Facial Emotion Analysis & Discriminating Real and Fake Smile Using Convolution Neural Network

RAJESH KUMAR G A

DISSERTATION

On

Facial Emotion Analysis & Discriminating Real and Fake Smile Using Convolution Neural Network

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

Master of Technology

In

Computer Science and Engineering

Submitted By

RAJESH KUMAR G A

Examination Roll number:15/CS/409

Under Guidance of

Prof.(Dr.) Goutam Sanyal

Department of Computer Science and Engineering

National Institute of Technology Durgapur

M. G. Avenue, A-Zone, Durgapur - 713209

West Bengal, India

2017

Department of Computer Science and Engineering

National Institute of Technology Durgapur, Durgapur - 713209

This work is dedicated to my guide, my parents, brother and friends.

National Institute of Technology
Durgapur – 713209

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of his **Master of Technology** in **Computer Science and Engineering**.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name (Block Letters) : RAJESH KUMAR G A

Exam Roll Number : 15/CS/409

Thesis Title : Facial Emotion Analysis & Discriminating Real and Fake Smile Using Convolution Neural Network

Signature with Date :

National Institute of Technology
Durgapur – 713209

CERTIFICATE OF RECOMMENDATION

This is to certify that the thesis entitled" **Facial Emotion Analysis & Discriminating Real and Fake Smile Using Convolution Neural Network "**, submitted by **RAJESH KUMAR G A** for the partial fulfillment of the requirements for the award of the degree of Master of Technology in Computer Science and Engineering with specialization in Computer Science and Engineering, is a bonded research work under the guidance of **Prof.(Dr.) Goutam Sanyal**. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma. In our opinion, this thesis is of the standard required for the partial fulfillment of the requirements for the award of the degree of Master of Technology.

Prof.(Dr.) Goutam Sanyal
(Project Supervisor)
Professor & Head
Department of CSE
National Institute of Technology
Durgapur-713209

National Institute of Technology
Durgapur – 713209

CERTIFICATE OF APPROVAL

The foregoing project work entitled **"Facial Emotion Analysis & Discriminating Real and Fake Smile Using Convolution Neural Network"** is hereby approved as a creditable study of engineering carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is to be understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein. But approve the thesis only for the purpose for which it is submitted.

Date-

BOARD OF THESIS EXAMINERS

1. …………………………… 4……………………………………………

2. …………………………… 5……………………………………………

3……………………………….. 6……………………………………………

Acknowledgement

First, I would like to thank my project supervisor **Prof.(Dr.) Goutam Sanyal (**Professor & Head, NIT Durgapur) for providing me with right balance of guidance and independence in my research. I am greatly indebted to him for his full support, constant encouragement and advice throughout the project. I consider him as my great advisers and will continue to seek his guidance in future accomplishments in my life.

It is my privilege to express my copious thanks and admiration to Ph.D. research scholar **Mr. Ravi Kant Kumar** for his ceaseless impetus and constructive advices.

I would like to express my gratitude to all **my teachers of the Computer Science and Engineering Department** for their valuable technical and moral suggestions and cooperation.

I would like to thank my batch mates Hridoy Sankar Dutta and Priyabrata Dash who have provided me many helpful suggestions time to time. Many other people have in some way contributed to my education, but they are too numerous to list here.

Above all, I am extremely grateful to my parents, my brother and all my friends for their unfailing support and motivation throughout my career.

Date-

RAJESH KUMAR G A
Department of CSE
National Institute of Technology
Durgapur-713209

Facial Emotion Analysis & Discriminating Real and Fake Smile Using Convolution Neural Network

ABSTRACT

Human emotions are mental states of feelings that arise spontaneously rather than through conscious effort and are accompanied by physiological changes in facial muscles which implies expressions on face. Some of critical emotions are happy, sad, anger, disgust, fear, surprise etc. Facial expressions play a key role in non-verbal communication which appears due to internal feelings of a person that reflects on the faces. In order to computer modeling of human's emotion, a plenty of research has been accomplished. But still it is far behind from human vision system. In this paper, we are providing better approach to predict human emotions (Frames by Frames) using deep Convolution Neural Network (CNN) and how emotion intensity changes on a face from low level to high level of emotion. In this algorithm, FERC-2013 database has been applied for training. The assessment through the proposed experiment confers quite good result and obtained accuracy may give encouragement to the researchers for future model of computer based emotion recognition system.

In our society, sometime we hide our genuine feeling and emotion and purposely express different emotion in front of our surrounding folks. But as it's not actually a natural emotion, hence, it is more or less, predictable by others. Human vision system has enormous capability to recognizing genuine and fake smile of an individual. Discriminating genuine and fake smile is very thought-provoking task and even though very smaller amount of research has been carried out in this topic. In this paper, we are exploring a method to distinguish real from fake smile with high precision by using convolution neural networks (CNN). System has been train with FERC-2013 dataset having seven types of emotions namely happy, sad, disgust, angry, fearful, surprised and neutral. Emotions percentages of real and fake face are recorded by the emotion detection system. Based on recorded score, we investigate the effect of various percentages of emotions presented on both faces and then we are going to classify the smile on the face is real or fake.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Introduction

What is an emotion? An emotion is a mental and physiological state which is subjective and private. It involves a lot of behaviors, actions, thoughts and feelings.

Primary research carried out in emotions can be outlined to the book ‘The Expression of the Emotions in Man and Animals’ by Charles Darwin. He thought emotions to be species specific rather than culture-specific [1],[2], but in 1969 after identifying a universality within emotions despite the social differences, worldwide classified basic emotions in to six kinds and are specified as: happiness, sadness, anger, disgust, surprise and fear [3],[4],[5],[6]. (Figure 1.1)

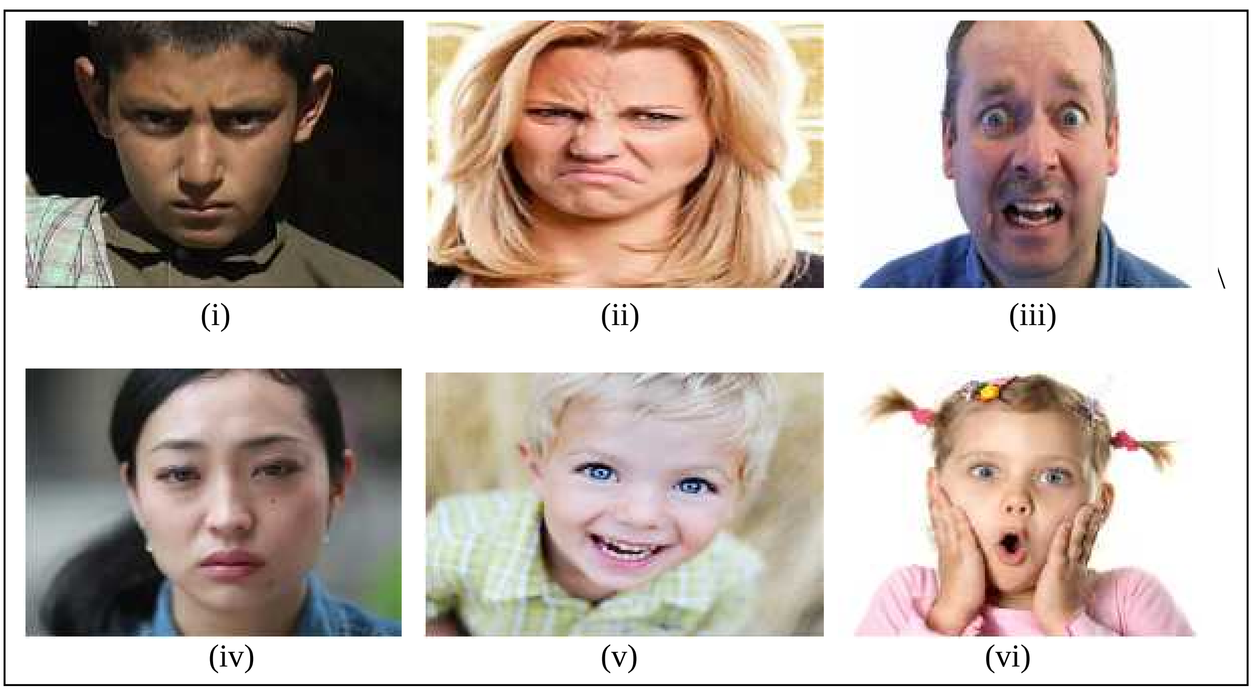


Figure 1.1: The six universal emotional expressions, (i)angry, (ii)disgust, (iii)fear, (iv)sadness, (v)happy, (vi)surprise

INTRODUCTION

Emotions and related fluctuations in the facial muscles are together known as facial expressions [7]. It gives us clue about the state of a person and enables to make conversation with the other person based on their mood. Furthermore, facial expressions also support to judge the existing state of emotion and mood [8] of a person. Facial expression plays an important role in non-verbal communication between people. Diverse classification of facial expressions might be used in numerous applications like; Human Behavior Predictor [9], Surveillance System [10] and Medical Rehabilitation [11].

Recognizing genuine and fake expression, seem on human face is one of the hardest job for once brain. Humans vision system, have a remarkable capacity to recognize genuine and fake smile of an individual. Even though, countless times our brain is also not talented enough to distinguish it clearly. But how computer vision system can differentiate between genuine and fake emotions? There is no appropriate reply for such questions, till date. Still, in order to discover solutions of such difficult problems at some extent, quite a few computational techniques have been demonstrated. To make these things understandable, primary challenge has been reserved by a well-known French physician named Guillaume Duchenne, from the 19th century to Distinguish genuine and fake smile founded on the muscles that are involved in generating facial expressions [13]. In [14], writers claim about eyes as an evidence for the finding of real and fake smiles. In order to find out social impact of truthful smiles a research has been conducted in [15]. In this research, examination discovered that; associated to involved and control members, excluded folks exhibited a better preference to work with folks displaying “actual” as opposed to “fake” smiles [16]. Some scientist proposed how the adaptive responses to social exclusion work and social refusal improves the finding of genuine and fake smiles. Eye movements based real and fake smiles judgment has been studied in [17]. The main aim of our proposed scheme is to find out the standardized parentages of several emotional states (happiness   sadness, disgust, anger, surprise, and fear) in a face. The emotion having the maximum parentages is projected as its resulting emotion on a specified face.

Likewise, founded on experimental outcomes, training and examination of various emotional phases (frame by frame) has also inspired us to develop a real-time facial expression recognition system. To attain such composite classification of images, an enormous and robust training is essential. Hence, in this proposed approach concept of deep learning using convolution neural network has been applied to train and test. The performance of a neural network mainly depends on numerous issues like initial random weights, activation function used, training data, and number of hidden layer and network structure of system. The convolutional neural networks use images directly as input. As a substitute of handcrafted intermediate features, convolutional neural networks are used to mechanically learn a pecking order of features which can further be used for classification.

1.1 Problem Overview

Utmost information that we have about facial expressions was molded by our parentages and our family. The faces of our parents and other people that took care of us as children were the first ones to see. Their faces may have been very communicative or very guarded in their facial expressions. Persons in our family may have exposed emotions in their faces as most other people do, or one of them might have developed an outlandish or unusual way of looking disgusted or fearful, and that may still effect our recognition of that emotion. As a kid, some people have been precisely taught not to look at the facial expressions of others, or at least certain facial expressions. For example, some kids are said never to look at somebody who is crying.

The number of the facial expressions that we use in our everyday life cannot be strictly specified due to the different surrounding and cultural background that each person has. Research on the facial expression analysis has focused more on the six basic emotional expressions (fear, anger, disgust, happiness, surprise and sadness). For these facial expressions conclusions about their universality have already been made. However, the emotional facial expressions used around the world are much more and some of them are combinations of more than one.

Based on the difficulty of recognition of several facial expressions from people of different backgrounds the focus of this thesis is to find a way to display a large variety of emotional facial expressions, validate them and produce rules in order to translate facial actions into emotional expressions and also gives an idea to distinguish between real and fake smile.

1.2 Motivation

Communication, may be in any forms i.e. verbal or non-verbal is vigorous to ample various daily repetitive tasks and plays a important part in life. Facial expression plays most important role in non-verbal communication and it affords a hint about emotional state, mentality and intention of a person. Facial expressions not only can alter the flow of conversation but also offers the spectators a way to communicate a means of information to the speaker without even expressing a single term. when the facial expression does not overlap with the other communication i.e. vocalized words, then the information conveyed by the person's face gets extra weight in interpreting information.

The computing atmosphere is affecting human-centered strategies instead of computer centered strategies, and people tend to interconnect wealth of information through sentimental states or terminologies. Outmoded Human Computer Interaction (HCI) based systems discounts majority of information interconnected through those affective states and just caters for user's deliberate input. As stated that paradigm is fluctuating towards human-centered designs, thus study of user affective states becomes expected. In near future people will not interact with machineries only over intentional inputs but also over their behavior i.e. affective states. hence, computer vision research community has publicized a proportion of attention in studying and automatically identifying facial expressions. There are tons of application zones that can profited from a system that can categorize facial expressions i.e. human-computer interaction, entertainment, medical applications e.g. pain detection, social robots, deceit detection, communicating video and behavior monitoring.

Facial expressions are studied since early times, as it is one of the most significant channel of non-verbal communication. Firstly, facial emotions were studied by great philosophers and thinkers like Aristotle and Stewart. With Darwin, the study of facial expressions became an experiential study.

In the computer vision municipal, the term “facial expression recognition” often denotes the classification of facial structures in one of the six so named basic or universal emotions: happiness, sadness, fear, disgust, surprise and anger, as announced by Ekman in 1971. We, the beings, appreciate various facial expressions every day without any additional effort. But for computer based systems on the other side, it is still tough to identify them automatically due to face appearance variations produced by pose differences, lighting differences and camera quality and angle changes.

1.3 Prospective Applications

A system that might allow fast and robust facial expression recognition would have numerous usages in both research and application zones as various behavioral science, education, entertainment, medicine, and security. Following are the list of applications that can be implemented using automatic classification of facial expressions.

1. *Avatars with expressions:* Simulated environments and characters have become extremely popular in the 21st era. Gaming engineering would get advantage extremely if the avatars were able to imitate their user’s facial expressions recorded by a webcam and evaluated by a facial emotion recognition system as the level of immersion and certainty in the cybernetic world. It can be utilized in many implications e.g. A game could adapt its difficulty level using information from the facial expressions of the operator.

2. *EmotiChat:* One more fascinating application has been established by Anderson and McOwen, termed the “EmotiChat”. It contains a chat-room application where users can log in and start chatting. The face expression recognition system is connected to this chat application and it automatically inserts emoticons based on the user’s facial expressions.

3. *Smart homes:* As stated earlier, computing atmosphere is traveling towards human-centered designs instead of computer centered designs and this paradigm change will have far reaching consequences, one of them being smart homes. The houses could be equipped with systems that will record dissimilar readings i.e. lighting conditions, type of music playing, room temperatures etc. and associate them with the facial expressions of the inhabitants over time. Thus, such system can later control different recorded environment parameters automatically.

4. *Social robots:* For social robots, it is too important that they can identify diverse expressions and act as per identified emotions in order to have real communications. The Social Robots projected at Carnegie Mellon University states its mission as “wanting robots to behave more like people, so that people do not have to behave like robots when they interact with them”. To attain such human-robot interaction, it is of paramount importance for the robot to understand the human’s facial expressions.

5. *Detection and treatment of depression and anxiety:* Study based on the FACS has exposed that facial emotions can predict the start and diminution of depression, schizophrenia, and other psychopathological diseases. FACS has also been talented to recognize patterns of facial activity involved in alcohol intemperance that viewers not trained in FACS failed to note. This proposes there are many applications for an automatic facial emotion recognition system based on FACS.

