Efficient Face Recognition through Sketch using CNN

Mubasheer Altaf Shaikh
Dept. Of Electronics and Telecommunication
AISSMS's IOIT
Pune,India

Nayan Suresh Pangare
Dept. Of Electronics and Telecommunication
AISSMS's IOIT
Pune, India

***Abstract* – Most of the time when a crime is committed there is usually a small set of witnesses that have seen the perpetrators face which is then used to find the assailant. This is highly useful in the event of a crime or a terrorist attacks taking place which can help the law enforcement to recognise the criminal through the sketch. For this purpose, there are multiple approaches that have been provided by the researchers. Most of these approaches utilize the machine learning paradigm. But the problem with these techniques is that most of them have a very high computational complexity that is observed. Some of the techniques also have an increased time complexity that is unacceptable in such a critical application. Therefore, this publication outlines an innovative technique that leverages the machine learning paradigm such as the Convolutional neural networks. The presented technique implements Sobel Edge Formation and Convolutional Neural Networks to achieve Facial Recognition through Sketch input.**

Keywords:*Sketch Images, Facial Recognition, Sobel Edges, Convolution Neural networks.*

I. INTRODUCTION

As it is can be noticed that there is an increased number of smartphones that are being manufactured at an alarming rate. There are already millions of smartphones in the market and the increasing affordability and the increase in the buying power of the masses have been catalyzing this sector enormously. With the introduction of an increasing number of smartphones, and continuous research and development has led to the introduction of cameras on almost every smartphone that is being manufactured. This allows the user to take a number of images and store them as memories from their smartphones. There has been a significant increase in the number of people utilizing the cameras of their smartphones regularly.

The increasing number of smartphones and the increasing population has been responsible for the majority of the smartphone to have an inbuilt camera. Most of these cameras are used for the purpose of taking images also some of the smartphones have deployed the paradigm of facial recognition in their smartphones with varying degrees of success. This is due to the fact that the facial image and feature recognition is a highly complex and difficult field. Facial recognition employs the paradigm of feature extraction that transforms the image into a set of features that can be easily correlated to authenticate the user accurately. The calculation of the features is a highly complicated process that needs to be improved significantly.

The facial images need to undergo the process of feature extraction to derive the various different, unique and defining features of a face to be able to effectively correlate with another image or a person. Therefore, this also allows the various law enforcement agencies to track and identify criminals using their facial features for matching with the security camera footage and other images. this allows law enforcement to accurately pinpoint and identify the criminal with high accuracy and reliability. The facial recognition paradigm plays an integral role in maintaining the law and order of a particular area.

The facial recognition is highly useful as it plays an important role in providing the identity to a particular person. The application of facial recognition in law enforcement agencies has been highly fruitful and has allowed to identify and incarcerate a variety of criminals. The various images gathered of the criminal committing a crime should have their face clearly visible to allow the extraction of the requisite features from the image. For the purpose of matching a viable database of the criminal's facial images must be used to train the system in recognition of those particular features in the incriminating evidence.

Most of the time, the police do not have proper and clear footage of the crime or do not have the resolution of the image high enough for feature extraction. Some of the times the criminals are highly motivated and intelligent to avoid detection and giving away their faces. This makes the process of identification and authentication of the criminal highly difficult. But if there is a witness that has seen the criminals face, the police have expert sketch artists that utilize the description of the criminal through the victim or the witness and draw a sketch matching their description. The problem with that is that most of the feature extraction technique cannot work on the sketch as it lacks any depth information.

This is one of the reasons why the identification and the recognition mechanism must be modified to better suit the requirements of the forensic investigation. A sketch must be analyzed and the matching of the facial features must be highly accurate and exact to allow the investigators to perform a fair and unbiased investigation. For the purpose of achieving this difficult task, there is a need to implement various machine learning and deep learning approaches that can shoulder the massive computational task of acquiring the facial features from a sketch. Therefore, the worthiest candidates for this task are Sobel Edge formation and Convolutional neural networks along with the inclusion of the Pearson Correlation.

