Robust Facial Expression Recognition using Local Binary Patterns and Gabor Filters

A Thesis Submitted In Partial Fulfillment Of The Requirements For The Degree Of

Master of Technology

in

Signal & Image Processing

by

Anusha Vupputuri

213EC6273

Under the supervision of

Prof.Sukadev Meher

Department of Electronics and Communication in Engineering

National Institute of Technology Rourkela

Odisha,India-769008

Dedicated to My Family...

DEPARTMENT OF ELECTRONICS AND
COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA
ODISHA, INDIA-769008

CERTIFICATE

This is to certify that the Thesis Report titled, **"Robust Facial Expression Recognition using Local Binary Patterns and Gabor Filters"**, submitted by **Ms. Anusha Vupputuri** bearing**Roll No. 213EC6273** in partial fulfillment of the requirements for the award of the degree of **Master of Technology in Electronics and Communication Engineering** with specialization in**"Signal & Image Processing"** during session 2013 - 2015 at National Institute of Technology Rourkela is an authentic work carried out by her under my supervision and guidance.

Prof. Sukadev Meher

DEPARTMENT OF ELECTRONICS AND
COMMUNICATION ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA
ODISHA, INDIA-769008

DECLARATION

I certify that

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

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

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

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

Anusha Vupputuri

Acknowledgment

I would like to express my gratitude to my guide **Prof. Sukadev Meher** for his guidance, advice and support throughout my thesis work. I am especially indebted to him for teaching me both research and writing skills, which have been proven beneficial for my current research and future career. Without his endless efforts, knowledge, patience, and criticism this research would have never been possible. The experimental methods and results presented in this thesis have been influenced by him in one way or the other. It has been a great honour and pleasure for me to do research under the supervision of Prof. Sukadev Meher.I would like to thank him for being my advisor here at National Institute of Technology, Rourkela.

Next, I want to express my respects to **Prof. K.K Mahapatra, Prof. S. K. Patra, Prof. Manish Okade,Prof. A. K. Sahoo, Prof. L.P.Roy,Prof. Samit Ari,Prof .A.K. Swain, Prof. D.P. Acharya** for teaching me and also helping me how to learn. They have been great sources of inspiration to me and I thank them from the bottom of my heart.

I would like to thank **Mrs.Sonia Das, Mr.Deepak Panda and Mr.Sananda Kumar** and staff of the Department of Electronics and Communication Engineering, N.I.T. Rourkela, for their generous help for the completion of this thesis.

I would like to thank my friends **Sobhan, Suraj, Sowjanya, Sailaja** and my labmates for thoughtful and mind stimulating discussions we had, which prompted to think beyond the obvious. I've enjoyed their companionship so much during my stay at NIT, Rourkela.

I am especially indebted to my parents **Sambaiah and Savitri** and my sister **Manaswini** for their love, sacrifice, and support. My parents are my first teachers, after I came to this world and I have set of great examples for me about how to live, study and work.

Anusha Vupputuri

Contents

ABSTRACT

Facial expressions and gestures provide intuitional cues for interpersonal communication. Imparting intelligence to computer for identifying facial expressions is a crucial task. Facial expressions and emotions are governed by identification of facial muscle movement by visual cortex and training a machine to identify these highly in-situ movements is our primary interest. This thesis presents robust facial expression analysis algorithms for static images as well as an efficient extension to sequence of images.

We present an efficient preprocessing method which eliminates the effect of illumination on the detected face images thus making them efficient for feature extraction. Robust Local Binary Patterns and Gabor filters are implemented for feature extraction which are known to provide efficient face representation and analysis.LBP facial features are represented in form of weighted histograms which are best classified using Kullback Leibler divergence measure .Artificial Neural Network classifier is also tested for classification of fused Gabor and LBP features.

Further expressions are rarely defined by static images as their complete essence lies in a sequence of images. So further exploration is concentrated on analyzing expressions from a sequence of images. To eliminate head pose variations in consecutive frames and register images to keep the spatial information intact which is necessary for LBP feature representation we adopted SIFT flow alignment procedure and further tested the resultant image classification with implemented algorithms. The classification accuracy resulted in 95.24% for static expression images and 86.31 % for sequence of images which is indeed appreciable when compared to other standard methods.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In present day scenario computers have become more ubiquitous and indispensable part of our lives. For this same reason Human Computer Interaction (HCI) has become an emerging area of research and it is a prior necessity for imparting intelligence to computers to understand and act according to human behaviour. Interpersonal communication is broadly classified into verbal communication and nonverbal communication. Verbal communication consists of only raw voice data input and nonverbal communication accounts for the tone and intensity of voice merged with facial expressions and gestures. The combined effect of these instils cognition to communicate. Recognising facial expressions is vital in human computer interfaces because it is a visible embodiment of a person's psychological state, intention and personality. Without vocal data facial expression cues combined with gestures efficiently elicit the internal meaning of speaker which forms the basis of facial expression recognition.

In communication 55 percent of perception is through facial expression, 38 percent through gestures and voice data carries 7 percent of information. Facial expressions are considered to be uniform universally among different human races. Facial expression can be defined as a temporal deformation of facial features like eyes, nose, lips, cheeks,etc which is a result of muscular activity aroused by internal feelings or events occurring in the surroundings. Extent of Opening of eyes, frowning of eyebrows,rise of eye brows, widening and shortening of mouth especially at the corners form an important aspect of expression classification as identified by the human visual cortex. Hence our facial expression recognition system should be designed such that even the slightest change in the movement of facial organs can be efficiently identified and adhere to exact classification.

1.2 Types of Emotion

According to Russel[1] based on the extent of activation and pleasure there are about 200 different emotions as listed below:

- **High activation:** aroused,astonished,stimulated,surprised,active,
- **Pleasant:** happy,delighted,glad,cheerful,pleased,warmhearted
- **Low activation:** idle,tranquil,still,passive,quiet
- **Unpleasant:** unhappy,sad,miserable,grouchy,blue
- **High activation+Pleasant :** enthusiastic,elated,excited,euphoric,lively
- **Low activation+Pleasant :** relaxed,contented,atrest,calm,serene
- **High activation+Unpleasant :** distressed,annoyed,fearful,nervous,jittery,anxious
- **Low activation+Unpleasant :** dull,tired,drowsy,sluggish,bored,droopy

All the above listed expressions vary with slightest change in the facial features , hence these are named as micro facial expressions . Detecting these changes is a crucial and difficult task and resultant emotion displayed by human is often a combined outcome of numerous microfacial expressions.So,classifying these micro expressions exactly is indeed not possible rather difficult. Paul Ekman a renowned psychologist classified expressions and emotions into seven universal classes namely[2]:

- **Anger**
- **Disgust**
- **Fear**
- **Happy**
- **Neutral**
- **Sad**
- **Surprise**

1.3 Applications of Facial Expression Recognition System

Application of Facial emotion recognition can be seen in different HCI areas such as:

1. **Treatment of Asperger's syndrome:** Asperger's syndrome or autism is a disorder where children are unable to recognise the words and emotions of the speaker which is a hindrance to interact with others. A Facial expression recognition system would assist them by recognising the speaker's emotion and help them with day to day communication.

2. **Driver state surveillance:** Driver state surveillance is an ultimate necessity for preventing unforeseen circumstances. When the driver is prone to an accident the foremost thing that happens is the flush of fear and anxiety in his face which if immediately identified can prevent all forthcoming happenings using automated preventive measures.

3. **Commercial survey:** As online shopping is gaining importance it is necessary to survey the amount of satisfaction the product replenishes on the customer. By recognizing the emotion on the customer's face the extent of satisfaction can be measured and hence product success can be estimated.

4. **Human computer interaction:** Human computer interaction is an emerging field where computers can interact with people by understanding the audio, video or combination signal input delivered and act accordingly. Recognising emotions in HCI platforms like gaming consoles for example Xbox Kinect,etc deals with cognitive and affective aspect of interaction where the machines changes according to the state of user.

5. **Affective computing:** Affective computing is the study and improvement of frameworks and gadgets that can perceive, decipher, handle, and recreate human influences. It is an interdisciplinary field crossing software engineering, brain science, and subjective science. The machine ought to translate the enthusiastic condition of people and adjust its conduct to them, giving a fitting reaction for those feelings.

1.4 Basis

Basic steps involving facial expression recognition are face detection, feature extraction followed by classification measures. Proficient expression recognition can only be achieved when these modules are chosen fittingly to distinguish the most differentiable features representing facial movements and have correspondence with one another in terms of bringing out accurate results while adjusting precision. This being an emerging topic under research many methods like PCA[3],LDA[3],Gabor filter based [4], and LBP [5] feature extraction have been proposed and various classification methodologies were combined with them. But each of these combinational algorithms has suffered either inaccuracy or hardware complexity.So there is a prior necessity to propose a facial expression recognition system which balances both accuracy and complexity while classifying emotions.Most of the above algorithms were unable to classify expressions exactly and there still lies confusion between almost similar classes. Further expression here was recognised from static images and not videos sequences. Here in this thesis an attempt was made to infer expression from series of images or video frames using SIFT flow[6] using registration of face.

1.5 Problem Description

The main aim of this thesis is to design a facial expression recognition system which can classify expression into predetermined set of classes accurately and efficiently from both static as well as a sequence of images. It further aims at recognising slightest change detection in facial features which results in varied emotion classes.

1.6 Organization of Thesis

The thesis is organized as follows:

- Chapter 2 provides basic information regarding step involved in a facial expression recognition system namely face detection, preprocessing, feature extraction followed by classification and throws light on the methods implemented till date in each module, comparing each with the next best.

- Chapter 3 gives a basic introduction to Local Binary Patterns [5] and Gabor filters[4] for feature extraction.
- Chapter 4 deals with the algorithms which have been proposed and comparison of their results.This work is also extended to sequence of images using SIFT flow algorithm.
- Chapter 5 concludes the work done with a knowledge into future extent of the work.

Chapter 2

OVERVIEW OF FACIAL EXPRESSION RECOGNITION

2.1 Generalised procedure of Facial Expression Recognition

The major procedural steps involved in facial expression recognition namely, face detection, feature extraction and classification. Dimensionality reduction can also be included based on the computational complexity and feature vector length. Different systems for perceiving human facial expressions from face pictures have been proposed and their execution has been assessed with databases of face pictures with varieties in expressions.Different modules involved in facial expression recognition and the approaches for implementation of each have been reviewed and their hindrances were tried to overcome in this project.

The generalized procedure followed for facial expression recognition is shown as a block diagram in figure 2.1.The images which are to be classified form the test image set and in order to train the system some predefined cues are to be given as input. This is implemented in the form of features extracted from the set of training images.The training set is a combination of images from different classes and distinguishable features are extracted for common classes to provide information cues to the classifier.

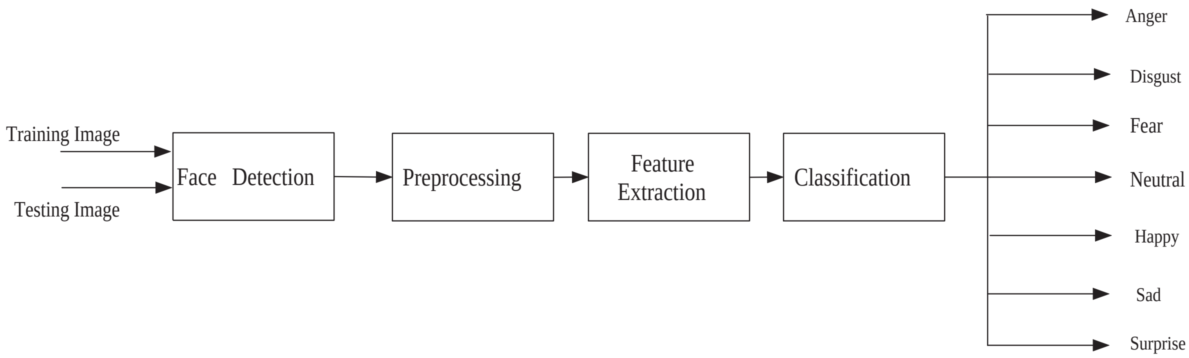


Figure 2.1: Generalised procedure of facial expression recognition

The training set is a combination of images from different classes and distinguishable features are extracted for common classes to provide information cues to the classifier.

