient-Based Technique for Facial Expression Recognition from Images with Poor Illumination", Multimedia Tools and Applications, 2020. [Submitted]