The Sobel Edge algorithm is utilized for the identification of the edges of the image. This allows the system to identify the boundaries of the image and various other features easily. The Sobel filter calculates the gradient of the image and the changing intensity within the image at the pixel level. This allows the pixels on the edge to be highly visible with a darker gradient from the rest of the image. The vector is highly useful for the identification of the edges of an image and for the purpose of Facial sketch image recognition it is irreplaceable.

Convolutional Neural Networks is one of the most widely used neural networks. The Neural Networks are computational networks that have been designed with inspiration from the human brain and its working. Therefore, the convolutional neural networks are neural networks that emulate the inner workings of a brain complete with neurons and along with their threshold values for activation. The convolutional neural network employs various different layers that allow for the computation of various tasks mainly image and sound processing tasks that are highly complex in nature. Therefore, the introduction of CNN for the purpose of facial sketch recognition is a valuable asset.

This research article is further classified as Literature Survey in the Section 2. The proposed methodology in the section 3. Results are evaluated in the section 4 and finally the research article is concluded in the section 5 with the traces of the future work.

II LITERATURE SURVEY

S. Shojaeilangari [1] presents the recognition of facial emotions using extreme sparse learning which is one of the fastest and interesting topics growing in upcoming years. It contains potential applications like human-computer interaction, automated tutoring systems, image and video retrieval, smart environments, and driver warning systems. Facial emotion is a very challenging task due to uncontrolled gestures. Sparse representation is a well-built tool for reconstruction, representation, and compression of high dimensional noisy data of images/videos and features due to its capacity to uncover important data. By using the regional covariance matrix face representation was proposed. A variable-intensity template was proposed to get a person-specific model for justifying various facial emotions. Thus the result of the proposed paper clearly defines the proposed emotion recognition system stable.

R. Ghasemi [2] claims in the interaction between man and computer —Facial expression recognition plays an important and powerful role. Thus the approaches were classified into six different facial expressions such as anger, disgust, fear, happiness, sadness, and surprise. There are three important steps of facial expression 1) face detection and image preprocessing, (2) feature extraction and feature selection, and (3) classification. Eyes, eyebrows and are the most important part of facial expression. By using an integral projection curve automatically these areas are achieved. Thus the proposed paper increases the accuracy of facial expression recognition and in very less time.

S.Mohseni [3] explains that human emotions are an important source of facial gestures. A large amount of information lies in human facial expressions with basic emotions such as anger, happiness, sadness, surprise, fear, and disgust are important facial gestures. In analyzing the facial activities, it contains two-level the first level is facial features tracking, which detects and tracks salient facial feature points and extracts face shape information and second is facial expression analysis attempts to recognize facial gestures that represent human emotion. The proposed method was tested on three different classifiers to demonstrate the efficiency the accuracy was 87.7% obtained.

B. Yang [4] estimates that face detection is the key issue in facial expression recognition for machines to perceive the emotional changes in human beings. Face detection, rotation rectification, and data augmentation several approaches used for preprocessing. In the proposed paper author has used a weighted mixture deep neural network (WMDNN) under tuff conditions. Hand-crafted features demonstrate a less recognition performance depends on facial expression recognition. Appearance, geometric, and motion features the main types of features used FER approaches. WMDNN that can process facial grayscale and LBP facial images simultaneously has been proposed by the author.

G. Yolcu [5] narrates deep learning is a subdomain of machine learning that uses hierarchical architectures to learn high-level abstractions in data. CNN is one of the classes in machine learning. Facial expression recognition is a region where deep learning methods have improved in problems. A facial component like mouth, eye, and eyebrow regions from the rest of the image are cascaded in a trained to segment in the first structure of CNN. Segmentation is formulated as a binary classification problem, where 16 × 16 blocks of the

image are categorized as a facial component. The area covered by either facial component in the block is above 80%. Thus, the accuracy of the proposed paper is achieved 93.43% facial expression recognition accuracy.

F. Ahmed [6] states that there is growth in development in user-specific interactions with consumer products and applications. There is a wide variety of applications in surveillance, human-computer interaction, data-driven animation, social robotics, and smart consumer products. Primitive filters on the original image are applied to the task for performing. The proposes paper introduces GP-based local texture feature fusion for facial expression recognition. A database such as CK and the JAFFE databases indicates the capability and it is compared with well-known facial feature descriptors.