2.2 Overview of approaches reviewed for different modules of recognition

2.2.1 Face detection

Face detection has two main approaches:

- **Skin color based segmentation :**[7] in which area and color of the skin are major parameters for classifying face from non face.The images are represented in YCbCr , RGB and HCI color models and a new model is proposed by combining them.As skin color becomes an important cue for the presence of face in an image Sanjay Kr. Singh et al. have proposed detection of face using skin color.This procedure becomes inefficient with movement and varying illumination and also when background is of the similar color of skin many false detections are present. Another hindrance of skin color is it is non uniform for all races and varies with ageing.

- **Boosted cascade of simple features :**[8] mostly known as 'Viola Jones algorithm' in which Haar like features are cascaded in each stage and face class is filtered out from non face classes using Adaboost algorithm.

2.2.2 Feature extraction

Feature extraction is categorised by two types:

- **Geometric feature :**based [9] which extracts features based on the movement , shape and location of facial muscles and of main organs like eyes, cheeks, nose lining of lips . The movement of these regions termed as action units(AUs) generalized based on their involvement in an expression and judgement for classification is done accordingly. Et al. have used fiducial points to track facial features in order to classify the expressions.This method is computationally heavy and real time implementation is difficult.

- **Appearance based features :**

  - *Linear Discriminant Analysis (LDA)*[3]

    Facial expression classes have some highly distinguished features separating classes from one another. Such describing features if identified can be mathematically utilized to solve a multiclass problem.Here images are projected over a subspace namely 'Fisher space'which usually deals with reducing feature dimensionality and classification of information.

  - *Gabor Filters*[4]

Facial expression are highly dependent on the direction of movement of facial organs and muscles.An excellent feature which is inspired by Human Visual Cortex and can detect the orientation of facial movement is Gabor filters.They extract facial features with varying orientations and dimensions when applied on image which can be further analysed.Thus Gabor filtered features are thus efficient in representing facial expressions.

- *Local Binary Pattern (LBP)*[5]

  Local binary patterns are predominantly used for texture identification and representation can be extended to face representation. Most importantly they can identify features which aim at representing varying texture of face with facial movements or expressions. The local binary pattern operator can be termed as an image operator transforming an image into an array of integer labels representing the image in small-scale. The statistics of these patterns use histogram which is further analysed. The LBP operator has many versions designed for analysis of monochrome and color still images and also volumetric data and videos.

2.2.3 Feature classification

Efficient approaches for classification are mentioned below:

- **Support Vector Machine (SVM) :**[30] is allocated administered learning models with related learning calculations that dissect information and perceive examples, utilized for classification. It is a binary classifier but can be extended to multiclass classification.SVM is a non-probabilistic binary classification method where classes are separated by boundary and patterns are shown as points in space.SVM being supervised learning method it is reliable tool for classification.

- **Distance classifiers :**like nearest neighbor classifier and KNN classifier which give the minimum distance class as detected class of test image. Efficiency of distance classifiers is less but their hardware implementation easy.
- **Artificial neural networks :**[10]are another efficient method to classify samples with large number of attributes. They take help of a back propagation algorithm such that every layer is trained efficiently with numerous iterations minimizing classification error in adhering to the correct class.

2.3 Preprocessing

Preprocessing of facial images is an essential step to enhance and condition facial expression recognition.Obtaining a pure facial expression image with normalized intensity accompanied by uniformity in shape and size[11] is the prime motive of preprocessing. It also eliminates the effect of uneven illumination and improper lighting conditions. The preprocessing procedure adopted here as a part of this system performs the following five steps in converting an image to a normalized pure expression image for feature extraction:

1. **Automatic detection of eye feature points**

   The real time images which we obtain are not facial images so we need to detect face from these images.Here we followed 'Viola Jones face detection algorithm' [8] .

   The highlights of this algorithm are[8] :

   - *Detection based on features not on pixels*
   - *Calculation of integral image*
   - *Only highly important features from various amount of features are detected using a variant or adaboost classifier or learning algorithm.*

- *Combination of several weak classifiers forming a strong classifier which enables the achievement of higher detection rates.*

Classification of images in this procedure is based on the value of simple features because feature based system is faster in operation when compared to pixel based system performance.Haar like basis functions are used as the describing features.

The difference between sum of pixels of two rectangular regions gives the two rectangle feature.The region on which feature is considered should have same size and are horizontally or vertically adjacent.Three rectangle features computing the sum of two outside rectangles and subtracting the same from the sum of pixels of center rectangle results in three rectangle feature.Four rectangle feature is computed by differencing the diagonal rectangular pair.

2. **Computation of 'integral image'**

The concept of integral image simplifies can be computation of Rectangle features extensively.The integral image at location (x,y) is contributed by summation of pixels residing above and to the left of (x,y) the same being included. M×N using DHT is given by

$$ii_m(x_m, y_m) = \sum_{\substack{(x_m)' \leq x \\ (y_m)' \leq y}} i(x'_m, y'_m) \tag{2.1}$$

where iim is the integral image and im is the actual image. Single iteration can compute Integral image using the equations

$$\begin{aligned} Si_m(x_m, y_m) &= Si_m(x_m, y_m - 1) + i(x_m, y_m) \\ ii_m(x_m, y_m) &= ii_m(x_m - 1, y_m) + Si_m(x_m, y_m) \end{aligned} \tag{2.2}$$

where Si_m(x_m,y_m) is cumulative row sum Si_m(x_m,-1)=0 and ii_m(-1,y_m)=0.Thus by obtaining integral image any rectangular sum for the haar features can be obtained by 4 array references only.For adjacent rectangles we can call by six references,eight and nine for two ,three and four rectangle features.

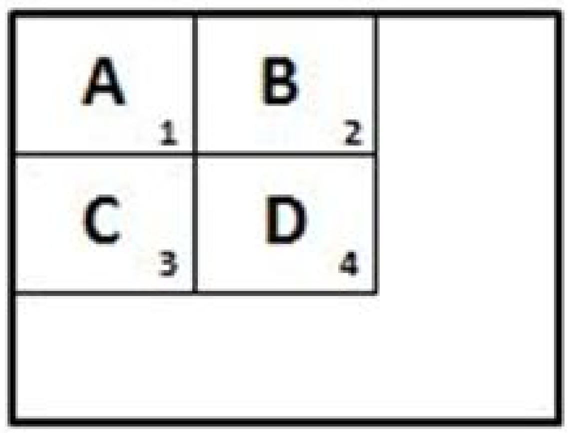


Figure 2.2: Referencing of integral image

With reference to Fig.2.2 integral image value at 1 is sum of pixels in rectangle A,at location 2 it is sum of both A and B, for C it is A+C and for location 4 is A+B+C+D. For sum of pixels in only rectangle D is 4+1-(2+3) which is a difference of diagonal sums.

3. **Identification of most efficient Haar features for classification**

   A learning algorithm is adopted here to identify face classes and non face classes and respectively train to identify the haar features which detect the face part in an image. In this algorithm extremely large number of features result in a small sub window. Hence to extract only the most distinguishing features a variant of Adaboost algorithm is applied which selects single rectangular features to differentiate face and non face classes.

4. **Face detection and removal of illumination effect**

   Referring to Fig.2.3 detection algorithm is applied to the images and it returns the center points of eyes as a structure. Based on the line joining these center points the detected face image is rotated to align with this line.

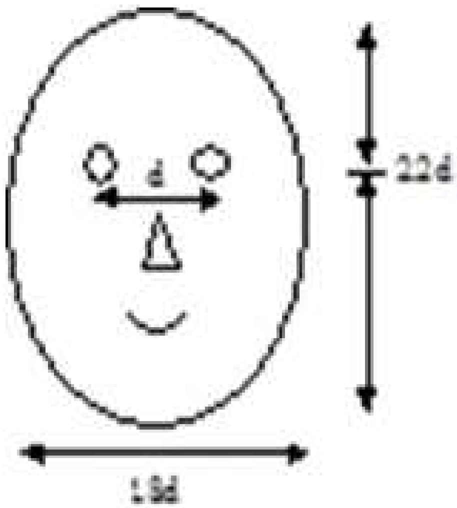
   Face region is located and cropped using a rectangle according to face model as shown in Fig. 2.3a.The distance between two eyes is assumed to be d, the cropped face rectangle will be 2.2d×1.8d;

   Real time images are not governed by uniform illumination and lighting conditions. As the feature extraction methods which we further utilize are based on intensity values non uniform illumination can highly vary the extracted feature values ,which in turn is responsible for determining minutely varying facial features.

   Hence we adopted illumination equalization and normalization as a part of preprocessing step. The series of steps followed for preprocessing is near to the process followed by mammalian visual cortex according to Tan et. al. [11] and are described as follows:

   - **Gamma correction**

     Gamma correction is a basic image processing operation which is nonlinear and helps in enhancing darker regions and decreasing the intensity of lighter regions. The transformation used here is I^{γ} where $\gamma = 0.2$ .Face part usually has higher intensity values which distinguish face from surrounding like hair and clothing.In order to enhance this lighter region we have chosen this



(a) Face model

(b) original face

(c) Detected eye pair

(d) Resultant Cropped face

Figure 2.3: Face detection and automatic cropping

particular value which keeps object information intact even with illumination effect.

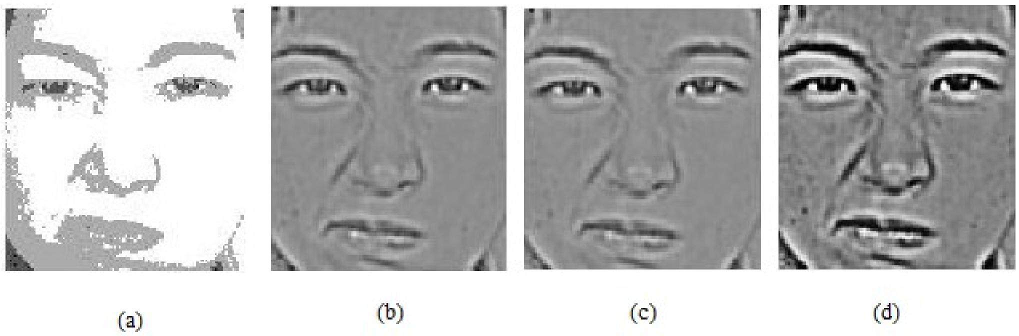


Figure 2.4: (a) gamma corrected image (b) Difference of Gaussian filtered image (c) contrast equalized image (d)final preprocessed image from JAFFE database.

- **Difference of Gaussian (DoG) filtering**

  Gamma correction cannot completely eliminate illumination non uniformity. Shadowing is another major drawback for feature extraction. High frequency components contain both essential information and shadows. So we follow bandpass technique to eliminate shadows while preserving incidental information.Fine Gaussian filtering is applied with two standard deviation values whose difference frames the pass band. The two Gaussian filters have a fine difference of 2 -3 pixels which eliminates shadow effect.Selected values are $\sigma_1 = 1$ and $\sigma_2 = 2$ for the two Gaussian filters.

