FACIAL EMOTION RECOGNITION USING LOCAL BINARY PATTERN

Master of Computer Application

In the Faculty of Engineering & Technology

Jadavpur University

BY

SHREYANNITA BALA

Exam Roll No: **MCA186023**

Registration No: **133686 of 2015-2016**

Under the Guidance of

Dr. Ram Sarkar

Department of Computer Science & Engineering

Jadavpur University

Kolkata - 700 032

2018

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

FACULTY OF ENGINEERING & TECHNOLOGY
JADAVPUR UNIVERSITY
CERTIFICATE

I hereby recommend that the thesis entitled **"FACIAL EMOTION RECOGNITION USING LOCAL BINARY PATTERN"** prepared under my supervision by **SHREYANNITA BALA,** Exam- Roll No: MCA186023 be accepted for the degree of **Master of Computer Application** of **Jadavpur University, Kolkata**.

Supervisor

Dr. Ram Sarkar

Department of Computer Science & Engineering

Jadavpur University,Kolkata - 32

Dr. Ujjwal Maulik

Head of the Department,

Computer Science & Engineering,

Jadavpur University, Kolkata -32

Dr. Chiranjib Bhattacharjee

Dean,

Faculty of Engineering & Technology,

Jadavpur University, Kolkata-32

JADAVPUR UNIVERSITY

FACULTY OF ENGINEERING & TECHNOLOGY

CERTIFICATE OF APPROVAL*

The foregoing thesis "**Facial Emotion Recognition using Local Binary Pattern**" at instance is hereby approved as a creditable study of an engineering subject carried out and presented in a manner of satisfactory to warrant its acceptance as pre-requisite to the degree for which it has been submitted. It is notified to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed and conclusion drawn there in but approve the thesis only for the purpose for which it has been submitted.

Final Examination for Evaluation of thesis **Board of Examiners**

(Signature of Examiners)

* Only in case thesis is approved

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. I owe my gratitude to all those people who have made this dissertation possible and because of whom my postgraduate experience has been one that I will cherish forever.

Foremost, I would like to express my profound gratitude and sincere thanks to my adviser, **Dr. Ram Sarkar**, **Professor, Department of Computer Science & Engineering**, **Jadavpur University**, for his valuable suggestions, guidance, constant encouragement and intent supervision at every stage of my work. I have been amazingly fortunate to have him as my project Supervisor who gave me the freedom to explore on my own, and at the same time guided me to recover when my steps faltered. It has been a great learning process for me.

I am grateful to her for allowing me to use the computing facilities of "Center for Microprocessor Application for Training Education and Research (CMATER)" laboratory, Department of Computer Science and Engineering, Jadavpur University.

I am grateful to **Mr. Dipayan Ghosh**, second year student of MCA, Department of Computer Science & Engineering, Jadavpur University, for his encouragement and practical advice. I am thankful to him, for the long discussions to make me understand the research work from the scratch. His patience and support helped me overcome many crisis situations and finish this dissertation.

I am grateful to **Mr. Showmik Bhowmik**, Research Scholar, Department of Computer Science & Engineering, Jadavpur University, for his encouragement and practical advice. I am thankful to him, for introducing me to Modified LBP and Feature Selection which helped in understanding the viewpoint of this dissertation.

I am deeply grateful to **Mr. Samir Malakar**, Asst. Professor, Department of Computer Applications, Asutosh College, Kolkata-26 and Research Scholars, Department of Computer Science & Engineering, Jadavpur University, for introducing me to Feature Selection which helped me in understanding the viewpoint of the dissertation.

I am thankful to **Dr. Ujjwal Maulik, Head of the Department of Computer Science & Engineering, Jadavpur University** for all the necessary help I got during my thesis work.

I am also thankful to the system staff who maintained all the machines in my lab so efficiently that I never had to worry about viruses, losing files, creating backups or installing software.

Many friends have helped me stay sane through these difficult years. Their support and care helped me overcome setbacks and stay focused on my postgraduate study. I greatly value their friendship and I deeply appreciate their belief in me.

Most importantly, none of this would have been possible without the love and patience of my family. My family, to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my family.

Date : --

Place : Kolkata

SHREYANNITA BALA

M.C.A. (C.S.E)

Roll No:MCA186023

INDEX

List of Figures

CHAPTER 1

INTRODUCTION

Emotions often mediate and facilitate interactions among human beings. In face-to-face communication, emotions are transmitted in proportion of 55% through facial expressions. The interaction between human beings and computers will be more natural if computers are able to perceive and respond to human non-verbal communication such as emotions.[1] That means that if the computer could capture and understand the emotions of its " interlocutor", communication would be more natural and appropriate, especially if we think of scenarios where a computer would play the role of a tutor.[5] Emotion can be recognized through a variety of means such as voice intonation, body language, and more complex methods such electroencephalography (EEG) [1]. However, the easier, more practical method is to examine facial expressions. The human face poses even more problems than other objects since the human face is a dynamic object that comes in many forms and colours [3]. However, facial detection and tracking provides many benefits. Facial recognition is not possible if the face is not isolated from the background. Human Computer Interaction (HCI) could greatly be improved by using emotion, pose, and gesture recognition, all of which require face and facial feature detection and tracking [2]. Here seven types of human emotions are considered. Anger, disgust, fear, happiness, normal, sadness, surprise. Although many different algorithms exist to perform face detection, each has its own weaknesses and strengths. Some use flesh tones, some use contours, and other are even more complex involving templates, neural networks, or filters. These algorithms suffer from the same problem; they are computationally expensive [2].

1.1 Overview of Facial Expressions

A common assumption is that facial expressions initially served a functional role and not a communicative one. I will try to justify each one of the six classical expressions with its functional initially role:

1. **Anger:** It involves two main features eyebrows down and inner side tightening, squinting eyes. The function is clear- preparing for attack.

2. **Disgust:** It involves wrinkled nose and mouth. This expression mimics a person that tasted bad food and wants to spit it out, or smelling foul smell.

3. **Fear:** It involves widened eyes and sometimes open mouth. The function- opening the eyes so wide is supposed to help increasing the visual field (though studies show that it doesn't actually do so) and the fast eye movement, which can assist finding threats.

4. **Surprise:** It is very similar to the expression of fear. Maybe because a surprising situation can frighten us for a brief moment, and then it depends whether the surprise is a good or a bad one. Therefore the function is a very similar one.