N. Song [7] introduces facial expression combined with sign language to an emotional speech conversion method to clear up the communication issues between people with speech disorders and healthy people. By using deep neural network (DNN) model sign language and the features of facial expression are featured. For acknowledging the text of sign language and emotional tags of facial expression support vector machine (SVM) is trained simultaneously hidden Markov model-based emotional speech synthesizer is trained by speaker adaptive training with a Mandarin emotional speech corpus. The recognition rate for static sign language is 90.7% as evidenced by the proposed paper.

S. Kim states that the paradigm of facial recognition has been gaining a lot of traction recently. This is due to the fact that the technology for capturing images has been getting increasingly affordable every day. an increasing number of smartphones have been introduced with the camera that can capture highly detailed images easily [8]. Therefore, in this paper, the authors have presented an innovative technique for the identification of facial expression through the facial action coding unit by utilizing the paradigm of deep learning. The proposed methodology has been experimented extensively to yield promising results. The main drawback of the proposed methodology is the increased computational complexity of the system.

T. Shen explains that there has been increased attention toward the paradigm of facial expression recognition. An increased number of academics and researchers have been performing researches and implementing this technique in various different applications such as medical assessment and HCI or Human-Computer Interface [9]. Therefore, the researchers in this publication have presented an innovative facial expression identification technique through the use of Depth map which provides an approximate estimation through the Light Field Camera. The experimental results indicate that the proposed methodology improves the existing techniques significantly. The major limitation of the proposed methodology is that the authors have not improved the estimation of light in the depth map.

S. Gaglio elaborates on the process of recognition of activities that are normally performed by humans. The recognition of the activities is done through the use of RGB-D cameras such as the cameras utilized in the Microsoft Kinect. The authors in this paper have utilized the Hidden Markov Model along with Support vector machines and K means Clustering [10]. The proposed methodology utilizes the variations in the poses of a human's body and the bending of the joints to constitute a motion or an activity. The proposed methodology has been experimented on for the purpose of performance evaluation which has resulted in satisfactory results. The main drawback of the proposed technique is that the authors have not improved the pose estimation process.

S. Happy states that the paradigm of extraction of facial expression features through the use of various different techniques. The detection of facial features in useful in a lot of different scenarios and applications. therefore, the researchers in this paper have proposed an effective technique for the automatic extraction of facial features through the use of salient facial patches. These patches allow the identification of a lot of facial features and create landmarks that make the identification of facial features very easy and effective [11]. The proposed methodology has been extensively tested to achieve effective results. The main drawback of the proposed technique is the increased space complexity that is observed.

K. Zhang explains that the process of facial expression recognition is one of the most difficult and challenging implementations, this is due to the fact that the process of identification of facial expression features is a highly complex task [12]. Therefore, the authors in this publication have proposed an innovative technique for the extraction of facial expression through the use of PHRNN or Part based hierarchical Bidirectional Recurrent Neural Networks. The proposed technique has been experimented on to evaluate the performance and the result has proved the superiority of the presented technique over the traditional approaches. The main limitation of the proposed technique is that the authors have not utilized powerful structures to model the motion of facial features.

III PROPOSED METHODOLOGY

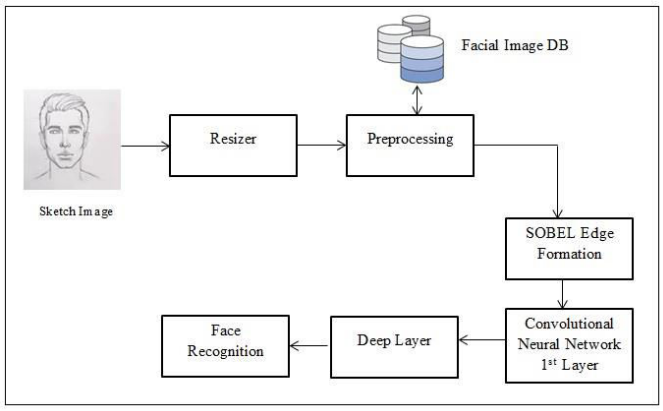


Figure 1: Proposed System overview Diagram for Sketch matching

The proposed methodology for Face recognition through sketch input is depicted in the figure1. The steps that are followed to achieve this methodology are being explained in detail below.