- **Contrast Equalisation**

  This is the final step of preprocessing which provides global scaling.An image is a mixture of highlight extremas, distorted image border values little dark regions for example here it is nostrils. This is a two stage process implemented by these equations:

$$Im(x_m, y_m) = \frac{Im(x_m, y_m)}{(mean(|Im(x'_m, y'_m)|^{\beta}))^{\frac{1}{\beta}}} \tag{2.3}$$

$$\mathrm{Im}(x_m, y_m) = \frac{\mathrm{Im}(x_m, y_m)}{(mean(\min(\tau, |\mathrm{Im}(x'_m, y'_m)|)^{\beta}))^{\frac{1}{\beta}}} \tag{2.4}$$

The influence of large valuesare reduced by a compressive component β and τ is a threshold truncating large values. The determined values of β is 0.1 and $\tau$is 10 Last but not the least , there may still remain some extrema values in the processed images which are removed using a nonlinear tan hyperbolic function given by

$$\mathrm{Im}(x_m, y_m) \leftarrow \tau \tanh(\mathrm{Im}(x_m, y_m)/\tau) \tag{2.5}$$

and limits the extreme values in the range of (-τ,τ).

Chapter 3

LOCAL BINARY PATTERNS AND GABOR FILTERS

3.1 Local Binary Patterns (LBP)

First proposed by Ojalaet. al.[5] local binary pattern (LBP) operator is an image operator capable of transforming an image into an image of integer labels describing small-scale local characteristic preserving appearance of the image . The statistics of these integer labels, especially the histogram, is further used for image analysis.LBP was first introduced for texture analysis as in [12] with features basing on nonparametric recognition and classification of textures at the same time taking rotation invariance into consideration. According to them texture is governed by a pattern and also by the extent of its intensity or strength. LBP operator has two main parameters (P_p,R_p) where P_p is representing the sampling points number and R_p is representing the radius. Figures 3.1 (a),(b),(c) represent LBP (8,1) (8,2) and (16,2) operators respectively where each cell represents a pixel value and radius pertains to number of blocks from the center pixel.If we consider a monochrome image where I(x,y) is the intensity value of center pixel then for every sampled point we can calculate the pixel position given as:

$$xm_p = xm_c + Rm_p \cos(2\pi p/P) \tag{3.1}$$

$$ym_p = ym_c - Rm_p \sin(2\pi p/P) \tag{3.2}$$

where $I\ (xm_{p,}ym_p)$ is the intensity value at the sampling point p $where\ p = 0, ..., P-1$

Based on the above formula the sampling point locations can be calculated for varying radius and sampling points.

LBP of a pixel is obtained by center pixel thresholding of the neighboring pixels based on the radius and sampling points specified according to the LBP operator. Referring to Fig 3.2 a 3×3 window is considered in which all surrounding neighbors of the center pixel greater than or equal to it are allotted a value 1 and less than that are allotted a zero and this binary data is read in a clockwise direction which gives the binary pattern for that pixel.

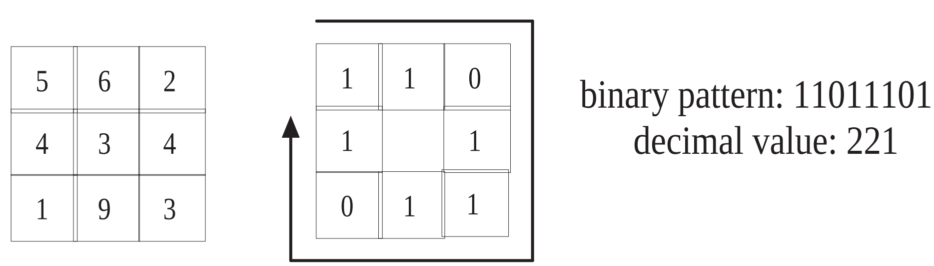
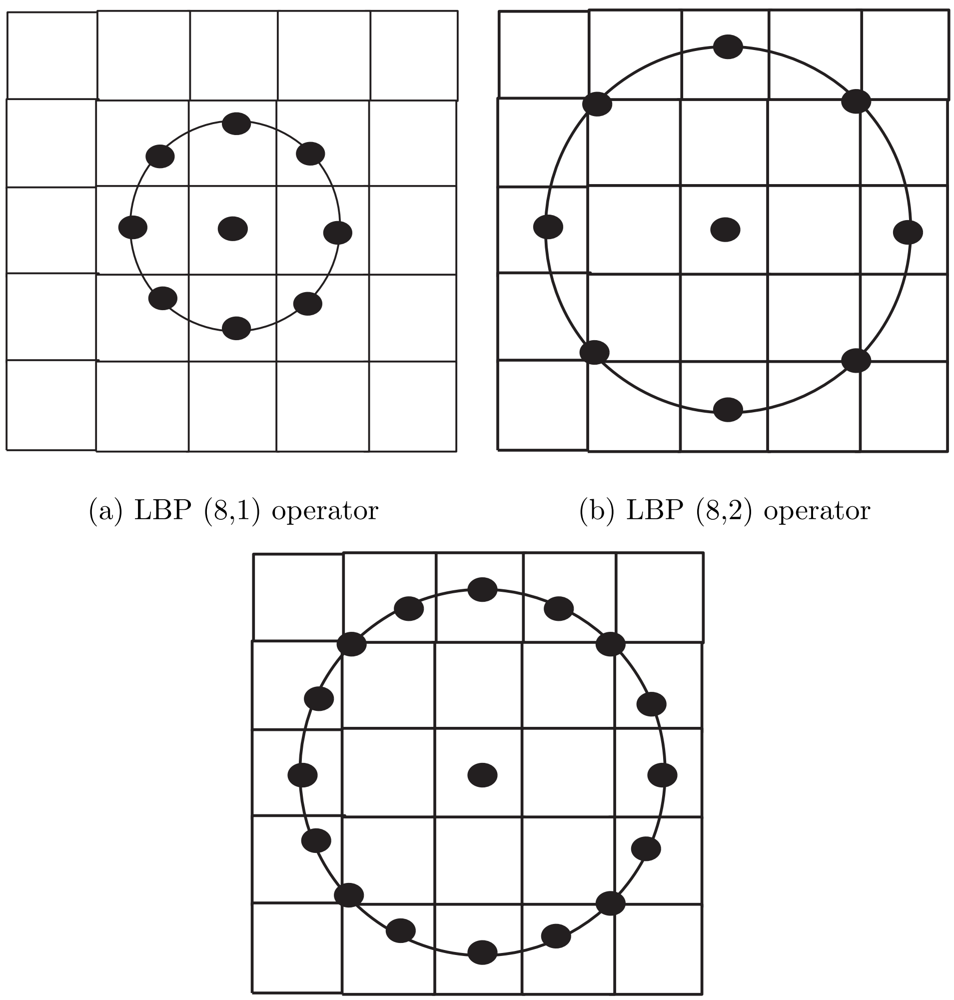


Figure 3.1: Center pixel thresholding in 3×3 window



(a) LBP (8,1) operator

(b) LBP (8,2) operator

(c) LBP (16,2) operator

Figure 3.2: LBP operators

Final decimal value of the LBP operator is given by converting the binary string to decimal with multiplication by powers of 2 where t_p is the thresholded pixel in the binary pattern from LSB to MSB each corresponding to p=0,. . . . . . ,P-1 respectively.

$$LBP_{P,R}(x_c, y_c) = \sum t_p.2^p \tag{3.3}$$

3.1.1 Uniform Local Binary Pattern(ULBP)

Uniform LBP is a variant of LBP where the binary LBP operator follows a uniform pattern. In uniform LBP the bitwise changes from zero to one considering the pattern circularly should be atmost two. For instance consider a 3×3 window where center pixel has 8 neighbours that means the binary values result in values from 0 to 255.Out of these values it was found that 58 patterns follow uniformity. For 8 bit patterns 00000000 (no transition), 00000100 (2 transitions), 11111101 (2 transitions), 11111111 (no transition) are some examples which follow uniform binary pattern.

Basis for selecting Uniform LBP

For facial expression analysis which is mainly based on feature texture of regions it is necessary that the features remain invariant to rotation because expression images are rarely static. Local patterns of natural images have less variations hence considered uniform. They also give robustness to statistical representations and hence mathematically very efficient in terms of comparison and computational complexity.According to Ojala et. al.ULBP has efficiently represented flat regions, edges,spots,corners and line ends.It

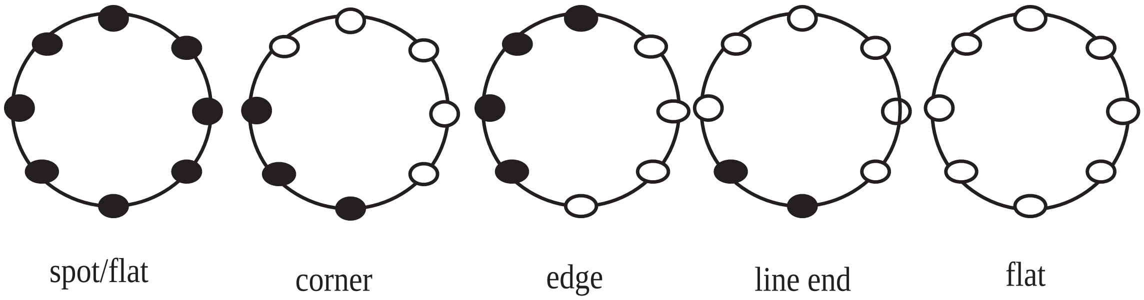


Figure 3.3: Various texture identification by ULBP

is also efficient in representing local facial features and their relation with the neighbours.

Representation of face in terms of facial features is a requisite for any recognition application.Uniform LBP is coherent in representing facial texture adhering to expression recognition. Basic LBP operator 3×3 is incompetent in representing facial features due to its small neighborhood. So we adopt a modified LBP called Uniform LBP which accounts for a major amount nearly 90 percent of the facial textures.

Algorithm 1: IDENTIFYING UNIFORMITY

let initial binary pattern be bin_1
bin_2=shift right bin_1
res= bit_sum($bin_2 \oplus bin_1$)
if res==2
return uniform
else
return nonuniform

For (8,1) LBP operator 58 uniform patterns are identified and these patterns are given labels counting from 1-58 and all other pixels following non uniformity are given a separate label 59.By this we obtain an LBP operated labelled face representation[12].

Statistical representation of LBP image

In order to represent data of LBP image statistically we adopt label count histogram.This LBP histogram representation is important because it replicates the information about the distribution of the local micro-patterns.These patterns include edges, spots and flat areas, over the whole image, which can statistically describe image characteristicsand their statistical representation together .As mentioned in the face representation the image has labels ranging from 1-59 so the histogram has 59 bins in total. LBP labeled histogram of image is given by eq.3.4

$$H_t = \sum_{p,q} I(f_l(p,q) = t), \qquad t = 0, ..., n-1 \tag{3.4}$$

where t signifies the label count produced in the LBP image and

$$I_n(Q) = \begin{Bmatrix} 1; \ Q \ is \ true \\ 0; \ Q \ is \ false \end{Bmatrix} \tag{3.5}$$

3.2 Gabor Wavelet description and Feature extraction

Facial expressions change according to the change in direction of movement in facial muscles. According to research mammalian visual system perception of these movement is very near to that of Gabor Filter responses in different orientations.The Gabor wavelets, are efficient in capturing spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationship properties.Consequently they are a decent close estimation to the channel reaction profiles experienced tentatively in cortical neurons[13].

The Gabor wavelets were discovered to specially suit image decomposition and representation when the objective is the deduction of nearby and separating highlights. Gabor wavelets here are utilized for picture investigation as a result of their natural pertinence and computational properties.The Gabor wavelets (kernels, filters) can be defined as follows[14]:

$$\varphi(xm, ym) = \frac{f_0^2}{\pi\gamma\eta} e^{-(\frac{f_0^2}{\gamma^2} x_p^2 + \frac{f_0^2}{\eta^2} y_p^2)} e^{i2\pi f_0 x_p} \tag{3.6}$$

$$\begin{aligned} x_p &= x\cos\theta + y\sin\theta \\ y_p &= -x\sin\theta + y\cos\theta \end{aligned} \tag{3.7}$$

$$f_{\max} = 0.25 \quad \gamma = \sqrt{2} \quad \eta = \sqrt{2} \tag{3.8}$$

$$f_u = \frac{f_{\max}}{\sqrt{2}^u} \quad \theta = \frac{\upsilon\pi}{8} \quad \alpha = \frac{f_u}{\gamma} \quad \beta = \frac{f_u}{\eta} \tag{3.9}$$

f_{max} is the maximum frequency, and f_u give the frequency selectivity relation between maximum frequency and scales. In most cases one would use Gabor wavelets of five different scales u = 0,..., 4 , and eight orientations v=0,..., 7. γ and ηgive the scaling relationship between the fundamental frequency and oreination selection.