5. **Sadness:** It involves a slight pulling down of lip corners, inner side of eyebrows is rising. Darwin explained this expression by suppressing the will to cry. The control over the upper lip is greater than the control over the lower lip, and so the lower lip drops. When a person screams during a cry, the eyes are closed in order to protect them from blood pressure that accumulates in the face. So, when we have the urge to cry and we want to stop it, the eyebrows are rising to prevent the eyes from closing.

6. **Happiness:** It usually involves a smile- both corner of the mouth rising, the eyes are squinting and wrinkles appear at eyes corners. The initial functional role of the smile, which represents happiness, remains a mystery. Some biologists believe that smile was initially a sign of fear. Monkeys and apes clenched teeth in order to show predators that they are harmless. A smile encourages the brain to release endorphins that assist lessening pain and resemble a feeling of well being

7. **Neutral:** Technically with a neutral face the eyes are open (i.e. the upper eyelid has been raised) so a more accurate answer is one where no facial actions have occurred. In neutral the lip corners appear to be pulled down. What is happening is the crease at the lip corners has a downward slant. This downward slant creates the appearance of a curve to the mouth

1.2 Problem Definition

It is often been said that the eyes are the "window to the soul." This statement may be carried to a logical assumption that not only the eyes but the entire face may reflect the "hidden" emotions of the individual. Darwin's research on facial expressions has had a major impact on the field in many areas; foremost, his belief that the primary emotions conveyed by the face are universal [7]. Darwin placed considerable emphasis on the analysis of the action of different muscle groups in assessing expression.

The research on the statement of Darwin was done by Ekman and Friesen. They hypothesized that the universal of facial expression are to -be - found in the relationship between distinctive patterns of the facial muscles and particular emotions (happiness; sadness; anger; fear; surprise, disgust and normal) [7]. This project is aimed at recognizing these 7 human emotions from static images. This is a system which is capable of classifying an facial image into one of the seven basic emotions using some texture based feature descriptor. We have approached by making a comparison with different existing standard techniques like LBP (Local Binary Pattern), ULBP (Uniform Local Binary Pattern), MLBP (Multiscale Local Binary Pattern) etc.

Figure1: Different types of emotion we are considering

1.3 Application

Research on facial expression recognition has been observed dramatic growth in recent years, thanks to the advancements in related fields, especially machine learning, image processing and human cognition. Accordingly, the impact and potential usage of automatic facial expression recognition have been expanded. Some of the significant applications of facial expression recognition system are:

1. Alert system for driving.
2. Social Robot emotion recognition system.
3. Medical Practices.
4. Feedback system for e-learning.
5. The interactive TV applications enable the customer to actively give feedback on TV Program.
6. Mental state identification.
7. Automatic counselling system.
8. Face expression synthesis.
9. Music as per mood.
10. In research related to psychology.

11. In understanding human behaviour.
12. In interview.

1.4 Motivation

Recently there has been a growing interest in improving the interaction between humans and computers. To achieve effective human-computer intelligent interaction, there is a need for the computer to interact naturally with the user, similar to the way humans interact. Humans interact with each other mostly through speech, but also through body gestures to emphasize a certain part of speech and display emotions. Emotions are displayed by visual, vocal and other physiological means. There is more and more evidence appearing that shows that emotional skills are part of what is called 'intelligence' [8]. One of the most important ways for humans to display emotions is through facial expressions. To achieve more effective human-computer interaction, emotional state of the human is recognised here from his or her face.

In general, texture based features have been used for facial emotion recognition since long. LBP is one of most popular and useful texture based features. Hence, in this thesis, different variants of LBP features vectors are calculated from the face image to identify the emotion displaying by the face.

1.5 Scope of the Present Work

It was mentioned earlier that research on facial expression recognition system has been seen a noticeable growth since last decade. In this thesis work, mainly a popularly used texture based feature vector called LBP has been applied. As the basic LBP has some limitations, hence some other variants of LBP namely ULBP and MLBP have been applied for the said purpose. These alternatives of LBP show some size and rotation invariant properties.

Apart from this, from the literature survey, it is observed that mostly researchers extract features from the entire face images. But to identify the emotion of the face, processing of entire image is unnecessary. So, in some recent works, it is seen that people use Viola-Jones object detection to estimate the required region from the face images. Even some researchers, focus on detecting more confined regions for example eyes, mouth, nose to get more informative features which can help the classifier in order to achieve better recognition accuracy. In doing so, they have estimated some rectangular grids enclosing eyes, mouth and nose. But to get the better estimation of these body organs, rectangular region would not be the suitable choice. To overcome this limitation, in this thesis work, a hypothetical ellipse is fitted on each of the important regions of the face namely eye, mouth and nose. These elliptical regions are finally used to extract the different LBP feature vectors. Exclusion of extraneous portion would help to estimate the features in a better way. The experimental outcomes shown in the result section (Chapter 4 of this thesis) justify the selection of the elliptical regions over the certain portions of a face image.

1.6 Organization of Thesis work

In this thesis work, an emotion recognition system is adopted on the basis of facial emotion. The work can be divided into mainly four parts, namely- . Image standardization, face detection, facial component detection and Decision function. The entire work reported here is organized as follows:

Chapter 1: Introduction

In introduction, a brief history of facial emotion detection and the need of the same are explained. Then the problem definition, application of the research work and characteristics of facial emotion are discussed. It also includes the motivation behind this work, scope of the work and the chapter ends with organization of the thesis.

Chapter 2: Related Work

This chapter describes some previous work regarding recognition of facial emotion using LBP.

Chapter 3: Working Methodology

This chapter has few subsections which are briefed as following:

3.1 Data pre-processing

JAFFE datasets have been considered for evaluating proposed method. Images are firstly pre-processed here before applying the feature extraction algorithms.

3.2 Feature Extraction

Features are extracted from eyes, lips and nose using Haar cascade feature extraction technique. Four steps for feature extraction are described in this chapter.

3.3 Basic LBP and all its variants are discussed in this chapter.

3.4 Description of the dataset is given here

3.5 Classification

In this chapter the working principle of the classifier used here called Sequential Minimal Optimization (SMO) has been discussed.

Chapter 4: Results and Discussion

In this chapter, a thorough discussion about the results obtained by the current techniques is made.

4.1 Comparison of the results

A comparative study is discussed with similar works.

4.2 Error Case Analysis

In this section, a study has been conducted over the misclassification word images.

Chapter 5: Conclusion and Future Scope

Finally, the thesis work is concluded in this chapter and the future scope of the work is specified.