6. *Pain monitoring of patients:* Pain monitoring of patients is a very difficult but very significant job. Presently, this is done manually but it is necessary to design such a system that can automate this duty. Physically monitoring of pain has some difficulties: First, pain cannot be logged endlessly. Secondly, some patients can under report the pain while other can do just opposite. Lastly, the individual recording the pain has to make decision of pain level, which could vary from person to person (subjectivity problem). An automatic facial emotion recognition system might resolve above cited problems. It has been shown that it is likely to derive a amount of pain and to differentiate between different types of pain from a patient's facial expressions. In this research work, we have projected a novel computer vision system that can identify expression of pain in videos by analyzing facial features.

1.4 Proposed Solution

In this work, a system which will efficiently recognize the six universal emotions from 2D color face images or from video. The work has been limited to the universal emotions since classification and identification of other marginal emotions is problematic. In order to computer modeling of human's emotion, a plenty of research has been accomplished. But still it is far behind from human vision system. We are providing better approach to predict human emotions (Frames by Frames) using deep Convolution Neural Network (CNN) and how emotion intensity changes on a face from low level to high level of emotion. In this algorithm, FERC-2013 database has been applied for training. The assessment through the proposed experiment confers quite good result and obtained accuracy may give encouragement to the researchers for future model of computer based emotion recognition system.

1.5 Thesis outline

The rest of the thesis is organized as follows. Chapter 2 contains a brief overview of the recent work carried out in the area of facial emotion recognition. The system model of the proposed solution and detailed descriptions on each of its stages (image pre-processing, deep convolutional neural networks and Emotion classification) can be found in chapter 3. An evaluation of the overall system has been described in chapter 4. In chapter 5, discussion and future Works have been explained briefly. Finally, the thesis is concluded in chapter 6.

Chapter 2

RECENT WORK

Recent Work

The recent work applicable to the study can be largely characterized into three steps: Face detection, Facial feature extraction and Emotion classification. The amount of research carried out in each of these categories is quite sizeable and noteworthy. These three categories are concerned with the central background pertaining to the issue of facial emotion recognition. Apart from them, another core area is the work carried out in forming an opposite facial database for such studies.

2.1 Face Detection

Given a picture, spotting the existence of a human face is a complex job due to the imaginable differences of the face. The dissimilar sizes, angles and postures a human face might have within the image can cause this variation. The emotions which are deducible from the human face and diverse imaging settings such as illumination and occlusions also affect facial appearances. In addition, the presence of eyeglasses, beard, hair and makeup have a significant effect in the facial appearance [18], [19].

The methods of the past few periods in face detection can be largely classified in to four segments: knowledge-based approach, feature invariant approach, template-based approach and appearance-based approach. (Table 2.1)

Knowledge-based Approach
Feature Invariant Approach -Facial Features -Texture -Color -Multiple Features
Template-based Approach -Predefined Face Templates -Deformable Face Templates
Appearance-based Approach -Neural Network -Support Vector Machine (SVM) -Eigenface -Distribution based -Nave Bayes Classifier -Hidden Markov Model (HMM) -Information-Theoretical Approach

Table 2.1: Categorization of face detection approaches

2.1.1 Knowledge-based Approach

Knowledge-based approach is built on guidelines resulting from the knowledge on the face geometry. A typical face used in this approach is shown in figure 2.1. The utmost common way of defining the guidelines is by the relative distances and positions of facial features. By smearing these rules faces are spotted, then a verification method is used to trim the improper detections. Decoding knowledge about the face geometry into actual rules is one trouble faced in this method, since strict rules might fail to identify faces but if the rules are too universal it can rise incorrect recognitions. This method is too imperfect since extending this to detect faces in all cases is impossible [20],[21],[22].

Figure 2.1: A typical face used in knowledge-based approach

2.1.2 Feature Invariant Approach

In feature invariant approach, facial features are identified and then assembled according to the face geometry. Choosing a set of suitable features is very critical [23]. This method is not appropriate for pictures with noise, decorations and obstructions since they can decline the feature limits. The main disadvantage of template-based approaches is the sensitivity to scaling and rotation. Meanwhile feature-based method is not affected by this sensitivity, it offers a better result to a facial recognition problem.

A face model which is articulated by means of features or a face texture model which can be expressed in terms of a set of dissimilarities can be used for face recognition in the feature invariant approach [24] [20] [21] [22]. Newly human skin color has grasps the attention of the scholars as an important feature, meanwhile skin color exists in a minor color range in diverse color spaces irrespective of the race. Therefore, modern studies on several skin color recognition techniques can be found [26] [25] [21] [22]. Tactics which combine many facial features have also being proposed.

2.1.3 Template-based Approach

A typical design of a human face can be used as the foundation for the template-based approach. The pixels inside a picture window are associated with the regular pattern to recognize the existence of a human face inside that window. However, the method is simple, the scaling mandatory for the picture is a model’s disadvantage. In addition, this method is unable to dealing with the dissimilarities of the human face. A predefined template based and deformable template based approaches are the two classes of the template-based approach [24] [20] [21] [22].

2.1.4 Appearance-based Approach

Appearance-based approach deliberates the human face in the form of pattern of pixel intensities. Since in a training process non-face patterns are not used of this method because it is not strong enough. Even the time engaged is long, as the number of patterns which requires to be verified is large. A neural network is the normally used answer for catching compound facial patterns from facial pictures. Both supervised and unsupervised learning methods have been used to train the network. Since identifying a sufficient training data set is doubtful, unsupervised neural networks are more better choice than supervised neural network. Apart from neural networks, Support Vector Machines (SVM), eigen-faces, Distribution based approaches, Nave Bayes classifiers, Hidden Markov Models(HMM) and Information theoretical approaches can also be used for face recognition in the appearance-based approach. Rather than diminishing the training error as in neural networks, SVM operate by minimizing the upper bound on the generalization error. Eigenfaces is a probabilistic visual learning method which uses eigenspace decomposition. Nave Bayes classifier provides a better assessment of the conditional density functions in facial sub-regions. The HMM does not require exact alignment as in template-based and appearance-based approaches. HMM usually see a face pattern as a sequence of observation vectors.

2.2 Facial Feature Extraction

Choosing a satisfactory set of feature points which signify the significant features of the human face and which can be extracted effortlessly is the foremost challenging task, a fruitful facial feature extraction method has to response (Figure 2.2). The luminance, facial geometry, chrominance and template based approaches, symmetry based approaches and Principal Component Analysis (PCA) based methods are the main classes of the methods available. Methods which syndicate two or more of the above-mentioned types can also be found [27] [28].

Figure 2.2: A selected set of facial feature points

2.2.1 **Geometric feature based methods**

As stated earlier, geometric features indicate the shape and positions of facial components (including eyes, brows, mouth, nose). Therefore, the encouragement for using a geometry-based technique is that facial expressions affect the relative location and size of many facial features, and that, by computing the movement of certain facial points, the original facial expression can be predicted. In order for geometric approaches to be effective, the positions of these fiducial points essentially be determined exactly; in real-time systems, they must also be found speedily. The exact type of feature vector that is extracted in a geometry-based facial expression recognition systems depends on

1. which points on the face are to be chased,
2. whether 2-D or 3-D positions are used,
3. the technique of changing a set of feature locations into the final feature vector.

Active shape models (ASMs) are statistical representations of the form of objects which repeatedly deform to fitting to a sample of the object in a new picture. One drawback of ASM is that it only uses shape restrictions (organized with some information about the picture structure near the landmarks), and does not take benefit of all the existing information: the texture across the target. Hence, active appearance model (AAM) which is associated to ASM is proposed for matching a statistical approach of entity based on both shape and appearance to a new picture. Similar to the ASM, AAM is also built during a training stage: on a set of pictures, composed with coordinates of benchmarks that appear in all of the pictures, is providing to the training supervisor. AAM could be observed as the hybrid approaches based on both geometric and appearance features.

The characteristic examples of geometric-feature-based approaches are those of Pantic and her contemporaries, who used a set of facial typical points around the eyebrows, nose, mouth, eyes and chin. Figure 2.3 displays the benchmarks in Pantic’s work, and through following these benchmarks, the motion information is attained to do expression recognition.

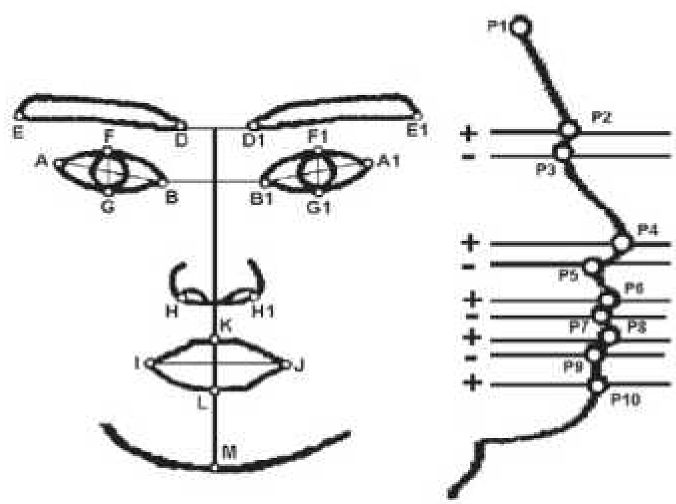


Figure 2.3: Demonstration Landmarks on the face.

Zhang et al. measured and tracked the facial motion by means of Kalman Filters. To attain the expression detection task, they have also demonstrated the temporal behaviors of the facial expressions by means of Dynamic Bayesian networks (DBNs). In authors have praposed Facial Action Coding System's (FACS) Action Unit (AU) recognition system by means of features considered from the "Particle Filter" tracked fiducial facial points. They trained the system on the MMI-Facial expression database and tested on the Cohn-Kanade database and accomplished detection rate of 84%. Bai et al. extracted only shape information using Pyramid Histogram of Orientation Gradients (PHOG) and showed the "smile" recognition accuracy as high as 96.7% on Cohn-Kanade database. PHOG is a spatial shape descriptor and it label object appearance and shape by the distribution of concentration gradients or edge directions.

Study has been done with victory in current days to combine features extracted by means of appearance-based approaches and geometric feature-based approaches. The key problem with geometric approaches is to precisely pinpoint the landmark and track it. In the real applications, due to the posture and lighting variations, small resolution input pictures, and the noise from the background, it is still very hard to exactly locate the benchmarks.

2.2.2 Template based Approach

Some scholars have created their methods on the unique shape of features in hope of gaining better-quality results. By defining each feature with a distinct template, Craw and others' [29] have put forth a resourceful method for extraction of facial features. They have used deformable models to extract the nose, eyes and face outline. A characteristic eye model used in most template-based methods is shown in figure 2.4.

Kuo and Hannah (30) have also proposed a extremely accurate and elastic approach for the extraction of eyes using basic shapes, eye corner detection and deformable templates. The iris is selected as the first element to extract, because of its spherical shape and high intensity difference with the adjoining areas. In detecting the eye epicenter using a PCA based approach, a deformable circle model is used to search across to get the finest. For a best fit, the area of the sphere should be dark and a distinguishable intensity dissimilarity should be existing along the circumference After iris extraction, a simple and actual eye corner detection method is used for the detection of eye corners. This method has shown 94% accuracy in iris extraction and 88% accuracy in eye corner and eyelid fitting. The errors in extraction are mainly due to the strong reflections in the eye areas.

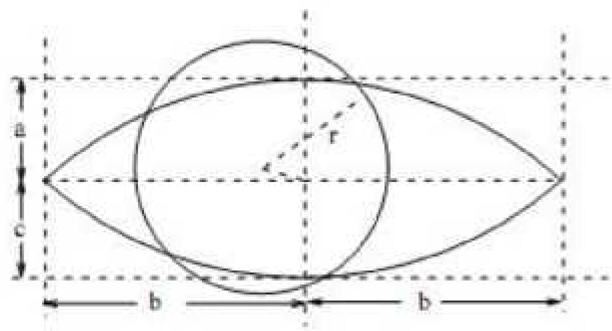


Figure 2.4: A template used for the extraction of single eye

2.2.3 PCA based Approach

For takeout features based on PCA, unsupervised PCA and global PCA based methods have been used. Though PCA based methods which consider the complete face picture have the drawback of being sensitive to the face position, size, background. Also, greater illumination differences tend to skew the PCA learning process. Thus, interest has risen over more precise and localized PCA based methods.

Such a localized PCA based learning method for facial feature extraction has being put forth by King and Xu (31). The focused and directed training in the localized PCA learning united with simple matching approaches has formed a resourceful method to extract features for matching face pictures. This method results several masks for localized features. Several of such masks attained for mouth are exposed in figure 2.5. A batch algorithm has been used to implement the localized PCA learning. Result of this method is a set of eigen features beneficial for face detection. A method for feature extraction and approximation of control points for facial picture warping based on a PCA based statistical face model, has being offered by Xue and others [32]. The PCA has being used to generate full face model with contour points and control points.

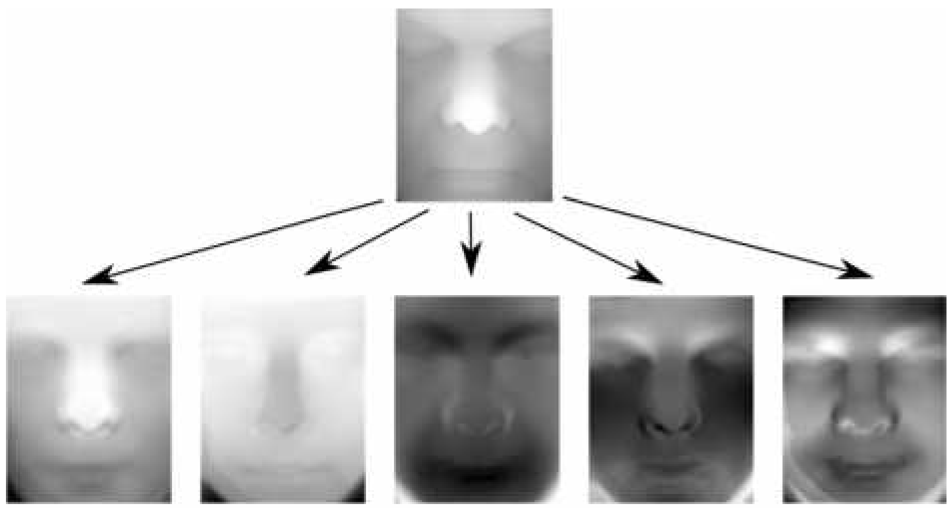


Figure 2.5: Mouth masks obtained from PCA learning of mouth images

The Luminance, chrominance, facial geometry and symmetry based methods, template based methods, Principal Component Analysis (PCA) based methods apart from these, other facial feature extraction methods also exist. Some of them include eigenfaces, wavelets and discrete cosine transform.

A wavelet network is a network which comprises of a set of wavelets and related networks. The geometric configurations of a wavelet network are defined with reverence to a single coordinate system. Feris and others' method [33] on two-level hierarchical wavelet networks restricts small features with the use of falling sets of GWN features. Besides, they have tested their outcomes with both one-level and two level hierarchical wavelet networks. In first level hierarchies the GWN features are trained for the entire face while GWN features are trained for each facial feature in the two level hierarchical networks.

