Study of Smile Detection Methods

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

Master of Engineering
in
Computer Science and Engineering

Submitted By
Prerna Menroy
(801732036)

Under the supervision of:
Dr. Karun Verma
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

AUGUST 2019

CERTIFICATE

I hereby certify that the work which is being presented in the thesis titled, *"Study of smile detection methods"*, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Karun Verma and refers other researcher's work which are duly listedin the reference section.

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

Prerna 15/10/19
Signature:
Prerna Menroy

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

Dr. Karun Verma
Assistant Professor
Computer Science and
Engineering Department
Thapar Institute of
Engineering and Technology

ACKNOWLEDGEMENT

No volume of words is enough to express my gratitude towards my guides **Dr. Karun Verma** Department of Computer Science & Engineering, Thapar Institute, Patiala who has been very concerned and has aided for all the materials essentials for the preparation of this thesis report. They helped me to explore this vast topic in an organized manner and provided me all the ideas on how to work towards a research-oriented venture.

I am also thankful to **Dr. S.S. Bhatia,** Dean of Academic Affairs, **Dr. Maninder Singh,** Head of Computer Science & Engineering Department and **Dr. Ashutosh Mishra,** P.G. Coordinator, for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank the staff members and my colleagues who were always there at the need of hour and provided with all the help and facilities, which I required, for the completion of my thesis work.

Most importantly, I would like to thank my parents and the almighty for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

Prerna
15/10/19

Prerna Mehroy

(801732036)

ABSTRACT

A happy face or smile is a common expression in our daily life. This brings out basic emotions which are hidden emotions, like satisfaction and happiness. This is the most powerful and challenging tasks in social communication. Smile detection is a feature of a digital camera that resists the image from being captured until and unless person is smiling. Smile detection is divided into two parts or phases. Firstly, it detects a face from image or video, then waits for smile. In more detail we can say, a motion detector splits the image or video into frames, then analyzes frameworks such as flash and balanced level for the facial region. When a person laughs or smiles, the camera detects a deformation by identifying multiple criteria, such as raised cheeks, upturned mouth visible teeth, and narrowed eyes, etc. Detection of smile is done in many applications like automatic image capturing, interactive systems, patient monitoring, product rating, video conferencing, etc. Effective representation of smile is important for smile detection and image retrieval applications. In this report we have studied various smile detection methods and proposed two smile-detection models. In the first model we use LBP and SOM classifier. The second model is extended version of first models. The second model has two consecutive actions: 1) amalgamation of Geometric Feature Extraction (GFE) and regional Local Binary Pattern (LBP) features extraction using autoencoders; 2) Self-Organizing Map (SOM) is adopted to classify smile based on these features. A comprehensive evaluation of the proposed models on a benchmark dataset GENKI-4K for smile detection shows a notable improvement in terms of performance measures on respective datasets as compared to the other popularly used smile detection methods.

Table of Contents

List of Figures

List of Tables

Abbreviations

HMM	Hidden Markov Model
ASM	Active Shape Model
FE	Feature Extraction
GF	Geometric Feature
LBP	Local Binary Pattern
HCI	Human-Computer Interface
SOM	Self-Organizing Map
PWD	Pair-wise distance Vector
EDV	Euclidean Distance Vector
PCA	Principal Component Analysis
LTP	Local Ternary Patterns
LDP	Local Directional Patterns
Aus	Action Units
LMeP	Local Mesh Patterns
LDTP	Local Directional Texture Pattern
LDN	Local Directional Number
SVM	Support Vector Machines
KNN	K- Nearest Neighbor
EOH	Edge Orientation Histograms
HOG	Histogram of Oriented Gradients
MCFs	Mouth-Corner Features
ELM	Extreme Learning Machine
CNN	Convolutional Neural Network

Chapter 1

Introduction

1.1 Motivation

The facial expression is nonverbal way to communicate and express emotion and intention of any human. It plays an important role in our daily social life. It also play an important part in recognizing the human expressions. They deliver nonverbal communication signs in face-to-face communication and serve a very natural and important means for humans to communicate their intentions, mindset, and emotions. Facial Expressions provides human beings for a way to complement speech and elicit the intended meaning of the speech. For human beings, face detection and interpreting facial expression under varying conditions is a day to day task which they perform effortlessly. We can simply interpret the gender, identity, age, etc. of a person by just looking once face. The smiling face is the most common and important facial expressions that indicate emotion like delight, satisfaction, and excitement in normal scenario. Smile detection has grabbed attention for researchers. Mehrabian says, effective communication include, 55 percent of the conversation is done by facial articulations, 38 percent by paralinguistic, and 7 percent by the words that are spoken at that instant of time. Our attitude towards the speaker changes depending on the impression we get by reading the facial expressions and it also affects our interpretation of the spoken words, i.e. expression gives a different meaning to the statement the speaker is saying. Facial expression is a source of essential information communicated by human beings; therefore it plays a vital role in our daily lives.

On contrary, for computer-based systems, smile detection still remains a very challenging and difficult task. The traditional systems which are based on Human-Computer Interaction (HCI) only cater the inputs which are intentionally given by users while they simply ignore the other that is communicated in bulk through those affective states. But with the advent in technology, the focus is being shifted on human-centered designs and therefore human beings are able to communicate with machines with not only the intentional inputs but also through affected states i.e. their behavior. Smile detection has therefore gathered much popularity in the computer vision research community. A system for automatic smile detection can benefit a lot of application areas. Smile detection has been studied extensively in the literature by the computer vision group but it fails to meet the certain standard considering camera quality, geometry of face and image quality, postures, lighting, angle change, and lot more.

1.2 Smile Detection

A facial smile detection system is a pattern recognition based application which detects or verifies that a person is smiling face or non-smiling face from a given image or video. This is generally done by comparing the facial features obtained from the image to that of various standard available databases. According to Dr. Paul Ekman who is a pioneer in the field of Facial Expression Recognition, universally categorized facial expressions. There are six universally accepted emotions which can be used as benchmark for facial expression recognition.

- **Happiness:** It is signified when the mouth corners are raised and the eyelids are tightened.
- **Sadness:** It is signified by dropping of jaws, arching of the eyebrows and widening of eyes.
- **Surprise:** It is signified by drooping eyelids, lowering of the mouth corners and when eyebrows descend to the inner corners.
- **Fear:** It is signified by lowering of the eyebrows, bulging of the eyes and when lips are pressed firmly.
- **Disgust:** It is signified by raising of cheeks, wrinkling of the nose bridge and raising of the upper lip.
- **Anger:** It is signified by stretching of the lips, the opening of the eyes and raising of the upper eyelids.

A seventh universal emotion called ‘Contempt’ was proposed to be added in the year 1990 which is a combination of anger and disgust but there was a dispute on whether or not it should be considered as a universally recognized emotion. Happiness emotion considers as smiling face remaining six emotions consider as the non-smiling face.

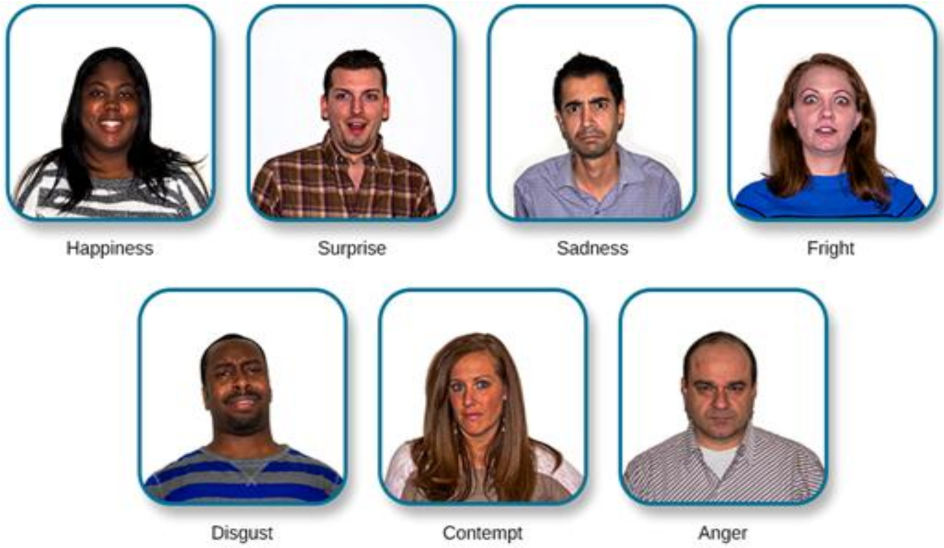

Fig. 1.1 The seven universal expressions

1.3 Framework for Smile Detection

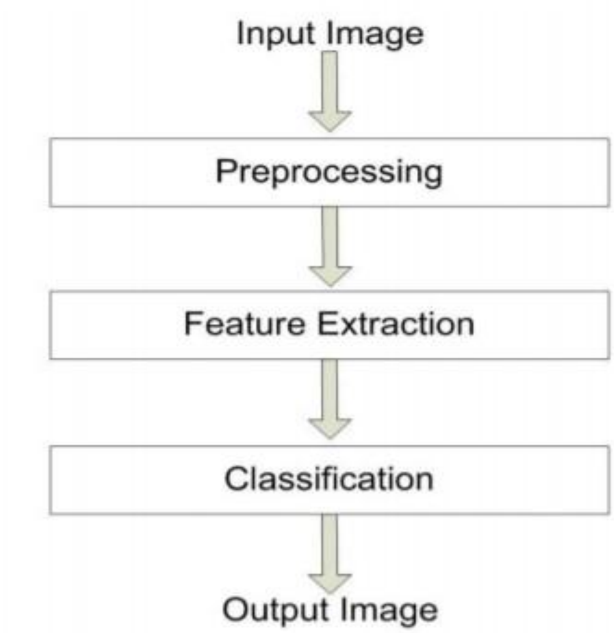

Fig. 1.2. Steps involved in Smile Detection

The analysis of smile expressions (automatic) involves the measurement of changes in facial muscles and the classification between smiling and non-smiling expressions. The process of smile detection involves the sequential processing of blocks, which is according to a standard pattern recognition problem. The main steps include image acquisition, pre-processing of image and feature extraction on the processed image, image classification, and post-processing on the image. The feature extraction is the most important and valuable step in facial expression recognition as if inadequate features are used to represent the face in an image then even the best classifier would not be able to obtain the accurate classification.

1.3.1 Face Detection

Face detection involves automatic trace face in an input image or video sequence. An image may be comprised of a variety of objects like fruits, cars, etc. Face detection, therefore, involves localization of human face within an image.

Face Detection can be done through 4 methods primarily:

- Facial Invariant
- Knowledge-Based
- Template Matching
- Appearance Based

1.3.2 Pre-Processing

Facial images obtained from the video or image sources are generally prone to noise. Thus in order to get complete facial images with normalized intensity, shape and size and pre-processing are applied on faces detected in the first step. This is done because the representation of the expressions can be influenced by scaling, translation, rotation of the head in a facial image. Therefore to suppress the damage caused by these unwanted transformations, a geometrical standardization of facial image is done before classification. Detection of feature points, rotation of the head and locating and cropping of the facial region i.e. segmentation using a bounding box according to the face models are the steps involved in the pre-processing of the faces in order to normalize them for feature extraction. Segmentation usually involves the division of the facial regions and is performed on the basis of shape, color, motion, texture, etc. of the face.

Table 1.1: Facial detection methods and its related works

Techniques	**Related Work**
Knowledge-based	Multi-resolution rule-based method
Feature invariant (Facial features, Texture, Skin color, Multiple Features)	Grouping of edges, Mixture of Gaussian and Integration of skin color, shape and size
Template Matching (Deformable templates, Predefined face templates)	Shape template, Active Shape Model(ASM)
Appearance Based (Hidden Markov Model(HMM), Information- Theoretical approach)	Higher-order statistics with HMM, Pullback relative information

1.3.3 Feature Extraction (FE)

Feature extraction aimed conversion of the pixel level data into a higher-level representation of texture, shape, color, motion, etc. of the components. This extracted information is further used for expression classification of expressions. Feature extraction also attempts to minify the dimensional of the input space. The procedure involved in reduction aims to retain the important information which possesses high stability and high discrimination power.

Geometric feature-based appearance feature-based approaches are two widely used approaches for feature extraction.

- **Geometric Features (GF) based Methods** In GF based approaches, differentiating facial features can be detected by utilizing facial properties like the relationships in between the facial region between eyes, eyebrows, nose and mouth on the face, structures of edge corresponding to various facial areas and the structural dimensions. This also includes the position and structure of the corners of the mouth, the nose and the eyes.
- **Appearance Based Methods** In appearance-based methods, image filters like Haar wavelet, Gabor wavelets, moments, etc. are used on the face or of the facial region for facial feature extraction.

Some hybrid approaches are also applied for facial feature extraction which is a combination of various techniques and does not exclusively fall into an either geometric based or appearance based approaches. After the feature vector is obtained, the next step is the classification of emotion into either of the smiling or non-smiling expressions. But for good classification accuracy, it is necessary to select discriminating and sufficient amount of features. The efficiency depends on the fact that how relevant the selected features are and also the quality of these selected features. In general, we extract a set of variables from huge sets of data to describe the data in simple and efficient way. It is therefore quite challenging to identify and select relevant features that will improve the accuracy of classification.

