Facial Emotion and Gender Recognition System

[1]Hira Khan
Department of Information Technology
M.H Saboo Siddik College of Engineering,
Mumbai,India
hira.6117015.it@mhssce.ac.in

[2]Juveriah Khan
Department of Information Technology
M.H Saboo Siddik College of Engineering,
Mumbai,India
juveriah.6117017.it@mhssce.ac.in

[3]Shaza Khot
Department of Information Technology
M.H Saboo Siddik College of Engineering,
Mumbai,India
shaza.6117026.it@mhssce.ac.in

Prof.Shrinidhi Gindi
Department of Information Technology
M.H Saboo Siddik College of Engineering,
Mumbai,India
shrinidhi.gindi@mhssce.ac.in

***Abstract*—Emotions plays a very important role in expressing our thoughts. Facial expressions can display emotions and indicate an individual's intentions within a situation. They are extremely important to the social interaction of individuals. Various research has proved that almost 90 percent of our communication can be non verbal. Our overall body language along with our emotions says a lot about our thought process. Hence recognizing emotions becomes a vital task. There a six types of emotions namely, happy, sad, anger, smile, neutral and fear. To effectively identify these would help to decipher the thinking. Hence the importance and need of recognizing emotions is necessary. Also, gender plays an equally important part. Recognizing and effectively analyzing these human characteristics is becoming important with time. This study helps in the field of medical as well as security and surveillance.It is also helpful in better customer service and future predictions.**

Keywords—Machine Learning(ML),Human Computer Interaction(HCI)

I. INTRODUCTION

Human–computer interaction (HCI) studies the design and use of computer technology, it is mainly focused on the interfaces between people (users) and computers. Researchers within the field of HCI observe the ways during which humans interact with computers and style technologies that permit humans interact with computers in novel ways. Emotional Computing [1] aims to enable machines to recognize and synthesize human emotions. As we all know, a change of user's emotion is one of the foundation of communication. Emotional states can motivate human's actions, and can also supplement the meaning of communication Our paper mainly focuses on achieving a successful Human-Computer Interaction and fulfilling our goal of doing project under the domain of Machine learning.Humans have different way of expressing their emotions.Some expresses them through their gestures while some with their words. But the most effective way is by using Facial expressions.Facial expressions says a lot about what a person is really thinking.It is the most visible and genuine way of showing emotion. Facial expression is a nonverbal way of communicating but it says a lot about a person's thoughts. So basically, Emotion plays an important role in human life.

Emotions are used by humans in their daily lives.Understanding and knowing how to react to people's expression greatly enriches the interaction. The field of psychology has played a crucial role in understanding human emotion and developing concepts which will aid these HCI technologies. This system is often significantly useful, nonverbal way for people to speak with one another. The important thing is how fluently the system detects or extracts the countenance from image. The

system is growing attention because this might be widely utilized in many fields like lie detection, medical assessment and human computer interface.

Eye contact is the important phase of communication which provides the mixture of ideas. Eye contact is helpful for maintaining a healthy connection with other person. Eyes express alot of person's feeling. Facial expressions when some unpredicted thing happens and is expressed with eye widening and jaw dropping and this expression can be easily identified. The expression of fear is expressed with surprise expression which is expressed with Outer eyebrow down, inner eyebrow up, mouth open[2].Thus we can see,eyes are the major expressing feature.

The proposed work aims to design a classification system to recognize gender and emotions i.e. expressions of the face images simultaneously. The system proposed in this paper is divided into two parts. The first part is about detection of face from the input frame, while the second part is about the rest of the activities comprising of emotion and gender recognition. The general approach to automatic facial expression analysis consists of three steps: face detection and tracking, Feature extraction and expression classification / recognition. Also, the data set used in this project is fer2013, an open-source data set that was made publicly available for a Kaggle competition. It consists of 48 X 48-pixel grayscale images of the face.