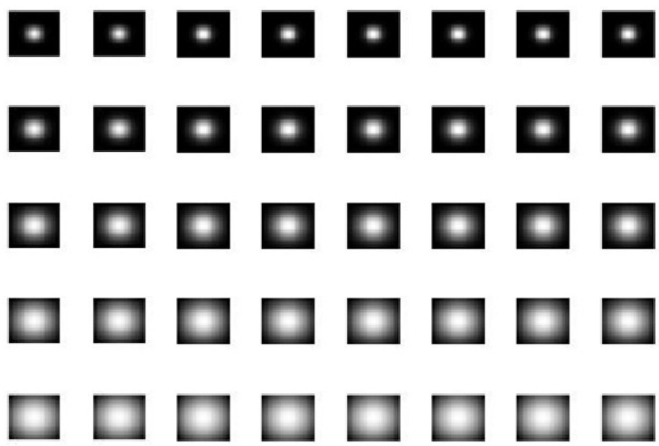


Figure 3.4: Real part of Gabor filter in different scales

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernel filters of different scale and orientation as defined by equation 3.7 . Let I(x, y) be a gray level input image to be converted into Gabor, the convolution output $GF_{u,v}(z)$ of image I and a Gabor kernel $\varphi_{u,v}$ is given as

$$GF_{u,v}(x,y) = I(x,y) \otimes \varphi_{u,v}(x,y) \tag{3.10}$$

where Z =(x,y) denotes the coordinates of the pixel and u,v denote the scale and orientation of the filter.

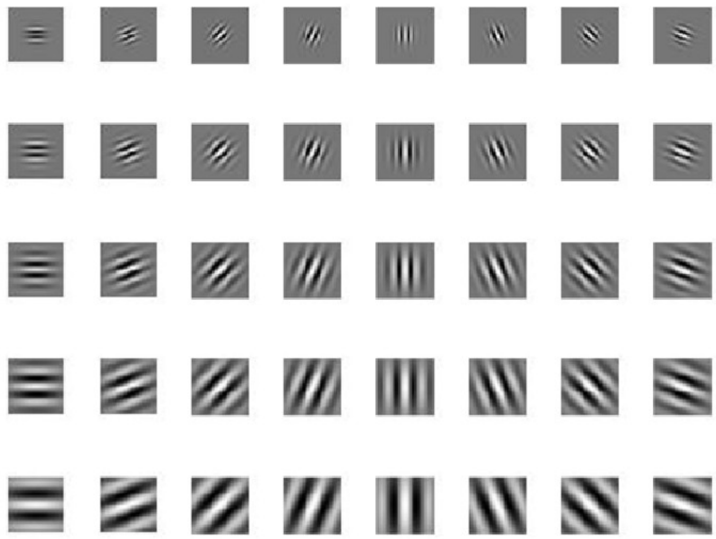


Figure 3.5: Real part of Gabor filter in different orientations

Gabor's multiple scale and multiple orientation disintegration makes the information measurement increase significantly, particularly when the picture size is too extensive[3]. Dimensionality decrease strategy must be embraced to evade the issue of measurement catastrophe. In the flexible chart coordinating works, Gabor wavelet change is utilized to concentrate highlight vectors on the expression fiducial focuses, yet this strategy obliges higher highlight focuses area exactness. The immediate lessening measurement is down sampling of data, and this may lose some critical information. Principal Component Analysis (PCA) can also be used for significant reduction of feature vector.

3.3 Classification methodologies

3.3.1 Distance based classifiers

Mathematical representation of LBP facial features is available in form of histogram which is best compared using distance classifier. Support Vector Machine (SVM) based classification involves complex training and is not implementation worthy. ANN classifier requires repeated experimentation due to its lack of consistency; different distance classifiers or similarity measures which can possibly yield better results for histogram comparison are as follows:

- **Weighted Histogram intersection (HI)** which is maximum distance classifier given by eq 3.9
$$D(S, M) = \sum_{i,j} w_j (S_i, M_i) \tag{3.11}$$
- **Euclidean Distance (ED)** which is minimum distance classifier given by equation 3.10
$$D(S, M) = \sqrt{\sum_{i,j} w_j (S_i - M_i)^2} \tag{3.12}$$
- **Chi Square (CS)** dissimilarity measure classifies with minimum dissimilarity value given by equation 3.11
$$\chi^2_w(S, M) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \tag{3.13}$$
- **Log Statistic (LS)** measure is maximum distance classifier given by equation 3.12
$$D(S, M) = -\sum_{i,j} w_j S_i \log M_i \tag{3.14}$$
- **Kullback Leibler (KL) divergence** It is a non symmetric measure of estimating the dissimilarity between two probability distributions here histograms. It gives the

information lost while mapping one distribution to other, therefore greater the loss higher is the divergence or dissimilarity between the distributions. KL divergence [15] is calculated using equation 3.13 shown below.

$$KL(S, M) = \sum_i S_i \log \frac{S_i}{M_i} \tag{3.15}$$

where are the two distributions under comparison for dissimilarity.

3.3.2 ANN Classifier

Neural Network

A neural network is a parallel, conveyed data handling structure comprising of preparing components (which can have a neighborhood update y and can complete restricted data preparing operations) interconnected together with unidirectional sign channels called associations. Every handling component has a solitary yield association which branches (”fans out”) into the same number of insurance associations as fancied (every conveying the same sign - the preparing component yield signal). The preparing component yield sign can be of any scientific sort sought. The greater part of the preparing that goes ahead inside of every handling component must be totally nearby: i.e., it must depend just upon the present estimations of the info signs touching base at the handling component through impinging associations and upon qualities put away in the handling component's neighborhood memory[16].

Artificial Neural Network (ANN) classifier[16]

- Artificial neural network (ANN) is a machine learning approach modeling human brain and consisting of a number of artificial neurons.
- Neurons in ANN's are usually having lesser number connections compared to biological neurons.

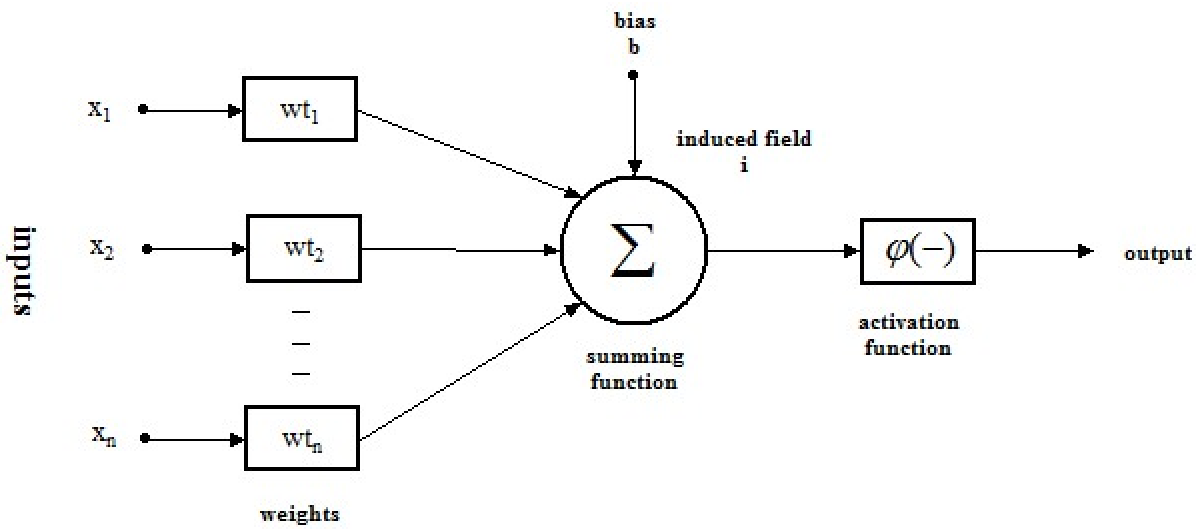


Figure 3.6: single neuron

- Every neuron in ANN is given a number of inputs.
- An activation function here sigmoid function is applied to these inputs resulting in the activation level of neuron i.e(output value of the neuron)
- Training examples in form of feature vector here are fed to the system in order to inculcate knowledge about the learning task.

Basic information processing unit of a NN the is neuron. It consists of:

- A feature vector bins describing the neuron inputs, with weights wt_1, wt_2,.....,wt_n
- A summing function (linear combiner) for computation of the weighted sum of the inputs:

$$u_{sum} = \sum_{j=1}^{n} wt_j x_j \tag{3.16}$$

- Activation function $y_{out} = \varphi(u_{sum} + b)$ which limits the amplitude of the neuron output for given set of inputs and 'b' denotes bias.

In back propagation neural network the activation function usually followed is the sigmoid function which is given by:

$$\phi(i) = z + \frac{1}{1 + \exp(-xi + y_{out})} \tag{3.17}$$

3.3.3 Japanese Female Facial Expression (JAFFE) Database

The database contains 213 images of 7 facial expressions (6 basic facial expressions and 1 neutral) which was posed by 10 Japanese female models. Each image was rated on 6 emotion variants by 60 Japanese people for subjective verification. The database was planned and assembled by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba[17].
Highlights of JAFFE database:

- Total of 213 TIFF images of 10 subjects
- Subjects display 7 basic emotions
- Each emotion class has 30 images
- Publicly available database for research use

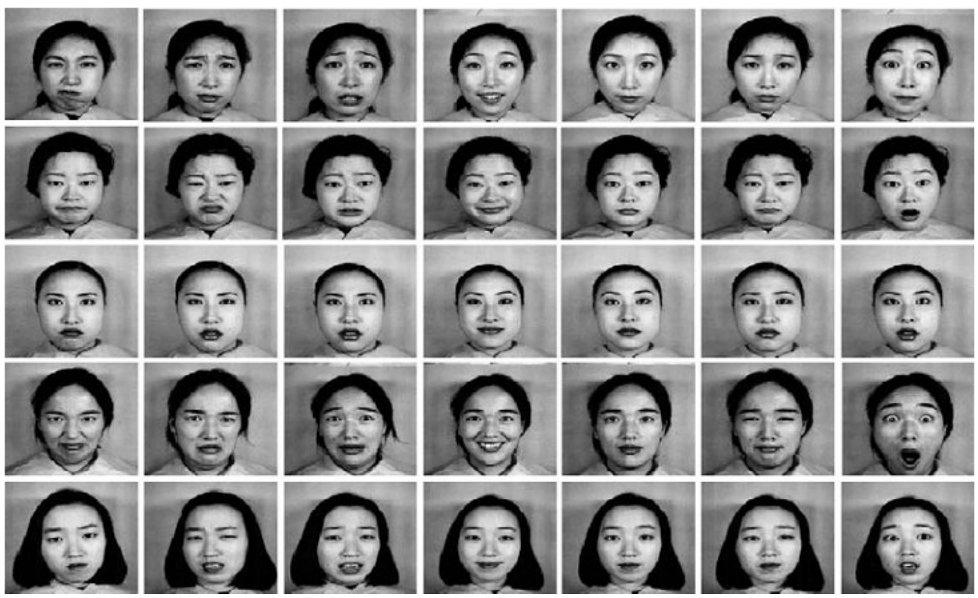


Figure 3.7: JAFFE Database

Chapter 4

Facial Expression Recognition Algorithms

4.1 Algorithm 1
Feature extraction using LBP and classification using Kullback Leibler (KL) Divergence

Face representation including spatial information

In section 3.1.2 we have dealt with labelling of an image using LBP and its statistical representation using histogram.Generally LBP for Texture description obtains holistic information over an area.Initial Histogram representation of LBP face image also returns holistic information.For normal textures this method is commendable because they remain invariant to rotation and translation.But coming to the case of face description averaged description of face results in loss of efficient information.In face description it is recommended to retain spatial information along with facial feature information.