CHAPTER 2

RELATED WORK

Chetverikov and P´eteri [11] placed the existing approaches of temporal texture recognition into five classes: methods based on optic flow, methods computing geometric properties in the spatiotemporal domain, methods based on local spatiotemporal filtering, methods using global spatiotemporal transforms and, finally, model-based methods that use estimated model parameters as features. The methods based on optic flow [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] are currently the most popular ones 11], because optic flow estimation is a computationally efficient and natural way to characterize the local dynamics of a temporal texture. P´eteri and Chetverikov [12] proposed a method that combines normal flow features with periodicity features, in an attempt to explicitly characterize motion magnitude, directionality and periodicity. Their features are rotation-invariant, and the results are promising. But they did not consider the multi-scale properties, multi-resolution histograms based on velocity and acceleration fields [18]. Velocity and acceleration fields of different spatiotemporal resolution image sequences were accurately estimated by the structure tensor method. This method is also rotationinvariant and provides local directionality information. Fazekas and Chetverikov compared normal flow features and regularized complete flow features in dynamic texture classification [22]. They concluded that normal flow contains information on both dynamics and shape. Saisan et al. [26] applied a dynamic texture model [23] to the recognition of 50 different temporal textures. Despite this success, their method assumed stationary DTs that are wellsegmented in space and time, and the accuracy drops drastically if they are not. Fujita and Nayar [24] modified the approach [25] by using impulse responses of state variables to identify model and texture. Their approach showed less sensitivity to non-stationarity. However, the problem of heavy computational load and the issues of scalability and invariance remain open. Fablet and Bouthemy introduced temporal co-occurrence [16], [17] that measures the probability of co-occurrence in the same image location of two normal velocities (normal flow magnitudes) separated by certain temporal intervals. Recently, Smith et al. dealt with video texture indexing using spatiotemporal wavelets [26]. Spatiotemporal wavelets can decompose motion into the local and global, according to the desired degree of detail. Otsuka et al. [27] assumed that DTs can be represented by moving contours whose motion trajectories can be tracked. They considered trajectory surfaces within 3D spatiotemporal volume data, and extracted temporal and spatial features based on the tangent plane distribution. The spatial features include the directionality of contour arrangement and the scattering of contour placement. The temporal features characterize the uniformity of velocity components, the ash motion ratio and the occlusion ratio. These features were used to classify four DTs. Zhong and Scarlaro [28] modified [27] and used 3D edges in the spatiotemporal domain. Their DT features were computed for voxels taking into account the spatiotemporal gradient. It appears that nearly all of the research on dynamic texture recognition has considered textures to be more or less 'homogeneous', i.e., the spatial locations of image regions are not taken into account. The dynamic textures are usually described with global features computed over the whole image, which greatly limits the applicability of dynamic texture recognition. Using only global features for face or facial expression recognition, for example, would not be effective since much of the discriminative information in facial images is local, such as mouth movements. In their recent work, Aggarwal et al. [29] adopted the AutoRegressive and Moving Average (ARMA) framework of Doretto et al. [30] for video-based face recognition, demonstrating that temporal information contained in facial dynamics is useful for face recognition. In this approach, the use of facial appearance information is very limited. We are not aware of any dynamic texture based approaches to facial expression recognition [31], [32], [33]

CHAPTER 3

WORKING METHODOLOGY

An image is only a collection of colour and/or light intensity values. Analysing these pixels for face detection is time consuming and difficult to accomplish because of the wide variations of shape and pigmentation within a human face. Pixels often require reanalysis for scaling and precision. Viola and Jones devised an algorithm, called Haar Classifiers, to rapidly detect any object, including human faces, using SMO classifier cascades that are based on Haar-like features and not pixels [4].

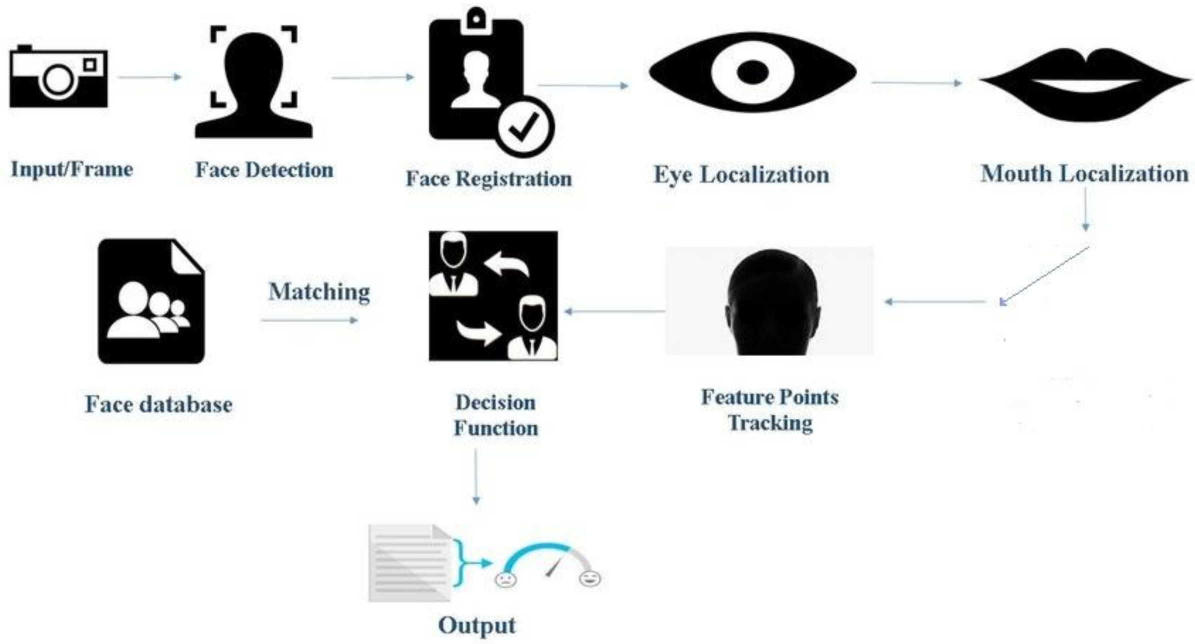

Figure 2. *Face detection and emotion recognition using geometric feature*

3.1 DATA PREPROCESSING

3.1.1 Viola–Jones object detection

3.1.1.1 Haar cascade algorithm

Object Detection using Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001.Face detection is used in a variety of applications, such as control and security applications, and identification systems. A face detector has to inform whether an image of arbitrary size contains a human face and if so, where it is. Face detection is a very essential biometric application in the region of image analysis and computer vision. The elementary face detection method is AdaBoost algorithm with a cascading Haar-like attributes classifiers based on the structure proposed by Viola and Jones. Haar features are composed of either two or three rectangles. Face candidates are scanned and searched for Haar features of the current stage.The weights are constants generated by the learning algorithm.