A neural network can also be trained to detect the wanted features from a facial picture. To attain this, either one neural network can be trained for all the features or several networks for each feature. The key challenge faced in using supervised neural networks is the achievement of the training set. In the attempt of Paul Debevec, [34] a neural network is trained for the extraction of mouth, eyes and nose. The back-propagation algorithm and a training set with 97 pictures have being used to train the network. The neural network recommended in the method consisted of a hidden layer with ten neurons. Four such neural networks have been trained for each feature (left and right eyes, mouth, nose). For the conception of the user, the neural network's proposal for the feature has being mapped in to a picture (also called a feature maps). Therefore, the accuracy of the neural network can be calculated by visual comparison of original image and the above-mentioned image. A effective neural network would produce a complete white image except for the black area at the corresponding location of the feature. (Figure 2.6)

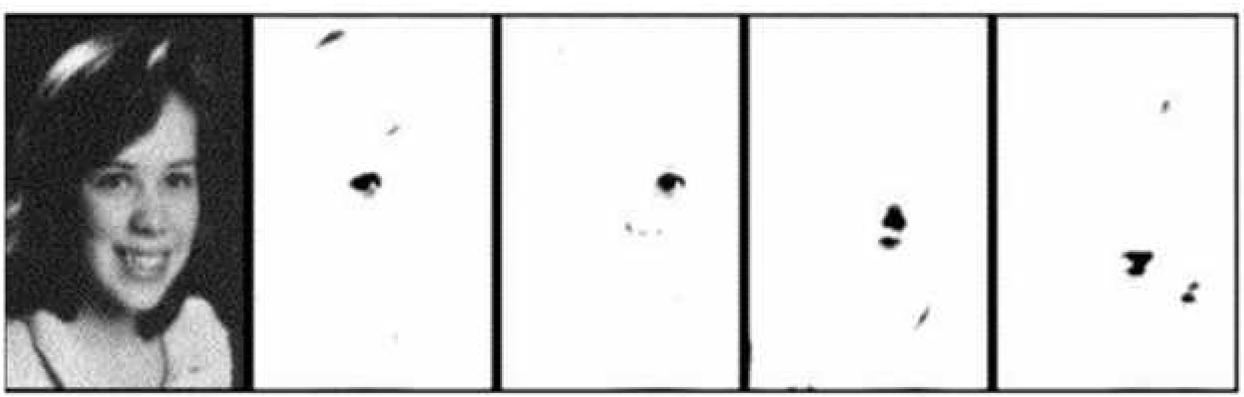


Figure 2.6: The neural network's corresponding feature maps for left eye, right eye, nose and mouth (12)

2.3 Emotion Classification

The study carried out by Ekman on emotions and facial expressions is the main motive overdue the attention in the topic of emotion classification. His effort has spellbound the scholars and advised them to study the emotions through pictures and video processing systems. First, face location determination and feature extraction was completed from pictures, then those extracted features were used as input to the classification system which in turn chooses a predefined emotion class. Over the past few periods numerous approaches have been announced for the classification of emotions. They differ only in the features extracted from the pictures and in the classification technique used to differentiate between the emotions [35] [36].

All these methods have attentive on classifying the six common emotions. The non-universal emotion classification for emotions like amusement, wonder, greed and pity is yet to be engaged into consideration. A good emotional classifier should be able to identify emotions independent of ethnic group, gender, age, glasses, hair styles, beard lighting conditions, pose, backgrounds and birth marks. In categorizing emotions for video sequences, pictures corresponding to each frame have to be obtained, and the features extracted from the early image have to be mapped between each frame. An adequate rate for frame processing has to be maintained as well.

The most universal technique for answering this problem is through the use of a neural network which has got over 85% accuracy for emotion classification. Current research has focused on techniques other than neural networks in hope of gaining higher accuracy [35] [36].

2.3.1 Support Vector Machines (SVMs)

why the use of SVM the classification problem can be viewed as a quadratic optimization problem. Since SVM categorize data with a set of support vectors by diminishing the structural hazard, the regular error among input and their target vectors is reduced.

With the use of the SVM package LIBSVM, Dumas [35] has projected an emotion classification method. The LIBSVM has being praposed by ChihChung Chang and Chih-Jen Lin. The objective of this study was to obtain the highest possible accuracy achievable with SVM to categorize the Images of Facial Affect (POFA) dataset. The POFA dataset is a static human facial image dataset. Rather than focused on all features some scholars have focused on one feature and discovered its ability to classify emotions. Li has confirmed that the mouth feature is enough to classify among sadness, happiness and surprise. But to differentiate between all basic emotions the information about the eyes and eyebrows should be attained in addition to the mouth. He has classified emotions in video sequences with a lip tracking rate of 25 frames per second. Seyedarabi and others' method [37] proves a novel FAP based method to animate the facial emotions from video sequences. 21 predefined facial key points are extracted from the initial frame (Figure 2.2) and the additional feature points are automatically extracted from the subsequent frames by the use of cross-correlation based optical flow (Figure 2.7). For animating purposes, they have created three object models for eyes, mouth and eyebrows, in which each vertex of the triangle were determined by the feature points extracted.

The use of trajectories in eigenspace in the classification is demonstrated by Schwerdt and others' method. After the face is tracked using color based techniques, the detailed movements of single features like eye brows were captured by eigenvectors with small eigenvalues. From these eigenvectors, those which do not contribute in distinguishing between emotions are removed and the remaining ones are used for the classification process. The observation that large eigenvalues are not the most reliable eigenvectors for the emotion classification was found from this study.

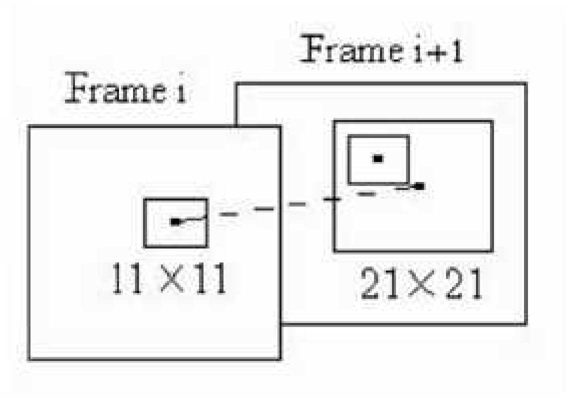


Figure 2.7: Use of cross-correlation based optical flow within frames

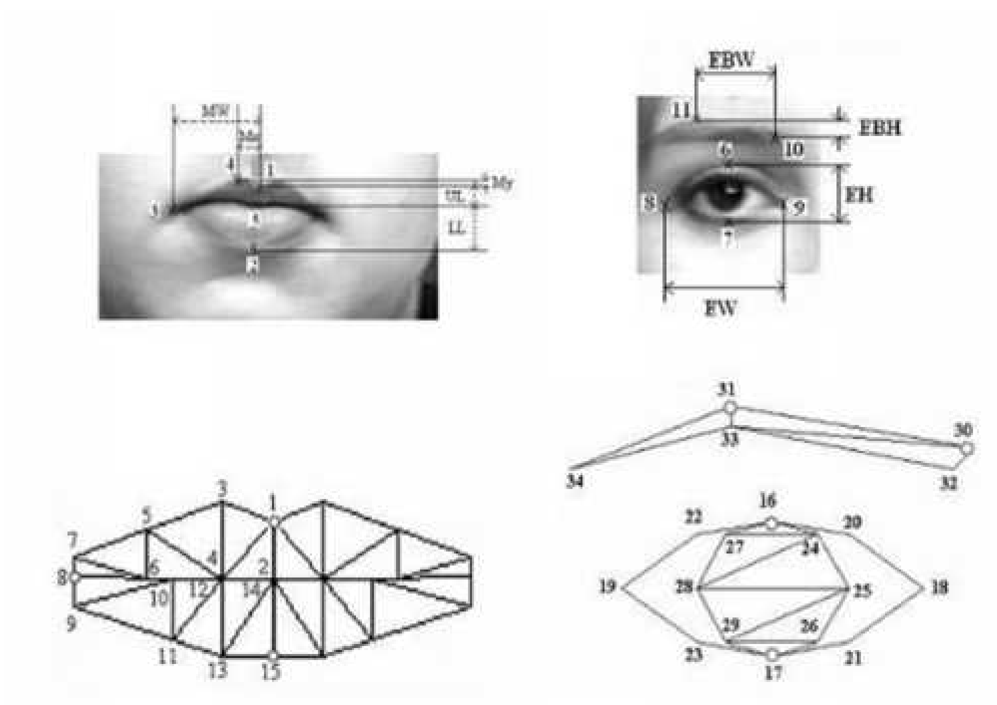


Figure 2.8: object models created for eyes, mouth and eyebrows

2.4 Facial Databases

Since a suitable database has to be found for the evaluation process, a survey on the available face databases were also performed. Due to the approach selected, certain constrains has to be met in selecting a database. The requirements were for the database to have color and frontal facial images. The subject, background variations and other variations like spectacles, beard and hair in images were favorable. After some research, we found the Caltech Faces, Georgia Tech Face Database and Facial Expressions and Emotion Database suitable for our study.

2.4.1 Caltech Faces

This database contains 450 frontal face images with 896 x 592 pixels. The images are of 27 or more subjects with different expressions, backgrounds and lighting conditions. This database is provided by Markus Weber at the California Institute of Technology.

2.4.2 Georgia Tech Face Database

This contains images of 50 people taken at the Center for Signal and Image Processing at Georgia Institute of Technology. All images are of 150 x 150 pixels each represented in 15 color JPEG images.

2.4.3 Facial Expressions and Emotion Database

This is a database created by the Technical University at Munich. It contains images of 18 different individuals expressing the six basic expressions. Each individual has expressed each emotion three times. The images contain a simple and defined background.

2.4.3 Experiment database

We are applying two databanks in this research namely FERC-2013[39] and Extended Cohn Kanade (CK+) database [38]. The datasets basically differ by image-quality, clearness and total number of images in database. In, FERC-2013 contains about 32000 low resolution face pictures of dissimilar age groups and having different degrees of angle are available. In adding to this, facial expressions have been exhibited very clearly in the CK+ database. (Because they are taken from similar distance and with high resolution images.), Whereas FERC-2013 database, demonstrates emotions in the wild (i.e. ’taken from random distance and are low resolution images’). Which made pictures from the FERC-2013 database are tougher to interpret. We have trained our system on FERC-2013 database. Since images are ’very clear’ and have well define expressions, they effortlessly classified for different emotions on a face. Therefore, the convolutional networks are trained with the FERC-2013 database.

The database holds of [48x48] pixels of grayscale pictures of human faces. The faces are automatically processed, so that it holds up round a comparably equivalent volume of face space in all images. The prime task is to place each face in view of the emotions of one of seven classes (0: Happy, 1: Sad, 2: Surprise, 3: Angry, 4: Disgust, 5: Fear, 6: Neutral). Thus, database exists in the form of emotion and its matching pixels array. Some examples of FERC dataset are shown in Figure 2.9.

Figure 2.9. Some Valid Samples of FERC-2013 Database

Training Data

Before training, we pre-processed the FERC-2013 database images, in the pre-processing, we used the Viola-Jones algorithm [18] [40] on the dataset, we used 28,709 samples for pre-processing and validation among them we got 11246 valid samples for training, Due to drawback of Viola-Jones algorithm, many samples fail in face detection task, some images are shown below:

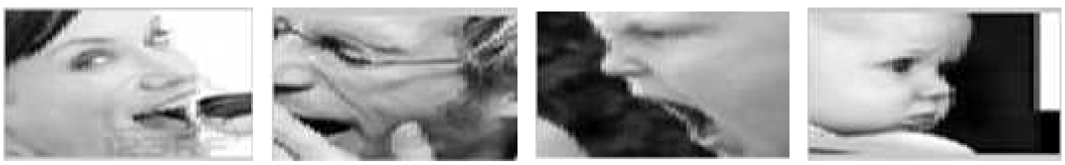


Figure 2.10. Failed images of Viola-Jones algorithm

Testing Data

Testing and validation purpose, we used FERC-2013 and Cohn Kanade (CK+) database and both the database we are getting the encouraging results. And some sample images of Cohn Kanade (CK+) database are shown below

Figure 2.11: images of Cohn Kanade (CK+) database

Chapter 3

EMOTION RECOGNITION SYSTEM

Emotion Recognition System

The system for emotion recognition is divided into 2 stages: testing and training. First, we need to train the networks to classify the emotions of given face. The first step of our algorithm is to face location determination, feature extraction and emotion classification. After locating the face with the use of a face detection algorithm, the knowledge in the symmetry and formation of the face combined with image processing techniques are used to process the face region to determine the feature locations. These feature areas are further processed to extract the feature points required for the emotion classification stage. From the feature points extracted, distances among the features are calculated and given as input to the neural network to classify the emotion contained. The neural network has been trained to recognize the 6 universal emotions.

Algorithm: Whole Algorithm flowchart

step1: if (trained database is not available)
step2: run Algorithm1
step3: run Algorithm 2
step4: save trained database
step5: else (load trained database)
step6: Get input image from webcam or system folder
step7: run Algorithm1
step8: run Algorithm2
step9 :(result 1) display the emotions with percentage of each emotion.

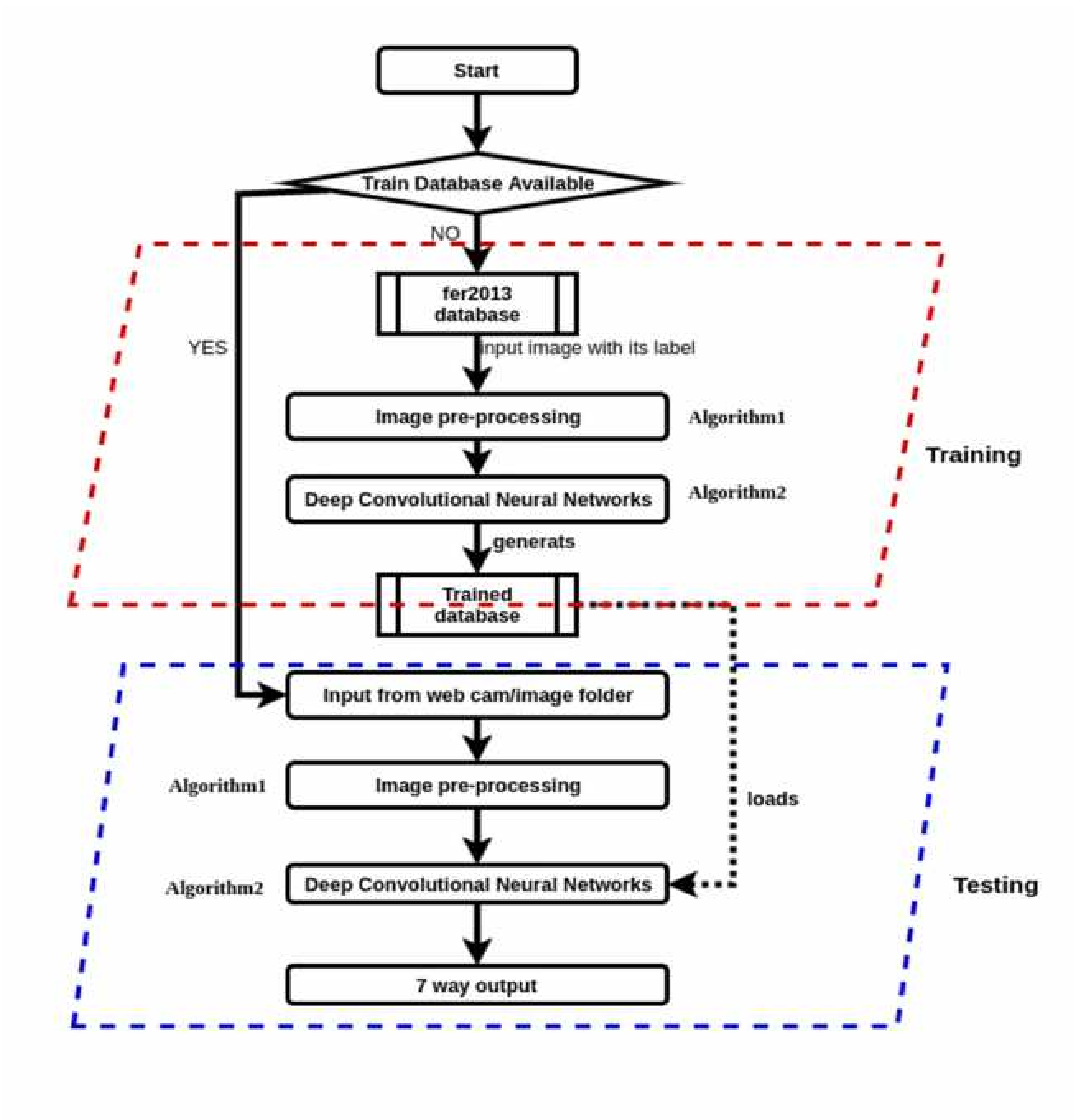


Figure 3.1: Complete System Workflow

3.1 Image pre-processing

Algorithm 1: Image pre-processing

step1: Get input from user.

step2: Face-detection using Viola Jones algorithm [18].

step3: Taking maximum area face among all faces.

step4: Crop the selected maximum area face from image.

step5: Resize the cropped face into 48x48 images.