Table 1.2: Feature Extraction Methods

Deformation Extraction	**Geometric Features**	**Appearance Features**
Image-Based	Gabor Filter	Local Gabor Filter Bank, Fisher's Linear Decomposition
Model-Based	Local Binary Pattern (LBP)	Feature point Tracking
Motion Extraction	**Geometric Features**	**Appearance Features**
Frame-Based	Active contour	Gabor Filter Bank
Sequence-Based	PCA, Gabor Filter Bank, AdaBoost	Haar-like feature

1.3.4 Classification

Once the feature vector is obtained, the classification of features is performed. For this purpose, various techniques that are published in literature like SVM, Bayesian classifiers, Neural Networks, fuzzy techniques, etc. can be used. The class labels usually comprise of the two expressions as identified by the smile and natural.

Classifiers used for expression recognition can either be frame-based classifiers or sequence-based classifiers.

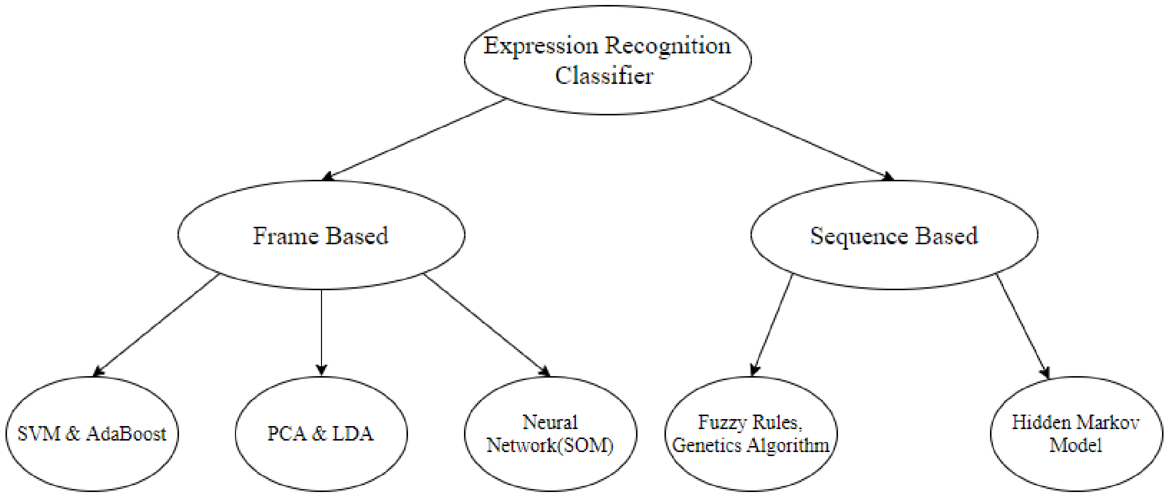

Fig. 1.3. Classification Approaches for Facial Expressions

1.3.5 Post-Processing

In order to improvise the classification accuracy, the domain knowledge is applied in the post-processing step to correct the classification errors, so that we can get correct results.

1.4 Applications

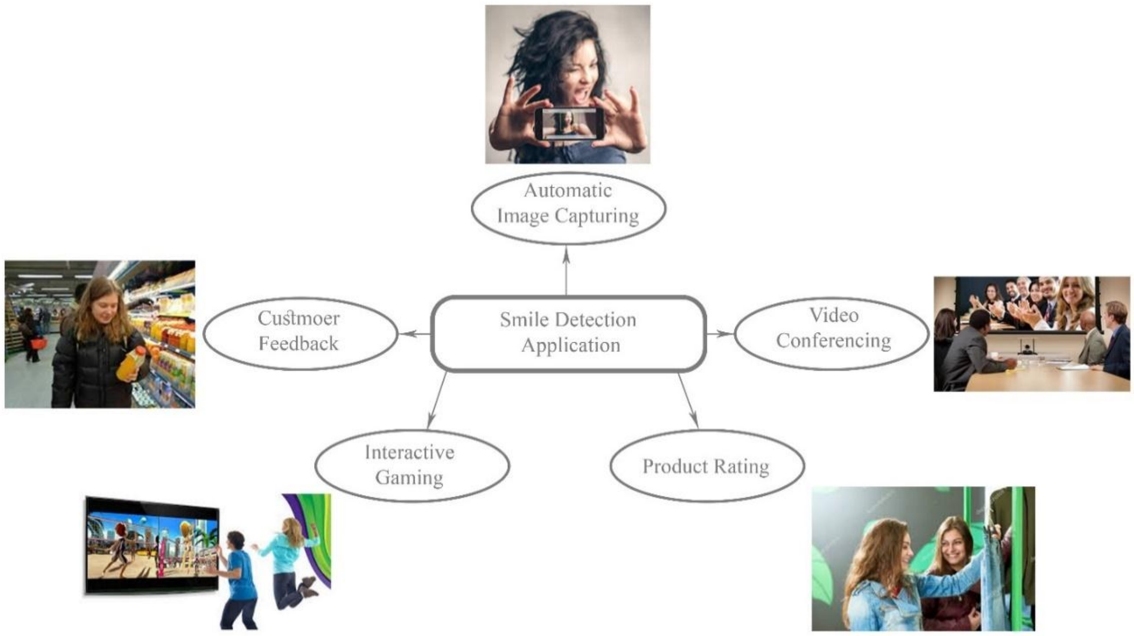

Fig. 1.4. Applications of Smile Detection

The recent advances in the field of automatic image capturing have made smart AI-based cameras are part of our day to day life. This demands a smart camera to become more intelligent and they must be able to capture the image when person or people smile. This is how Smile Detection Systems will help in capturing better image with natural smile.

Face plays an important role in social interactions. Among all the mediums of communication, a smile is the one that is always acting according to the researchers of social psychology. A smile can, therefore, be used for measurement of customer satisfaction i.e. the customers can be evaluated based on their smile at the counter. The multimedia-based approach serves as an alternative to the traditional ways of collecting customer's feedback.

Smile detection can also play a vital role in HCI (Human-Computer Interface) based applications like alertness interactive gaming, product rating, video conferencing, etc. Researchers have demonstrated the practical applications of these systems time and again.

1.5 Challenges

Although many existing approaches for smile detection have shown good results, still there remain many issues which need to be addressed by the researchers.

- Most of the methods which are proposed in the literature are computationally expensive. They require large-sized feature vectors in order to achieve good classification results. Therefore, they are not so feasible for real-time applications despite producing good results on standard databases.
- Very few of the existing approaches give good classification results on low-resolution images which is a must for numerous application areas like video conferencing and smart meetings, security surveillance, etc.
- If a person already notices about capturing the image or video then their expression that they will be displaying are manufactured emotions and not the natural ones, so capturing natural smile is a major challenge.
- It is hard to catch expressions when people’s wear spectacles or goggles and wrap or scarfs (facial region obstacle).
- Another important challenges, capturing smile, who does not have lost their spontaneous expressions due to some medical challenges. Like some of these diseases which cause

loss of their spontaneous expressions: Asperger Syndrome, Autistic Disorder, Bell's palsy, Depression, Depressive Disorders, Facial Paralysis, Facial weakness, Hepatolenticular Degeneration, Major Depressive Disorder, Parkinson Disease, Scleroderma and Wilson Disease.

1.6 Thesis Outline

In the chapter-2 various state of the art techniques and their contribution to the literature have been discussed. In chapter-3 we discuss problem statement. In chapter-4 the proposed method for smile detection using LBP and SOM classifier are presented. The experiments and results of smile detection using data amalgamation on a benchmark dataset namely GENKI-4K are presented in chapter-5. Finally, in chapter-6 the thesis is concluded and the future scope of this work.

1.7 Discussion and Summary

In this chapter, a limited introduction about the thrust research area with its salient features and limitations is discussed. It also contains some wide research issues with the various threats in the ongoing scheme with an eye on future applications too. Smile detection framework is also presented in this chapter. Further, identified research gaps are discussed following by the dissertation contributions in brief.

Chapter 2

Literature Review

An extensive study on various feature descriptors for facial feature extraction needed for smile detection has been done and presented in the sections below. Many benchmark methods for extraction of facial features and many new approaches for effective feature representation are discussed. A detailed survey on smile detection has been given in this chapter.

2.1 Related Work

Geometric features and appearance-based methods are two commonly used approaches for the extraction of facial features. Geometric based methods represent the face by encoding the shape and location of different components of the face and combining them into a feature vector. An example of these methods is Graph-based methods which use different parts of the face to create and process a facial representation. Also, the Local-Global Graph algorithm which segments the local features using Delaunay triangulation and Voronoi tessellation to build a graph for face and expression recognition in an interesting approach. The features then obtained are mixed into local graphs which are then interrelated to obtain the global graph which represents the face topology. As a result of the work done by Dongshun Cui, Guang-Bin Huang and Tianchi Liu Pair-wise Distance Vector and Extreme Learning Machine (ELM) were developed which uses facial features to identify smile expression. This system was only on mouth region so computation work is very less.

However, GF based approaches generally expect reliable and accurate facial region feature detection and tracking. Also, a lot of pre-processing is necessary to localize various facial components before facial features are extracted.

On the other hand, facial textures can easily be represented using appearance-based methods which are also relatively easy to implement. In order to obtain holistic features, appearance-based methods use image filters, either on the entire-face or some specific face-region so as to extract the appearance changes in the face image and create local features. They perform very well in the constrained environment else their performance degrades.

For the holistic class, many methods had been studied in the literature like Eigenfaces, Fisher faces which are based on the concept of Principal Component Analysis (PCA). 2D PCA and Linear Discriminate Analysis (LDA) are types of holistic methods.

Despite these methods being studied widely, lately, local methods have gained much popularity due to their pose variations and robustness to illumination. In the local feature methods, the information is gathered from parts of the face to compute local descriptor which is then gathered into one descriptor. Examples of these methods include Local Feature Analysis, Gabor features, Local Binary Patterns (LBP), Elastic Bunch Graph Matching, etc. Gabor wavelets and Local Binary Patterns are the two most popular-appearance based methods which are widely used for facial expression recognition.

LBP has strong texture discrimination capability and can quickly extract expression features as compared to Gabor wavelet transform. LBP is widely used in various pattern recognition problems like texture classification, facial recognition, expression recognition, and many other areas. But, the traditional LBP operator which compares the gray value of a center pixel with neighboring pixels does not accurately describe the texture of a facial image accurately especially under conditions of noise and data input variability mainly because it only considers the sign of the pixel difference whereas the magnitude is totally ignored.

The newer LBP variants like Local Ternary Patterns (LTP), Local Directional Patterns (LDP) attempted to deal with the drawbacks of LBP. LDP encodes the directional information in place of intensity information. Higher-order local derivatives Patterns (LDeP) were also explored by Zhang *et al.* to achieve good results compared to LBP. These methods used other information in place of intensity to overcome problems related to illumination variation and noise. But they still suffer random noise, non-monotonic illumination and changes in pose, age and expression conditions. A few methods like Gradient faces have a better result under illumination variation; it still has lower recognition rates for age and expression variation conditions. Different features like infrared, near-infrared and phase information, etc. have also been explored to overcome the problem of illumination variation but they still have lower recognition capabilities under conditions of age and expression variability.

In the last decade many smile detection techniques proposed which is mention below:

In 2011, C. Shan developed an efficient method to smile detection for face picture captured in real-world unconstrained scenarios. In their method, they compared the pixel intensities in the grayscale facial image and the Intensity Differences (ID) are used as a features extraction technique. They developed a powerful classifier using a combination of AdaBoost and Intensity Differences.

In 2012, Y. Zhang *et al.* proposed a method in which they take only Mouth Feature (MF), which is efficient for memory utilization. For feature extraction, they used Intensity Difference (ID). Maximum Feature Difference (MFD) algorithm is defined to minimize the ID features. AdaBoost used as a strong classifier.

In 2012, Yadappanavar, H *et al.* focused on smile detection which is part of facial expression recognition. They proposed a method based on ML technique for detecting a smiling face in real-time. ML method includes training a classifier and using it in real-world unconstrained scenarios to classify the smile.

In 2012, C. Shan developed an efficient method to smile detection for face picture captured in the real-time image. In their method, they compared the pixel intensities in the grayscale facial image and the Intensity Differences (ID) are used as a features extraction technique. They developed a powerful classifier using a fusion of AdaBoost and Intensity Differences.

In 2013, Jain, V. *et al.* employed Multi-Scale Gaussian Derivative (MGD) and SVM for smile detection, and obtained an accuracy of 92.97% with compare to Gabor Energy Filters (GEF) on GENKI4K database. Expression recognition in social life is very important for evaluating their impressive communication in the group. Among all expressions of a human smile is one of common expression. It helps to understand

In 2013, Nergui, M. *et al.* developed a model for automatic detecting of laughter with the help of respirator sensor data. They fused sensor data with smile degree and extracts feature. HMM is used as a classifier.

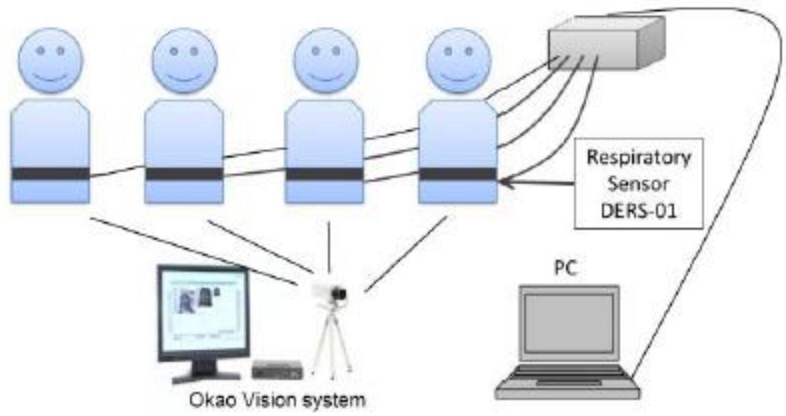

Fig.2.1. Smile detection using Respiratory Sensor Data with Smile Degree

In 2014, T. George *et al.* developed an embedded model using a Raspberry Pi board. They extract features from mouth and eye pair using Haar-Cascade classifier. For classification KNN matching algorithm's used.