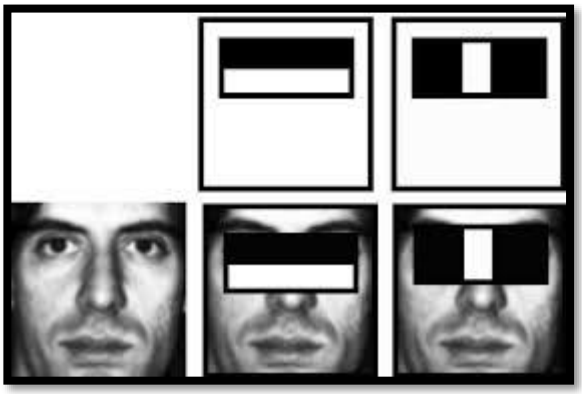

Figure 3:Examples of Haar features. Areas of white and black regions are multiplied by their respective weights and then summed in order to get the Haar feature value.

Each Haar feature has a value that is calculated by taking the area of each rectangle, multiplying each by their respective weights, and then summing the results. The area of each rectangle is easily found using the integral image. The coordinate of the any corner of a rectangle can be used to get the sum of all the pixels above and to the left of that location using the integral image Since L1 is subtracted off twice it must be added back on to get the correct area of the rectangle. The area of the rectangle R, denoted as the rectangle integral, can be computed as follows using the locations of the integral image: L4-L3-L2+L1.

- **Haar Feature Classifier** A Haar feature classifier uses the rectangle integral to calculate the value of a feature. The Haar feature classifier multiplies the weight of each rectangle by its area and the results are added together. Several Haar feature classifiers compose a stage. A stage comparator sums all the Haar feature classifier results in a stage and compares this summation with a stage threshold. The threshold is also a constant obtained from the AdaBoost algorithm. Each stage does not have a set number of Haar features. For example, Viola and Jones" data set used 2 features in the first stage and 10 in the second. All together they used a total of 38 stages and 6060 features [39]. Our data set is based on the OpenCV data set which used 22 stages and 2135 features in total.

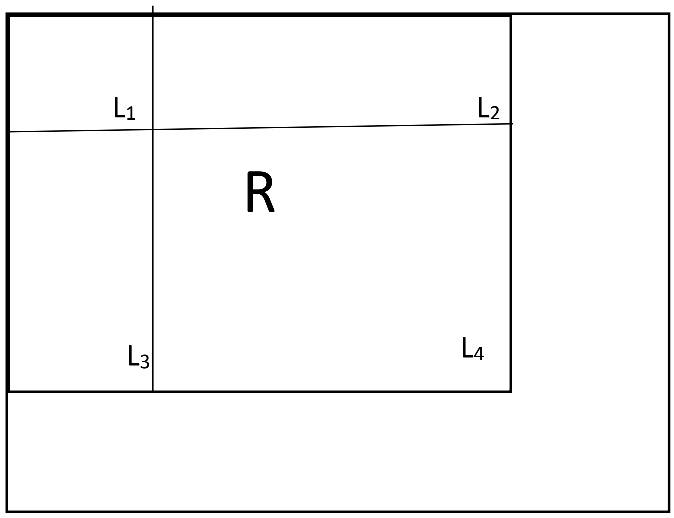

Figure 4: Calculating the area of a rectangle R is done using the corner of the rectangle: L4-L3-L2+L1

Cascade:

The Viola and Jones face detection algorithm eliminates face candidates quickly using a cascade of stages. The cascade eliminates candidates by making stricter requirements in each stage with later stages being much more difficult for a candidate to pass. Candidates exit the cascade if they pass all stages or fail any stage. A face is detected if a candidate passes all stages.

3.2 Feature Extraction

The geometric feature-based approach needs 4 steps to perform facial emotion recognition, as shown in Figure 2 [36].

3.2.1. Image standardization: It includes various sub-processes such as making all the images uniform in size .This makes the image data available for image analysis.

3.2.2. Face detection: This phase involves detecting of a face in the given image data. It aims to remove all the unwanted things from the picture, such as background, and to keep only relevant information, the face, from the data. This phase employs various methodologies such as face segmentation techniques and curvature features. Some of the algorithms that are used in this step include edge detection filters such as Sobel, Prewitt, Laplacian, and Canny[36].

3.2.3. Facial component detection: Here, regions of interests are detected. These regions vary from eyes to nose to mouth, etc. The primary step is to localize and track a dense set of facial points. This step is necessary as it helps to minimize the errors that can arise due to the rotation or the alignment of the face. [36]

3.2.4. Decision function: After the feature point tracking of the face using parameters it is the decision function responsible for detecting the emotion of the subject. These functions make use of classifiers such as AdaBoost and SMO for facial emotion recognition [36].

3.3 Algorithms

- **Basic Local Binary Pattern:**

During the past few years, local binary patterns (LBPs) [8] have aroused increasing interest in image processing and computer vision. As a nonparametric method, LBP summarizes local structures of images efficiently by comparing each pixel with its neighboring pixels. The most important properties of LBP are its tolerance regarding monotonic illumination changes and its computational simplicity. LBP was originally proposed for texture analysis [9], and has proved a simple yet powerful approach to describe local structures. With the LBP operator, the occurrences of the LBP code for an image are collected into a histogram.

Ojala et al. [9] introduced the LBP texture operator, which originally works with the 3×3 neighborhood. The pixel values of eight neighbors are thresholded by the value of the center pixel, then, the so-thresholded binary values are weighted by powers of two and summed to obtain the LBP code of the center pixel. In fact, let g_c and $g_0,\ldots, g_7$ denote respectively the gray values of the center and its eight-neighbor pixels, then the LBP code for the center pixel with coordinate (x, y) is calculated by

$$\mathrm{LBP}(x, y) = \sum{}^{7}_{p=0} s(g_c - g_p)^{2p}$$

where $s(z)$ is the threshold function

$$s(z)=1, \qquad z\leq 0;$$
$$s(z)=0, \qquad z<0;$$

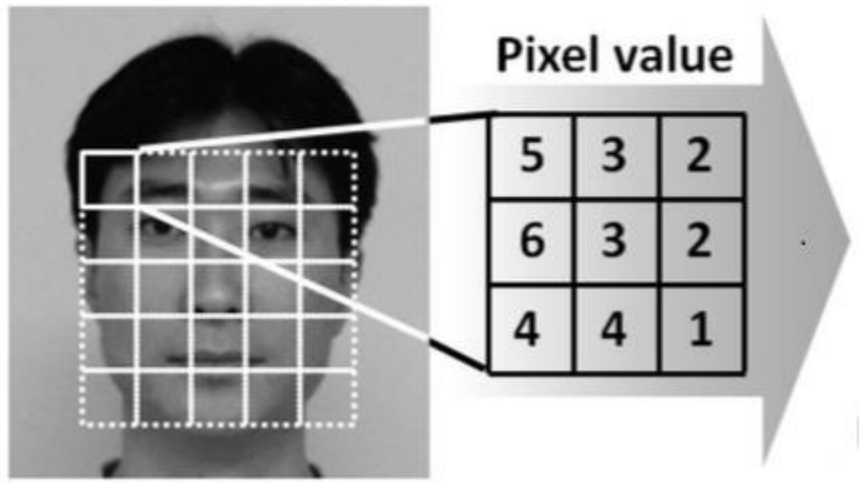

Figure 5: An example of pixel value extraction from the original image.