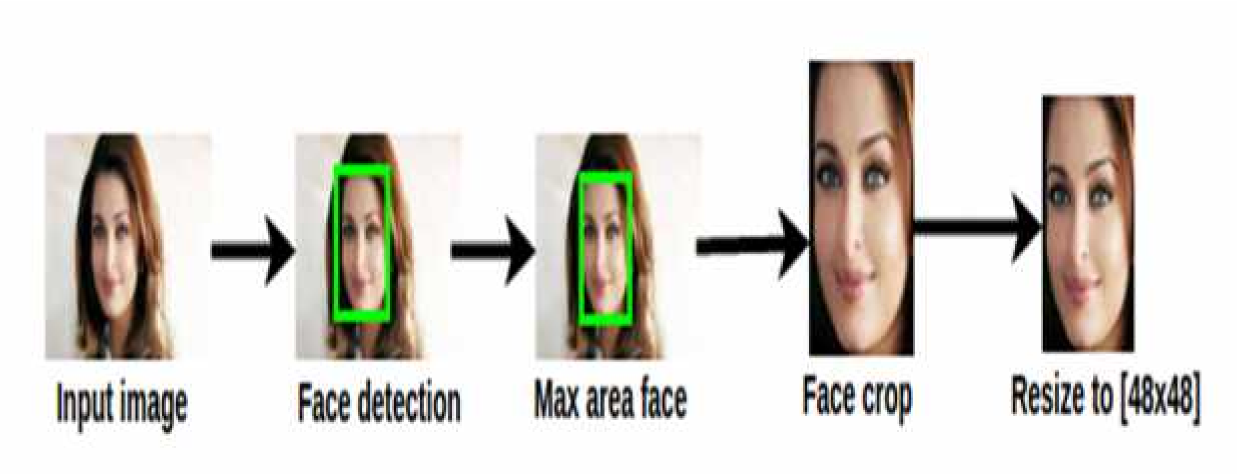


Figure 3.2. Image Pre-processing

As shown above the figure and flow chart the first we get the input image from user as either in form of video frames or images. then next impotent thing is face detection. The system offers two methods for face detection. Though various knowledge based and template based techniques can be developed for face location determination, a feature invariant approach based on the skin color, we selected template based techniques as the first method due to its flexibility and simplicity.

The important elementary phase of algorithm is finding the faces in the picture. We used the existing well known face detector Viola-Jones algorithm. A face identifier must able to state whether a picture containing a human face or not. It is usually the preprocessing step for the face emotion detection system. The Viola-Jones algorithm [18,19], is the most robust face detection algorithm. The algorithm employed mostly three most important stages for fast and flawless face detection. For feature computation we use integral images, Adaboost method is used in feature selection from picture and to boost performance and an attentional cascade for resource allocation on pictures. The initial phase of Viola-Jones algorithm find out the Haar-Like Features [19], which are advance features and are exploited for object classification. The concept of integral image is used for the time reduction in the computational task. The value at pixel (x, y) is the totality of pixels above and to the left of (x, y), all-encompassing. Since the Haar function produces more than 160,000 features, out of which all are not related for face localizing, hence, the AdaBoost algorithm is used to remove the irrelevant features. A set of relevant features is named as weak classifier. The weak classifiers are linearly unified to make a strong classifier. The last step is producing a cascade classifier which is collection of phases. At each phase, it is determined whether the given sub-window contains a face or a non-face. If it fails, it is considered as a non-face.

3.2 Deep Convolutional Neural Network

3.2.1 Algorithm 2: Deep Convolutional Neural Network

- ***Phase1:*** We initialize all filters and weights with random values.

- ***Phase2:*** The training image is input to the network and goes over the forward propagation phases (i.e. convolution layer, ReLU layer and pooling layer actions along with forward propagation in the Fully Connected layer) and detects the gives output probabilities for all class. Let's assume the output probabilities for the first given image are [0.5, 0.2, 0.3, 0.3, 0, 0, 0]. Since weights are randomly assigned for the first training image, therefore output probabilities are also random.

- ***Phase3:*** Calculating the entire error at the output layer is given as (Summation over all 7 classes).

$$\text{Total Error} = \sum (\text{target probability} - \text{output probability})^2$$

- ***Phase4:*** Using Back propagation we compute the gradients of the error for all weights in the network and use gradient descent to update all filter values / weights and parameter values to minimize the output error. The weights are updated in proportion to put their influence to reducing the total error. When the same image is imputed again, output probabilities might now be [0.1, 0.1, 0.7, 0.1, 0, 0, 0] which is closer to the target vector [0, 0, 1, 0, 0, 0, 0]. This implies now the network has learnt to categorize this particular picture correctly by altering its weights / filters, so that the output error is reduced. Factor like architecture of the network, number of filters used, filter sizes etc. have all been fixed before Step 1 and do not change during training process – only the values of the filter matrix and connection weights search out updated during the process.

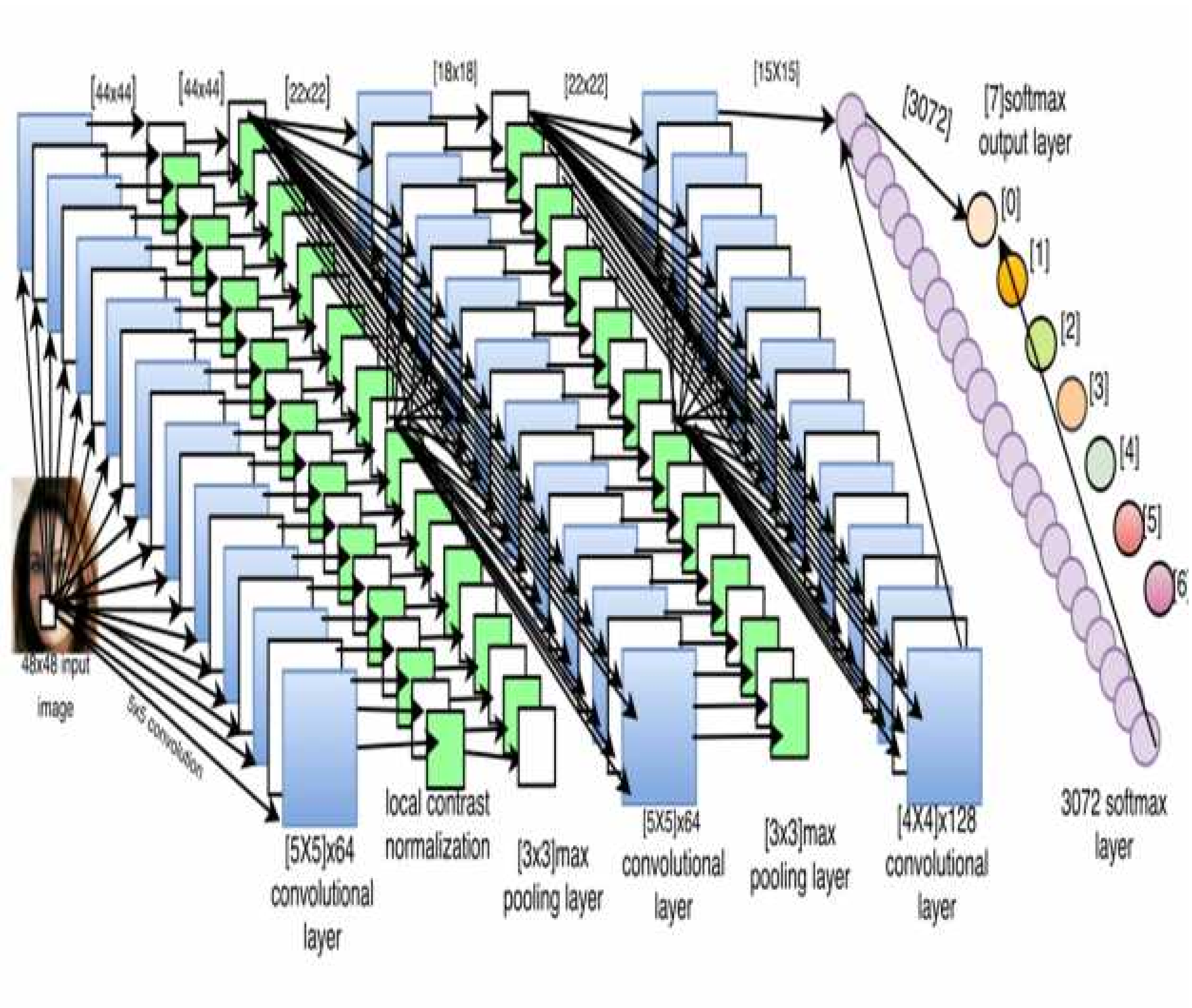


Figure 3.3. Architecture of Deep Convolutional Neural Network

3.2.2 Layer by Layer Explanation of CNN

- ***Layer 0*: *Input layer*** Input[48x48x1] contains the pixel values of the input image. In this case, an image of width 48, height 48, and with one color channel is considered.
- **Layer 1:** *Convolutional layer* calculates the output of all neurons that are associated to local regions in the input layer, each calculating a dot product among their weights and a small region they are associated to in the input volume. This might result in volume such as [44x44x64] if we decided to use 64 filters with 64 filters of size 5*5, stride 1and padding 0 .so now Total Size is [44 x 44 x 64] and (48-5)/1 + 1 = 44 is the size of the outcome and 64 depths because 1 set denotes 1 filter and there are 64 filters.
- **Layer 2:** *RELU layer* will apply on elementwise activation function, such as the *max* (0, *x*) zero. This leaves the size of the volume unchanged ([44x44x64]), and batch normalization is done.
- **Layer 3:** *Max pool layer* will perform a down sampling operation along the spatial dimensions (width, height), resulting in volume such as [22x22x64]. Max-Pooling with 3×3 filter and stride 2, gives size [22x22x64], i.e. (44-3)/2+1=22 is output size, depth is same as before, i.e. 64 because pooling is done independently on each layer.
- **Layer 4:** Convolution with 64 filters, size 5×5, stride 1, now size is [18x18x64], i.e. (22-5)/1+1=18; is size of output 64 depths because of 64 filters.
- **Layer 5:** Max Poling Layer with 64 filters, size 5×5, stride 1, now size is [18x18x64], i.e. (18+2*1-3) +1=18 original size is restored.
- **Layer 6:** Convolution with 128 filters of size 4x4 and stride 1and we used padding 0, therefore now size is given as [15x15x128], i.e. (18-4)/1+1=15, is size of output 64 and depths of 128 filters.

- **Layer 7:** Fully connected with 3072 neurons. In this layer, each of the 15x15x128=28800 pixels is fed into each of the 3072 neurons and weights determined by back-propagation.
- **Layer 8:** Fully-connected layer calculates the class scores, resultant volume of size [1x1x7], where each of the seven numbers correspond to a class score, such as among the seven classes of emotions. As with normal neural networks and as the name implies, each neuron in this layer will be linked to all the numbers in the previous volume and soft max layer with 3072 neurons.
- **Layer 9:** Soft max layer with 7 neurons to predict 7 classes output.

3.2.3 Mathematical interpretation of Deep Convolutional Neural Network

Feedforward Pass

In the derivation that trails, we will consider the squared-error loss function. For a multiclass problem with c classes and N training examples, this error is given by

$$E^N = 0.5 \sum_{n=1}^{N} \sum_{k=1}^{c} (t_k^n - y_k^n)^2$$

Here is the k-th dimension of the n-th pattern's corresponding target (label), and y k n is similarly the value of the k-th output layer unit in response to the n-th input pattern. For multiclass classification problems, the targets will typically be organized as a "one-of-c" code where the k-th element of t n is positive if the pattern x n belongs to class k. The rest of the entries of t n will be either zero or negative depending on the choice of your output activation function (to be discussed below). Because the error over the whole dataset is just a sum over the individual errors on each pattern, we will consider backpropagation with respect to a single pattern, say the n-th one:

$$E^n = 0.5 \sum_{k=1}^{c} (t_k^n - y_k^n)^2 = 0.5||t^n - \mathrm{y}^n||_2^2$$

With ordinary fully connected layers, we can compute the derivatives of E with respect to the network weights using backpropagation rules of the following form. Let ` denote the current layer, with the output layer designated to be layer L and the input "layer" designated to be layer 1. Define the output of this layer to be

$$x^{\,l} = f\,(u^{\,l}), \quad w \;\; h \quad u^l = W^{\,l} x^{\,l-1} - 1 + b^{\,l}$$

where the output activation function f (·) is commonly chosen to be the logistic (sigmoid) function f (x) = (1 + e −βx) −1 or the hyperbolic tangent function f (x) = a tanh(bx). The logistic function maps [−∞, +∞] → [0, 1], while the hyperbolic tangent maps [−∞, +∞] → [−a, +a]. Therefore, while the outputs of the hyperbolic tangent function will typically be near zero, the outputs of a sigmoid will be non-zero on average. However, normalizing your training data to have mean 0 and variance 1 along the features can often improve convergence during gradient descent [5]. With a normalized dataset, the hyperbolic tangent function is thus preferable. LeCun recommends a = 1.7159 and b = 2/3, so that the point of maximum nonlinearity occurs at f (±1) = ±1 and will thus avoid saturation during training if the desired training targets are normalized to take on the values ±1 [5].

Backpropagation Pass

The "errors" which we propagate backwards through the network can be thought of as "sensitivities of each unit with respect to perturbations of the bias 1. That is to say,

$$\frac{\partial}{\partial} = \frac{\partial}{\partial}\frac{\partial}{\partial} = \delta$$

since in this case $\frac{\partial}{\partial} = 1$. So, the bias sensitivity and the derivative of the error with respect to a unit's total input is equivalent. It is this derivative that is backpropagated from higher layers to lower layers, using the following recurrence relation:

$$\delta^{\,l} = (W^{\,l+1})^T \, \delta^{\,l+1} \quad \cdot \quad f'(u^l)$$

where "." denotes element-wise multiplication. For the error function (1), the sensitivities for the output layer neurons will take a slightly different form:

$$\delta^{\,L} = f'(u^L) \cdot (y^n - t^n).$$

Finally, the delta rule for updating a weight assigned to a given neuron is just a copy of the inputs to that neuron, scaled by the neuron's delta. In vector form, this is computed as an outer product between the vector of inputs (which are the outputs from the previous layer) and the vector of sensitivities:

$$\frac{\partial}{\partial W^l} = \delta^l (x^{l-1})^T$$

$$\Delta W^l = -n \frac{\partial}{\partial W^l}$$

with analogous expressions for the bias update given by (3). In practice there is often a learning rate parameter η ij specific to each weight (W) ij .

Convolution Layers

Let's move forward with deriving the backpropagation updates for convolutional layers in a network. At a convolution layer, the previous layer's feature maps are convolved with learnable kernels and put through the activation function to form the output feature map. Each output map may combine convolutions with multiple input maps. In general, we have that

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} * k_i^l + b_j^l\right)$$

Where M_j represents a selection of input maps, and the convolution is of the "valid" border handling type when implemented in MATLAB. Some common choices of input maps include all-pairs or all triplets, but we will discuss how one might learn combinations below. Each output map is given an additive bias b, however for a particular output map, the input maps will be convolved with distinct kernels. That is to say, if output map j and map k both sum over input map i, then the kernels applied to map i are different for output maps j and k.