In 2014, I. K. Timotius *et al.* proposed a smile detection model using EOH for lips image-based smile detector. Their proposed model works only on lips region which differentiates smiling and non-smiling expressions. By dividing the lips image into 2×4 cells, and using 5^o histogram bin size, we obtained an 87.8% arithmetic means of efficiency.

In 2014, H. Liu *et al.* proposed a new approach for smile detection. Their proposed approach divided into three parts. In the first one, eyes-mouth alignment strategy is better than popular eyes alignment for the image registration procedure. The second one, for feature extraction they proposed Self- Similarity of Gradients (GSS) which gives better output compare to baselines methods. Last one, feature fusion, and multi-classifier fusion strategies are used in experiments and the outstanding outcome achieved.

In 2014, C. Chang *et al.* represented a low-complexity method for multilevel smile intensity analysis based on Mouth Corner Features (MCFs). In developed MCFs method mouth region is used and the value of enhanced grayscale pixel's along the horizontal axis is accumulated. To ahead measurement of mouth region, local-maximum information of the collected values is obtained to find the width and height of an opening mouth. At last, the normalized threshold technique is used to measure the smile intensity.

In 2015, Le An *et al.* developed a smile detection algorithm based on the Extreme Learning Machine (ELM). In this algorithm first, faces are detected and a holistic flow-based face registration is used which does not require any key point detection or manual labeling. Then Extreme Learning Machine is applied to train the classifier. The proposed smile detection model tested on many publicly available databases. ELM gives better performance compare to popular classifiers like SVM and LDA.

In 2015, K. Zhang *et al.* proposed a smile detection model in which high-level features are extracted with the help of well-designed Deep Convolution Neural Networks (CNN). The main contribution of their work is they use verification and recognition signals as a supervised algorithm to learn expression features, which is helpful for normalization and gives better performance in different real-world scenarios.

In 2015, JeffreyM.Girard *et al.* developed smile detection algorithm and proved theoretically and practically decision-value-as-intensity heuristic gives imprecise results.

In 2015, A. Tsai *et al.* proposed a method for a smile and frown expression for multiple people at a time. They used grayscale analysis to recognize multiple people expressions in real life.

In 2016, D. Cui *et al.* proposed a smile detection method based on PWD vector. This is the effective feature extraction method. PWD is an extended version of the EDV. Basically used for smile detection. These method extract feature on the bases of landmark around mouth as mouth area convey most of the detail, the face is a smiley or not. 68 facial points are used to extract feature this is further divided into 7 face areas. In the PWD, two different images of one person are considered a pair. Difference between neutral or non-smiling face and smiling face facial points are calculated. Most of smile expression conveyed through the mouth is shown in Figure 2.2.

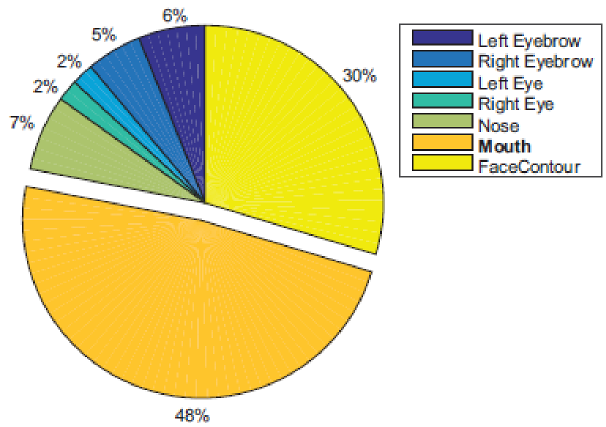

Fig. 2.2. 7 areas for smile face compare to natural face

There are different pre-processing methods are available like CFAN, SAN and many local SAN. These methods are used to detect facial landmarks. After pre-processing, 68 points are landmarked on the face, but 20 points are taken for smile detection training phase only for mouth area. After that PWD is used for feature extraction. Properties like rotation invariance, scaling invariance and translation invariance are included.

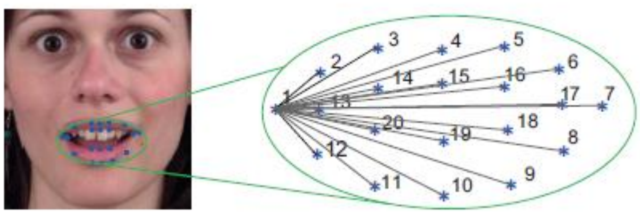

Fig. 2.3. 20 points around the mouth area

In 2016, J. Li *et al.* proposed an efficient smile detection method with a hierarchical visual feature. In their method feature extraction divided into three steps. In the first step, Gabor filters with multiscale, multi-orientation are first used to extract facial structure and called Gabor's faces. In the second step, HOG is applied to encode extracted Gabor faces to take and represents the facial appearance features. At last, convert the multiple HOG feature into a global visual feature with the help of max-pooling called Gabor-Hog. For classification, SVM is used.

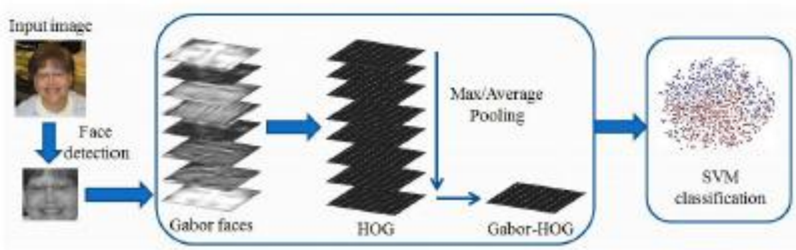

Fig. 2.4. Proposed model of J. Li et al. using SVM

In 2016, Gao, Y *et al.* developed a model for smile detection. Their main focus is on smile detection in unconstrained scenarios. They proposed a new descriptor called Self-Similarity of Gradients (GSS), which is inspired by Self-Similarity on Color (CSS) channels feature in the pedestrian detection area. There are many similar things between GSS and HOG feature map, while these similar things are useful and helpful for building a powerful smile detection system. They use a combination of HOG31, GSS and Raw pixel for feature extraction and use combination AdaBoost with linear Extreme Learning Machines (ELM) for classification, which gives outstanding results over the state of the arts in real-time.

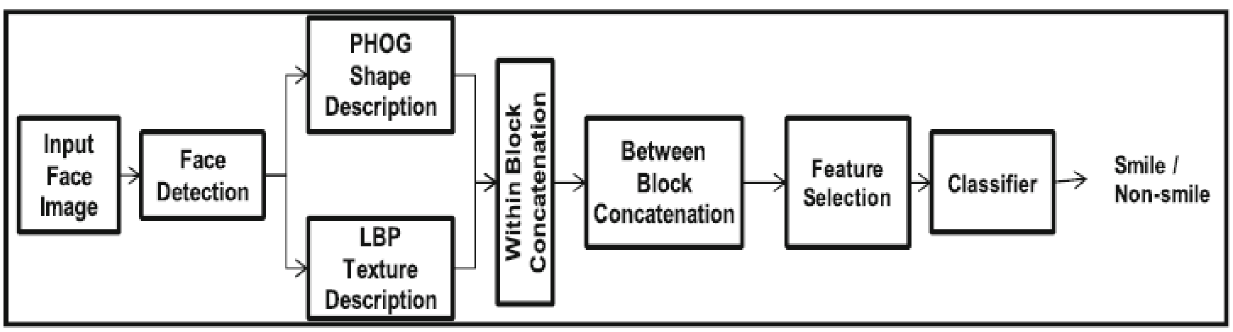

Fig. 2.5. Proposed model of Arigbabu

In 2016, Arigbabu, O. A. *et al.* proposed a smile detection algorithm which deals with unconstrained real-world scenarios. They divide feature extraction into two parts. First, a locally weighed multi-block shape-texture descriptor is proposed to extract local and global features from the face with different variations such as shapes, pose, different lighting and obstacle situation. The proposed algorithm merges the robustness of pyramid-histogram of oriented gradient and LBP for a facial feature showing using an adaptive local weight assignment. Second, the correlation-based feature choosing method is adopted to minimize the redundancy and extract the important facial feature. For classification kernel-based classifier is used such as SVM and KELM.

In 2016, Zhang, K. *et al.* developed a model based on deep learning which composed SNet and GNet for smile detection. They leverage the general to specific fine-tuning scheme and multitask learning to increase the performance of their model. Their model accomplishment person identification, smile detection and find the gender. They also proposed tasks aware face cropping method to extract feature for a specific area.

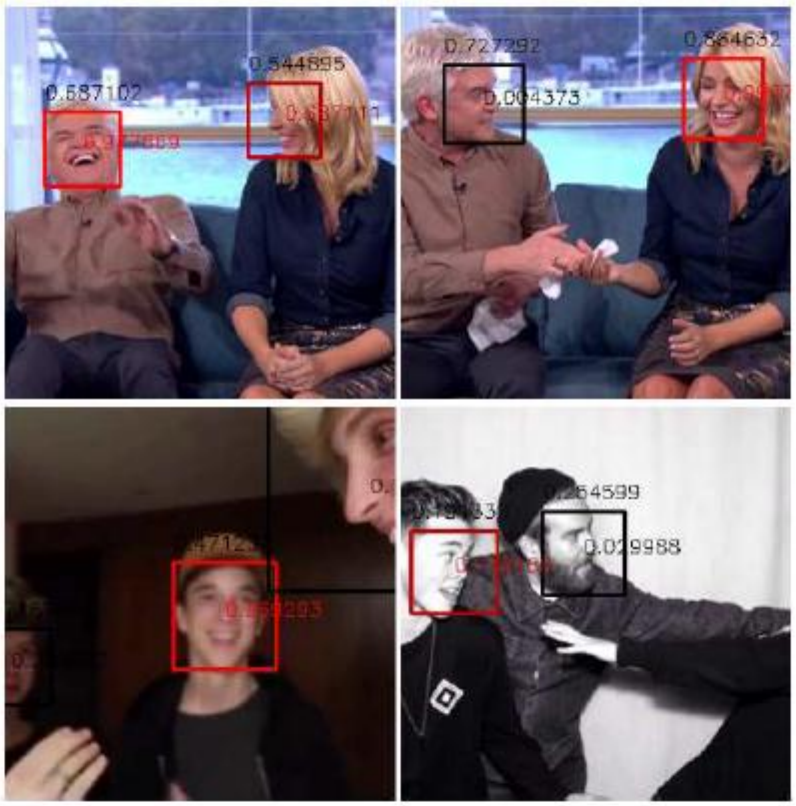

Fig. 2.6. Smile detection in wild area

In 2017, Jang, Y. *et al.* presented a new smile detection scheme called SmileNet for detecting face and smile in wild area. They use a Fully Convolution Neural Network (FCNN) to recognize many people smiling face in a given image of varying resolution. Their contributions are threefold. First, their model does not need preprocessing such as normalization and face detection. Second, the proposed model is ordinary and single FCNN structure cumulatively executing smile detection and face detection. Third, SmileNetestablishes 21.15 FPS real-time processing speed.

In 2017, Chen, J. *et al.* represented a powerful framework for smile detection in the wild-area based on deep learning. They apply the deep convolutional network to build a smile detection model called Smile-CNN to execute smile detection and feature learning simultaneously. For classification, they used Support Vector Machine (SVM) and AdaBoost.

In 2017, D. V. Sang *et al.* presented a powerful framework of Convolutional Neural Networks (CNNs) to recognize smile in unconstrained real-world scenarios with high accuracy. The proposed model is inspired by the VGG network called BKNet contains four convolutional blocks. All convolutional layer followed by Rectified Linear Unit (ReLU) activation function.

In 2017, R. Ranjan*et al.* developed an efficient model for multiple purposes such as age estimation, face detection, face alignment, face identification, gender recognition and smile detection using a single deep convolution neural network (CNN). Their method applies a multiple task learning scheme that distributes the shared parameters of CNN and builds synchronization among all domains and takes.

In 2018, P. Ghazi *et al.* proposed an asynchronous multithreading scheme for parallelizing the pipeline for smile detection called ImageNet. The Proposed model used binary classifier using the sigmoid activation function.

In 2018, Cui, D. *et al.* presented a new smile detection system in which a novel feature extraction technique used for feature extraction and ELM used for classification. They observe that the mouth region can effectively reflect a smile, a new and snappy group of information from a few of facial points around the mouth region are extracted.

In 2019, Liu, L. *et al.* in 2019 proposed a real-time smile detection framework based on conditional random regression forests. Since the connection between image patches and smile intensity is modeled limited to head posture, the proposed smile detection method is not sensitive to head posture. To obtain outstanding performance, methods including multi-label datasets augmentation, regression forest, and non-informative patch re-movement are employed.

In 2019, Vo, T. *et al.* developed a smile detection algorithm using a convolutional neural network (SD-CNN) to increase accuracy in the balanced data set, and then hybrid deep learning scheme (HF-SD) that uses an extended version of SD-CNN model to learn and then perform feature extraction from the dataset. For imbalanced data, extreme gradient boosting scheme is used for training.

2.2 Feature Extraction (FE) Methods

Feature extraction methods are used to identify unique properties from image or video. Detect features like ears, eyes, eyebrows, nose, mouth. There are many FE methods available. In this section, we discuss various feature extraction methods.