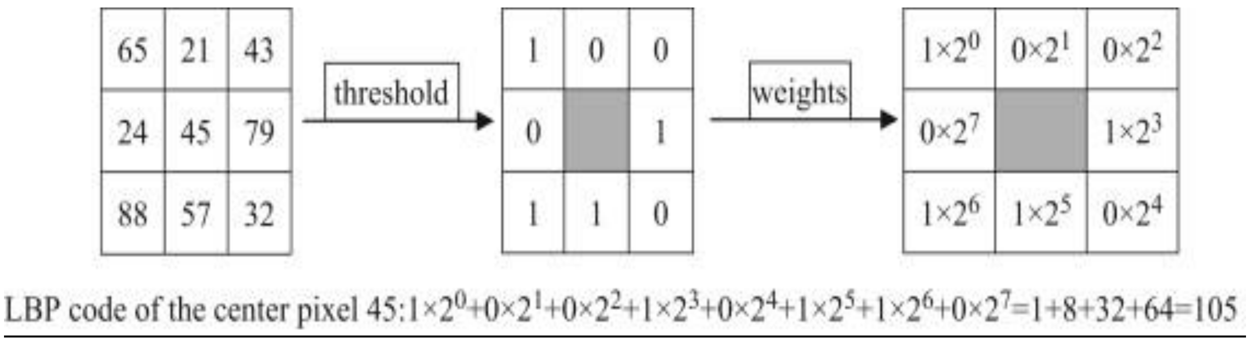

Figure 6: An example of the original LBP operator.

In practice the neighbouring pixels are sampled on a circle, such that the grey values of neighbours which do not fall exactly in the centre of pixels are estimated by interpolation. The co-occurrence of the comparison results is recorded in LBP(x, y), by a unique string of binary numbers, where the sign function s() ensures that the LBP code is invariant against any monotonic transformation of image brightness.

Given an N*M texture image, a LBP pattern LBP(x, y), can be the computed at each pixel c, such that a textured image can be characterized by the distribution of LBP patterns, representing a whole image by a LBP histogram vector h:

$$h(k)=\sum_{i=0}^{N}\sum_{j=0}^{M}\delta(LBP(i,j)-k)$$

where, $0 \leq k < d = 2^p$ is the number of LBP patterns. By altering r and p, one can compute LBP features for any quantization of the angular space and for any spatial resolution. The distinctive advantages of LBP are its ease of implementation, invariance to monotonic illumination changes, and low computational complexity.

Despite these merits, the original LBP has significant disadvantages:

(1) Producing rather long histograms, overwhelmingly large even for small neighbourhoods, leading to decreased distinctiveness and large storage requirements.

(2) Capturing only the very local structure of the texture and failing to detect large-scale textural structures.

(3) Being sensitive to image rotation.

(4) Being highly sensitive to noise: the slightest fluctuation above or below the value of the central pixel is treated the same way as a major contrast.

(5) Losing local textural information due to the use of hard, fixed and coarse quantization and only the signs of differences of neighbouring pixels are utilized.[35]

- **Multiscale Local Bineary Pattern**

A significant limitation of the original LBP operator is its small spatial support area. Features calculated in a local 3×3 neighborhood cannot capture large-scale structures that may be the dominant features of some textures. However, adjacent LBP codes are not totally independent of each other.

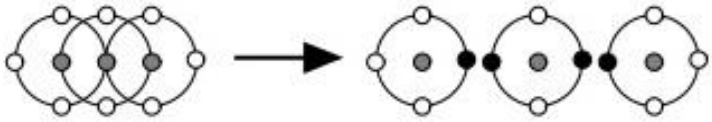

Figure 7:Three adjacent LBP4,R neighborhoods and an impossible combination of codes. A black disk means the gray level of a sample is lower than that of the center

the above image shows three adjacent
four-bit LBP codes [37]. Assuming that the first bit in the leftmost code is zero, the third bit in the code to the right of it must be one. Similarly, the first bit in the code in the center and the third bit of the rightmost one must be either different or both equal to one. The right half of the figure shows an impossible combination of the codes. Each LBP code thus limits the set of possible codes adjacent to it, making the "effective area" of a single code actually slightly larger than 3×3 pixels. Nevertheless, the operator is not very robust against local changes in the texture, caused, for example, by varying viewpoints or illumination directions. An operator with a larger spatial support area is therefore often needed. A straightforward way of enlarging the spatial support area is to combine the information provided by N LBP operators with varying P and R values. This way, each pixel in an image gets N different LBP codes. The most accurate information would be obtained by using the joint distribution of these codes. However, such a distribution would be overwhelmingly sparse with any reasonable image size. For example, the joint distribution of $LBP_{8,1}$, $LBP^{u2}_{16,3}$, and $LBP^{u2}_{24,5}$ would contain $256 \times 243 \times 555 \approx 3.5 \times 107$ bins. Therefore, only the marginal distributions of the different operators are considered even though the statistical independence of the outputs of the different LBP operators at a pixel cannot be warranted. For example, a feature histogram obtained by concatenating histograms produced by rotation-invariant uniform pattern operators at three scales (1, 3 and 5) is denoted as: $LBP^{riu2}_{8,1+16,3+24,5}$.