There are seven categories of emotions(0=Angry,1=Disgust,2=Fear,3=Happy,4=Sad,5=Surprise,6=Neutral)present within the info .

Fig.1:Exampleimagesoffer2013dataset

II. PROPOSED SYSTEM

1. Feature recognition system

There are different main stages included in this system such as face image preprocessing, feature extraction, and classification of different emotions includes smile, sad, anger, surprise, and fear. A smile on face shows their happiness and it is expressed through eyes with a curve, cheeks raised and lips corner pulled. The sad expression is that when a person is feeling down which is generally expressed through Outer eyebrow down, inner eyebrows raised, eyes closed and lips Cornered down. The surprise expression is expressed

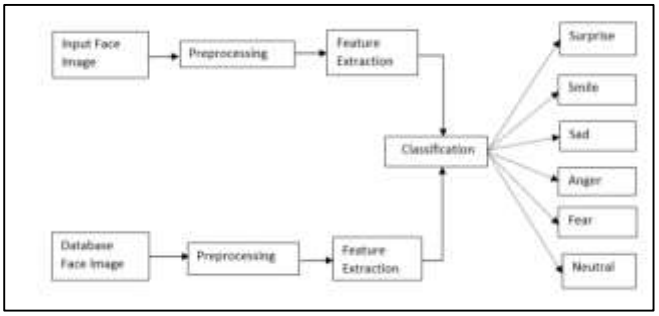


Fig.2:Block Diagram of the features recognition system

Firstly, the input image is taken and the face is detected from the image. When the face is detected important features are extracted from the facial image like eyes, eyebrows, lips, etc. After extracting these important features, the expression is classified by comparing the image with the images in the training dataset using some algorithm.

1.1 INPUT FACE IMAGE

As in the Fig shown above In this stage, The image can be taken as an input. Here, We can take any Human image whether it is in a group or a single person image. In our system, we have taken different actors and random people image which is further shown in the results

1.2 DATABASE FACE IMAGE

In this system, a Dataset name FER2013 is used. It consists of 28,709 examples. This dataset was prepared by Pierre - Luc Carrier and Aaron Courville[10], as part of an ongoing research project. This dataset is stored in a CSV file. Here the target labels are integer encoded in the CSV file. They are mapped as follows:

0=Angry, 1=Fear, 2=Happy, 3=Sad, 4=Surprise, 5=Neutral

It contains the raw pixel values of the images.

1.3 PREPROCESSING

Preprocessing is a process that is used to improve the performance of the system and it is carried out before the feature extraction process [3]. This stage includes different types of processes such as image scaling, contrast adjustment, and additional enhancement processes[4] to improve the

expression frames[5]. Here, the cropping and scaling processes were performed on the face image where the nose of the face was taken as midpoint whereas other important facial components are included physically[6]. Normalization is one of the preprocessing methods which issue for reduction of illumination and variations of the face images with the median filter and to achieve an improved face image[7]. The normalization method is also used for the extraction of eye positions in the FER system as it provides more clarity to the input images. Localization is also a preprocessing method and it uses the Viola-Jones algorithm[8] for detecting the facial images from the input image.

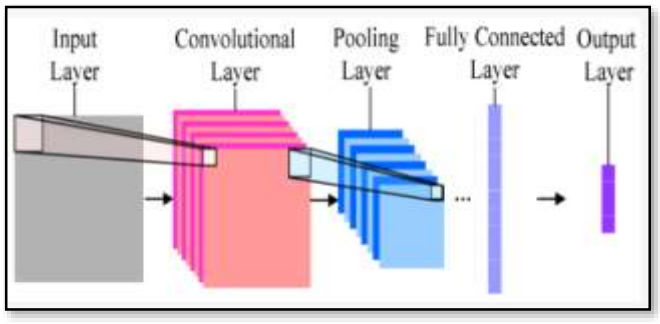


Fig.3:Convolutional Neural Network

1.4 FEATURE EXTRACTION

Feature extraction is the next stage of this system. It is finding and depicting positive features of concern within an image for further processes. In image processing, feature extraction is one of the significant stages, as it spots the move from graphic to implicit data depiction. Then the data depiction can be used as an input to the classification. The feature extraction methods can be categorized into five different types such as texture edge-based method, feature-based method, global and local feature-based method, geometric feature-based method, and patch-based method[9].