2.2.1 Local Mesh Patterns

Local Mesh Patterns (LMeP) proposed by Subrahmanyam Muralaet *al.* explores the relationship among the neighboring pixels in a local neighborhood. The calculation of LMeP is as shown in the equations below:

$$LMeP^{p,r}(a,b) = \sum_{i=1}^{p-1} 2^{i} \times f(x_{\omega} - x_{i}) \tag{2.1}$$

where,

$$\begin{cases} \omega = mod((i+p+k-1),p) \\ \quad \forall k = 1,2,\dots\dots,p/2 \end{cases} \tag{2.2}$$

$$f(x) = \begin{cases} 1 & if, x \geq 0 \\ 0 & otherwise \end{cases} \tag{2.3}$$

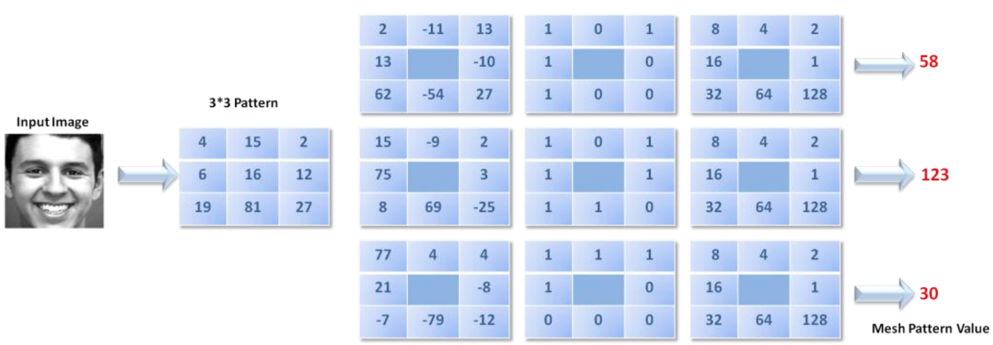

Fig. 2.7. Calculation of LMeP operator

Where k is the LMeP index, r is the radius of the local neighborhood, p is the number of neighbors at radius r and $f(x)$ can be calculated as given in Eq.(2.3).

For a local 3×3 neighborhood, the calculation of LMeP is as shown in Fig.2.7.

2.2.2 Local Directional Patterns

The facial features extracted by LBP are highly capable of conditions of illumination variation and random noises, etc. Local Directional Patterns (LDP) tries to overcome the issues of insufficient information for facial expression recognition as in the case of LBP and it is also less prone to conditions of varying illumination and random noise.

In order to obtain an LDP image, a convolution operation is performed using eight Kirsch masks $[M_0 \ldots . M_7]$ to obtain the edge response values in eight directions $[m_0 \ldots . m_7]$ for a reference pixel. The calculation of LDP is as shown in the equations below:

$$\underset{M_0}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}} \underset{M_1}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \underset{M_2}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \underset{M_3}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}}$$

$$\underset{M_4}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}} \underset{M_5}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \underset{M_6}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \underset{M_7}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}}$$

m_3	m_2	m_1
m_4	X	m_0
m_5	m_6	m_7

Fig. 2.8. Kirsch's Masks and eight directional edge response values [24]

$$LDp^k(a,b) = \sum_{i=1}^{p-1} 2^i \times f(m_i - m_k) \tag{2.4}$$

$f(x)$ can be calculated as given in Eq. (2.3)and m_k is k^{th} largest edge response value. The value of k is mostly chosen as 3.

For a local 3×3 neighborhood, the calculation of LDP is as shown in Fig.2.9.

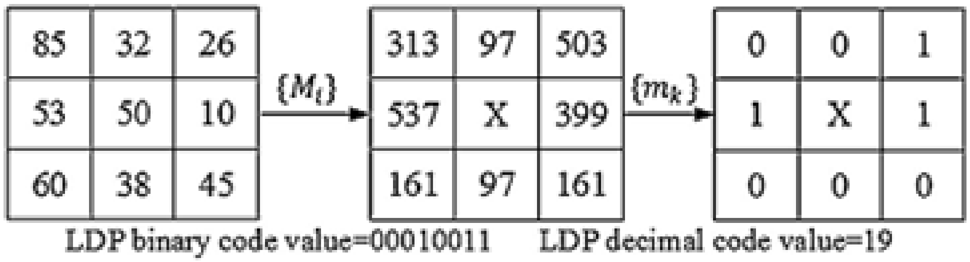

Fig. 2.9. Calculation of LDP

2.2.3 Local Directional Texture Patterns

Local Directional Texture pattern (LDTP) is an image descriptor which can be used for applications like scene recognition, facial expressions recognition, smile detection, etc. LDTP consists of a mixture of structural as well as contrast information which is extracted by analyzing the direction information and also for first and second maximum edge response values, the difference in intensity values is taken into account. Therefore, LDTP is quite robust in the conditions of noise and illumination variation and performs better than other competitive methods like LBP, LDP, LTP, etc.

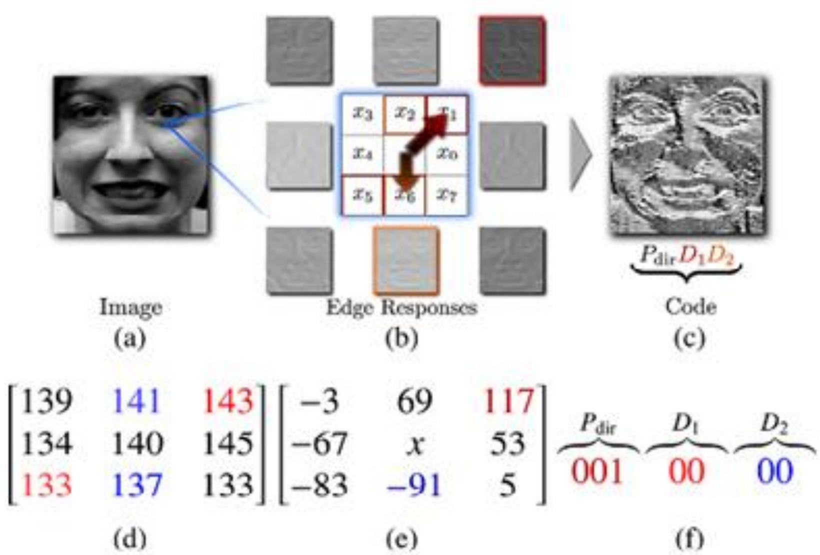

Fig. 2.10. Calculation of LDTP operator

To obtain an LDTP encoded feature image, the principal directions for each pixel are extracted within a local neighborhood. Then the difference in intensities in these directions is coded to generate the pattern. The eight principal direction numbers in a local neighborhood are calculated by performing a convolution operation with eight Kirsch masks $[\mathrm{M}_0 \ldots . \mathrm{M}_7]$ on each pixel.

The principal direction numbers can be generated from the equations below:

$$P_{dir}^{1} = argMAX(I_i | 0 \leq i \leq 7) \tag{2.5}$$

Where P_{dir}^{1} refers to the principal direction number and I_i is the absolute edge response obtained by convolution operation of the image with Kirsch's mask M_i and can be defined as follows:

$$I = |I_i \times M_i| \tag{2.6}$$

For LDTP, only maximum (P_{dir}^{1}) and second maximum edge response (P_{dir}^{2}) is taken into account and intensity differences of opposing pixels in the local neighborhood along these directions are calculated.

The final LDTP code is generated by concatenation of binary values of the first principal direction and the two different values.

$$LDTP(a,b) = P_{dir}^{1(a,b)} + 4D_f\left(d_1^{(a,b)}\right) + D_f\left(d_2^{(a,b)}\right) \tag{2.7}$$

Here, D_f is the encoded difference in intensities and is given by:

$$D_f(d) = \begin{cases} 0, & if, -\epsilon \leq d \leq \epsilon \\ 1, & if, d < -\epsilon \\ 2, & if, d > \epsilon \end{cases} \tag{2.8}$$

Where, d is the actual difference in the intensities, D_f is the encoded difference in the intensities and is the selected threshold value.

For a local 3×3 neighborhood, the calculation of LDTP is as shown in Fig.2.10.

2.2.4 Local Directional Number Patterns

LDN is a novel feature descriptor for face, facial expressions recognition, and smile detection. The directional information of face's texture is encoded using LDN in a compact fashion and thus the resultant code is much more discriminative as compared to the other methods. Compass masks are used to compute the micro-pattern structure which encodes the directional information using the main direction numbers and sign based information which helps in distinguishing between similar structures having different transitions in intensity.

In order to obtain an LDN image, a convolution operation is performed using eight Kirsch masks $[\mathrm{M}_0 \ldots . \mathrm{M}_7]$ to calculate the edge response values in eight directions $[m_0 \ldots . m_7]$ for a reference pixel. The calculation of LDP is as shown in the equations below:

$$LDN(a,b) = 8i_{a,b} + j_{a,b} \tag{2.9}$$

Where, (a,b) is the reference (central) pixel within a neighborhood which is being encoded, $i_{a,b}$ gives the direction number for maximum (positive) edge response and $j_{a,b}$ gives the direction number for minimum (negative) edge response, It can be defined as:

$$i_{a,b} = argMAX(I^i | 0 \le i \le 7) \tag{2.10}$$

$$j_{a,b} = argMAX(I^j | 0 \le i \le 7) \tag{2.11}$$

Where, I^i is absolute edge response is obtained by convolution operation of the image with i^{th} Kirsch's mask M^i and can be defined as follows:

$$I = |I^i \times M^i| \tag{2.12}$$

For a local 3×3 neighborhood, the calculation of LDN operator is as shown in Fig.2.6.

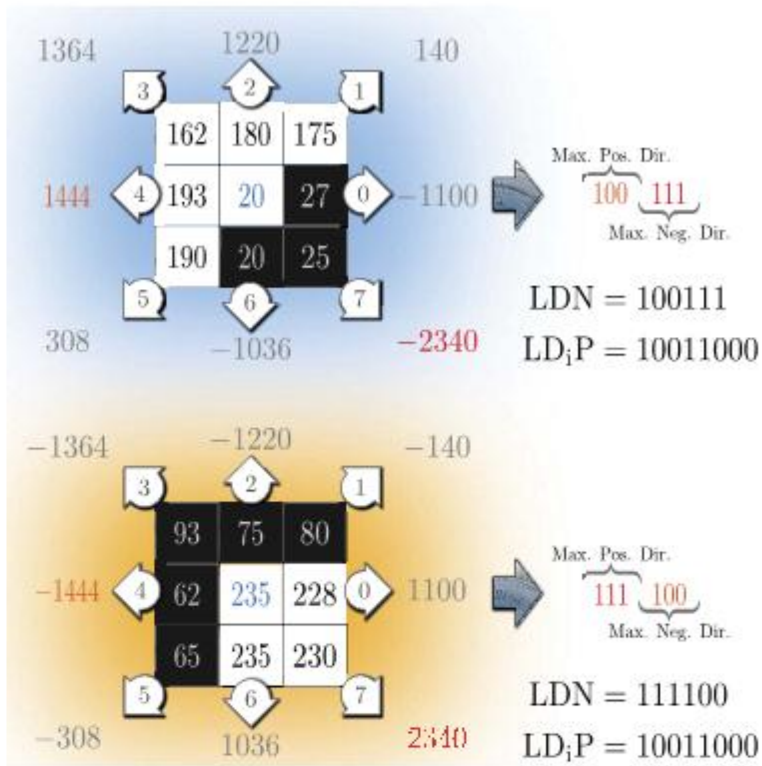

Fig. 2.11. Calculation of LDN operator

2.3 Classifier

After obtaining the feature vector for facial images, the next important task is the classification of the facial expressions. There are many classifiers are popular. Example:

2.3.1 Classification using Support Vector Machines (SVM)

SVM is a popular classification approach heavily used in pattern recognition problems. Experimental results have proven that SVM has much better performance accuracy as compared to other classifiers. Therefore, many researchers verify the effectiveness of their method they classify the obtained expression features using SVM. It is based on the concepts of Statistical Learning Theory and Structural Minimization Principle. It effectively obtains the optimal separating hyperplane which provides a maximum margin between positive and negative samples. For a set of given labelled training samples $(x_i, y_i), i = 1, \ldots\ldots, l$ where $x_i \in C_n$ and $y_i \in +1, 1$, a new test sample x can be classified by:

$$f(x) = sign\left(\sum_{1}^{1} \phi_i y_i Z(x_i, x) + \rho\right) \tag{2.13}$$

Where, $Z(a,b)$ is the kernel function, ϕ_i denotes the Lagrange's multipliers for a dual optimization problem and ρ is the threshold or bias parameter.

2.3.2 Classification using AdaBoost

AdaBoost is the first practical boosting algorithm proposed by Freund and Schapire in 1996. AdaBoost is a short form of Adaptive Boosting. It is mainly used for classification problems and objective of convert set of weak classifier into a strong one. The equation for AdaBoost classifier is:

$$F(x) = sign\left(\sum_{m=1}^{M} \phi_m f_m(x)\right) \tag{2.14}$$

Where, f_m is used for the m^{th} weak classifier and Θ_m is corresponding weight. Combination of M weak classifiers is exactly weighted.

2.3.3 Classification using Hidden Markov Model (HMM)

Hidden Markov Models (HMM) is a very popular classifier. HMM is applied in multiple areas of ML, such as face detection, facial expression recognition, smile detection, natural language processing, handwritten letter recognition, and speech recognition. HMMs are a set of statistical models applied to characterize the statistical properties of signals. HMM can be defined by the following parameters:

$$N = |S| \tag{2.15}$$

$$S = \{s_1, s_2, \dots, s_N\} \tag{2.16}$$

Where, N is the numbers of states in the model and S is a set of possible states, s_i. The state of the model at time t is given by

$$q_t \in S, \quad 1 \leq t \leq T \tag{2.17}$$

Where T is the length of the observation sequence.

$$M = |V| \tag{2.18}$$

$$V = \{v_1, v_2, \dots, v_M\} \tag{2.19}$$

Where M is a number of varied observation symbols and V is the set of all the possible observation symbols v_i. The observation symbol at time t is

$$o_t \in V, \quad 1 \leq t \leq T \tag{2.20}$$

HMM is define as:

$$\lambda = (A, B, \pi) \tag{2.21}$$

Where A is the state transition probability matrix, B is is the observation symbol probability matrix and π is the initial state distribution.