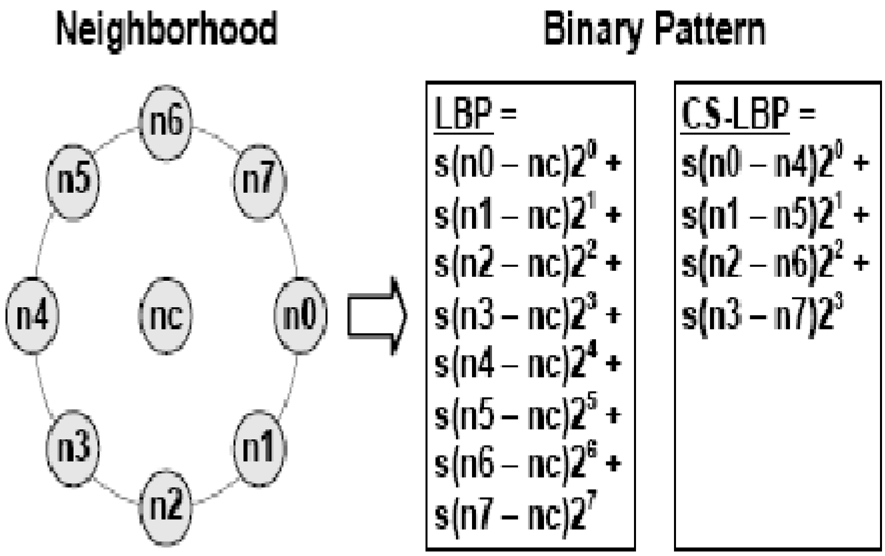

Figure 8: LBP and CS-LBP features for a neighbourhood of 8 pixels

The aggregate dissimilarity between a sample and a model can be calculated as a sum of the dissimilarities between the marginal distributions

$$L_N = \sum_{n=1}^{N} L(S^n, M^n),$$

where S^n and M^n correspond to the sample and model distributions extracted by the *n*th operator [38]. Of course, the chi square distance or histogram intersection can also be used instead of the log-likelihood measure. Even though the LBP codes at different radii are not statistically independent in the typical case, using multi-resolution analysis often enhances the discriminative power of the resulting features. With most applications, this straightforward way of building a multi-scale LBP operator has resulted in very good accuracy.

- **Uniform Local Binary Pattern**

A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa. In a matter of fact this means that a uniform pattern has no transitions or two transitions. Only one transition is not possible, since the binary string needs to be considered circular. The two patterns with zero transitions, with for example eight bits, are 00000000 and 11111111. Examples of uniform patterns with eight bits and two transitions are 00011100 and 11100001. For patterns with two transitions are $P(P - 1)$ combinations possible. For uniform patterns with P sampling points and radius R the notion $LBP^{u2}_{P,R}$ is used. Ojala et al. in [9] observed that some LBP patterns occur more frequently than others, therefore the uniform LBP preserves only the uniform patterns and groups all information contained in the non-uniform patterns. In particular, the uniformity measure

$$LBP^{u2}_{P,R} = \sum_{i=1}^{p} |s(g_{mod(i,p)} - g_c) - s(g_{i-1} - g_c)|$$

All patterns with U>2 are called non-uniform patterns and are classified under a single group, Using only uniform Local Binary Patterns has two important benefits. The first one is that it saves memory. With non-uniform patterns there are 2P possible combinations. With $LBP^{u2}_{P,R}$ there are P (P − 1) + 2 patterns possible. The number of possible patterns for a neighborhood of 16 (interpolated) pixels is 65536 for standard LBP and 242 for LBP^{u2}. The second benefit is that LBP^{u2} detects only the important local textures, like spots, line ends, edges and corners. See figure 1.5 for examples of these texture primitives.

3.4 Experimental Dataset

For the practical work over facial expression recognition, there is a need to have scientific research datasets for different subjects. The dataset which is having more uncontrollable conditions, it is a good choice. The uncontrollable conditions are like pose, occlusion, expression, illumination, expression variation, etc. There are many datasets presented to test methods of facial expression recognition. Some datasets are paid basis and some are publically available online. In some datasets pre-processed images are given for learners. In datasets, one person has different samples images. The datasets like FERET, Extended YaleB, CMU-PIE, AR, Cohn Kanade, ORL, and Indian Face database, Japanese Female Facial Expression JAFEE, etc.[34] Here the JAFFE dataset is used to detect the facial emotion.

- **JAFFE database**

  The JAFFE database (Lyons et al., 1998; Zhang et al., 1998) is used in this project contains 213 images of female facial expressions. Each image has a resolution of 256×256 pixels. The number of images corresponding to each of the 7 categories of expression (neutral, happiness, sadness, surprise, anger, disgust and fear) is almost the same. [6]

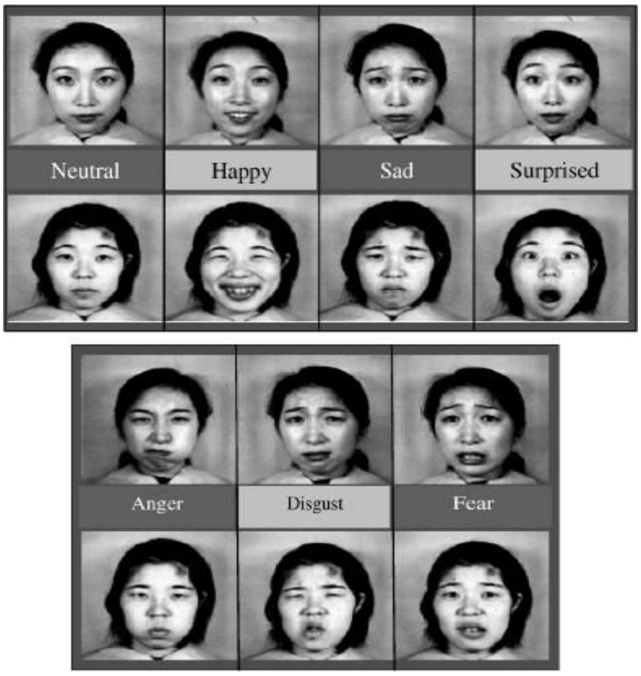

Fig. 9. Sample expressions of two expressers from the JAFFE database.

The images in the database are grayscale images in the tiff file format. The expression expressed in each image along with a semantic rating is provided in the database that makes the database suitable for facial expression research. The heads in the images are mostly in frontal pose.[6]

3.5 Classifier Selection

Sequential Minimal Optimization:

SMO quickly solves the SVM QP problem without using numerical QP optimization steps at all. SMO decomposes the overall QP problem into fixed size QP sub-problems, similar to the decomposition method [41]. SMO chooses to solve the smallest possible optimization problem at each step. For the standard SVM, the smallest possible optimization problem involves two elements of $\bar{\alpha}$ because the $\bar{\alpha}$ must obey one local equality constraint. At each step, SMO chooses two α_i to jointly optimize, finds the optimal values for these α_i and updates the SVM to reflect these new values. The advantage of SMO lies in the fact that solving for two α_i can be done analytically. Thus, numerical QP optimization is avoided entirely. The inner loop of the algorithm can be expressed in a short amount of C code, rather than invoking an entire QP library routine. By avoiding numerical QP, the computation time is shifted from QP to kernel evaluation. Kernel evaluation time can be dramatically reduced in certain common situations, e.g., when a local SVM is used, or when the input data is sparse (mostly zero). The result of kernel evaluations can also be cached in memory [40]. There are two components to SMO: an analytic method for solving for the two α_i, and α heuristic for choosing which multipliers to optimize. Pseudo-code for the SMO algorithm can be found in [42, 41], along with the relationship to other optimization and machine learning algorithms.