1.5 CLASSIFICATION

Classification is the final stage of this system where the classifier categorizes the expression such as happy, sad, surprise, anger, fear, disgust, and neutral. The dimensionality of data obtained from the last process which is the feature extraction method is very high so it is reduced using classification. Features should take different values for objects belonging to different classes so that classification can be done. Here emotions are classified as happy, sad, angry, surprise, neutral, disgust, and fear with 28,709 images for the training dataset, and here Each emotion is expressed with different facial features like eyebrows, opening the mouth, Raised cheeks, wide-open eyelids, wrinkles around the nose and many others. Training the large dataset for better accuracy and result is the object class for an input image. Based on those features convolution layers is performed.

2. Convolutional Neural Network

CNN's are the category of Neural Networks that are very effective in the recognition and classification of images. CNN's are made of many layers. CNN's consist of neurons or kernels that have learnable weights or parameters. Each filter takes some of the inputs and performs convolution and may follow it with a non-linearity. This structure of CNN contains Convolutional, pooling, Rectified Linear Unit (ReLU), and Fully Connected layers.

A. Convolutional Layer:

The convolutional layer performs here the core building block of a Convolutional Network which does most of the computational heavy lifting. The primary purpose of the Convolution layer is to extract features from the input data which is an image and it plays an important role to preserves the spatial relationship between pixels by learning the image's features using small squares of the input image. The input image can be convoluted by using a set of learnable neurons. This produces a feature map in the output image and after this, the feature maps are fed as input data to the next convolutional layer.

B. Pooling Layer:

The pooling layer helps to reduce the dimensionality of each activation map but it continues to have the most important information. In this, the input images are divided into a set of non-overlapping rectangles. Each region of images is down-sampled by a non-linear operation such as average or maximum. This layer achieves better generalization, faster convergence, robust to translation and distortion, and is usually placed between different convolutional layers.

C. ReLU Layer:

ReLU is a non-linear operation and includes units that employ the rectifier. It is applied to each and every pixel and it reconstitutes all negative values in the feature map by zero. Here, to understand how the ReLU operates, we may assume that there is a neuron input given as x, and from that, the rectifier is defined as $f(x)= \max(0,x)$ in the literature for neural networks.

D. Fully Connected Layer:

Fully Connected Layer (FCL) term refers to all the filter in the previous layer that is connected to every filter in the next layer. The output from the convolutional, pooling, and ReLU layers are high-level features of the input image. The goal of employing these layers is to employ these features for classifying the input image into various classes based on the training dataset.

3. Gender Recognition

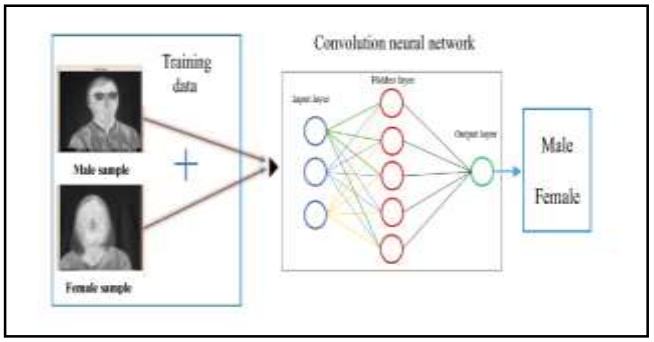


Fig.4:Gender Recognition diagram

For Gender Recognition a trained dataset of IMDb-wiki is used. This dataset was created automatically by crawling IMDb (Internet Movie Database) and Wikipedia. The authors searched the most popular actors on IMDb getting the date of birth and gender along with all the images associated with the actor. They searched for 20,284 celebrities and 460,723 labeled images, 57% of which are male faces. Similarly, they searched for Wikipedia pages of famous people and got the same data. The Wikipedia subset contains 62,328 images, 75% of which are male faces. The number of images included in the IMDB-WIKI dataset is 523,051. This is the largest open-source data set of the order, which gives the person in the image of sex and birth time, photoshoot time, and other information.