Chapter 4

EVALUATION

Evaluation

During the evaluation phase a naive bayes classification and a manual classification was used as a benchmark evaluation for the system.

4.1 Successfully detected emotions

Figure 4.1.1 : Successfully detected Happy emotion with its percentage of happy in given face.

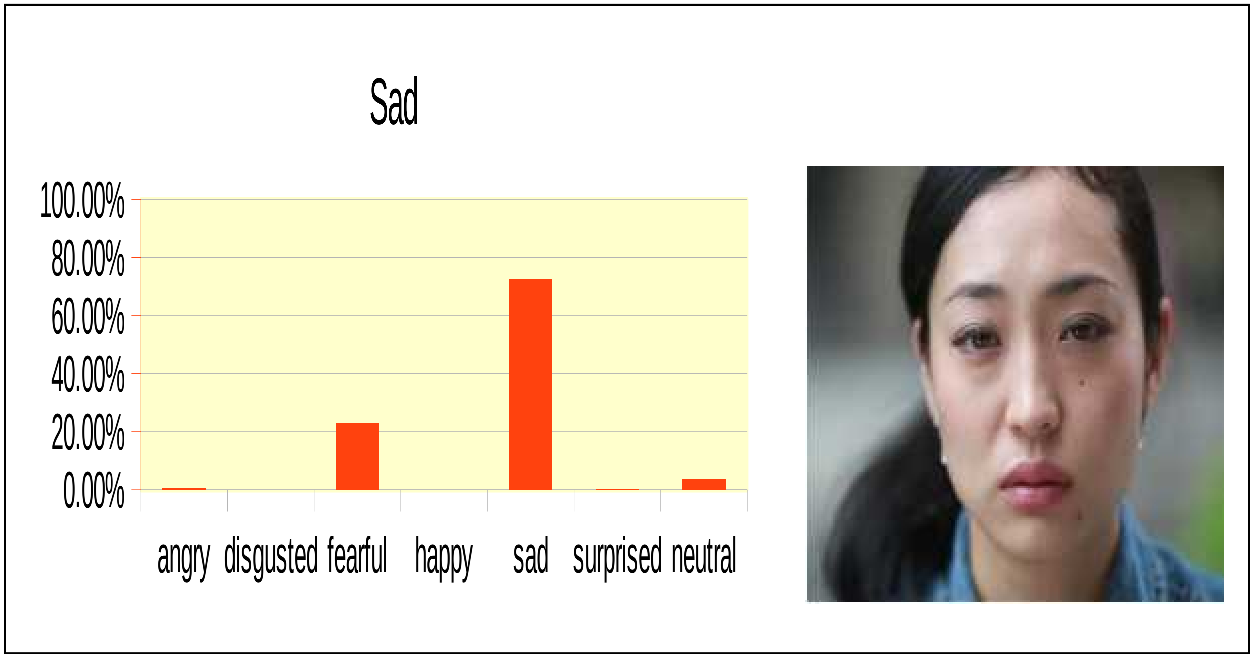


Figure 4.1.2 : Successfully detected Sad emotion with its percentage of Sad in given face.

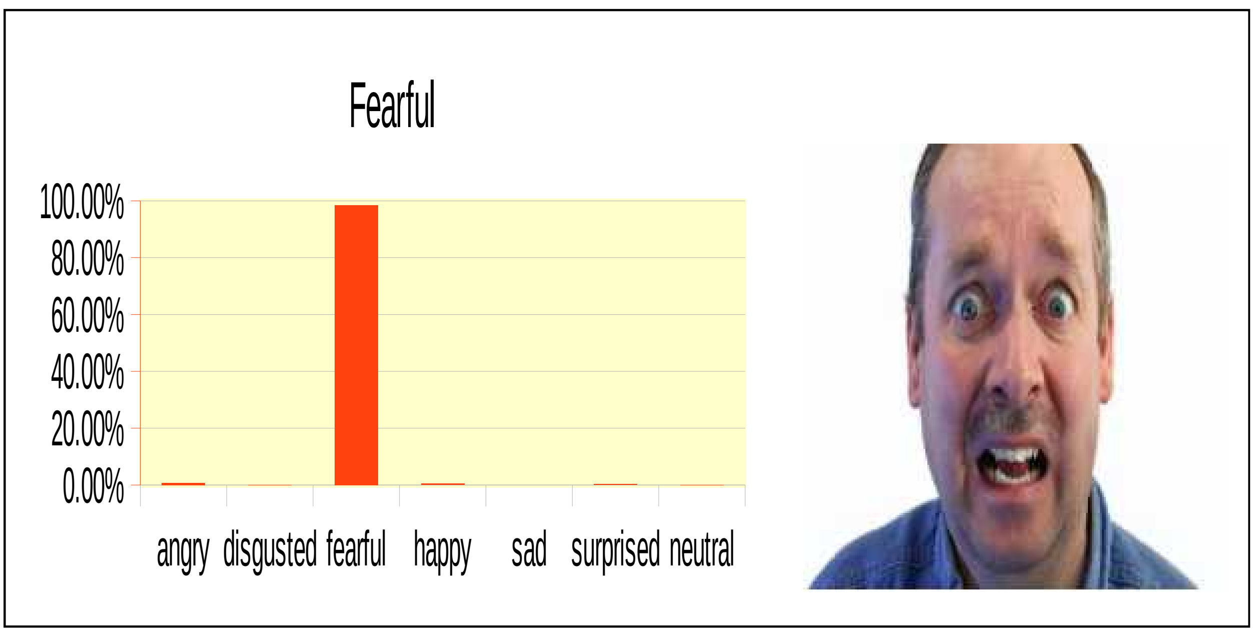


Figure 4.1.3 : Successfully detected Fearful emotion with its percentage of fear in given face.

Figure 4.1.4 : Successfully detected surprised emotion with its percentage of surprise in given face.

Figure 4.1.5 : Successfully detected Angry emotion with its percentage of angry in given face.

Figure 4.1.6: Successfully detected Disgusted emotion with its percentage of disgusted in given face.

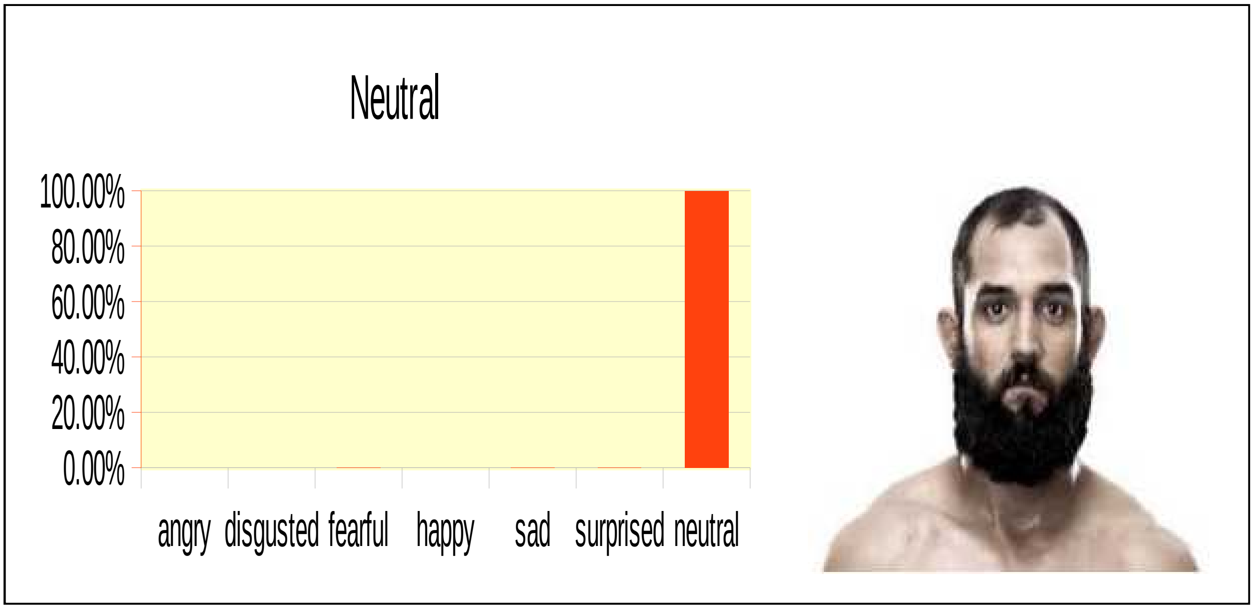


Figure 4.1.7 : Successfully detected Neutral emotion with its percentage of neutral in given face.

4.2 Some failure test cases

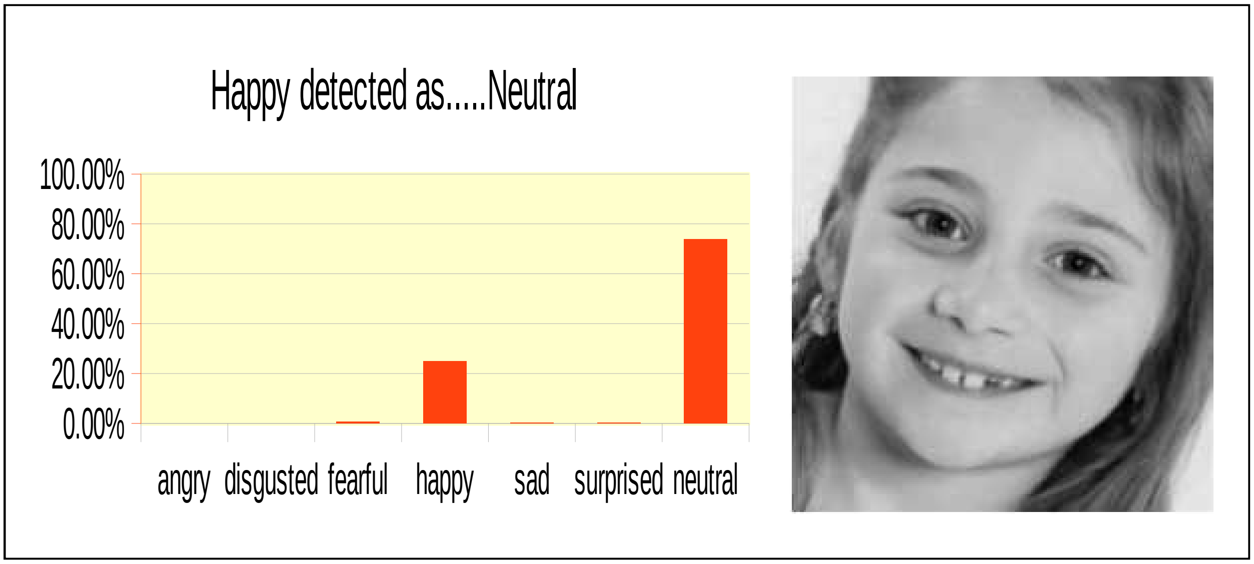


Figure 4.2.1 : Happy detected as Neutral and amount of percentage variations in mismatch.

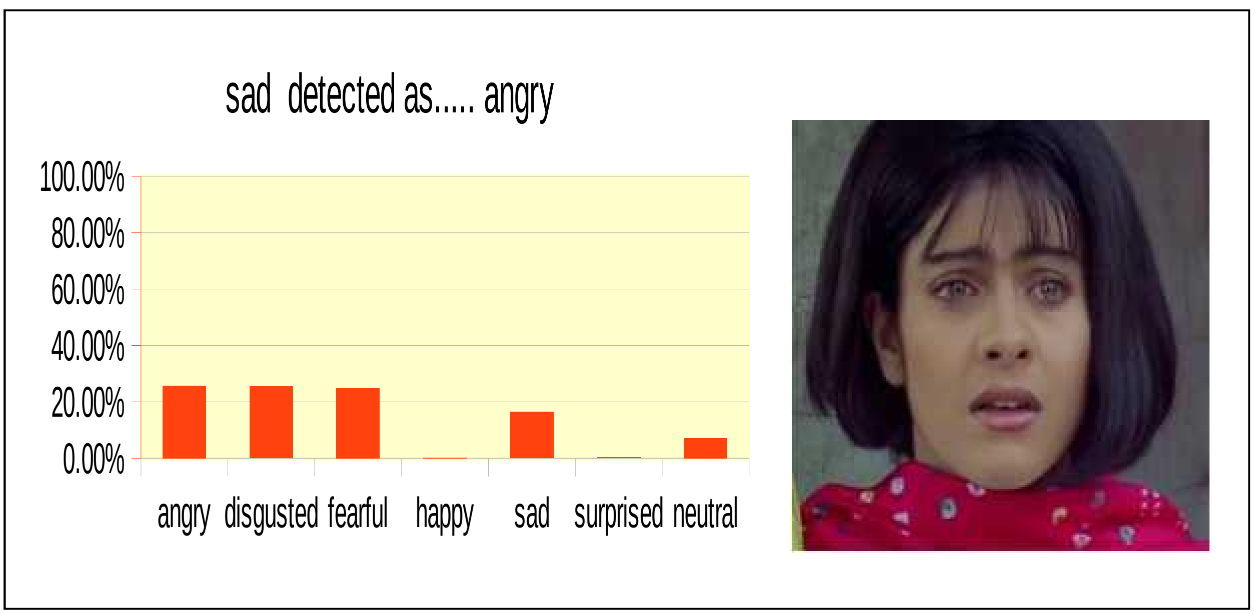


Figure 4.2.2: Sad detected as Angry and amount of percentage variations in mismatch.

Figure 4.2.3: Disgusted detected as Angry and amount of percentage variations in mismatch.

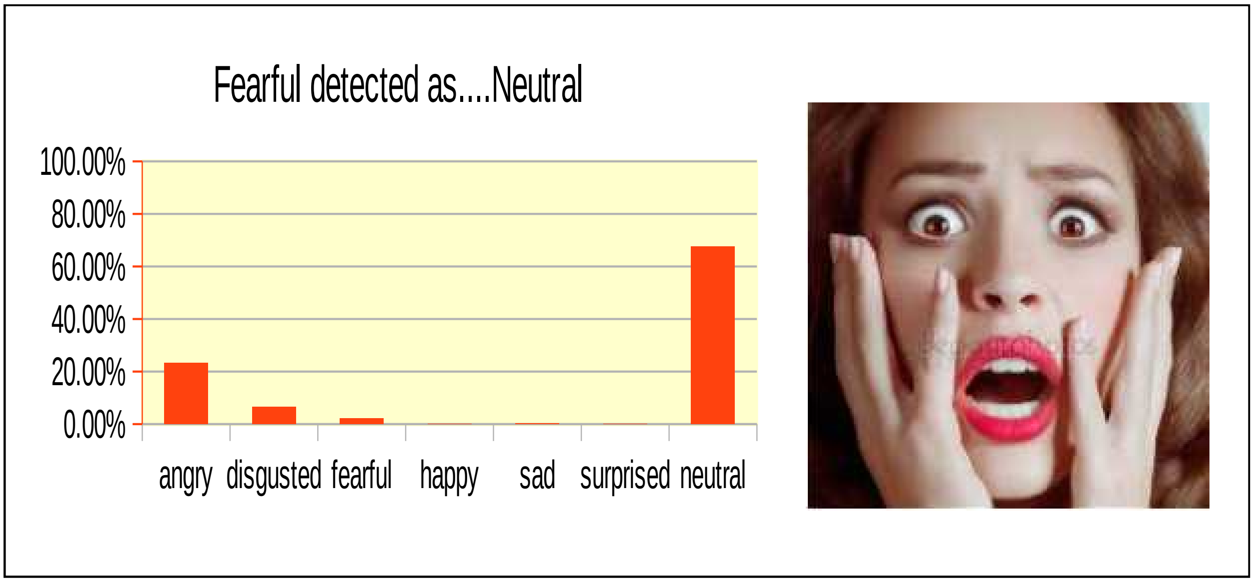


Figure 4.2.4 : Fearful detected as Neutral and amount of percentage variations in mismatch.