CHAPTER4

RESULT AND DISCUSSION

Here, each image in the database represents one of the seven emotions said earlier. The entire image set is divided into test and training sets. From each image either in training or test set, the basic LBP, ULBP and MLBP features are extracted. Before extracting the features, elliptical regions enclosing eyes, lips and nose are estimated and from these regions only features are computed. Finally, these feature vectors are fed to SMO for recognition of emotions.

4.1Comparision between results

The seven emotion labels on which subjects rated the faces are displayed on the *x*-axis.The confusion matrix are given for different varients of LBP.

=== Confusion Matrix (FOR BASIC LBP) ===

a	b	c	d	e	f	g	<-- classified as
68	48	9	16	20	31	7	a = ANGRY
46	69	21	19	19	15	10	b = DISGUST
13	21	46	32	18	23	46	c = FEAR
22	39	28	43	25	23	19	d = HAPPY
22	25	31	35	36	26	24	e = NORMAL
28	23	30	22	18	64	14	f = SAD
14	20	60	17	25	13	50	g = SURPRISE

=== Confusion Matrix (FOR Multiscale LBP) ===

a	b	c	d	e	f	g	<-- classified as
47	39	33	20	18	26	16	| a = ANGRY
41	45	21	31	24	23	14	| b = DISGUST
24	23	37	25	38	22	30	| c = FEAR
29	41	26	31	26	31	15	| d = HAPPY
23	26	45	26	34	28	17	| e = NORMAL
40	38	18	34	29	25	15	| f = SAD
30	22	42	23	20	18	44	| g =SURPRISE

=== Confusion Matrix (FOR Uniform LBP) ===

a	b	c	d	e	f	g	<-- classified as
163	3	2	0	31	0	0	| a = ANGRY
6	139	30	2	3	1	18	| b = DISGUST
4	43	148	2	0	0	2	| c = FEAR
1	2	1	192	0	3	0	| d = HAPPY
30	8	4	0	156	1	0	| e = NORMAL
1	4	4	4	0	181	5	| f = SAD
0	18	7	0	0	3	171	| g = SURPRISE

	Entire Image	Selected Region
	Correctly Classified	Correctly Classified
Basic Local Binary Pattern	**13.8593 %**	**26.9921 %**
Multiscale Local Binary Pattern	**14.6447 %**	**28.8801 %**
Uniform Local Binary Pattern	**30.2943 %**	**82.5556 %**

Table1: Comparison between correctly classified by different LBP feature vector while recognizing the emotion

As can be seen in the confusion matrices the results of classifying the emotion in seven emotions.According to this result we can see uniform local binary pattern is giving better result than any other LBP.

Fig. 10. Sample detection of left eye region from the JAFFE database.

Fig. 11. Sample detection of right eye region from the JAFFE database.

Fig. 12. Sample detection of nose region from the JAFFE database.

Fig. 13. Sample detection of lips region from the JAFFE database.

4.2 Error Case

According to the result displayed by weka there are errors in classifying the images into preper emotion.When the whole image is considered to calculate the histogram the occurrence of error is more than the special parts. Consideration of eye nose lips giving better results and elimination errors upto a limit.

	Entire Image	Selected Region
	Incorrectly Classified	Incorrectly Classified
Basic Local Binary Pattern	**86.1407 %**	**73.0079 %**
Multiscale Local Binary Pattern	**85.3553 %**	**71.1199%**
Uniform Local Binary Pattern	**69.7057 %**	**17.4444 %**

Table 2: Comparison among different variants of LBP feature vector while recognizing the facial emotions.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

In this thesis, we present a comprehensive empirical study of facial expression recognition based on Local Binary Patterns features. Different classification techniques are examined on several databases. This work describes a real-time automatic facial expression recognition system using static image input. Our work focuses on initially detecting the human face in the image, on classifying the human emotion from facial features and on visualizing the recognition results. The key issues of this work can be summarized as follows:

1. Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. We empirically evaluate LBP features to describe appearance changes of expression images. Extensive experiments illustrate that LBP features are effective and efficient for facial expression recognition.

2. One challenge for facial expression recognition is recognizing facial expressions at low resolutions, as only compressed low-resolution video input is available in real-world applications. We investigate LBP features on low-resolution images, and observe that LBP features perform stably and robustly over a useful range of low resolutions of face images.

One limitation of this work is that the recognition is performed by using static images without exploiting temporal behaviors of facial expressions. The psychological experiments by Bassili [43] have suggested that facial expressions are more accurately recognized from a dynamic image than from a single static image. We will explore temporal information in our future work. Recently LBP from three orthogonal planes have been introduced for dynamic texture recognition [44], showing promising performance on facial expression recognition in video sequences. Another limitation of the current work is that we do not consider head pose variations and occlusions, which will be addressed in our future work. We will also study the effect of imprecise face location on expression recognition results.

REFFERENCES

1. Bradski, G. Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, 2nd Quarter, 1998.
2. Muller, N., Magaia, L. and Herbst B.M. Singular value decomposition, eigenfaces, and 3D reconstructions. SIAM Review, Vol. 46 Issue 3, pp. 518–545. Dec. 2004.
3. Viola, P. and Jones, M. Rapid object detection using boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition, 2001.
4. Andreea Pascu and Prof. Ross King University of Manchester, School of Computer Science, Facial Expression Recognition System, pp 1-44 Apr 2015
5. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J., 1998. Coding facial expressions with Gabor wavelets. In: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition.

6. S. Bashyal , G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Eng. Appl. Artif. Intell.*, vol. 21, no. 7, pp. 1056–1064, 2008.