Hence we adopt the following method. The detected and preprocessed face image

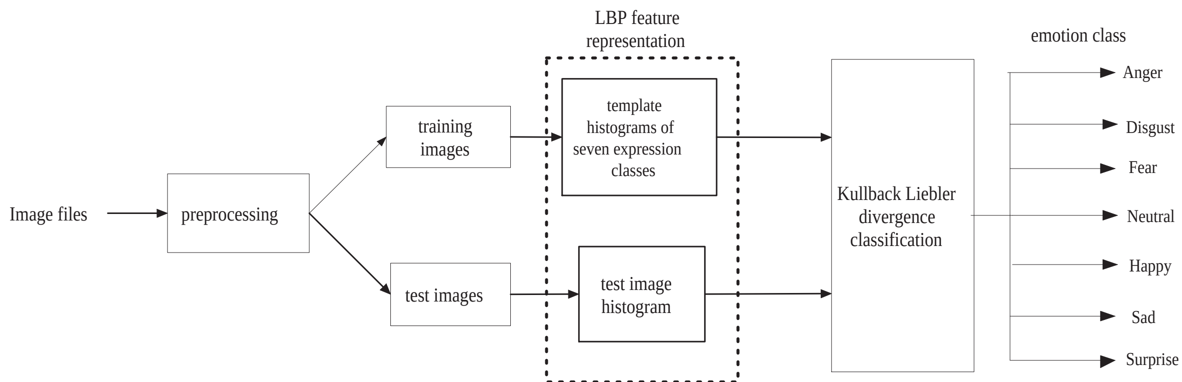


Figure 4.1: Block diagram for Algorithm 1

is divided into subregions and LBP feature descriptors are extracted for each region separately.The idea of generalized LBP histogram is slightly modified giving rise to a spatially enhanced histogram [lb rf 1] in which spatial information is combined with appearance data. The LBP facial image is divided into blocks or subregions named as $R_1, R_2, \ldots\ldots, R_n$ where each subregion is of specified size let it be p×q.

The histogram of each subimage is calculated and are spatially combined including the spatial information giving rise to a global histogram. If an LBP (R,P) operator is being used then we have 2^P values out of which m values might be following uniformity. Then we have a local histogram of one subregion with (m+1) bins. Such single histograms of single regions are combined to form a global histogram of size n×(m+1).

In the face image regions like cheeks , higher parts of forehead do not carry most defining information. Hence these regions are of lesser importance whereas facial parts near corner of mouth and corner of eyes account for most of the expressional feature changes. So primarily these are to be given higher importance which is executed by allotting weights in order of importance to the subregions as shown in fig. The colour code for the allotted weights is indicated by the data beside.

The histogram of this subdivided image including spatial information is given by equa-

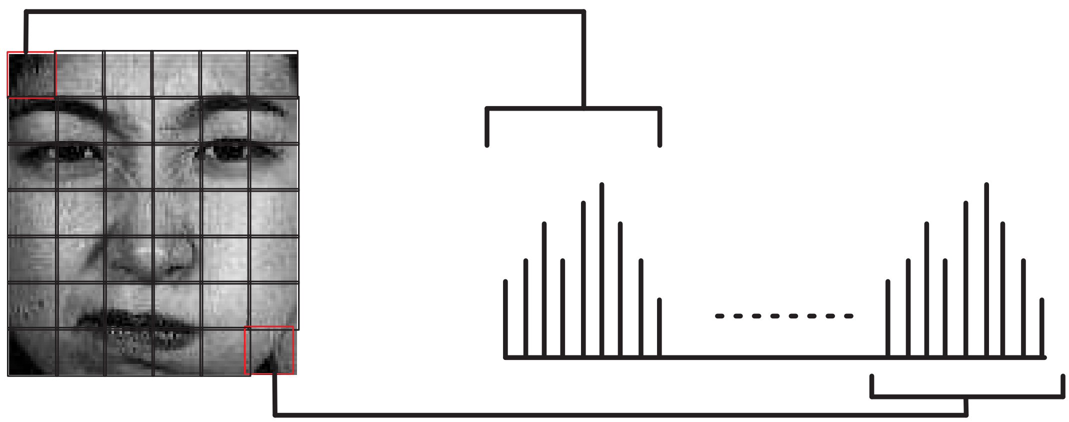


Figure 4.2: spatially enhanced histogram

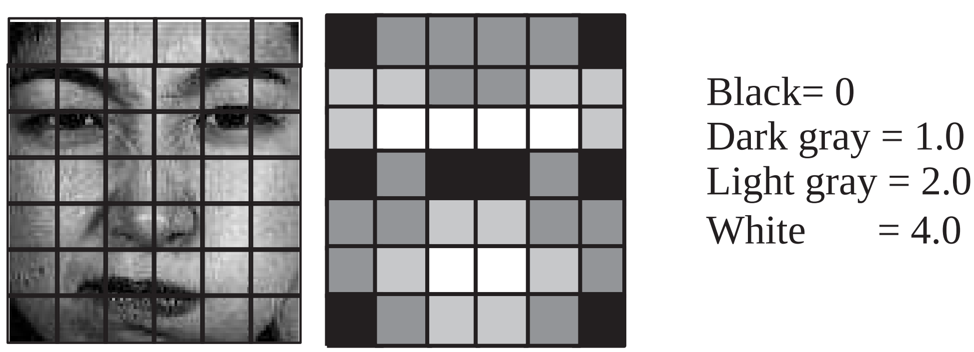


Figure 4.3: weight allotment according to ROI

tion 3.4 and the generation of LBP feature histogram from the combination of sub regions is given by eq 4.1

$$H_{p,q} = \sum_{x',y'} w_q.\, Im(f_l(x', y') = p) Im((x', y') \in R_q) \quad (4.1)$$

where p= 0,. . . . . ,m-1 (number of bins) and q= 0,. . . . .,n-1(subregions), w_q being weight for region R_q.

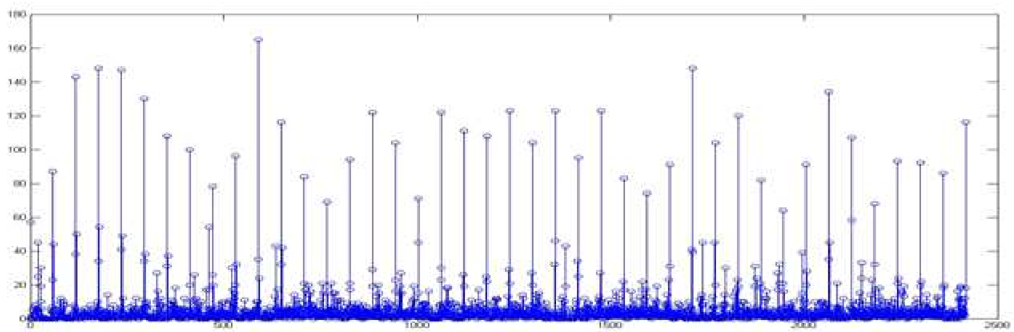


Figure 4.4: Generated LBP spatially enhanced histogram

4.1.1 *Implementation Details*

1. Uniform LBP (8,1) operator is considered
2. Training and testing images are divided into 7×6 subregions
3. Resulting bins in spatially enhanced histogram is 7×6×59=2478 bins as shown in fig.4.4
4. For expression classification the database is segregated into training and testing image sets ,test image set not inclusive in training
5. The seven expression images of training are grouped manually and training algorithm (refer fig 4.4) is applied.

6. Spatially enhanced LBP histograms of each expression are averaged to form a template histogram for one expression.

7. Test image LBP histogram is also obtained which is compared with all the seven template histograms of anger, disgust,fear,happy,neutral,sad and surprise using weighted KL divergence measure given by eq 4.2

$$KL(templ(r), test(r)) = \sum_{r} templ(r) \log \frac{templ(r)}{test(r)} \tag{4.2}$$

where templ(r) is the template histogram and test (r) is the test image histogram., r is the number of bins.

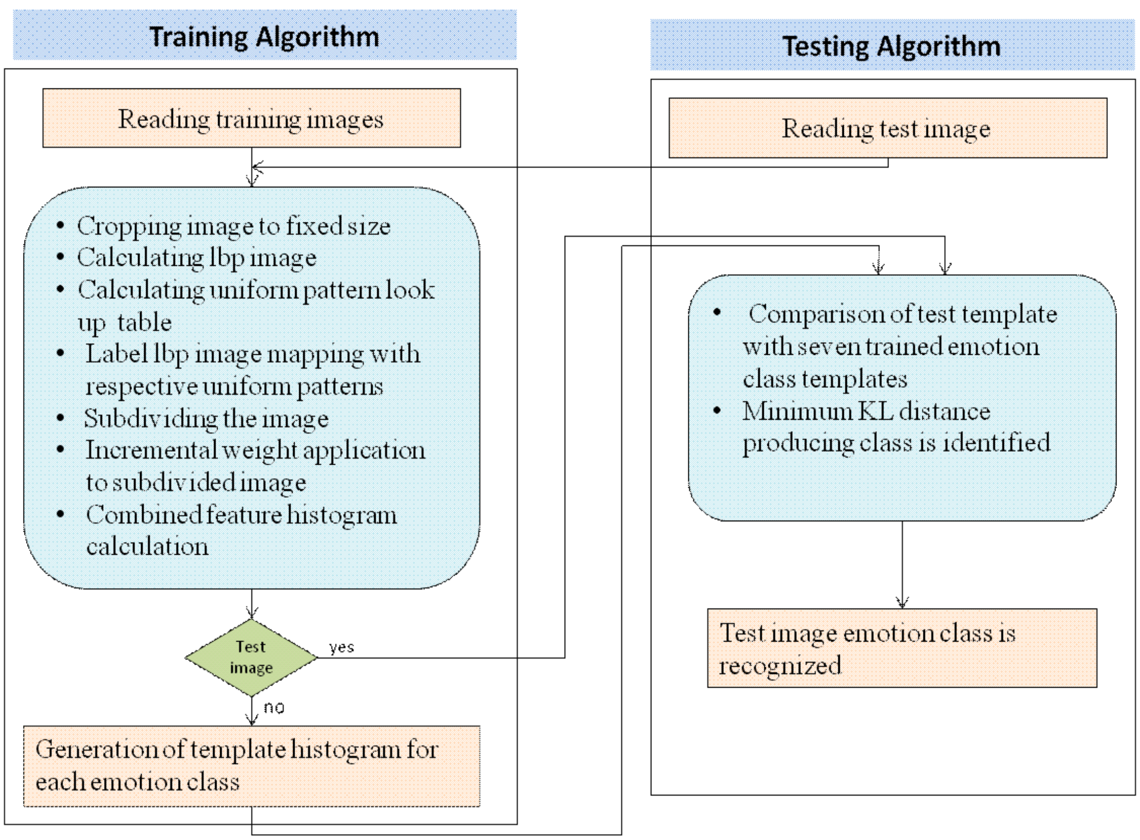


Figure 4.5: Flow diagram of algorithm 1

Japanese Female Facial Expression (JAFFE) database is used for experimentation .In The JAFFE database consists of images of 10 persons. Each person has 3 images each of seven types of facial expressions: angry, disgust, fear, happy, neutral, sadness, and surprise. There are about three samples corresponding to each facial expression of each person. All the detected face images are preprocessed and automatically cropped to a fixed size of 122×112 pixels. The training is performed in three groups by taking 8 training and 2 test subjects in first group,six test and four training subjects in second group and 5 each for training and testing in group 3. The results averaged over all 3 distinct groups to obtain overall classification accuracy.