This trained model contains two classes M which indicates the male and F which indicates the Female.

The following data augmentation methods are performed to train the data:

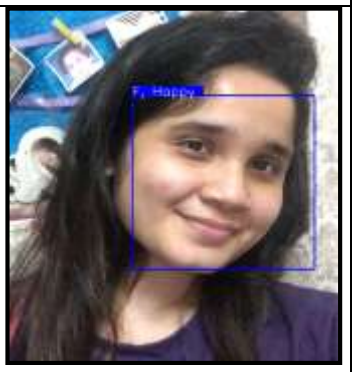
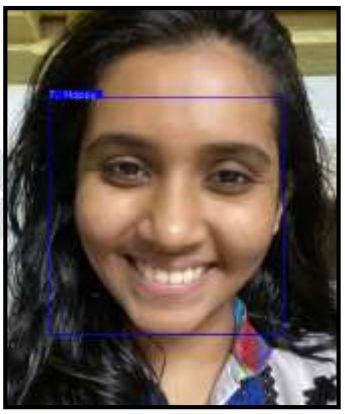
Rescale —This will rescale the image to provide a format understood by TensorFlow, and it uses this to convert our image to a matrix of floats our model will expect

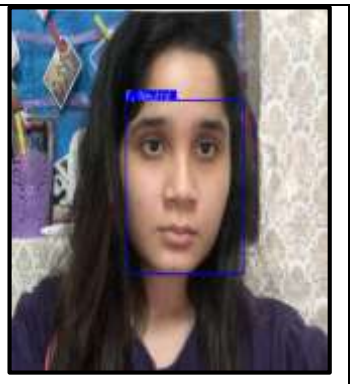
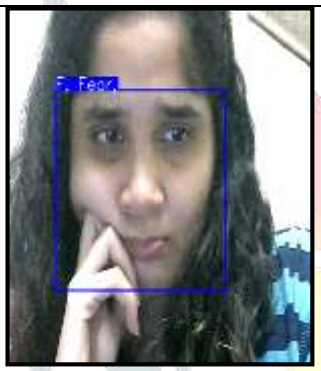
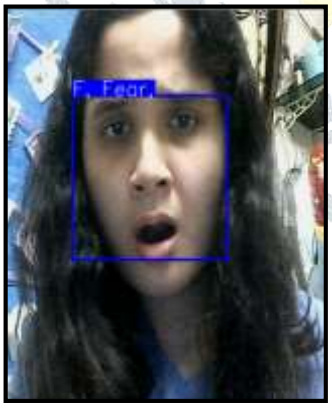
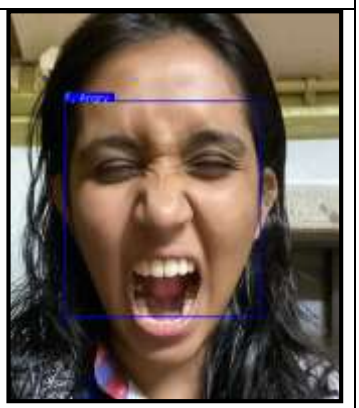
Shear Range — The shear angle in a counter-clockwise direction in degrees

Zoom Range — The range for random zoom

Horizontal Flip — This randomly flip inputs horizontally

III. RESULTS

Emotion	Motion of Facial parts	Output
Happy	Open eyes, open mouth, lip corner pulled, cheeks raised	 
Surprise	Eyebrow up, open eyes, jaw dropped	
Sad	Outer eyebrow down, inner eyebrow raised, eyes slightly closed, lip corner down	