2.4 Facial Action Coding Unit

Recent works follow two different kinds of approaches. The first classifies the expressions into one of the six universal emotions as defined by Dr. Paul Ekman based on the prior definitions of emotion. However, these six emotions are not defined objectively and some interpretation is left to individuals as well. Therefore, it is not easy to classify expressions based on these six emotions. Also, many other expressions like pain, boredom, etc. exist which cannot be generalized based on just six emotions. Another way is to classify emotions in term of Action Units (AU) as defined in FACS (Facial Action Coding System). It's a form of objective measurement and the AUs can be classified as emotions or no-emotions. Researchers, therefore, focus on classifying expressions in terms of AUs as shown in Fig 2.1.

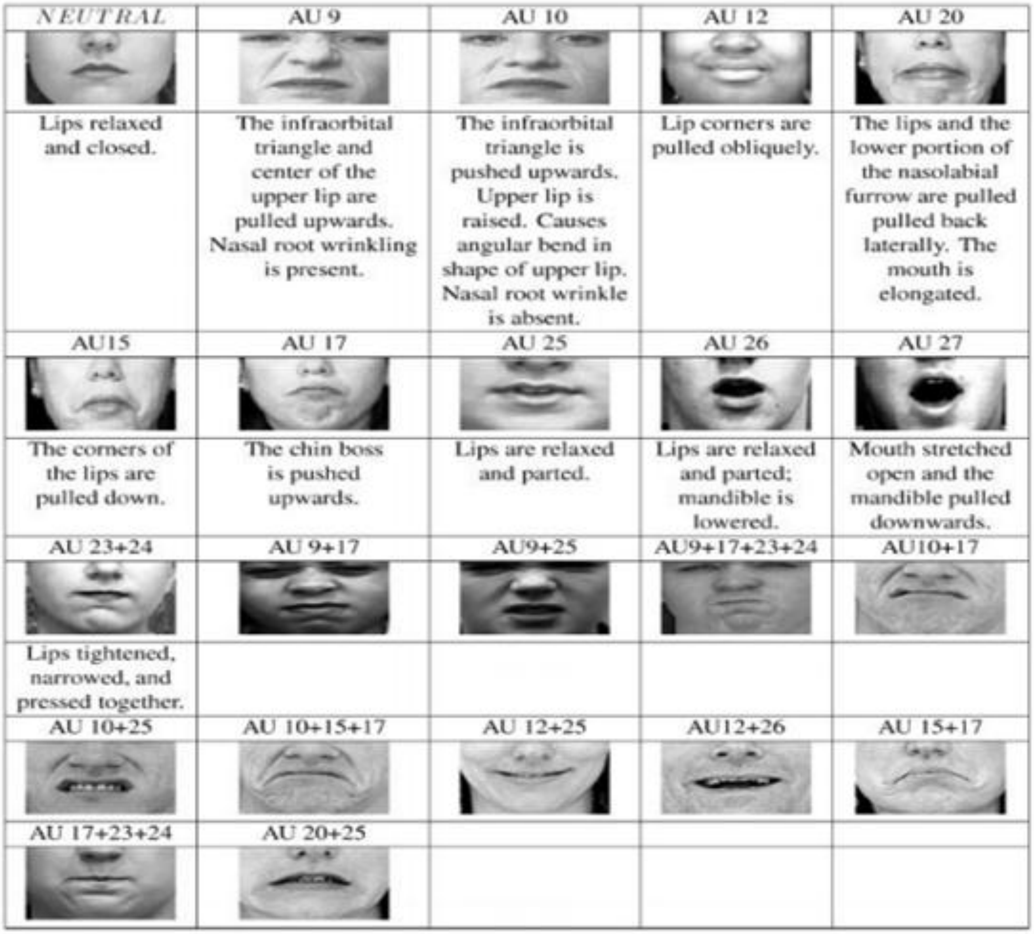

Fig.2.12. Lower face AUs and their combinations

FACS developed by Friesen and Ekman is an approach based on muscle movements. The facial muscles when collectively or individually cause change in facial behaviour which are identified using FACS. The facial variation and the variation in underlying muscles of the face which leads to these variations are called AU's. FACS comprises of many such AU's. For example:

- AU 1 represents a raising of the Inner Brow which is caused by the movement of the Part Medialis and Frontalis and muscles.
- AU 2 represents the raising of the Outer Brow which is caused by the movement of Part Lateralis and Frontalis muscles.

AUs are of two types-additive and non-additive. AUs are additive if each AU appears independently and the AUs are non-additive when they tend to modify appearances of one another. With the definition of AUs, the representation of smile is quite easy. Every facial expression is represented in the form of a combination of one or more AUs (additive/non-additive). For example, a combination of AU 12+25 and 15+26 represents ‘smile’.

2.5 Literature Summary

The summary of various feature extraction and classification methods proposed for smile detection systems is as given in Table 2.1.

Table 2.1: Recently Proposed Model

S. No.	Techniques	Data-base	Feature Extraction	Classifiers	Result /Accuracy
1.	C. Shan [27] in 2011	GENKI-4K	PixelComparisons	Ada-Boost	89.70 ± 0.45%
2.	Y. Zhang *et al.* [28] in 2012	GENKI-4K	Intensity Difference	AdaBoost	88%
3.	Yadappanavar, H *et al.*[29] in 2012	GENKI	Machine learning mythology	Haar Classifier	82.2%
4.	C. Shan *et al.* [30] in 2012	GENKI-4K	Pixel Differences	AdaBoost	85% (for 20 pair of pixel) and 88% (for 100 pair of pixel)
5.	Jain, V. *et al.* [31] in 2013	GENKI and Cohn-Kanade	Gabor Energy Filters	Support vector machine (SVM)	90.78% (for GEF) and 92.97% (for MGD with PCA)
6.	Nergui, M. *et al.* [32] in 2013	Random Datasets	Smile degree	Hidden Markov Model (HMM)	94.19%

7.	T. George *et al.* [33] in 2014	Random Datasets	KNN	Haar-cascade	66.6%
8.	I. K. Timotius*et al.* [34] in 2014	VISiO lab lip image dataset	Arithmetic means	EOH	87.8%
9.	H. Liu *et al.* [35] in 2014	GENKI-4K	Self- Similarity of Gradients (GSS) and Histogram of Oriented Gradients (HOG)	AdaBoost + SVM	95.13±0.95%
10.	C. Chang *et al*. [36] in 2014	Random Datasets	Mouth Region Segmentation	MCFs	87.5% (single-level smile measurement)and 80%(multilevel)
11.	Le An *et al.* [37] in 2015	MIX-database (FEI, Multi-PIE, CAS-PEAL, CK+) & GENKI-4K	The holistic flow-based face registration method	ELM	94.4% (MIX-database) and 88.2%(GENKI-4K)
12.	K. Zhang *et al.* [38] in 2015	GENKI-4K	Deep convolutional networks (CNN)	Softmax	94.6±0.29%

13.	Jeffrey M. Girard *et al.* [39] in 2015	BP4D	Gabor Wavelets or SIFT Descriptors and Laplacian Eigenmap or Principal Components Analysis (PCA)	Two-class SVM or Multiclass SVM or SVM Regression	83.36%
14.	A. Tsai *et al.* [40] in 2015	Random Dataset	Open CV	Simplified Mouth-Corner Features (MCFs)	72.7%
15.	D. Cui *et al.* [41] in 2016	GENKI-4K	Pair-wise Distance Vector	ELM	93.42±1.46%
16.	J. Li *et al.* [42] in 2016	GENKI-4K	Gabor filters with multi-scale + HOG	SVM	90.10±1.40% using Gabor-Hog (Max pooling) + SVM and 91.60% using Gabor-Hog (Avg. pooling) + SVM
17.	Gao, Y *et al.* [43] in 2016	GENKI-4K	HOG31 + GSS + Raw pixel	AdaBoost + Linear ELM	94.61±0.53%
18.	Arigbabu, O. A. *et al.* [44] in 2016	GENKI-4K	Robustness of pyramid histogram of oriented gradient + LBP	SVM and KELM	88.5% using SVM and 87.3% using KELM

19.	Zhang, K. *et al.* [45] in 2016	ChaLearn 16	VGG-Faces and fine-tune model	Two convolutional neural networks (CNNs) GNet and S Net.	88.79%
20.	Jang, Y. *et al.* [46] in 2017	GENKI-4K	Convolutional neural network (CNN)	Sigmoid	95.76±0.56%
21.	Chen, J. *et al.* [47] in 2017	GENKI-4K	Raw pixel and Learned features	AdaBoost and SVM	84±0.91% using Raw pixel and SVM, 80±0.76% using Raw pixel and Adaboost, 92.4±0.59% using Raw pixel and SVM and 91.8±0.95% using Learned features and Adaboost
22.	D. V. Sang *et al.* [48] in 2017	GENKI-4K	Convolutional neural network (CNN)	Softmax	95.08±0.29%
23.	R. Ranjan *et al.* [49] in 2017	Large-scale CelebFaces Attributes (CelebA) and ChaLearn Faces of the World	Multi-task Learning(CNN)	Convolutional neural networks (CNNs) GNet and SNet.	Celeb A: 93% and Faces of the World: 90.83%

24.	P. Ghazi *et al.* [50] in 2018	GENKI-4K, CelebFaces Attributes, and ChaLearn Looking At People (LAP) 2016	Convolutional neural network (CNN)	Sigmoid	GENKI-4K: 93.6% Celeb A: 88.5%
25.	Cui, D. *et al.* [51] in 2018	GENKI-4K	CFAN + Pair-wise Distance Vector	ELM	93.42±1.46%
26.	Liu, L. *et al.* [52] in 2019	Labeled Faces in the Wild (LFW) dataset and CCNU-Class datasets	Conditional random forest and	Regression forests	LFW – 95.80% and CCNU – 93.94%
27.	Vo, T. *et al.* [53] in 2019	GENKI-4K	Learned Feature and Raw pixel	Softmax and XGB	92.7% using Raw pixel + softmax and 93.6% using Learned Feature and XGB

2.6 Discussion and Summary

In this chapter a comprehensive survey on smile detection. From 2011 almost all smile detection methods are included along with challenges, future work, and their comparison. Some feature extraction and classification techniques are discussed in this chapter.

Chapter 3

Problem Statement

In this section, we need to specify the objective of our work based on the various research gaps and problem areas that were identified based on the literature.

3.1 Problem Definition

A happy face or smile is a common expression in our day to day life. This brings out basic emotions which are hidden emotions, like satisfaction and happiness. There are different applications in which Smile detection is used like automatic image capturing, patient monitoring, distance learning systems, video conferencing, interactive systems, and product rating. It is quite challenging and therefore has pulled a lot of attention because of its wide range of usage in different applications. But, despite the progress that has been made over the years, achieving high accuracy of recognition is still a difficult task due to the complexity, delicacy, and variability of smile.

Smile detection takes the flow of a classical pattern recognition problem which includes image acquisition, pre-processing, feature extraction, classification, and post-processing (Fig 1.2). Feature extraction plays a vital role in facial expression recognition. The feature descriptor used for the representation of features from given facial images determines the effectiveness of the feature extraction method. These feature descriptors should allow a high variation outside the classes (between different expressions or textures) while none or a very little variation outside the classes (same expression or texture under different conditions). It is very difficult to obtain a single suitable representation for an image especially under various conditions of input data variability like noise, illumination challenges, etc. which makes feature extraction very difficult. These challenges still pull the researchers' interest in various pattern recognition applications.

3.2 Research Gap

In the last decades, many techniques have been proposed for the analysis of smile detection with the help of numerous image processing, computer vision, and ML techniques. Researchers have utilized numerous ML and image processing techniques for developing such systems but one of its major drawbacks is that mostly the systems are designed and tested on the datasets which comprise images of people giving poses. These smiles are generally artificial. In Real-time environment, the systems which are trained with training features derived from such input images will have a shortcoming. Their accuracy decreases by a significant amount. These smiles aren't spontaneous and despite the fact that some of these artificial smiles may have some similarities to that of a natural spontaneous smile, it is very difficult for them to be similar to spontaneous ones. Therefore the accuracy reduces in real-time. Because of different reasons, it is very difficult to make a dataset with natural and spontaneous smile images so we should also look for alternative ways to deal with this problem.

The effectiveness of smile detection is largely dependent on the accurate representation of the Facial features as if inadequate features are used to represent the facial image then even the best of classifiers would fail to get a good recognition rate.

3.3 Dissertation Objective

To perform smile detection using data amalgamation based LBP and GFE that helps us in achieving optimal classification using SOM classifier which results for a smile and to overcome various challenges in smile detection. This goal will be achieved with the following objectives:

- To study the various feature extraction methods used and developed by the researchers to build a smile detection systems.
- To propose a new method for feature extraction that helps in achieving accurate recognition of smile.

Chapter 4

Smile Detection Using LBP and SOM

In this chapter, we have proposed a system model using LBP and SOM. We use mouth area for feature extraction as mouth area conveys most of the detail, the face is a smiley or not. Finally, we discuss the experimental result and performance of the proposed model for smile detection.