4.1.2 *Experimental Results*

The training was performed in three rounds, choosing varied test subjects each time. Individual emotion class accuracy and overall classification accuracy with and without preprocessing are compared and is shown in following Table 4.1. The abbreviations AN,

Table 4.1: CLASSIFICATION ACCURACY OF PROPOSED METHOD WITH AND WITHOUT PREPROCESSING

Classification method	Emotion class						
	AN	DG	FR	HA	NT	SD	SU
without preproc.	66.7	66.7	66.7	100	100	33.3	100
With preproc.	100	100	100	100	100	66.7	100

DG, FR, HA, NT, SD, SU represent anger, disgust, fear, neutral, sad and surprise respectively.

KL divergence used for classification in the proposed method is compared with other distance and dissimilarity measures in terms of overall classification accuracy. KL divergence yielded a highest accuracy of 95.24 % which is considered to be greater than other distance based classification methods mentioned in Section 3.3.1 . KL distance measures for weighted person independent classification is shown in Table II. Images from test set are randomly chosen and were observed that image belonging to a particular emotion class gave minimum KL distance against that particular class.

Table 4.2: CLASSIFICATION ACCURACY OF PROPOSED METHOD WITH AND WITHOUT PREPROCESSING

I/P	KL divergence values						
Emotion class	AN	DG	FR	HA	NT	SA	SU
AN	**1.2173**	1.3424	1.4496	1.4423	1.4912	1.3746	1.6481
DG	1.2565	**1.2438**	1.2947	1.3423	1.4488	1.2788	1.5139
FR	1.4094	1.3125	**1.2358**	1.387	1.3424	1.2366	1.3638
HA	1.8282	1.74	1.8477	**1.2755**	1.9625	1.7982	2.0466
NT	1.4629	1.5394	1.4506	1.4836	**1.3105**	1.3805	1.4499
SD	1.6229	1.4161	1.3870	1.5304	1.5673	**1.3272**	1.6165
SU	1.6341	1.7622	1.5730	1.8652	1.5442	1.5688	**1.2594**

Table 4.3: COMPARISON OF VARIOUS DISTANCE CLASSIFIERS

Emotion class	Distance classifier				
	HI	ED	CS	LS	KL
Anger	66.7	66.7	66.7	100	100
Disgust	66.7	66.7	33.3	100	100
Fear	33.3	66.7	33.3	66.7	100
Happy	100	100	100	100	100
Neutral	66.7	66.7	66.7	66.7	100
Sad	33.3	33.3	33.3	66.7	66.7
Surprise	66.7	100	100	100	100
Overall accuracy	62	71.44	62	85.72	95.24

Further KL divergence is comparison with other distance classifiers is shown in Table 4.3. From this we can infer that our proposed method which is a proper combination of preprocessing technique, LBP feature representation with KL divergence comparison is accurate and efficient for expression classification.

4.2 Algorithm 2

Feature extraction using Gabor followed by LBP and classification using ANN classifier

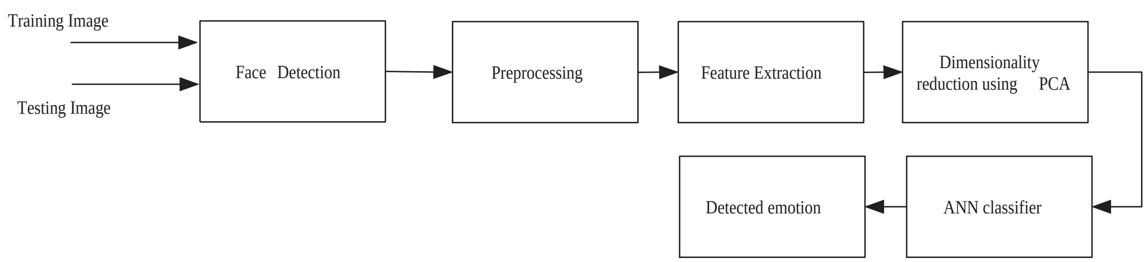


Figure 4.6: Block diagram of algorithm 2

4.2.1 *Implementation details of algorithm 2*

- Gabor filtered images are obtained in 8 orientations and 5 scales resulting in 40 images.To decrease the feature vector size which is currently ($122 \times 112 \times 40$) we averaged the gabor filtered images in different orientations into a single image.The Gabor filtered images are obtained by convolution of training image and 8×5 gabor filters using equ ( ) of section

- Averaging of orientations for 5 scales is given by the eq ()

$$GF_v = \sum_{u=0}^{7} GF_{v,u} \quad where\ v = 0, 1, 2, 3, 4 \tag{4.3}$$

- LBP operator is applied to the gabor filtered orientation averaged images resulting in fused LGBP(Local Gabor Binary Pattern) features extracted.

- Both Gabor and LBP are efficient in extracting local features adopting different orientations and scales. Therefore using them together is considered to give robust feature extraction to determine multiclass classification. The feature vector length obtained here 68320 ($122 \times 112 \times 5$)which increases the computational time .So we apply Principal component Analysis for obtaining the most distinct features for comparison.

The important steps involved in PCA are described below:

1. Feature vector for each image for one class is obtained and represented as each column in matrix X.

2. Features are normalized by subtracting mean and dividing by standard deviation

3. Covariance matrix for the feature matrix is obtained by $C = X^T X$ and singular value decomposition is applied on the covariance matrix to obtain Eigen values and corresponding Eigen vectors.

4. Variance of 65%-90% is selected and corresponding k Eigen values λ from m are selected

5. Depending on the selected Eigen values reduced Eigen vector matrix is obtained.

6. Eigen vector matrix U (1: k) is multiplied with feature matrix to obtain reduced data.

- Unlike the former algorithm which classifies using KL divergence, here the extracted features are classified using both Artificial Neural Network classifier and KL divergence and their results are compared.

4.2.2 Experimental Results

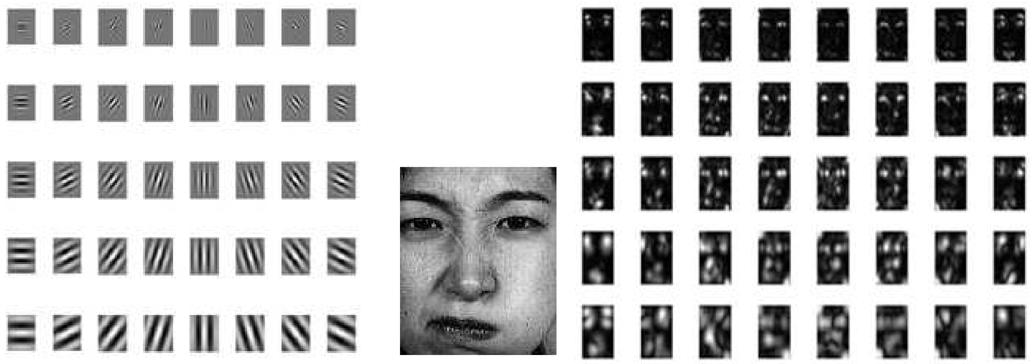


Figure 4.7: Result of Gabor filter on the example face image

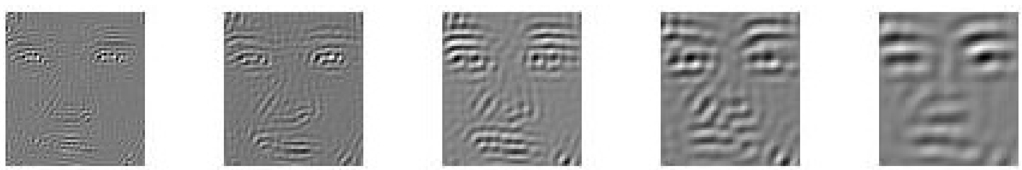


Figure 4.8: Gabor filter images averaged in 8 orientations

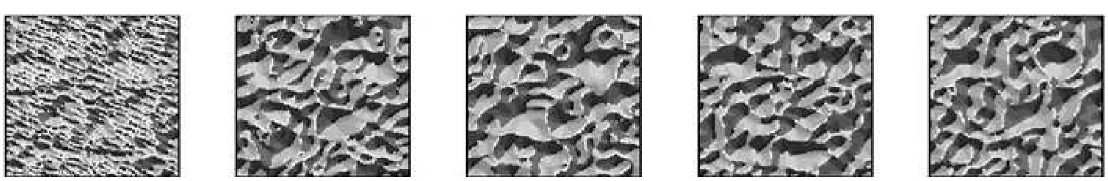


Figure 4.9: Fused Gabor and LBP averaged images

Table 4.4: ANN classifier accuracy measure 1

Output class	Target class									
	anger	disgust	sad	surprise	neutral	happy	fear	classification accuracy		
anger	5	0	3	1	0	0	0	55.6		
disgust	4	23	0	0	0	2	3	71.9	training samples: (60%)	121
sad	5	2	19	1	0	4	2	57.6	testing samples: (35%)	71
surprise	2	1	1	21	0	0	1	80.8	validation samples: (5%)	10
neutral	2	0	0	2	27	0	1	84.4	number of neurons	10
sad	9	2	5	2	0	21	2	51.2		
surprise	3	1	3	3	0	0	19	65.5		
Overall classification accuracy	66.8									

Feature vector is derived from this fused LGBP similar to the LBP weighted feature histogram. For the set of training images LGBP histograms are obtained and these are fed to the ANN classifier .The training and testing samples are varied in terms of number of sample data.The algorithm is also verified for different number of neurons in the classifier network. The tables shown below validate the classification result.

It was observed from the simulation that by increasing the number of neurons above 20 the was no change in the classification accuracy.That means the back propagation network stabilizes after passing through an average of 20 neurons i.e error minimizes and validation data error increases which prevents the training data from overfitting the network.

Table 4.5: ANN CLASSIFIER ACCURACY MEASURE 2

output class	target class									
	anger	disgust	sad	surprise	neutral	happy	fear	classification accuracy		
anger	21	3	0	1	0	0	2	77.8		
disgust	1	20	4	3	0	0	3	64.5	Training samples: (60%)	121
sad	2	4	22	1	0	5	3	59.5	Testing samples: (35%)	71
surprise	3	0	1	22	0	1	4	71	Validation samples: (5%)	10
neutral	0	0	0	1	27	0	0	96.4	Number of neurons:	20
sad	1	1	4	2	0	21	1	70		
surprise	2	1	0	0	0	0	15	83.3		
Overall classification accuracy	73.3									

Table 4.6: ANN CLASSIFIER ACCURACY MEASURE 3

output class	target class									
	anger	disgust	sad	surprise	neutral	happy	fear	classification accuracy		
anger	24	3	0	1	0	0	2	77.8		
disgust	5	20	4	3	0	0	3	64.5	Training samples: (70%)	141
sad	0	4	22	1	0	5	3	59.5	Testing samples: (25%)	51
surprise	0	0	1	22	0	1	4	71	Validation samples: (5%)	10
neutral	0	0	0	1	27	0	0	96.4	Number of neurons:	10
sad	1	1	4	2	0	21	1	70		
surprise	0	1	0	0	0	0	15	83.3		
Overall classification accuracy	80.7									

Table 4.7: ANN classifier accuracy measure 4

output class	target class									
	anger	disgust	sad	surprise	neutral	happy	fear	classification accuracy		
anger	27	2	1	1	0	0	0	87.1		
disgust	0	25	0	1	0	0	0	96.2	Training samples: (70%)	141
sad	0	0	29	1	0	0	2	90.6	Testing samples: (25%)	51
surprise	2	0	0	25	0	3	0	83.3	Validation samples: (5%)	10
neutral	1	0	0	1	27	0	0	93.1	Number of neurons:	10
sad	0	1	0	0	0	24	0	96		
surprise	0	1	1	1	0	0	26	89.7		
Overall classification accuracy	90.6									