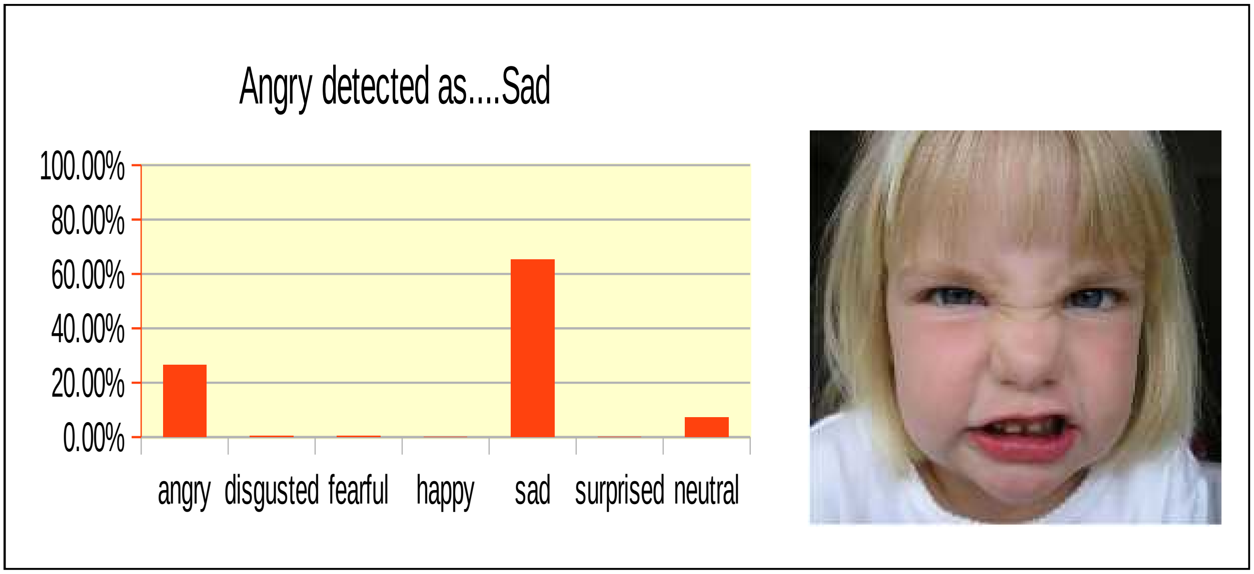


Figure 4.2.5 : Angry detected as Sad and amount of percentage variations in mismatch.

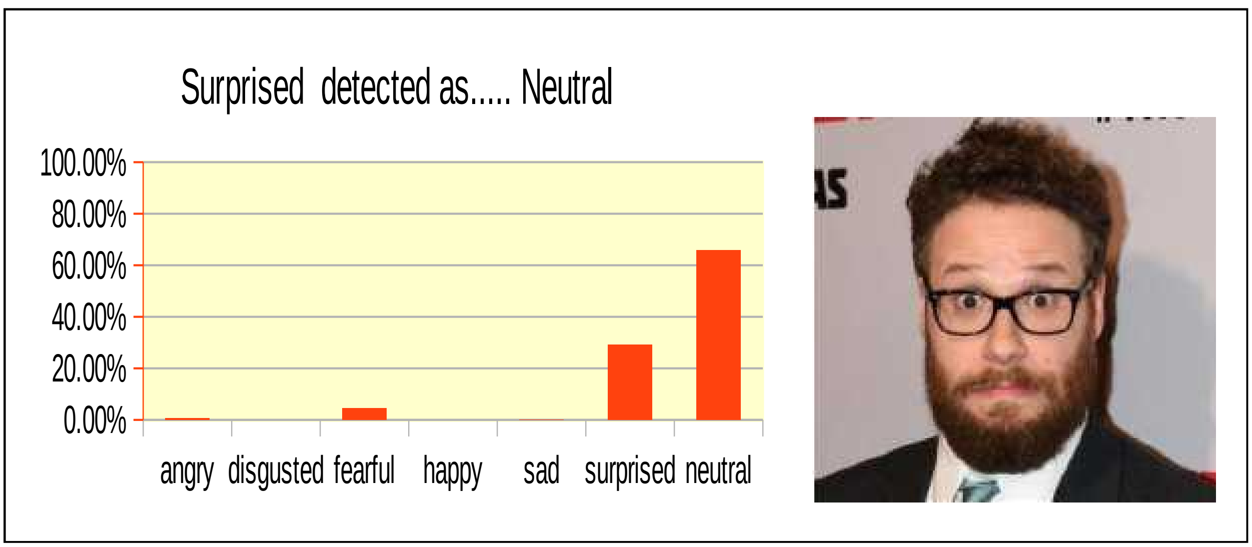


Figure 4.2.6 : Surprised detected as Neutral and amount of percentage variations in mismatch.

The above failures may be due to the dataset imbalance, the (FERC-2013) data set contains non-uniform number of images to different emotions in training set is shown in figure 10. Among 28,709 samples after pre-processing and validation among them we got 11246 valid samples for training Due to drawback of Viola-Jones algorithm [5], [6] most the samples fail during validation.

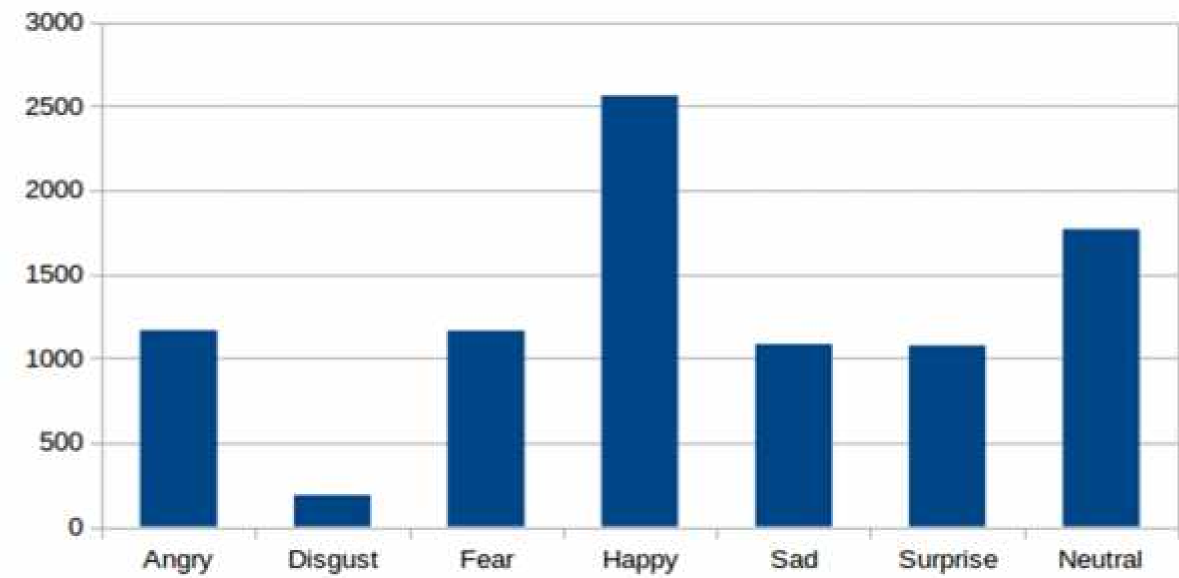
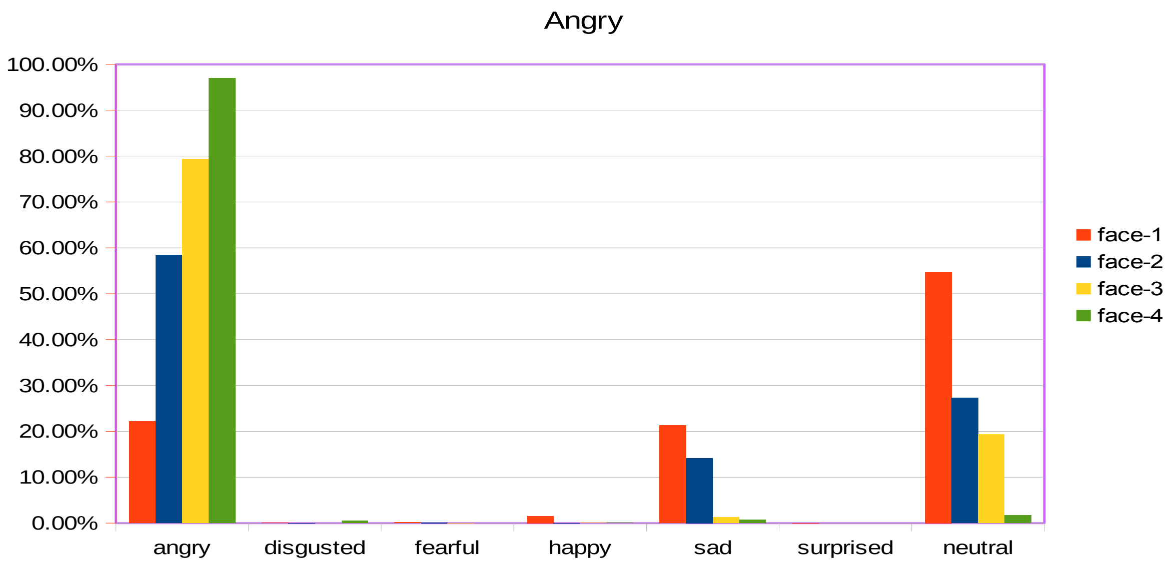


Figure 4.2.7 : Number of sample images for each emotion in FERC-2013 database.

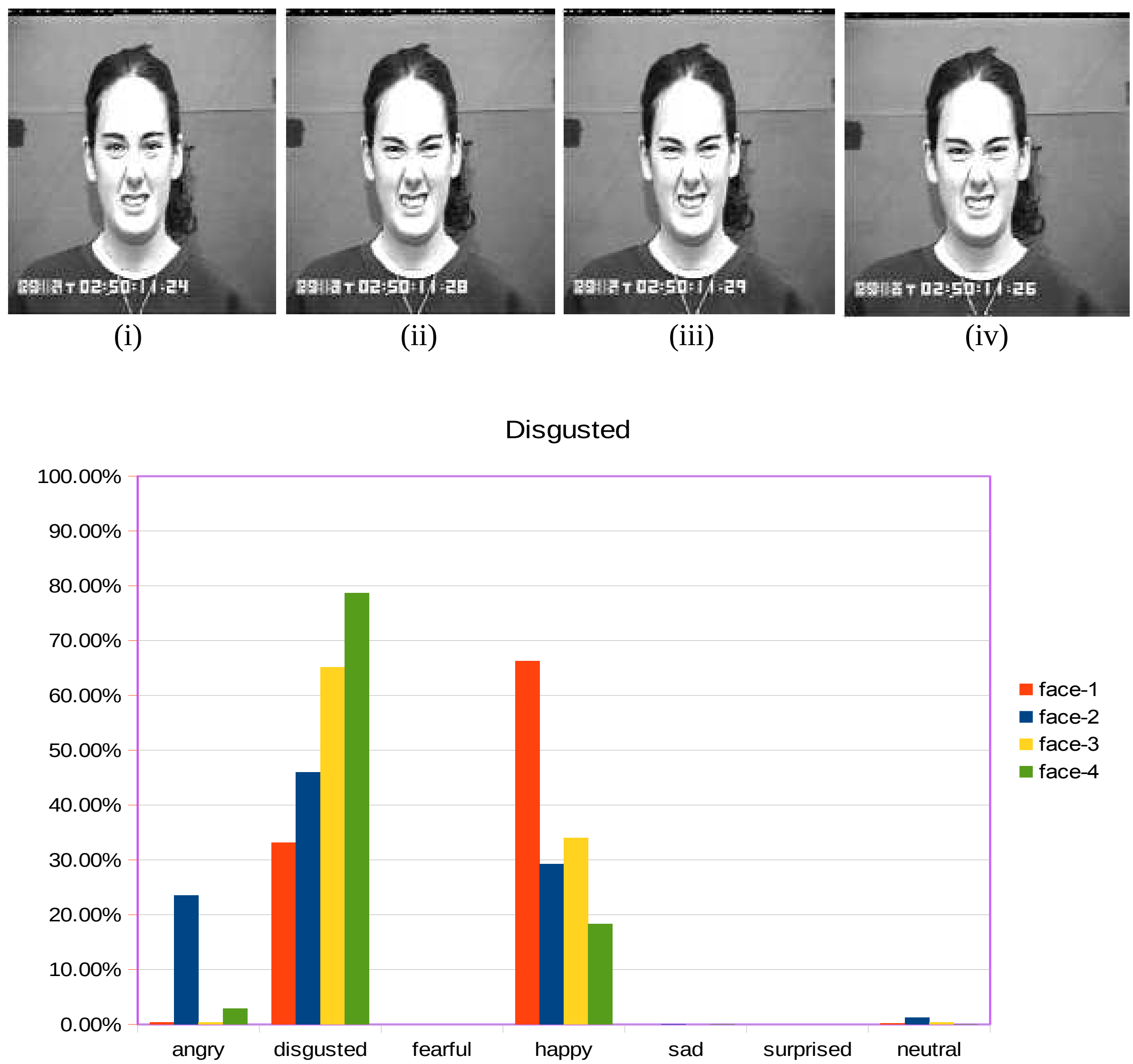
4.3 Emotions at different rate of intensity

(i) (ii) (iii) (iv)



(v)

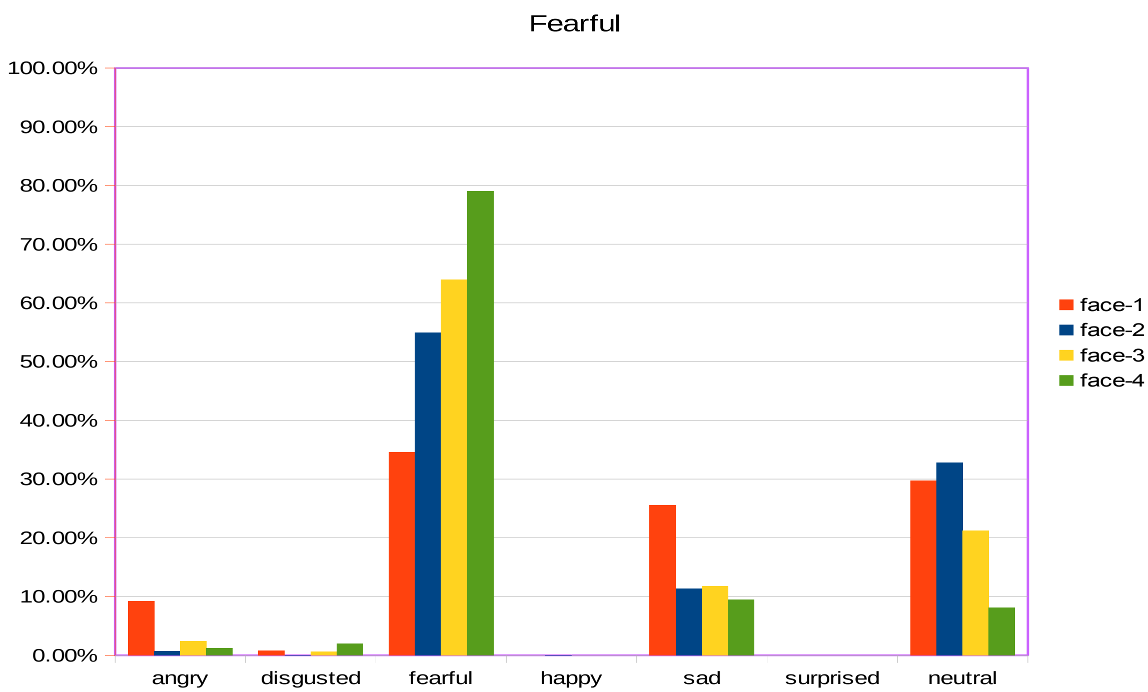
Figure 4.3.1. (i)(ii)(iii)(iv) are showing angry face from low level to its extreme level of angry emotion, and (v) graphical representation of emotion percentages and how other emotions are influencing while emotion level changing from low level to high level.