4.1 Database Description

Fig.4.1. **Smiling Face Images**

Fig.4.2. Natural Face Images

There are many datasets available for smile detection like BP4D, Chalearn 16, Cohn-Kanade, CCNU, FEI, GENKI-4K, Multi-PIE, LFW, etc. For our proposed method we are using GENKI-4K dataset. GENKI4K is very popular dataset widely used for smile detection in many research undertakings. It is a subset of the image used in. GENKI4K contains 4000 images in which 2162 images are smile face images and 1838 non-smile images. Some examples of smile face are shown in Fig. 4.1 and non-smile face are shown in Fig. 4.2. This dataset image span a wide range of age, color, ethnicity, facial hair, gender, glasses, indoors and outdoors, etc. This dataset gives us the ground truth on “smile” and “non-smile” images. There are two major challenges in smile detection. Firstly the oblique face and the second is the occlusion. For handling oblique face problem proposed technique extract feature only from the mouth region.

4.2 Proposed System Model

Fig. 4.3 summarizes our proposed model. This model includes the following steps: pre-processing, feature extraction and classification. In pre-processing, we detect the mouth region. After pre-processing feature extraction performs on mouth region using LBP. Finally, we perform classification using SOM classifier.

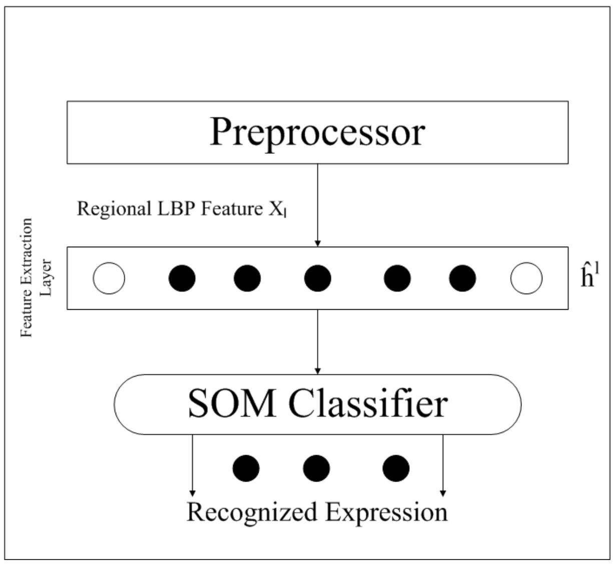

Fig 4.3. Proposed System Model

4.2.1 Pre-processing

For filter and optimized training, first of all, image is preprocessed to reduce noise. Pre-processing also includes image alignment, converting an image into grayscale and resizing it into predefined dimensions. With the help python library, I detect mouth. After detecting mouth we will perform further operations.

4.2.2 Local Binary Pattern (LBP)

Local Binary Patterns have strong texture discrimination capability and can quickly extract expression features as compared to Gabor wavelet transform. Local Binary Patterns are widely used in various pattern recognition problems like texture classification, facial recognition, expression recognition, and other areas.

The pixels are marked by the binary operator by differentiating the value of middle pixel and the 3×3 neighboring pixels. While valuing if the pixel is more prominent than center pixel then we write 1 else 0. These pixels follow circular and framed LBP code.

Now this 8-bit code (Byte) is converted into decimal number. Appropriate position of LBP is, that it figure out code generally to the center pixel, regardless of the light or darkness in the picture, and code will be same to a certain extent. So it is robust against illumination environment. For a local 3×3 neighborhood the calculation of LBP is as shown in Fig. 4.4.

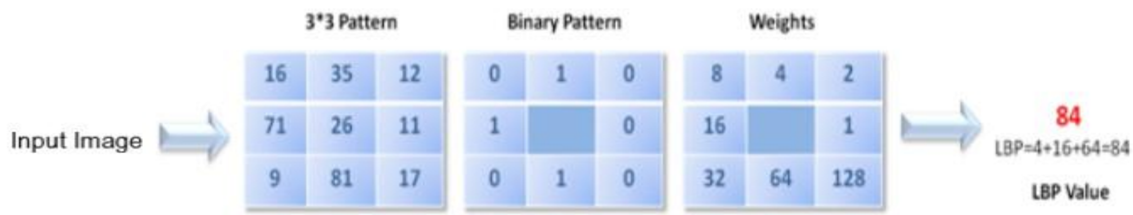

Fig.4.4. Calculation of LBP operator

4.2.3 Self-Organizing Map (SOM) Based Classifier

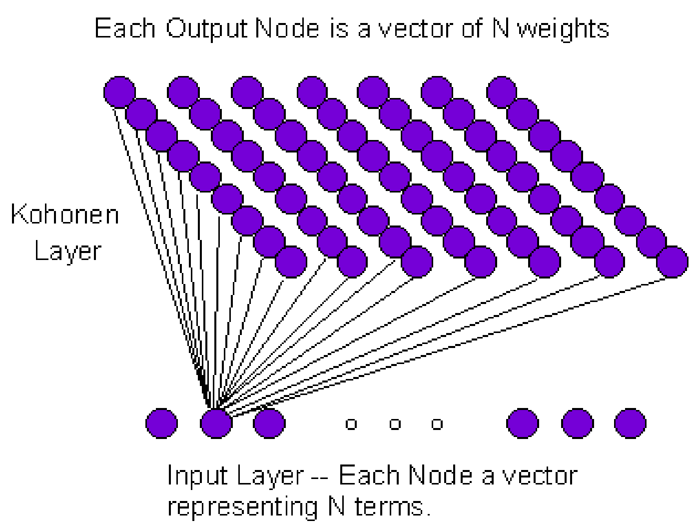

Fig. 4.5 SOM Classifier

SOM is part of a neural network. SOM algorithm is based on unsupervised learning. It is an extended version of the kohonen feature map. It uses for maps features from multilayer network and includes associate layer. SOM classification has two variant, the first is hard classification and the second is soft classification. For our proposed model we use soft classification.

Given LBP feature input x with respect to a given facial image, the output of the classifier y shows the expressions class, such as smile and non-smile. y is thus a 2-D vector. Smile is defined as $y = [1 - 1]^{\mathrm{T}}$ where the first term is ON while second term is OFF. Non-smile is represented in a similar manner. The input-output mapping from the facial LBP feature space to smiling and non-smiling expression categories. This mapping is highly nonlinear. With the help of first-order Taylor series expansion around every neuron, we can convert into linear mapping. In this technique, the overall response model is expressed using locally valid linear models linked with each neuron γ. This technique gives two benefit – fist one is fast learning and the second one is better accuracy. The proposed model assures that smiling or non-smiling one of output node will be ON and others will be OFF. It has been practically studied that parameter updates based on the minimization of least square error provide a better result.

4.3 Performance Measures

In order to demonstrate the effectiveness of our proposed methods, we use certain performance measures. The performance of smile detection is correct rate or the accuracy of classification which is by the following relation:

$$\text{Correct Rate} = \frac{Number\ of\ Correctly\ Classified\ Samples}{Total\ Number\ of\ Classified\ Samples} \tag{4.1}$$

4.4 Result and Performance

In order to demonstrate the effectiveness of the proposed model for smile detection, the experiments were conducted on GENKI-4K dataset. The proposed smile detection method in section 4.2 using LBP and SOM classifier. The proposed method gives an outstanding outcome. SOM learning rate initiate from $\eta_{\mathrm{I}} = 0.005$ and end at $\eta_F = 0.9$. The size of the SOM-based network is $10\ X\ 8$. To validate the proposed method, we will compare with different methods which are mention in Table 4.1.

Table.4.1. Comparison of the proposed method to related work

Technique	Feature Extraction	Classification	Accuracy (%)
Kahou et al.[57]	LBP	SVM	93.2±0.92
L. An el al. [58]	LBP	ELM	94.2
L. An el al. [58]	LBP	LDA	92
Proposed Method	LBP	SOM	94. 3±0.22

4.5 Discussion and Summary

In this chapter, we propose a smile detection scheme using LBP and SOM, which will give result in the form of accuracy. As compared to existing techniques simulated result is better in the term of accuracy.

In the next chapter, smile detection using data amalgamation is discussed which helps to obtain a more precise result**.**

Chapter 5

Smile Detection Using Data Amalgamation

In this chapter, we have proposed a system model for smile detection using data amalgamation. The proposed method has two consecutive actions: 1) Amalgamation of Geometric Feature Extraction (GFE) and regional Local Binary Pattern (LBP) features extraction using autoencoders; 2) Self-Organizing Map (KSOM) is used to classify smile based on these features. The proposed method is mathematics more dynamic and performance-wise more precise. We use mouth region for feature extraction as mouth area conveys most of the detail, the face is a smiley or not. In this chapter, we will discuss used dataset for our work and proposed model in details. Finally, we discuss the experiment result and performance of the proposed model for smile detection.

5.1 Database Description

Fig.5.1. Smiling Face Images

Fig.5.2. Natural Face Images

For our proposed method we are using GENKI-4K dataset. GENKI4K is very popular dataset widely used for smile detection in many research undertakings. It is a subset of the image used in. GENKI4K contains 4000 images in which 2162 images are smile face images and 1838 non-smile images are shown in Fig. 5.1 and non-smile face are shown in Fig. 5.2.

5.2 Proposed System Model

Fig. 5.3 summarizes our proposed model. The propounded model includes the following methodical steps: Pre-processing, GFE, LBP, data amalgamation using autoencoder, and KSOM-based classifier. In pre-processing, we detect the mouth region. After pre-processing feature extraction performs on mouth region using GFE, LBP, and autoencoder. Finally, we perform classification using SOM classifier.

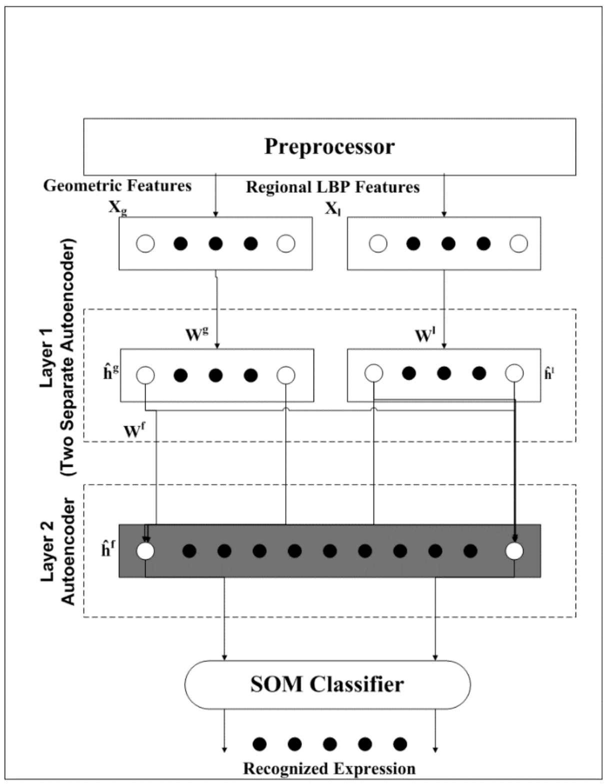

Fig.5.3. Proposed System Model

5.2.1 Pre-processing

Facial images obtained from the video or image sources are generally prone to noise. Thus in order to get complete facial images with normalized intensity, shape and size, pre-processing is applied on faces detected in the first step. This is done because the representation of the expressions is highly sensitive to scaling, translation, and rotation of the head in a facial image. Therefore to suppress the damage caused by these unwanted transformations, a geometrical standardization of facial image is done before classification. Detection of feature points, rotation of the head and locating and cropping of the facial region i.e. segmentation using a bounding box according to the face model et al. are the steps involved in the pre-processing of the faces in order to normalize them for feature extraction. Segmentation usually involves the division of the facial regions and is performed on the basis of shape, color, motion, texture, etc. of the face.

5.2.2 Geometric Feature Extraction

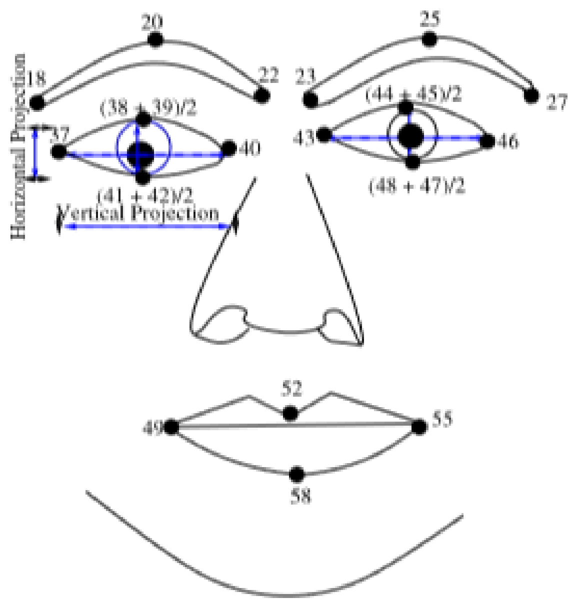

Fig. 5.4. Geometric Feature Extraction (GFE)

A geometric feature input set $x_g \in \mathbb{R}^{22}$ comprising of eye interpolation proportions and directional uprooting data of facial points is utilized in Fig. 4.4 demonstrates an identifying the regions of eyes, eyebrow, nose and lips regions. utilizing facial geometric data. There are few steps for geometric features extraction for an input image:

- Identify face, Get estimate face height H_{face}.
- Identify both the eyes.
- Find eyes' centers coordinates $(x_1, y_1)$and (x_2, y_2). Measure the eyes' area which additionally incorporates eyebrow areas. Obtain the eye area estimate height.
- With the help of the eye’s center point and face height, we obtain the mouth and nose areas.