Table 4.8: ANN classifier accuracy measure 5

output class	target class									
	anger	disgust	sad	surprise	neutral	happy	fear	classification accuracy		
anger	30	0	0	0	0	1	0	96.8		
disgust	0	28	3	1	0	2	0	82.4	Training samples: (85%)	172
sad	0	1	26	0	0	1	0	92.9	Testing samples: (10%)	20
surprise	0	0	0	27	0	0	0	100	Validation samples: (20%)	10
neutral	0	0	0	1	27	0	0	96.4	Number of neurons:	15
sad	0	0	0	0	0	23	0	100		
surprise	0	0	2	1	0	0	28	90.3		
Overall classification accuracy	93.6									

Table 4.9: ANN classifier accuracy measure 6

output class	target class									
	anger	disgust	sad	surprise	neutral	happy	fear	classification accuracy		
anger	26	2	0	1	0	0	0	89.7		
disgust	1	25	0	0	0	0	0	96.2	Training samples: (85%)	172
sad	0	1	30	0	0	1	0	96.8	Testing samples: (10%)	20
surprise	1	1	0	28	0	0	1	90.3	Validation samples: (20%)	10
neutral	0	0	0	1	27	0	0	96.4	Number of neurons:	20
sad	2	0	0	0	0	26	1	89.7		
surprise	0	0	1	1	0	1	26	92.9		
Overall classification accuracy	93.1									

Table 4.10: KL divergence measure on Fused LBP and Gabor

I/P Emotion class	KL divergence values						
	AN	DG	FR	HA	NT	SD	SU
AN	1.3546	1.4339	1.4989	1.5123	1.5222	1.4337	1.6865
DG	1.4078	1.3764	1.4123	1.4456	1.4985	1.3956	1.5755
FR	1.4527	1.4012	1.3658	1.5524	1.3957	1.3715	1.5243
HA	1.956	1.7478	1.6678	1.3668	1.9875	1.8976	1.9985
NT	1.5154	1.5123	1.574	1.5436	1.3987	1.4958	1.5879
SD	1.6874	1.6321	1.3985	1.5874	1.5978	1.3856	1.7457
SU	1.7548	1.8526	1.7730	1.9524	1.7695	1.7142	1.4394

4.3 Expression recognition from sequence of images

In previous sections we have dealt with images which have a static expression. In real time expressions are always accompanied by head movement and gestures. when the consecutive frames in a sequence of images vary by a significant amount in terms of features, rotation, translation, etc there is a necessity to find the correspondence among these features . moreover in the Facial Expression recognition algorithms implemented there is a incremental weighted measure allotted to the facial blocks of interest which varies with head pose change or rotation.To eliminate this difficulty there is a necessity to align facial images prior to recognizing expression. SIFT Flow alignment[18] introduced by Liu.et.al for dense scene correspondence and matching is adopted here for facial image correspondence.

4.3.1 Scale Invariant Feature Transform(SIFT)

This methodology changes a picture into a vast gathering of nearby component vectors, each of which is invariant to picture interpretation, scaling, and affine, and in part invariant to brightening changes and relative or 3D projection[19]. SIFT transforms data in an image to freatures which are invariant to scale.These features are named as keypoints.

Steps involved in Keypoint identification using SIFT:

- Representation of image in different scales

  Different scales are nothing but different levels of smoothing. Here Gaussian function is applied for convolution with the image with different values of σ. Gaussian function given by

  $$G(x_p, y_p, \sigma) = \frac{1}{2\pi\sigma^2} e^{-{-(x_p^2+y_p^2)} \big/ {2\sigma^2}} \tag{4.4}$$

  By using Laplacian of Gaussian i.e $\nabla^2 G$ we can easily identify the keypoints which are edges and corners in an image.But this involves high computation.So we follow

Difference of Gaussian(DoG) which produces approximately the same result and also it is scale invariant i.e it does not depend on blurring(σ).

According to Lowe SIFT scale space in one octave requires 5 scaled images which results in 4 DoG images.Each scale is defined by σ, σ^2 ,. . . .. multiplied by factor k in each scale.This also makes DoG scale inavariant because by selecting σ=k we can get rid of the smoothing factor σ^2.

Scale space is created by convolving image with Gaussian function

$$I_c(x_p, y_p, \sigma) = G(x_p, y_p, \sigma) * \mathrm{Im}(x_p, y_p) \tag{4.5}$$

Difference of Gaussian function is used to select scale space extrema points

$$\begin{aligned} DoG(x_p, y_p, \sigma) &= (G(x_p, y_p, k\sigma) - G(x_p, y_p, \sigma)) * \mathrm{Im}(x_p, y_p) \\ &= I_c(x_p, y_p, k\sigma) - I_c(x_p, y_p, \sigma) \end{aligned} \tag{4.6}$$

The image for second octave is obtained by subsampling i.e removing alternate row and column Scale space construction is shown in figure.

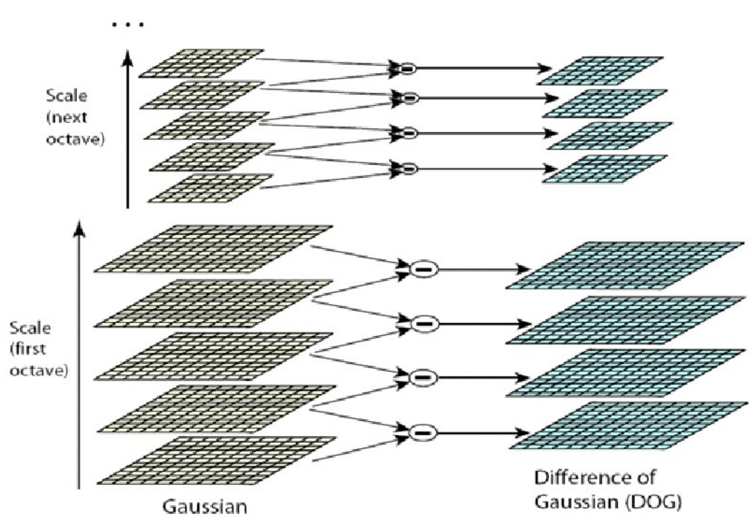


Figure 4.10: scale space construction

- Identification of keypoints

  In a Dog image in each octave a pixel is compared with its eight neighbors in same scale ,a scale above and a scale below and if it a maxima or minima it can be termed as a keypoint as shown in figure.

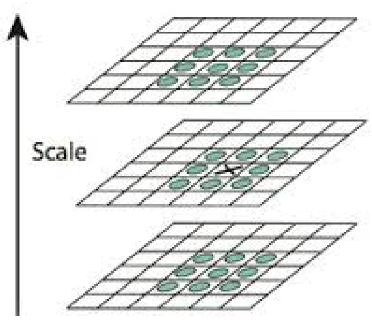


Figure 4.11: maxima/minima keypoint selection

- Elimination of unnecessary keypoints

  Some keypoints or descriptors may have low contrast or localization at edges which are eliminated using a threshold.

- Representing keypoint as SIFT descriptor or feature

  Each keypoint is given a magnitude and direction using the equ

$$mag(x_p, y_p) = \sqrt{(I_c(x_p+1, y_p) - I_c(x_p-1, y_p))^2 + (I_c(x_p, y_p+1) - I_c(x_p, y_p-1))^2}$$
$$\theta_g(x_p, y_p) = \tan^{-1}((I_c(x_p+1, y_p) - I_c(x_p-1, y_p))/(I_c(x_p, y_p+1) - I_c(x_p, y_p-1))) \tag{4.7}$$

  For every key point we choose a 16x16 neighborhood which is divided into 4x4 blocks

  Each 4x4 block has an 8 bin quantized orientation giving to a 4x4x8 SIFT descriptor

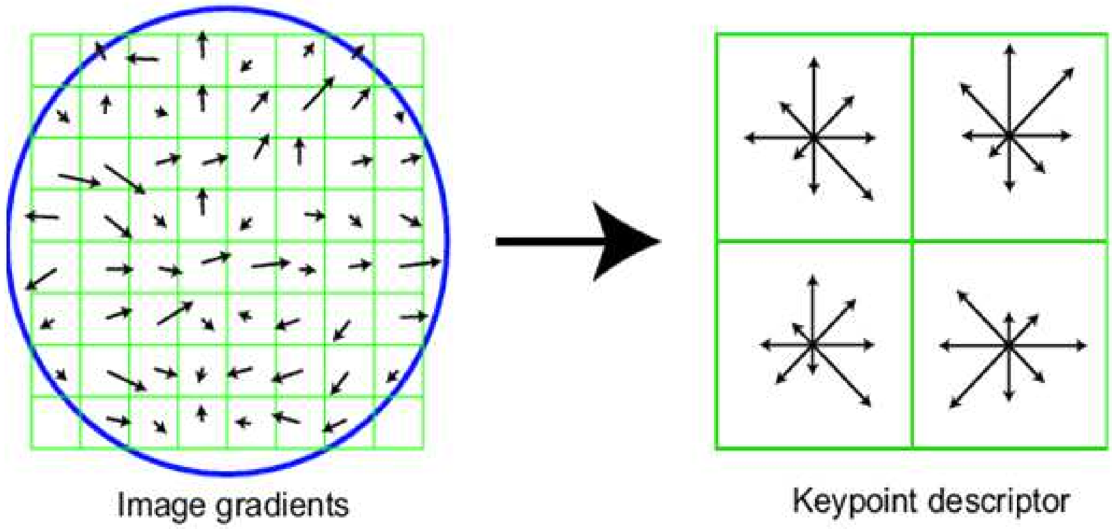


Figure 4.12: maxima/minima keypoint selection

4.3.2 Sift flow alignment

For two images with varying features we can align them using the energy flow function of their highly matching SIFT descriptors and minimizing the flow fuction. SIFT flow energy function is given by the equation

$$\begin{aligned} Ef(w) = \sum_z \min(\|si_1(z) - si_2(z + w(z))\|) + \sum_z \eta(|uf(z)| + |vf(z)|) \\ + \sum_{(z1,z2)\in\varepsilon} \min(\alpha\,|uf(z1) - uf(z2)|) + \min(\alpha\,|vf(z1) - vf(z2)|) \end{aligned} \tag{4.8}$$

si_1 , si_2 : Two sift images to be aligned

z(x,y) : grid coordinate of image

w(z) = (uf(z),vf(z)) SIFT flow vector at p(x,y)

ε : 4 point neighborhood around the key point

To optimize this function dual layer loopy belief propagation is applied depending on the 4 neighbors in images one and two.

Algorithm 2: ALIGNED EXPRESSION IMAGES

Input: M (1,. . . . . . .,m) : total count of image sequences;

N_m : total frame count in sequence 'm';

q : number of iteration levels (here we assume q=3) ;

$Initialise \quad A_0^{ref} = \frac{1}{\sum_{m=1}^{M} N_m} \sum_{m=1}^{M} \sum_{n=1}^{N_m} I^{(m,n)}$

for $q \leftarrow 1$ ***to*** Q **do**

  for $m \leftarrow 1$ ***to*** M **do**

    for $n \leftarrow 1$ ***to*** N_m **do**

      $I_{align}^{(m,n)} = SIFT\,flow(I^{(m,n)}, A_{i-1}^{ref})$

    $AI_i^m = \frac{1}{N_m} \sum_{n=1}^{N_m} I_{align}^{(m,n)}$

  $A_i^{ref} = \frac{1}{\sum_{m=1}^{M} N_m} \sum_{m=1}^{M} EAI_i^m$

$return \ AI_Q^m$

4.3.3 Experimental results

Alignment of rotated head using SIFT flow

Refering to fig we can see that image1 is the cue image and image2 is the one to be warped.we can also find that image2 is having a head pose variation or rotation.Based on flow visualisation and their colou code in RGB as depicted by fig. we obtain the SIFT image.These SIFT images describe the most matching descriptors according to which features of image2 are matched to cue image by energy minimisation function of SIFT flow.