(v)

Figure 4.3.2. (i)(ii)(iii)(iv) are showing Disgusted face from low level to its extreme level of angry emotion, and (v) graphical representation of emotion percentages and how other emotions are influencing while emotion level changing from low level to high level.

(i) (ii) (iii) (iv)



(v)

Figure 4.3.3. (i)(ii)(iii)(iv) are showing Fearful face from low level to its extreme level of fear emotion, and (v) graphical representation of emotion percentages and how other emotions are influencing while emotion level changing from low level to high level.

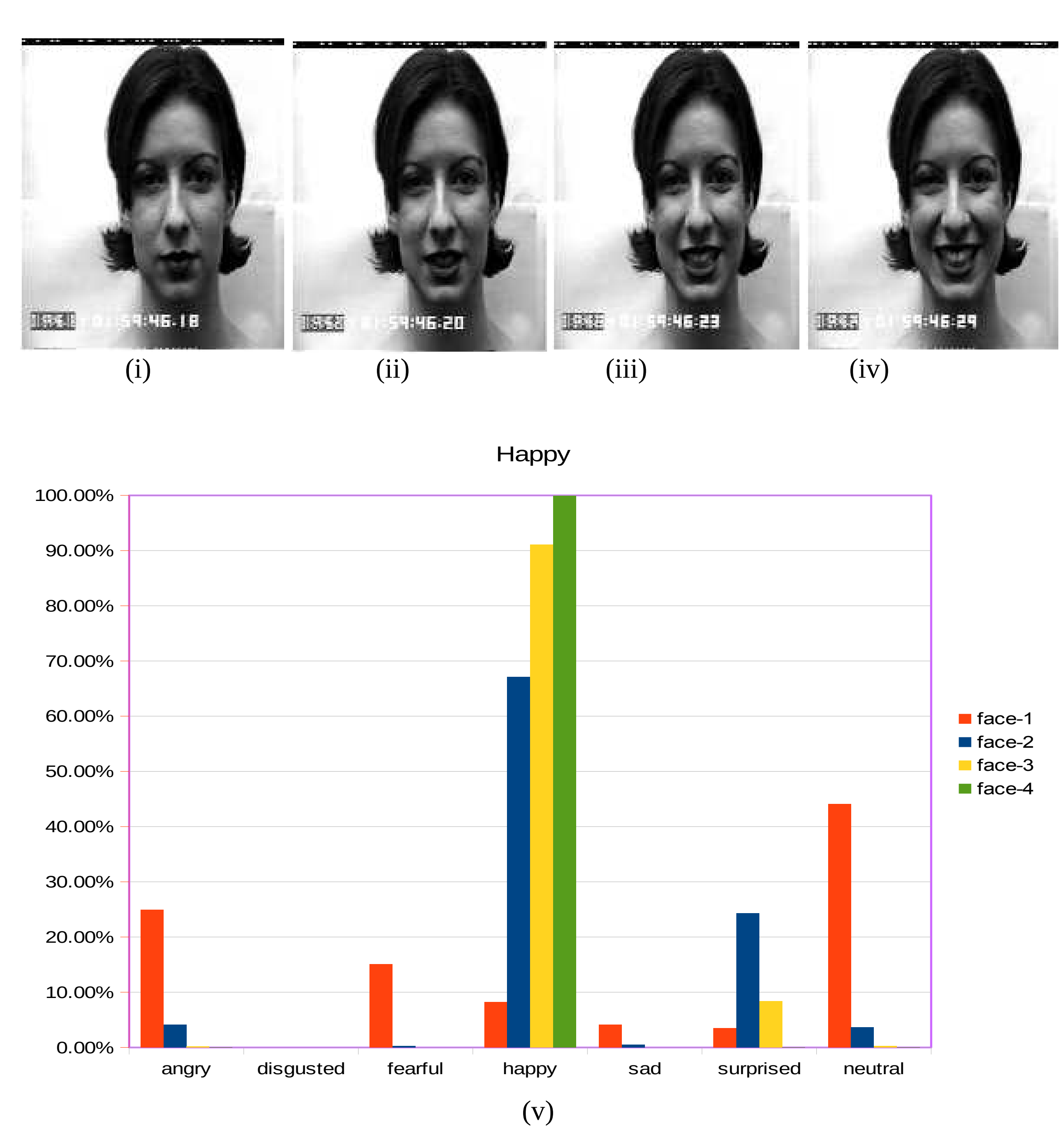
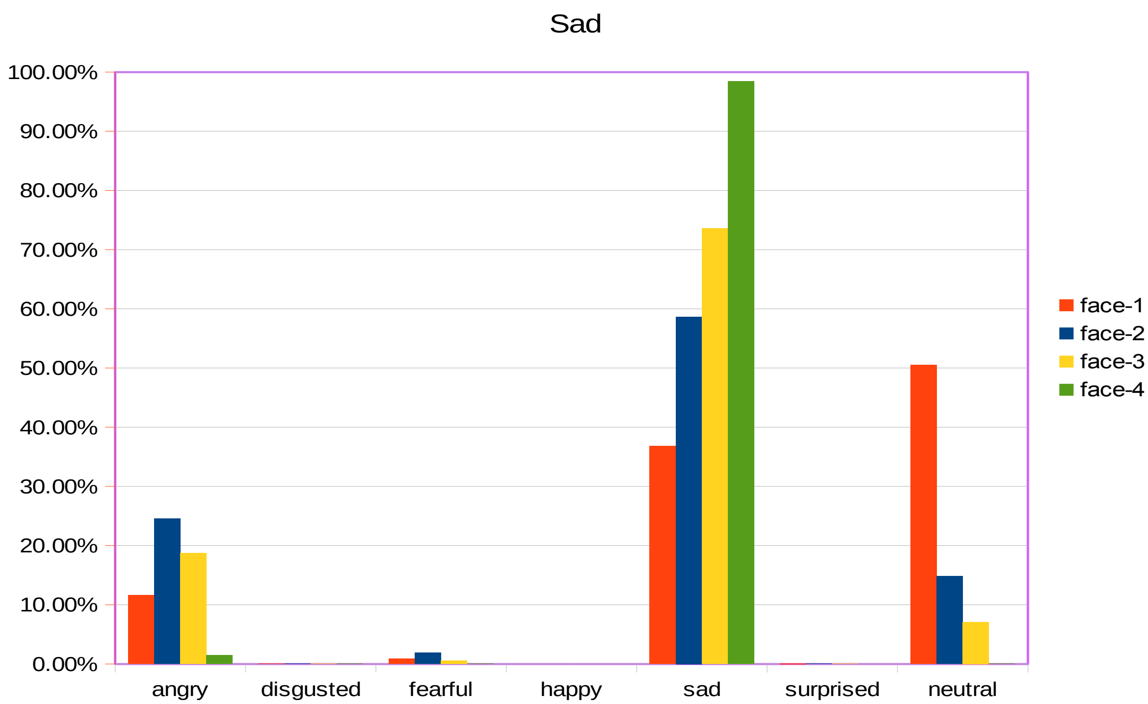


Figure 4.3.4. (i)(ii)(iii)(iv) are showing Happy face from low level to its extreme level of happy emotion, and (v) graphical representation of emotion percentages and how other emotions are influencing while emotion level changing from low level to high level.

(i) (ii) (iii) (iv)



(v)

Figure 4.3.5. (i)(ii)(iii)(iv) are showing Sad face from low level to its extreme level of Sad emotion, and (v) graphical representation of emotion percentages and how other emotions are influencing while emotion level changing from low level to high level.

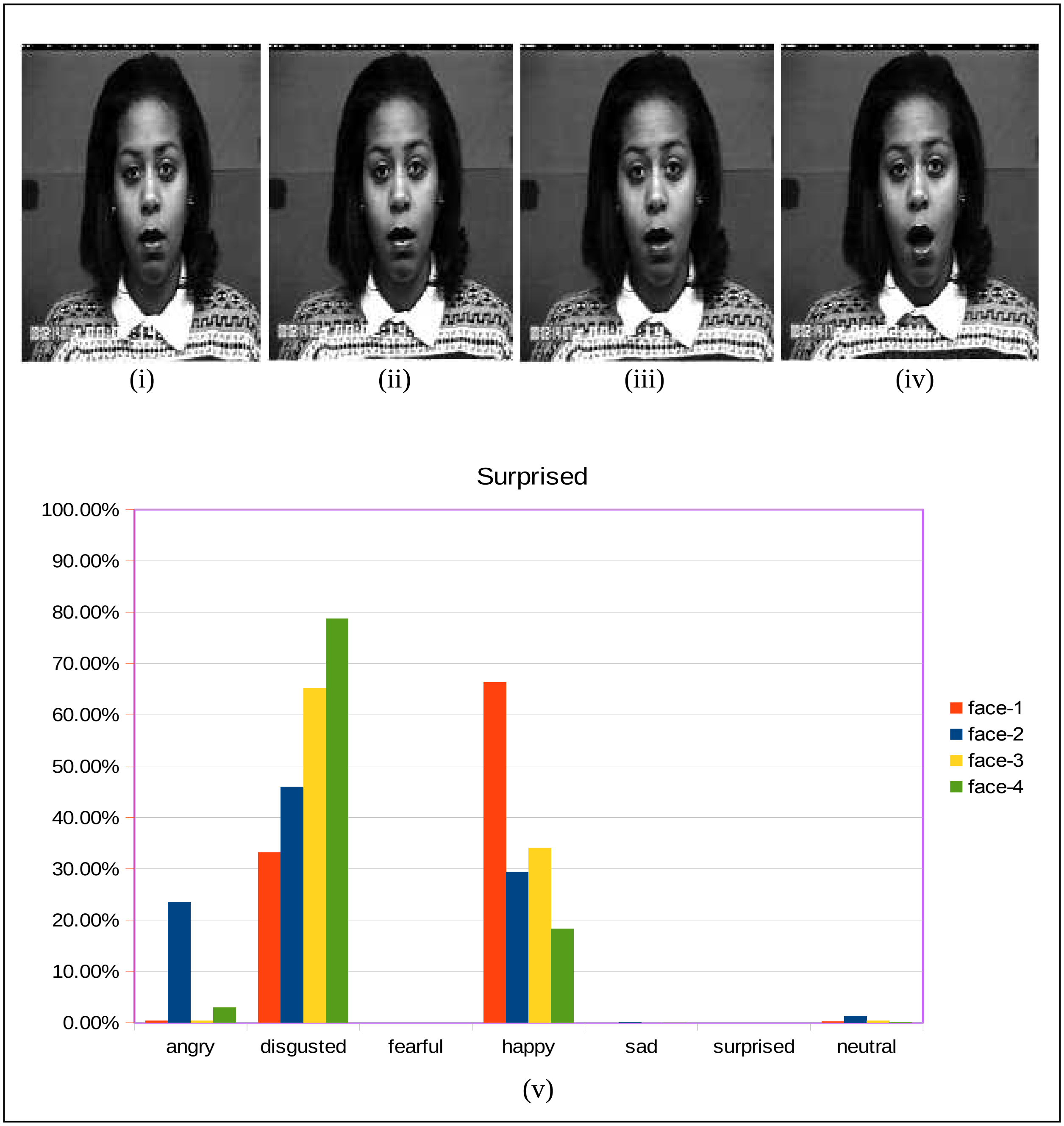


Figure 4.3.6. (i)(ii)(iii)(iv) are showing Surprised face from low level to its extreme level of Surprised emotion, and (v) graphical representation of emotion percentages and how other emotions are influencing while emotion level changing from low level to high level.

4.4 Discriminating real and fake smiles

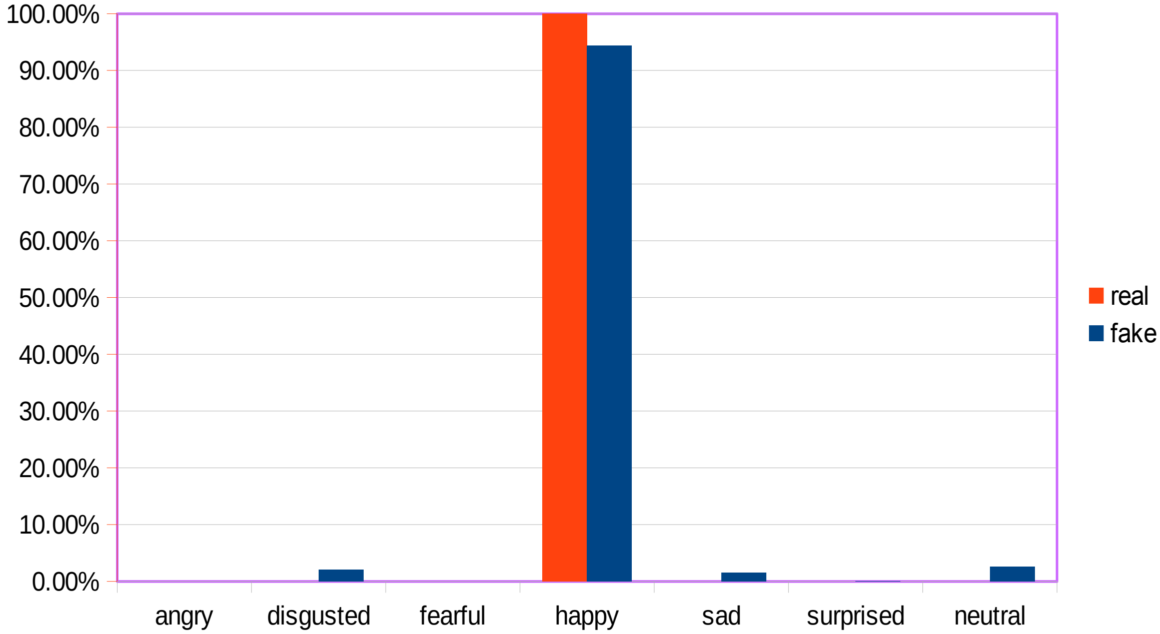
It’s very hard task for human to discriminate genuine and fake smile, but some expert person discriminates genuine and fake smiles observing at some facial muscles and dissimilarities in them. All smiles (Whether a real or a fake one) need that; we bend muscles around the mouth. But the main difference is that, we involve the synchronization of muscles around our eyes (Called, orbicularis oculi) with the mouth area. In a genuine smile, we contract those muscles, dragging in the skin subsequently to our eyes. Teeth or without teeth, does not determine the genuinely of simile. It’s highly dependent on contraction of the muscles around his eyes. That only happens with smiles that reflect true, happy emotions. On the other side, in the fake smile synchronization between these muscles doesn’t work out. When giving a smile deliberately, we exercise a muscle in each cheek, to drag our lips into the right shape, but the eye muscles don’t pact. In the fake smile, there are wrinkles on his cheeks but not around his eyes. The ‘orbicularis oculi’ muscles are not contracted. The skin near the eyes is not tow in tightly as it is in the real image. It is the indication of a fake smile. The dissimilarity in muscle contraction in actual versus fake smiles point up the separation between their prototypes in the brain. When a smile comes naturally to us, one set of muscles is activated. When we use our conscious control to pretend a smile, we change the pattern of muscle establishment, and people around us can catch our forcefully altered emotions.