In subsequent frames, the facial areas are highlighted and the characteristics are taken out from each identified location. When the four major areas are identified, the lips, nostrils, eyes, and eyebrows are fragmented. After fragmenting the area, the characteristic is taken out from four key regions. The features are mined by taking the proportion of two-dimensional projection of the fragmented area. The contrasts between (x, y) two-dimensional coordinates of marked points on the standard outline and progressive outline are determined for eyebrows and lips identifying features.

5.2.3 Local Binary Pattern (LBP) Feature Extraction

LBP operator was proposed in 1996 by *Ojala et al* as an important binary operator. It plays a powerful role in texture classifier. LBP is a fundamental tool for recognition of feature. It generally used in the aspect of its strength and simplicity.

The pixels of the image are labeled by the binary operator by thresholding the central pixel esteem and 3×3 neighborhood pixels. Then the scattered labels can be used as texture indicators. The operator explains specific pixel by the relative gray levels of its surrounding pixels, see Fig. 4.5 for an image with corresponding LBP image. If the gray level of the surrounding pixels is equal or higher than a central pixel, the value is assigned to 1, otherwise to 0. At that point, these pixels followed circularly and framed LBP code. Grayscale of a given image is represented by $I(x_p, y_p)$, with intensity value $g_c = I(x, y)$ at the point x and y.

$$g_p = I\left(x_p, y_p\right), \quad p = 0, \ldots.., P-1 \tag{5.1}$$

Where x_p and y_p are the correlative value of sampling points.

$$S\left(g_p - g_c\right) = \begin{cases} 1, & g_p \geq g_c \\ 0, & otherwise \end{cases} \tag{5.2}$$

Then feature extract using $LBP_{P,R}(x, y)$

$$LBP_{P,R}(x, y) = \sum_{P=0}^{N-1} S\left(g_p - g_c\right) 2^p \tag{5.3}$$

Where P is sampling point and R is a radius surrounding the central pixel.

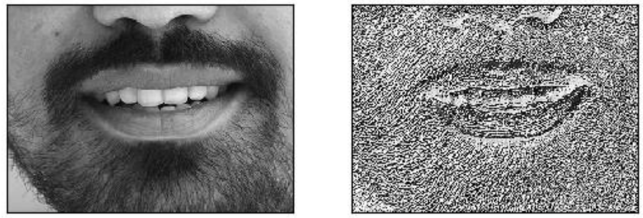

Fig.5.5. LBP image of the mouth region

5.2.4 Autoencoder

An autoencoder is an algorithm based on unsupervised schooling which uses BPN technique to determine the input feature vector. The primary aim of the encoder is to cognize an improved presentation in the compressed form of feature. For an assigned input vector $x = x_1 \; and \; x_2$, the encoder trains function and give network output h(x) for given input vector x. The middle layer represents using ĥ(x). Middle layer also is known as the hidden layer.

$$LBP_{P,R}(x,y) = \sum_{P=0}^{N-1} S\big(g_p - g_c\big)2^p \tag{5.4}$$

Where $f(.)$ is an activation function (Sigmoid function). Using decoding function we obtain reconstructed output$\hat{x}$.

$$\hat{x} = g(\hat{a}_k) \quad where \quad \hat{a}_k = c_k + \sum_j W^*_{jk}\,\hat{h}_j(x) \tag{5.5}$$

Where $g(.)$ Are activation function and assign value one to $b_j \; and \; c_k$.

In dissertation two isolated encoders are used for both extraction techniques. The output of both isolated encoders is combined. The combined output is used as an input for next layer encoder. The second layer represents the amalgamated features. In the initial step GF $x_g \in \mathbb{R}^{22}$ and LBP feature $x_l \in \mathbb{R}^{236}$ are individually passed using two distinct autoencoders. In the second step combining the output of the hidden layer of two autoencoders and passed to another autoencoder.

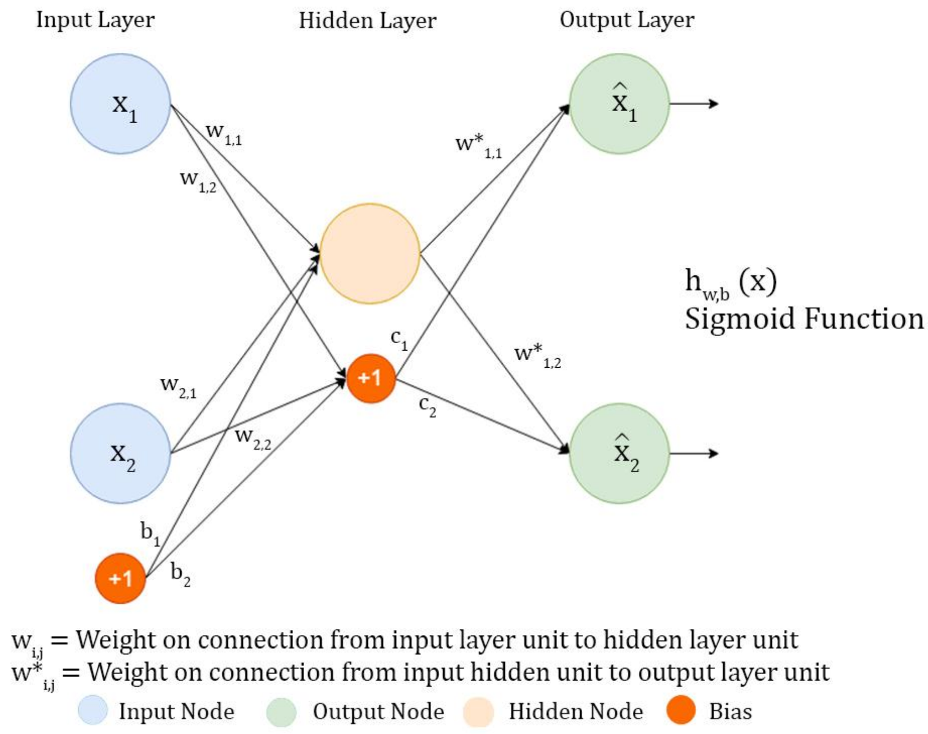

Fig.5.6. Structure of an Autoencoder

5.2.5 Self-Organizing Map (SOM) Based Classifier

Self-Organizing Map is a part of a neural network. It is based on the unsupervised network. It is an extended version of kohonen feature map. It uses for maps features from multilayer network and includes associate layer. SOM classification has two variant, the first one is hard classification and the second one is soft classification. For our proposed model we use soft classification. Fig.5.7. The emotional feature space is discretized in terms of Kohonen 2-D lattice of size $M \times N$.

Given an amalgamated feature input$x$ corresponding to a given facial image, the output of the classifier $y$shows the expressions class, such as smile and non-smile. y is thus a 2-D vector. Smile is defined as $y = [1 - 1]^{\mathrm{T}}$ where the first term is ON while the second term is OFF. Non-smile is represented in a similar manner. The amalgamated feature vector as an extract from the autoencoder is a 230-D vector, i.e., $x \in \mathbb{R}^{230}$. The input-output mapping from the facial amalgamated feature space to smiling and non-smiling expression categories can be mathematically represented as

$$y = f(x), \;\; y \in \mathbb{R}^2; \;\; x \in \mathbb{R}^{230} \tag{5.6}$$

This mapping is highly nonlinear. With the help of first-order Taylor series expansion around every neuron, we can convert into linear mapping. Given that each γ^{th} neuron is linked with a vector $w^{\gamma} \in \mathbb{R}^{230}$ in the lattice, the linear model linked with this neuron can be represented as

$$y_{\gamma}^{out} = y_{\gamma} + A\left(x - w_{\gamma}\right)$$

$$\text{where, } A_{\gamma} = \frac{\partial f}{\partial x} | x = w_{\gamma} \in \mathbb{R}^{2 \times 230}$$

$$y = f\left(w_{\gamma}\right) \; (ideally) \tag{5.7}$$

In this classifier, neighbors partake in decision making even if winning neuron has the bigger say. The overall response of the network given an input x is given as

$$y^{out} = \frac{\sum_{\gamma=1}^{M \times N} h_{i,\gamma} y_{\gamma}^{out}}{\sum_{\gamma=1}^{M \times N} h_{i,\gamma}} \tag{5.8}$$

where $h_{i,\gamma}$ is the neighborhood function for the best matching unit (BMU) i and the neuron γ. The BMU i is obtained by

$$i = \arg\min_{\gamma} \parallel x - w_{\gamma} \parallel \tag{5.9}$$

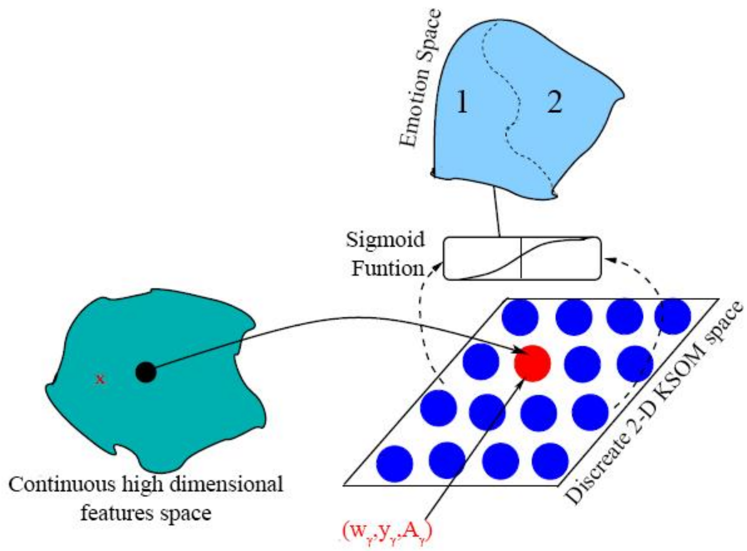

Fig. 5.7. SOM Classifier

In this technique, the overall response model (using equation 5.8) is expressed using locally valid linear models linked with each neuron γ. This technique gives two benefit – fist one is fast learning and the second one is better output. The learning problem is to update parameters $w_\gamma, y_\gamma, A_\gamma$ given training pairs $\{x, y^d\}$. A sigmoid activation function is applied at each node of the network to classify the input feature vector into smiling and non-smiling expression. Let the sigmoid function output of thek^{th} class is denoted by y_k^{sig}.

$$y_k^{sig} = \frac{e^{y_k^{out}} - e^{-y_k^{out}}}{e^{y_k^{out}} + e^{-y_k^{out}}} \tag{5.10}$$

$$\begin{gathered} if \quad y_k^{sig} = G_r, then\; y_k = 1 \\ else \quad y_k = -1 \end{gathered} \tag{5.11}$$

Where, G_r is greatest value collective sigmoid function.

The proposed model is shown in equation 5.11, the model assures that smiling or non-smiling one of output node will be ON and others will be OFF. It has been practically studied that parameter updates based on the minimization of least square error provide better result. Let least square error function is:

$$E = \sum_{k=1}^{2} E_k = \frac{1}{2}\sum_{k=1}^{2}\left(y_k^d - y_k^{sig}\right)^2 \tag{5.12}$$

For update parameter A_γ and y_γ with respect to the parameters at each output node k are given as:

$$\frac{\partial E}{\partial y_\gamma} = -\frac{\left(y_k^d - y_k^{sig}\right)h_{i,\gamma}}{\left(\sum_{\gamma=1}^{M \times N} h_{i,\gamma}\right)}\left(1 - \left(y_k^{sig}\right)^2\right) \tag{5.13}$$

$$\frac{\partial E}{\partial a_\gamma} = -\frac{\left(y_k^d - y_k^{sig}\right)h_{i,\gamma}(x - w_\gamma)}{\left(\sum_{\gamma=1}^{M \times N} h_{i,\gamma}\right)}\left(1 - \left(y_k^{sig}\right)^2\right) \tag{5.14}$$

5.3 Performance Measures

In order to demonstrate the effectiveness of our proposed methods, we use certain performance measures. The performance of smile detection is correct rate or the accuracy of classification which is by the following relation:

$$\text{Correct Rate} = \frac{Number\ of\ Correctly\ Classified\ Samples}{Total\ Number\ of\ Classified\ Samples} \tag{5.15}$$

5.4 Result and Performance

In order to demonstrate the effectiveness of the proposed model for smile detection, the experiments were conducted on GENKI-4K datasets. The proposed data amalgamation method in section 5.2 is applied to obtain the amalgamated feature vector. In the first layer,20 hidden units are used for GF vector and 170 hidden units for LBP vector. For the next autoencoder layer,230 hidden units are used. At each layer selection of hidden units based on hyperparameter optimization. The proposed method gives an outstanding outcome on the above-mentioned values. A SOM learning rate initiate from $\eta_{\mathrm{I}} = 0.005$ and end at $\eta_F = 0.9$. The size of the SOM-based network is $10\ X\ 8$. To validate the proposed method, we will compare with different methods which are mention in Table 5.1.

Table.5.1. Comparison of the proposed method to related work on GENKI-4K

Technique	Feature Extraction	Classification	Accuracy (%)
Arigbabu et al. [44]	Weighted (PHOG + LBP) + $mRMR_{spearman}$	KELM	88.5
Liu, Mengyi, et al. [59]	HOG (labelled + unlabeled)	SVM	92.3±0.81
M. Nergui et al.[32]	Pair-wise Distance Vector	ELM	93.42±1.46
Jain, Anil K. et al.[60]	Smile Degree	HMM	94.19
H. Liu et al. [35]	GSS + HOG	AdaBoost + SVM	95.13±0.95
Y. Jang et al. [46]	CNN	Sigmoid	95.76±0.56
R. Ranjan et al. [49]	CNN	Softmax	95.08±0.29
Proposed Method	GFE + LBP	SOM	98. 89±0.46

5.5 Discussion and Summary

In this chapter, we proposed data amalgamation method is applied to obtain the amalgamated feature vector. We use LBP and GF extraction for feature extraction. We use autoencoder for data amalgamation.