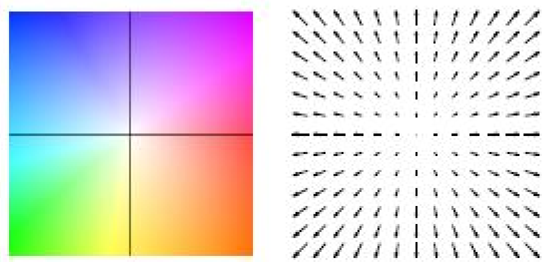


Figure 4.13: flow field visualisation

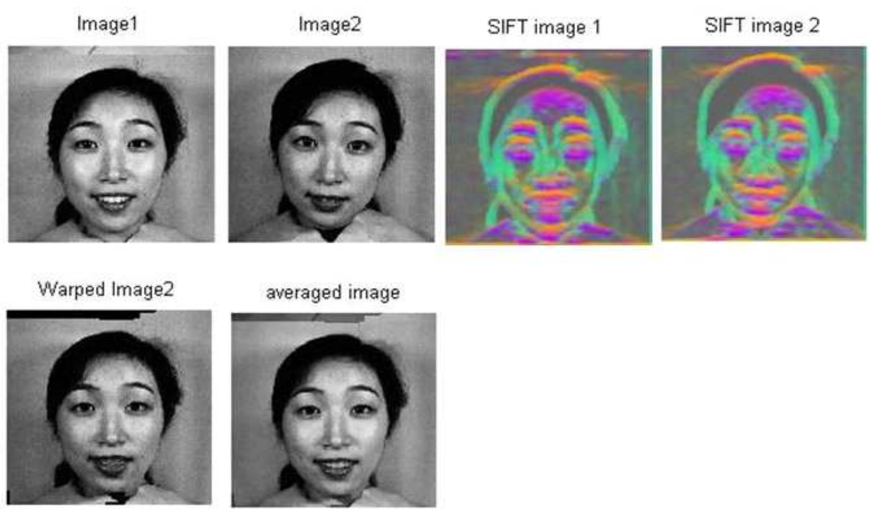


Figure 4.14: warping image2 using image1 as cue

Alignment results for various subjects for six expressions and 3 levels

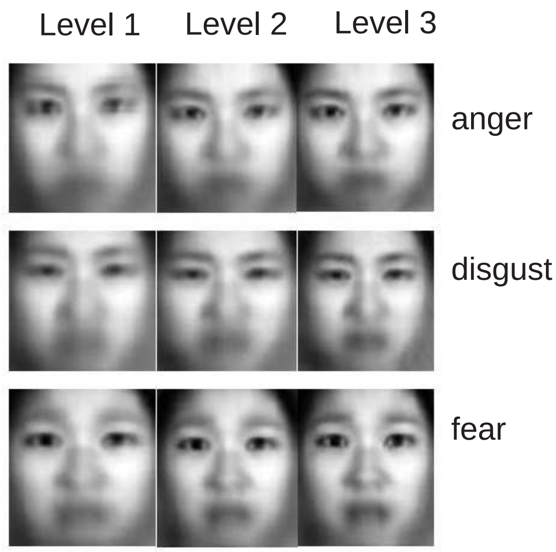


Figure 4.15: aligned images of various subjects of similar expression I

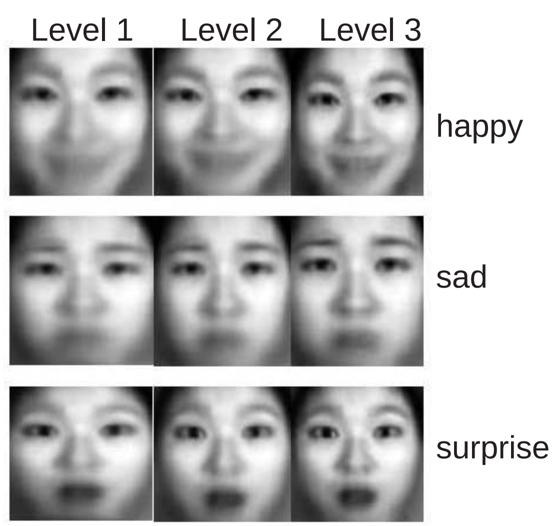


Figure 4.16: aligned images of various subjects of similar expression II

For every expression one subject image is taken as the cue and the remaining images are aligned basing on the cue image using sift flow alignment algorithm.Then the aligned images are followed through algorithm 1 for feature extraction and classification.

Table 4.11: CLASSIFICATION ACCURACY FOR SEQUENCE OF IMAGES USING SIFT FLOW AND ALGORITHM 1

classified emotion class							
Ground Truth	Anger(%)	Disgust(%)	Fear(%)	Happy(%)	Neutral(%)	Sad(%)	Surprise(%)
anger	100	0	0	0	0	0	0
disgust	3.3	85	2.5	0	0	2.5	0
fear	0	0	61.7	0	8.3	30	0
happy	0	0	0	100	0	0	0
neutral	0	0	1.5	0	95	3.5	0
sad	2.5	0	2.5	0	20	75	0
surprise	7.5	0	0	5	0	0	87.5
Overall accuracy	86.31						

Table 4.12: CLASSIFICATION ACCURACY FOR SEQUENCE OF IMAGES USING SIFT FLOW AND ALGORITHM 2

classified emotion class							
Ground Truth	Anger(%)	Disgust(%)	Fear(%)	Happy(%)	Neutral(%)	Sad(%)	Surprise(%)
anger	95	2.5	2.5	0	0	0	0
disgust	2.5	87.5	7.5	0	0	2.5	0
fear	2.5	2.5	65.7	0	4.3	25	0
happy	0	0	0	100	0	0	0
neutral	0	0	5	0	90.5	4.5	0
sad	2.5	0	5	0	20	73.5	0
surprise	5	0	0	5	0	0	90
Overall accuracy	86.028						

Chapter 5

Conclusion

5.1 Conclusion

This thesis presents two facial expression recognition algorithms which are efficient in recognizing and differentiating seven expression classes.An efficient facial expression recognition method is proposed in algorithm 1 which uses robust Local Binary Patterns for facial feature extraction and representation along with Kullback Leibler divergence for classification. Preprocessing improves the classification accuracy of KL divergence measures. Experimental results show that preprocessing combined with KL divergence have less confusion for the correct emotion class recognition.

They also show that the proposed method gives a classification accuracy of 95.24 which is better than other distance based classification methods whose accuracy is ranging between 62-85.72 depicting classification improvement of 9.99 to maximum of 34.9 over existing HI and LS classifiers. Algorithm 2 includes fusion of Gabor with LBP feature extraction and resulting a classification with ANN classifier which is further compared with KL Divergence also. These two methods proves their efficiency in terms of classification accuracy compared to other methods such as [25] facial movement analysis of features by Zhang .et. al which gives 92.2 accuracy.

Although proposed method proves in terms of accuracy, there still lies confusion with classifying sad and fear classes because single emotion is a combination of many intensions whose recognition becomes simple accompanied with a combination of speech and gesture recognition. The proposed method can be extended to dynamic facial expression recognition from video sequences and implemented for real time situations as both preprocessing and LBP are robust to illumination variations. Our proposed method combining LBP with KL divergence with high classification accuracy delineates its contribution in the field of facial expression classification.

Unlike other methods this thesis also throws light on dealing with expression recognition from a sequence of images by aligning images removing pose variation and averaging them to obtain resultant expression image which can finally be classified using either of the algorithms.It can be concluded that for static images algorithm 1 gives a better classification accuracy than algorithm 2.In case of image sequences both of them give an almost equal classification accuracy algorithm 1 slightly better but in classifying fear and sad classes confusion is better eliminated by algorithm 2.

5.1.1 Future scope

- Emotions are an admixture of facial expressions, gestures, and vocal data.There is a necessity to explore data from all these domains in order to analyse a complete emotion of a person.So apart from facial expression significant work has to be done on gesture and speech recognition and their compatible combination.

- Expressions are a combination of many micro expressions.Hence a single expression class is never enough to determine state of one's expression.Differentiating these microfacial expressions and increasing the standard emotion class number demands higher scope for research
- Real time implementation of the expression analysis algorithms and test the same on larger databases of different races and ages.

Bibliography

[1] J. A. Russell, "Is there universal recognition of emotion from facial expressions? a review of the cross-cultural studies.," *Psychological bulletin*, vol. 115, no. 1, p. 102, 1994.

[2] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues.* Ishk, 2003.

[3] H.-B. Deng, L.-W. Jin, L.-X. Zhen, and J.-C. Huang, "A new facial expression recognition method based on local gabor filter bank and pca plus lda," *International Journal of Information Technology*, vol. 11, no. 11, pp. 86–96, 2005.

[4] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 200–205, IEEE, 1998.

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

[6] S.-S. Liu, Y. Zhang, K.-P. Liu, and Y. Li, "Facial expression recognition under partial occlusion based on gabor multi-orientation features fusion and local gabor binary pattern histogram sequence," in *Intelligent Information Hiding and Multime-*

dia Signal Processing, 2013 Ninth International Conference on, pp. 218–222, IEEE, 2013.

[7] S. K. Singh, D. Chauhan, M. Vatsa, and R. Singh, "A robust skin color based face detection algorithm," *Tamkang Journal of Science and Engineering*, vol. 6, no. 4, pp. 227–234, 2003.

[8] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

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

[10] S.-C. Wang, "Artificial neural network," in *Interdisciplinary Computing in Java Programming*, pp. 81–100, Springer, 2003.

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

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

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

[14] G. Bai, W. Jia, and Y. Jin, "Facial expression recognition based on fusion features of lbp and gabor with lda," in *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*, pp. 1–5, IEEE, 2009.

[15] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, pp. 79–86, 1951.

[16] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks, 1989. IJCNN., International Joint Conference on*, pp. 593–605, IEEE, 1989.

[17] M. Kamachi, M. Lyons, and J. Gyoba, "The japanese female facial expression (jaffe) database," *URL http://www. kasrl. org/jaffe. html*, vol. 21, 1998.

[18] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 5, pp. 978–994, 2011.

[19] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.

[20] R. C. Gonzalez, *Digital image processing.* Pearson Education India, 2009.

[21] A. C. Bovik, *Handbook of image and video processing.* Academic press, 2010.

[22] J. Ilonen, J.-K. Kämäräinen, and H. Kälviäinen, *Efficient computation of Gabor features.* Lappeenranta University of Technology, 2005.

[23] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of face recognition*, pp. 247–275, Springer, 2005.

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

[25] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," *Affective Computing, IEEE Transactions on*, vol. 2, no. 4, pp. 219–229, 2011.

[26] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, pp. II–370, IEEE, 2005.

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

[28] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, "Local binary patterns for still images," in *Computer Vision Using Local Binary Patterns*, pp. 13–47, Springer, 2011.

[29] Q.-Y. Zhao, A.-C. Pan, J.-J. Pan, and Y.-Y. Tang, "Facial expression recognition based on fusion of gabor and lbp features," in *Wavelet Analysis and Pattern Recognition, 2008. ICWAPR'08. International Conference on*, vol. 1, pp. 362–367, IEEE, 2008.

[30] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," in *Proceedings of the 5th international conference on Multimodal interfaces*, pp. 258–264, ACM, 2003.

[31] V. P. Lekshmi and M. Sasikumar, "Analysis of facial expression using gabor and svm," *Int. J. Recent Trends Eng*, vol. 1, no. 2, pp. 47–50, 2009.

[32] T. Ahsan, R. Shahriar, and U. Chong, "Application of completed local binary pattern for facial expression recognition on gabor filtered facial images.," *International Journal of Digital Content Technology & its Applications*, vol. 7, no. 12, 2013.

[33] M. Lahbiri, A. Fnaiech, M. Bouchouicha, M. Sayadi, and P. Gorce, "Facial emotion recognition with the hidden markov model," in *Electrical Engineering and Software Applications (ICEESA), 2013 International Conference on*, pp. 1–6, IEEE, 2013.

Publication

Anusha Vupputuri,Sukadev Meher,"Facial expression Recognition using Local Binary Patterns and Kullback Liebler Divergence", *International Conference on Communication ans signal Processing*,ICCSP 2015