In this experiment, we are proposing a system which can distinguish real and fake smile based on their variations in percentage of emotions. We are getting an encouraging accuracy in detection of genuine and fake smiles, the main difficult in this project, is to finding the large dataset with real and fake smile. Till date, there is no such databases are available for working for modeling real and fake emotions. We took some available images from the open sources for testing and we are getting comparatively good results. These results may be helpful in detection systems, human understandability, visual surveillance and customer satisfaction etc.

Some of results for differentiating genuine and fake smile are shown below:

(i)

(ii)

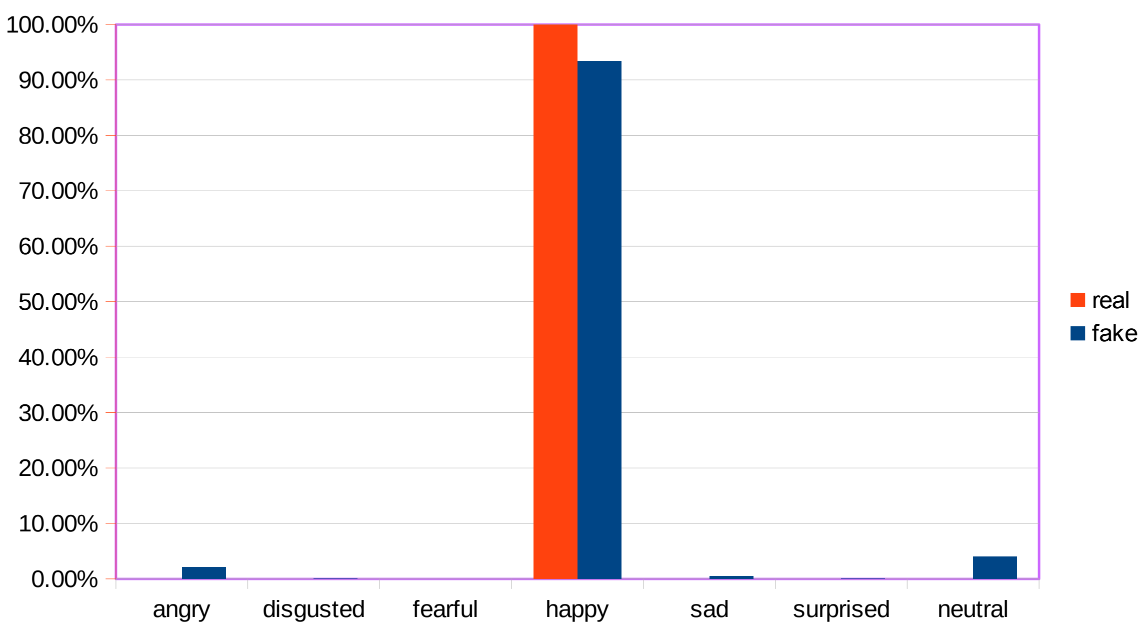


(iii)

Figure 4.4.1 : (i)Genuine smile face (ii) Fake smile face and (iii) Graphical representation of emotion percentages on both faces.

(i)

(ii)

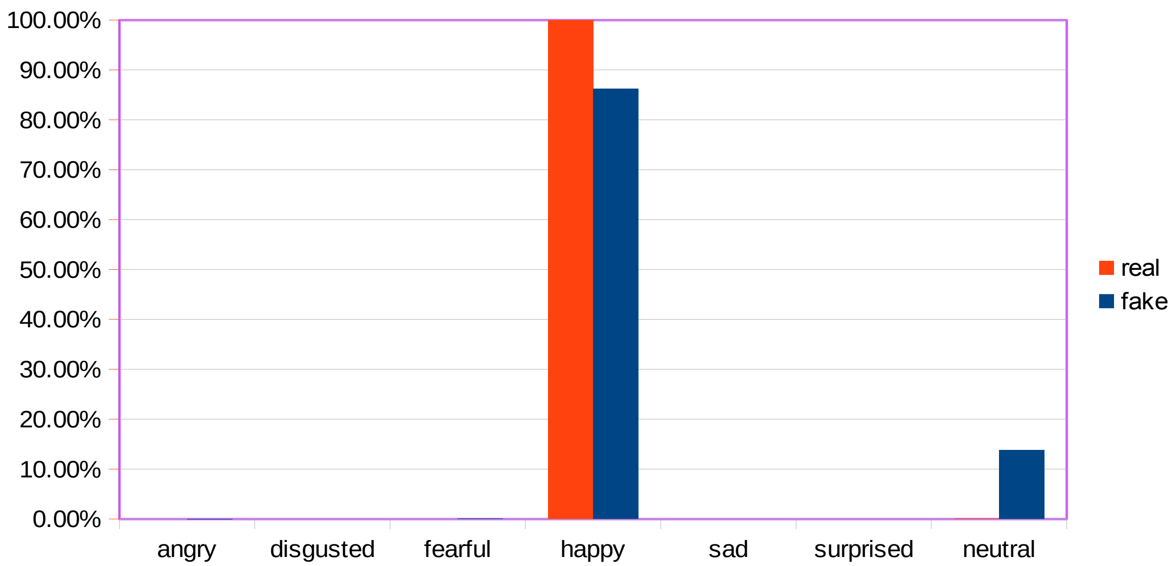


(iii)

Figure 4.4.2: (i)Genuine smile face (ii) Fake smile face and (iii) Graphical representation of emotion percentages on both faces.

(i)

(ii)



(iii)

Figure 4.4.3: (i)Genuine smile face (ii) Fake smile face and (iii) Graphical representation of emotion percentages on both faces.

Chapter 5

DISCUSSION AND FUTURE WORKS

Discussion and Future Works

Accuracy of facial emotion recognition systems using convolutional neural networks is much better than arrived at in the project. Considering that the algorithm takes raw images rather landmark points or FACS labels as input, it performs fairly well. The dataset used in the project was not clean and most of the images failed during training prohibits any general claim about the success or failure of deep learning methods. It is expected that a larger dataset would improve the accuracy of the algorithm and better features would be learned. This comprises a major portion of our future work in this project.

Observing the features one may say that algorithm is able to extract some meaningful features. In the absence of any principled way of discriminating the receptive fields learned by the model it becomes difficult to argue about the ’goodness’ or’ badness’ of a feature other than evaluating the classification accuracy that the feature facilitates. As observed increasing number of hidden layers resulted in a slight improvement in classification, but further increase in hidden layers however deteriorated the results. The number of hidden units in each layer was one of the hyperparameters which wasn’t satisfactorily investigated but an important and somewhat counter-intuitive observation that came up was that the number of hidden units in the first layer should be less than the number of visible units which in other words means that there should be a significant reduction in the amount of information from the visible layer to the first hidden layer. This is appealing because something very similar happens in our visual system where a lot of information is thrown out in successive layers of processing. What this does is that it forces the hidden units to learn the most important features. Led by this observation, we thought that sparsity constraints might lead to even better features and accuracy but as it turned out that there was not any improvement.

Again, this might be attributed to the small dataset we are working with. One of the important results coming out of this project is the observation that low resolution images had better classification accuracy than higher resolution images. Various psychological experiments done on human beings suggest that we make use of mid spatial frequency band for recognizing emotions rather than the high spatial frequency band. Although here, we do not present any quantitative similarities for spatial frequency versus classification accuracy, the few experiments that we performed suggest that lower spatial frequency information is more useful for recognizing emotions which speaks for the cognitive relevance of the model. In our future work, we would like to work quantitative ways of evaluating cognitive importance of features which would help argue for DBNs as a very good model of our visual system.

For real and fake smile discrimination, we are not having the appropriate data to train the networks which results in the very low accuracy of the system in our future work we would like to train the networks with appropriate data and we will extend this project to discriminate all emotions real and fake discrimination.

Chapter 6

CONCLUSIONS

This work try to address the problem of Facial Emotion Analysis and discriminating real and fake smile from an image based and a neural network based approach. A facial expression of emotions determines the state, mood and current feeling of a person through nonverbal communication. We can understand a person emotion if we analyze it in various stages. In different stages the percentages of emotions are significantly varying. In this paper, we have used convolution neural network with 9 layers, for training and classification of 7 types of standard emotions. For better analysis and interpretation of micro expressions, percentages of emotions in various stages have also been measured with our proposed method. FERC-2013 and Extended Cohn Kanade (CK+) databases have been used in this experiment. For detecting the faces Viola Jones algorithm has been applied prior to recognizing emotion. The normal accuracy rates for people prior to training in Matsumoto & Hwang's (Studied Based on American Physiological Association) study were 48%. A real-time emotion recognition system using face data is proposed and developed using convolution neural networks and the accuracy of the system we are getting around 90+ %.

For different reasons, we smile a lot to hide our discomfort, to react to pain or grief or disgust, or sometimes to show that we are sad. There's only one type of smile that's used to convey happiness i.e. genuine smile. A genuine happy smile is naturally one that covers not just the eyes, but the skin around the eyes and the formation of crow's feet. When someone is giving a fake smile, he often concentrates too much on his mouth and its surrounding muscles, but not changing in our experiment, CNN (Convolutional Neural their muscles near to eyes, synchronously.

Networks) has been used for the training and the testing purpose. FERC-2013 database has been taken for the training. Testing has been performed on selected images of FERC-2013 dataset and few taken from the WWW.(World Wide Web). Proposed System is giving encouraging results, even though the system is trained with all standard types of emotions. Till date the specified real and fake databases are not available. Proposed experiment may achieve excellent accuracy if we can train the networks with the real and fake database.

It even has some capability in accurate emotion classifications when poses and lighting conditions vary and also in cases where glasses are present.

LIST OF PUBLICATIONS ON THIS THESIS

[1] **Rajesh Kumar G A**, Ravi Kant Kumar and Goutam Sanyal, "**Facial Emotion Analysis using Deep Convolution Neural Network**", IEEE International Conference on Signal Processing and Communication (ICSPC'17), Karunya university, Tamilnadu, India(Accepted).

[2] **Rajesh Kumar G A**, Ravi Kant Kumar and Goutam Sanyal, "**Discriminating real from fake smile using convolution neural network**", International Conference on Computational Intelligence in Data Science ICCIDS 2017, SSN College of Engineering, Tamilnadu, India (Accepted).

BIBLIOGRAPHY

[1] Darwin, Charles, Paul Ekman, and Phillip Prodger. *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.

[2] Darwin, Charles. *The expression of the emotions in man and animals*. Vol. 526. University of Chicago press, 1965.

[3] Russell, James A., and José Miguel Fernández-Dols. *The psychology of facial expression*. Cambridge university press, 1997.

[4] Russell, James A. "Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies." *Psychological bulletin* 115.1 (1994): 102.

[5] Izard, Carroll E. *Human emotions*. Springer Science & Business Media, 2013.

[6] Izard, Carroll E. "Basic emotions, relations among emotions, and emotion-cognition relations." (1992): 561.

[7] Ekman, Paul, and Harriet Oster. "Facial expressions of emotion." *Annual review of psychology* 30.1 (1979): 527-554.

[8] Forgas, Joseph P. "Mood and judgment: the affect infusion model (AIM)." *Psychological bulletin* 117.1 (1995): 39.

[9] Simon, Herbert A. "Invariants of human behavior." *Annual review of psychology* 41.1 (1990): 1-20.

[10] Lemos, Jakob. "System and method for determining human emotion by analyzing eye properties." U.S. Patent Application No. 11/522,476.

[11] Borod, Joan C. *The neuropsychology of emotion*. Oxford University Press, 2000.

[12] Dimberg, Ulf, Monika Thunberg, and Kurt Elmehed. "Unconscious facial reactions to emotional facial expressions." *Psychological science* 11.1 (2000): 86-89.

[13] Essa, Irfan A., and Alex Paul Pentland. "Coding, analysis, interpretation, and recognition of facial expressions." *IEEE transactions on pattern analysis and machine intelligence* 19.7 (1997): 757-763.

[14] Mai, Xiaoqin, et al. "Eyes are windows to the Chinese soul: Evidence from the detection of real and fake smiles." *PloS one* 6.5 (2011): e19903.

[15] Bernstein, Michael J., et al. "A preference for genuine smiles following social exclusion." *Journal of Experimental Social Psychology* 46.1 (2010): 196-199.

[16] Bernstein, Michael J., et al. "Adaptive responses to social exclusion: Social rejection improves detection of real and fake smiles." *Psychological Science* 19.10 (2008): 981-983.

[17] Calvo, Manuel G., et al. "Attentional mechanisms in judging genuine and fake smiles: Eye-movement patterns." *Emotion* 13.4 (2013): 792.

[18] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." *International journal of computer vision* 57.2 (2004): 137-154.

[19] Hsu, Rein-Lien, Mohamed Abdel-Mottaleb, and Anil K. Jain. "Face detection in color images." *IEEE transactions on pattern analysis and machine intelligence* 24.5 (2002): 696-706.

[20] T. Pham and M. Worring. Face detection methods: A critical evaluation. ISIS Technical Report Series, University of Amsterdam, 11, 2000.

[21] M. H. Yang, N. Ahuja, and D. Kriegman. A survey on face detection methods, 1999.

[22] M. H. Yang, D. J. Kriegman, S. Member, and N. Ahuja. Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24:34{58, 2002.

[23] H. Gu, G. Su, and C. Du. Feature points extraction from faces. 2003.

[24] R. Brunelli and T. Poggio. Face recognition: features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10):1042{1052, 1993.\

[25] T. M. Mahmoud. A new fast skin color detection technique. In Proceedings of world academy of Science, Engineering and Technology, 33, 2008.

[26] M. R. Girgis, T. M. Mahmoud, and T. Abd-El-Hafeez. An approach to image extraction and accurate skin detection from web pages. International Journal of Computer Science and Engineering, 1(2).

[27] H. Gu, G. Su, and C. Du. Feature points extraction from faces. 2003.

[28] C. Sanderson and K. K. Paliwal. Fast feature extraction method for robust face verication.Electronics Letters, 8:1648 { 1650, 2002}.

[29] I. Craw, D. Tock, and A. Bennett. Finding face features. In ECCV '92: Proceedings of the Second European Conference on Computer Vision, pages 92{96, London, UK, 1992. Springer-Verlag.

[30] A. Raouzaiou, S. Ioannou, K. Karpouzis, N. Tsapatsoulis, S. Kollias, and R. Cowie. An intelligent scheme for facial expression recognition. O. Kaynak (ed.), Articial Neural Networks and Neural Information Processing, Lecture notes in Computer Science 2714, Springer, pp. 1109 1116, 2003., 2003.

[31] I. King. Localized principal component analysis learning for face feature extraction. In Proc. Workshop 3D Computer Vision, pages 124{128, 1997.

[32] Z. Xue, S. Z. Li, and E. K. Teoh. Facial feature extraction and image warping using pca based statistic model. In In Proceedings of 2001 International Conference on Image Processing. Thessaloniki, pages 7{10, 2001.

[33] R. S. Feris, J. Gemmell, and K. Toyama. Hierarchical wavelet networks for facial feature localization. In In ICCV'01 Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, pages 118{123, 2002.

[34] P. Debevec. A neural network for facial feature location. 1992.

[35] M. Dumas. Emotional expression recognition using support vector machines. Technical report, Machine Perception Lab, Univeristy of California, 2001.

[36] N. Sebe, Y. Sun, E. Bakker, M. S. Lew, I. Cohen, and T. S. Huang. Towards authentic emotion recognition.

[37] H. Seyedarabi, A. Aghagolzadeh, and S. Khanmohammadi. Facial expressions animation and lip tracking using facial characteristic points and deformable model. International Journal of Information Technology, (4).

[38] Lucey, Patrick, et al. "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression." *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010.

[39] FERC 2013, Form 714 – Annual Electric Balancing Authority Area and Planning Area Report (Part 3 Schedule 2). 2006–2012 Form 714 Database,Federal Energy Regulatory Commission (2013)

[40] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, 2001.