8. Thelehaeswary Appalasamy , Article: "REAL-TIME FACIAL EXPRESSION RECOGNITION SYSTEM" A thesis submitted in partially fulfilment of the requirements for the award of degree of Bachelor of Computer Science (Software Engineering) ; Faculty of Computer System & Software Engineering Universiti Malaysia Pahang (IJMP); pp. 2-14 , May 2011
9. T. Ojala, M. Pietikinen, T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, Jul. 2002.
10. T. Ojala, M. Pietikinen, D. Harwood, "A comparative study of texture measures with classification based on featured distribution", *Pattern Recog.*, vol. 29, no. 1, pp. 51-59, 1996.
11. D. Chetverikov and R. P´eteri, "A Brief Survey of Dynamic Texture Description and Recognition," Proc. Int'l Conf. Computer Recognition Systems, pp. 17-26, 2005.
12. R. P´eteri and D. Chetverikov, "Dynamic Texture Recognition using Normal Flow and Texture Regularity," Proc. Iberian Conference on Pattern Recognition and Image Analysis, pp. 223-230, 2005.
13. R. Polana and R. Nelson, "Temporal Texture and Activity Recognition,"Conf- Motion-based Recognition, pp. 87-115, 1997.
14. R.C. Nelson and R. Polana, "Qualitative Recognition of Motion Using Temporal Texture," Computer Vision, Graphics, and Image Processing, vol. 56, no. 1, pp. 78-99, 1992.
15. P. Bouthemy and R. Fablet, "Motion Characterization from Temporal Co-occurrences of Local Motion-based Measures for Video Indexing," Proc. Int'l Conf. Pattern Recognition, vol.1, pp. 905-908, 1998.
16. R. Fablet and P. Bouthemy, "Motion Recognition Using Spatio-temporal Random Walks in Sequence of 2D Motion-related Measurements," IEEE Int'l Conf. Image Processing, pp. 652-655, 2001.

17. R. Fablet and P. Bouthemy, "Motion Recognition Using Nonparametric Image Motion Models Estimated from Temporal and Multiscale Cooccurrence Statistics," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, pp. 1619-1624, 2003
18. Z. Lu, W. Xie, J. Pei, and J. Huang, "Dynamic Texture Recognition by Spatiotemporal Multiresolution Histogram," Proc. IEEE Workshop Motion and Video Computing, vol. 2, pp. 241-246, 2005.
19. C.H. Peh and L.-F. Cheong, "Exploring Video Content in Extended Spatiotemporal Textures," Proc. 1st European Workshop Content-Based Multimedia Indexing, pp. 147-153, 1999.
20. C.H. Peh and L.-F. Cheong, "Synergizing Spatial and Temporal Texture," IEEE Trans. Image Processing, vol. 11, pp. 1179-1191, 2002.
21. R. P´eteri and D. Chetverikov, "Qualitative Characterization of Dynamic Textures for Video Retrieval," Proc. Int'l Conf. Computer Vision and Graphics, 2004
22. S. Fazekas and D. Chetverikov, "Normal Versus Complete Flow in Dynamic Texture Recognition: A Comparative Study," 4th Int'l Workshop Texture Analysis and Synthesis, pp. 37-42, 2005.
23. M. Szummer and R.W. Picard, "Temporal Texture Modeling," Proc. IEEE Conf. Image Processing, vol. 3, pp. 823-826, 1996.
24. K. Fujita and S.K. Nayar, "Recognition of Dynamic Textures Using Impulse Responses of State Variables," Proc. Third Int'l Workshop Texture Analysis and Synthesis, pp. 31-36, 2003.
25. P. Saisan, G. Doretto, Y.N. Wu, and S. Soatto, "Dynamic Texture Recognition," Proc. Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 58-63, 2001.
26. J.R. Smith, C.-Y. Lin, and M. Naphade, "Video Texture Indexing Using Spatiotemporal Wavelets," Proc. IEEE Int'l Conf. Image Processing, vol. 2, pp. 437-440, 2002.
27. K. Otsuka, T. Horikoshi, S. Suzuki, and M. Fujii, "Feature Extraction of Temporal Texture Based on Spatiotemporal Motion Trajectory," Proc. Int'l Conf. Pattern Recognition, vol. 2, pp. 1047-1051, 1998.
28. J. Zhong and S. Scarlaro, "Temporal Texture Recognition Model Using 3D Features," Technical report, MIT Media Lab Perceptual Computing, 2002.
29. G. Aggarwal, A.R. Chowdhury, and R. Chellappa, "A System Identification Approach for Video-based Face Recognition," Proc. 17th Int'l Conf. Pattern Recognition, vol. 1, pp. 175-178, 2004.
30. G. Doretto, A. Chiuso, S. Soatto, and Y.N. Wu, "Dynamic Textures," International Journal of Computer Vision, vol. 51, no. 2, pp. 91-109, 2003.
31. M. Pantic and L.L.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no.12, pp. 1424-1455, 2000.
32. B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," Pattern Recognition, vol. 36, pp. 259-275, 2003.
33. T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 46-53, 2000.
34. Shyna Dutta1, V.B. Baru ," REVIEW OF FACIAL EXPRESSION RECOGNITION SYSTEM AND USED DATASETS" . International Journal of Research in Engineering and Technology, Volume: 02 Issue: 12 |pp. 644, Dec-2013,

35. Li Liu , Paul Fieguth , Yulan Guo , Xiaogang Wang , Matti Pietikäinen , "Local binary features for texture classification: Taxonomy and experimental study" , Pattern Recognition 62 (2017) 135–160
36. Dhwani Mehta, Mohammad Faridul Haque Siddiqui and Ahmad Y. Javaid, " Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality" , EECS Department, The University of Toledo, Toledo, OH 43606, USA; |pp. 4-24 Published: 1 Febuary 2018
37. Mäenpää, T.: The local binary pattern approach to texture analysis—extensions and applications. PhD thesis, Acta Universitatis Ouluensis C 187, University of Oulu (2003)
38. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
39. P. Viola and M. Jones, "Robust real-time object detection," International Journal of Computer Vision, 57(2), 137-154, 2004.
40. c. J. C. Burges. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2), 1998.
41. J. C. Platt. Fast training of SVMs using sequential minimal optimization. In B. Scholkopf, C. J. C. Burges, and A. J. Smola, editors, Advances in Kernel Methods - Support Vector Learning, pages 185-208. MIT Press, 1998.
42. J. C. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR- TR-98-14, Microsoft Research, 1998. Available at http://www.research.microsoft.comrjplattlsmo.html.
43. J.N. BassiliEmotion recognition: the role of facial movement and the relative importance of upper and lower area of the face Journal of Personality and Social Psychology, 37 (11) (1979), pp. 2049-2058
44. G. Zhao, M. PietikäinenDynamic texture recognition using local binary patterns with an application to facial expressions IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (6) (2007), pp. 915-928

*