Neutral	Eyebrows straight, lips straight, eyes opened	 
Fear	Outer eyebrow down, inner eyebrow up, open eyes, mouth opened or closed	 
Anger	Eyebrows pulled down, closed eyes, mouth opened	

IV.LIMITATIONS

The limitation of this system is, is not fully automated .We do not get full accuracy from this system.This system is limited to detect some of basic facial expression like fear ,anger, joy, surprise.Totally based on physical expression.One of it's drawback is,it is unable to recognize transgender people.

V.FUTURE WORK

We are working towards a machine with emotions. A machine or a system, which can think like humans, can feel the warmness of heart; can judge on events, prioritized between choices, and with many more emotional epithets. To make the dream reality first we need the machine or system to understand human emotions, ape the emotion and master it. We just started to do that. Though there is some real example exists these days. Some features and services are getting popularity like Microsoft Cognitive Services but still, there are a lot of works required in the terms of efficiency, accuracy, and usability. Therefore, in the future Emotion Recognition is an area that requires great intentness**.** Different problems would require different network architecture and a lot of trial and errors to produce desirable validation accuracy.

Having examined techniques to cope with expression variation, in future it may be investigated in more depth about the face classification problem and optimal fusion of color and depth information. Further study can be laid down in the the direction of the allele of gene matching to the geometric factors of the facial expressions. The genetic property evolution framework for a facial expressional system can be studied to suit the requirement of different security models such as criminal detection, governmental confidential security breaches etc.

VI.CONCLUSION

The human expression recognition system presented in this research work contributes a resilient face recognition model based on the mapping of behavioral characteristics with physiological biometric characteristics. The physiological characteristics of the human face with relevance to various expressions such as happiness, sadness, fear, anger, and surprise are associated with geometrical structures which are restored as the base matching template for the recognition system. The gene training set evaluates the expressional uniqueness of individual faces and provides a resilient expressional recognition model in the field of biometric security.

Experimental analysis and study show that the hierarchical security structures effective in geometric shape identification for physiological traits.

REFERENCES

[1] R. Picard. Affective computing. MIT Press, 1997

[2] I.Michael Revina, W.R. Sam Emmanuel.A Survey on Human Face Expression Recognition Techniques,Journal of King Saud University - Computer and Information Sciences,2018

[3] A. Poursaberi, H.A. Noubari, M. Gavrilova, S.N. Yanushk evich Gauss – Laguerre wavelet textural feature fusion with geometrical information for facial expression identification EURASIP J. Image Video Process. (2012)

[4] S. Bashyal, G.K.V. Venayagamoorthy Recognition of facial expressions using Gabor wavelets and learning vector quantization Eng. Appl. Artif. Intell., 21 (2008)

[5] P. Taylor, M.H. Siddiqi, R. Ali, A. Sattar, A.M. Khan, M.H. Siddiqi, R. Ali, A. Sattar, A.M. Khan, S. Lee Depth camera-based facial expression recognition system using multilayer scheme IETE Tech. Rev., 31 (2014)

[6] L. Zhang, S. Member, D. Tjondronegoro Facial expression recognition using facial movement features IEEE Trans. Affect. Comput., 2 (2011)

[7] Y. Ji, K. Idrissi Automatic facial expression recognition based on spatiotemporal descriptors Pattern Recognit. Lett., 33 (2012)

[8] S. Noh, H. Park, Y. Jin, J. Park Feature-adaptive motion energy analysis for facial expression recognition Int. Symp. Vis. Comput. (2007)

[9] L. Zhang, S. Member, D. Tjondronegoro Facial expression recognition using facial movement features IEEE Trans. Affect. Comput(2011)

[10] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. Technical Report arXiv:1206.5538, U. Montreal (2012)