In the next chapter, the research outcomes of the dissertation are concluded.

Chapter 6

Conclusion

Smile detection attracts a swarm of researchers in the present scenario, but the smile detection research community has transferred its aim to the detection of natural smile expression. The main focus of this dissertation is to obtain better accuracy in smile detection. To fulfill the requirements of the proposed objectives, two new solutions have been proposed and implemented. One solution is used LBP and SOM which gives better accuracy compare to many methods. The second one, the proposed method has represented a smile detection system, where the first and second layers of auto-encoder successfully amalgamate geometric and appearance-based features for a more accurate representation of mouth region data. GFE and LBP have been used for data amalgamation. In the third layer of the proposed approach, SOM-based soft classifier has been used which gives the benefit of supervised and unsupervised learning techniques. The experiments have been performed on GENKI4K dataset and it has been shown that the results of amalgamated data perform better than earlier proposed techniques. The proposed smile detection system proves to be effective with high accuracy.

Still, there is a lot of scope of improving the efficiency of the presently used methodology. The proposed technique can be ameliorated by improving feature extraction phase, for example, replacing GEF and LBP with some other improved feature extraction techniques. Future expansion of smile detection perusal could be the research on slant faces along with micro-expressions. Currently, the techniques which exist to deal with slant faces along with micro-expressions are myopic in their approach. Hence, a large amount of work needs to be pursued in order to improve upon these studies. Another key challenge is capturing the smile of a person, who has lost their spontaneous expressions because of the medical and accidental challenges. Because of these 12 diseases, people lose their spontaneous expressions: Asperger Syndrome, Autistic Disorder, Bell's Palsy, Depression, Depressive Disorders, Facial Paralysis, Facial weakness, Hepatolenticular Degeneration, Major Depressive Disorder, Parkinson’s Disease, Scleroderma and Wilson’s Disease. Finally, both the challenges and opportunities in this field of work are limitless but the uses tend to be so vast that smile detection has not only survived but has flourished through decades.

References

[1] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," in *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 1-12, 1 Jan.-March 2015.

[2] Mehrabian, Nonverbal Communication. New Brunswick, NJ, USA: Aldine, 2007.

[3] S. Z. Li, A. K. Jain, Y. L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," *in Handbook of Face Recognition. New York: Springer-Verlag*, pp. 247-275, 2005.

[4] X. Xie and K.-M. Lam, "Facial expression recognition based on shape and texture," *Pattern Recognition*, vol. 42, no. 5, pp. 1003–1011, 2009.

[5] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, Oct. 1999.

[6] Hai Hong, H. Neven, and C. von der Malsburg, "Online facial expression recognition based on personalized galleries," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, 1998, pp. 354-359.

[7] P. Ekman and W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *Palo Alto, CA: Consulting Psychologists Press*, 1978.

[8] P. Ekman, "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. New York," *Times Books*, 2003.

[9] P. Kakumanu and N. Bourbakis, "A Local-Global Graph Approach for Facial Expression Recognition," *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*, Arlington, VA, 2006, pp. 685-692.

[10] C. Shan, S. Gong, and P. W. McCowan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[11] J. Yang, D. Zhang, A. Frangi, and J.-Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[12] I. Kotsia and I. Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines," in *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172-187, Jan. 2007.

[13] N. Bourbakis and P. Kakumanu, “Skin-based Face Detection-Extraction and Recognition of Facial Expressions,” *Applied Pattern Recognition Studies in Computational Intelligence*, pp. 3–27, 2008.

[14] N. Bourbakis, A. Esposito, and D. Kavraki, “Extracting and Associating Meta-features for Understanding People’s Emotional Behaviour: Face and Speech,” *Cognitive Computation*, vol. 3, no. 3, pp. 436–448, 2010.

[15] A. Cheddar, D. Mohamad, and A. A. Manaf, “Exploiting Voronoi diagram properties in face segmentation and feature extraction,” *Pattern Recognition*, vol. 41, no. 12, pp. 3842–3859, 2008.

[16] Y.-L. Tian, “Evaluation of Face Resolution for Expression Analysis,” *2004 Conference on Computer Vision and Pattern Recognition Workshop*.

[17] M. Pantie and L. Rothkrantz, “Automatic analysis of facial expressions: the state of the art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[18] M. Turk and A. Pentland, “Eigenfaces for recognition,” *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[19] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[20] K. Etemad and R. Chellappa, “Discriminant analysis for recognition of human face images,” *J. Opt. Soc. Amer. A*, vol. 14, no. 8, pp. 1724–1733, 1997.

[21] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[22] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Jianzhuang Liu, "Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor," in *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533-544, Feb. 2010.

[23] A. R. Rivera, J. R. Castillo, and O. Chae, “Local Directional Texture Pattern image descriptor,” *Pattern Recognition Letters*, vol. 51, pp. 94–100, 2015.

[24] Yuan Luo, Tian Zhang, Yi Zhang. “A novel fusion method of PCA and LDP for facial expression feature extraction.” *Optik, Elsevier*, 2014.

[25] S. Murala and Q. M. J. Wu, "Local Mesh Patterns Versus Local Binary Patterns: Biomedical Image Indexing and Retrieval," in *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 929-938, May 2014.

[26] A. Ramirez Rivera, J. Rojas Castillo, and O. OksamChae, "Local Directional Number Pattern for Face Analysis: Face and Expression Recognition," in *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1740-1752, May 2013.

[27] C. Shan, "An efficient approach to smile detection," *Face and Gesture 2011*, Santa Barbara, CA, 2011, pp. 759-764.

[28] Y. Zhang, L. Zhou, and T. Sun, "A Novel Approach to Detect Smile Expression," *2012 11th International Conference on Machine Learning and Applications*, Boca Raton, FL, 2012, pp. 482-487.

[29] H. Yadappanavar and S. S. S, “Machine Learning Approach for Smile Detection in Real Time Images,” *International Journal of Image Processing and Vision Sciences (IJIPVS)*, vol. 1, no. 1, 2012.

[30] C. Shan, "Smile detection by boosting pixel differences," *in IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 431-436, Jan. 2012.

[31] V. Jain and James L, “Smile Detection Using Multi-scale Gaussian Derivatives,” *12th WSEAS International Conference on Signal Processing, Robotics and Automation*, 2013.

[32] M. Nergui, L. Pai-hui, G. Nagamatsu, M. Waki, and M. Otake, “Automatic Detection of Laughter using Respiratory Sensor Data with Smile Degree,” *Proceedings of the International Conference on Advances in Computing, Communications, and Informatics*, vol. 1, no. 1, pp. 8–11, 2013.

[33] T. George, S. P. Potty and S. Jose, "Smile detection from still images using KNN algorithm," *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, Kanyakumari, 2014, pp. 461-465.

[34] I. K. Timotius and I. Setyawan, "Evaluation of Edge Orientation Histograms in smile detection," *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, 2014, pp. 1-5.

[35] H. Liu, Y. Gao and P. Wu, "Smile detection in unconstrained scenarios using self-similarity of gradients features," *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, 2014, pp. 1455-1459.

[36] C. Chang *et al*., "Multi-level smile intensity measuring based on mouth-corner features for happiness detection," *2014 International Conference on Orange Technologies*, Xian, 2014, pp. 181-184.

[37] L. An, S. Yang, and B. Bhanu, "Efficient smile detection by Extreme Learning Machine," *Neurocomputing*, vol. 149, pp. 354–363, 2015.

[38] K. Zhang, Y. Huang, H. Wu, and L. Wang, "Facial smile detection based on deep learning features," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, 2015, pp. 534-538.

[39] J. M. Girard, J. F. Cohn, and F. D. L. Torre, "Estimating smile intensity: A better way," *Pattern Recognition Letters*, vol. 66, pp. 13–21, 2015.

[40] A. Tsai, T. Lin, T. Kuan, K. Bharanitharan, J. Chang and J. Wang, "An efficient smile and frown detection algorithm," 2015 *International Conference on Orange Technologies (ICOT)*, Hong Kong, 2015, pp. 139-143.

[41] D. Cui, G. Huang, and T. Liu, "Smile detection using Pair-wise Distance Vector and Extreme Learning Machine," *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, 2016, pp. 2298-2305.

[42] J. Li, J. Chen, and Z. Chi, "Smile detection in the wild with hierarchical visual feature," *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, 2016, pp. 639-643.

[43] Y. Gao, H. Liu, P. Wu, and C. Wang, "A new descriptor of gradients Self-Similarity for smile detection in unconstrained scenarios," *Neurocomputing*, vol. 174, pp. 1077–1086, 2016.

[44] Arigbabu, O. A., Mahmood, S., Ahmad, S. M. S., & Arigbabu, A. A., “Smile detection using hybrid face representation,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 3, pp. 415–426, 2016.

[45] K. Zhang, L. Tan, Z. Li, and Y. Qiao, “Gender and Smile Classification Using Deep Convolutional Neural Networks*,” 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2016.

[46] Y. Jang, H. Gunes, and I. Patras, “SmileNet: Registration-Free Smiling Face Detection In The Wild,” 2017 *IEEE International Conference on Computer Vision Workshops (ICCVW),* pp. 1581–1589, 2017.

[47] J. Chen, Q. Ou, Z. Chi, and H. Fu, “Smile detection in the wild with deep convolutional neural networks,” *Machine vision and applications,* vol. 28, no. 1-2, pp. 173–783, 2017.

[48] D. V. Sang, L. T. B. Cuong, and D. P. Thuan, "Facial smile detection using convolutional neural networks," *2017 9th International Conference on Knowledge and Systems Engineering (KSE),* Hue, 2017, pp. 136-141.

[49] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, "An All-In-One Convolutional Neural Network for Face Analysis," *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017),* Washington, DC, 2017, pp. 17-24.

[50] P. Ghazi, A. P. Happonen, J. Boutellier, and H. Huttunen, “Embedded Implementation of a Deep Learning Smile Detector,” *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018.

[51] D. Cui, G.-B. Huang, and T. Liu, “ELM based smile detection using Distance Vector,” *Pattern Recognition*, vol. 79, pp. 356–369, 2018.

[52] L. Liu, W. Gui, L. Zhang, and J. Chen, “Real-time pose invariant spontaneous smile detection using conditional random regression forests,” *Optik,* vol. 182, pp. 647–657, 2019.

[53] T. Vo, T. Nguyen, and C. T. Le, “A hybrid framework for smile detection in class imbalance scenarios,” *Neural Computing and Applications,* 2019.

[54] Y. Ito and S. Omatu, “Category classification method using a self-organizing neural network,” *Int. J. Remote Sens*., vol. 18, no. 4, pp. 829–845, 1997.

[55] F. Giacco, C. Thiel, L. Pugliese, S. Scarpetta, and M. Marinaro, “Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3769–3779, Oct. 2010.

[56] Z. Li and J. R. Eastman, “Commitment and typicality measures for the self-organizing map,” *Int. J. Remote Sens.*, vol. 31, no. 16, pp. 4265–4280, 2010.

[57] Kahou, S.E., Froumenty, P. and Pal, C, "Facial expression analysis based on high dimensional binary features." *European Conference on Computer Vision*. Springer, Cham, 2014.

[58] L. An, S. Yang, and B. Bhanu, “Efficient smile detection by Extreme Learning Machine,” *Neurocomputing*, vol. 149, pp. 354–363, 2015.

[59] Liu, Mengyi, et al. "Enhancing expression recognition in the wild with unlabeled reference data." *Asian Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2012.

[60] Jain, Anil K., and FarshidFarrokhnia. "Unsupervised texture segmentation using Gabor filters." *Pattern recognition* 24.12 (1991): 1167-1186.

URKUND

Urkund Analysis Result

Analysed Document:	Thesis.docx (D55192581)
Submitted:	9/2/2019 12:11:00 PM
Submitted By:	karun.verma@thapar.edu
Significance:	14 %

Sources included in the report:

karuppanan survey journal 28052019.docx (D53495947)
22God-Thesis-19-7-19FINAL-6PM.pdf (D54492114)
1424489847-TS.pdf (D34250572)
https://arxiv.org/pdf/1802.02185.pdf
https://www.alliedjournals.com/download_data/IJEMS_V4IS90027.pdf
https://github.com/truongnmt/multi-task-learning
https://link.springer.com/article/10.1007%252Fs00521-019-04089-w
https://www.cl.cam.ac.uk/~hg410/JangEtAl_ICCVW2017.pdf
https://www.semanticscholar.org/paper/Enhancing-Expression-Recognition-in-the-Wild-with-Liu-Li/061e29eae705f318eee703b9e17dc0989547ba0c
http://freepaper.me/downloads/abstract/10.1016/j.patcog.2018.02.019
https://mafiadoc.com/developing-a-practical-smile-detector-semantic-scholar_5a0b6dbd1723dd69cf978aae.html
https://www.researchgate.net/publication/221292666_An_efficient_approach_to_smile_detection
https://rd.springer.com/article/10.1007/s12559-018-9601-x
47a54064-65e3-4758-a7ed-3d96e8c441a1
3dfc264a-2542-4bd8-a223-5702c5942635
75d05cef-2684-4bb5-9686-fa507a22284d
e7156340-5efb-4552-8d93-e018e2b45548
dc6c8804-a3a4-44a1-a97f-6327170c2d66

Instances where selected sources appear:

56