Step 1: Inputting Sketch and Resizing- This is the preliminary step of the proposed model, where the input sketch image file is being fed to the system. On receival of this sketch image the proposed model read this sketch file in the form of the Image object belongs to the BufferedImage class of the Java. This Image object is being resized using the Graphics2D Class of the Java to the desired size, so that the next step of the authentication can be carried out in a proper way.

Step 2: Preprocessing the Database image- As the user feed the sketch image to the proposed model, and then with respect to this all the original images in the database are read in the image object array. Each of these images is resized with respect to the size of the input image's height and width. Each of the resized image objects of the database are now subjected to match with the given input sketch image as explained in the next step.

Step 3: Sobel Filter- This is one of the Core step of the proposed model, where the database image is subjected to convert in a sketch image using the Sobel Filter Algorithm. According to this algorithm, each of the Database image pixels is estimated for the gray scale object using the equation1.

$G_V = 0.2126 * R + 0.7152 * G + 0.0722 * B$

Where ,

R=Red, G=Green, B=Blue

G_V= Gray Scale value

This gray scale conversion is applied to each of the pixel's subsequent pixels in 8 directions like NORTH, SOUTH, EAST, WEST, NORTH EAST, NORTH WEST, SOUTH EAST and SOUTH WEST positions. Once these gray scale value is estimated , then it is utilized to calculate the gradient of the each of the pixels. These gradient values are then applied to get the max gradient using the equation 2.

$$\boldsymbol{Mg} = \sqrt{(\boldsymbol{gx} - \boldsymbol{gx})^2 + (\boldsymbol{gy} - \boldsymbol{gy})^2} \quad (1)$$

Where

M_g = Maximum Gradient
g_x= Gradient value of the X position pixels
g_y= Gradient value of the Y position pixels

Then the obtained maximum gradient values are applied to the respective pixels to get the proper edges to form a sketch image object. This process of Sobel filter or edge formation is shown in the below algorithm 1.

ALGORITHM 1: SOBEL FILTER

```
// Input: Image Frame IMG_F
// Output: Sobel Filter Edge Image SFE_I
sobelFilter(IMG_F)
1:  Start
2:          mg= -1 [ Max gradient]
3:          edge[][]= ∅
4:    P_SET={N,E,W,S,NE,NW,SE,SW,C}
[ Position Set ]
5: for i = 0 to size of Width of IMG_F
6: for j=0 to size of Height of IMG_F
7:    P_SIGN = IMG_F PSET [i] RGB
8:    R= P_SIGN >> 16 & H_D
9:    G= P_SIGN >> 8 & H_D
10:     B= P_SIGN >> 0 & H_D
11:   value[ ]=getGrayModelValue(R,G,B)
12:                  gx=  ∑ value→i
13:           gy=  ∑ value→j
14:         gval=√ (gx* gx)+ (gy* gy)
15:          if mg<gval  THEN
16:          mg=gval
17:          END if
18:       edge[i][j]=gval
19:  End for
20:  End for
21:  scale= 255/mg
22:  for i = 0 to size of Width of IMG_F
23:  for j=0 to size of Height of IMG_F
24:           E_C=  edge[i][j]
25:         E_C=  EC * scale
26:         E_C<<16 | E_C<< 8 | E_C
```

27: $SFE_{I[i,j]} = E_C$
28: ***End for***
29: ***End for***
30: **return** SFE_I

Step 4: CNN First Layer- This step of the proposed model is the initial step in face authentication. Here both the input image from the user and the instance database image are divided into 48 blocks each. 8 blocks according to width and 6 blocks according to height.

Step 5: Deep Layer- The each and every block of both the input and the stored instance images are subject to evaluate their respective average block brightness. Then the respective block, brightness is measured for their absolute difference between them. If the measured difference is less than the value of 25, then it is considered as the matched block with the respective block of the one another and it is counted for the matching process. In the same way all the respective blocks are counted. The obtained respective count is then attached with the respective database image. And then the maximum count image is selected as the best matched image for the given input image.

IV RESULT AND DISCUSSIONS

The presented technique for facial image recognition through sketch input has been achieved in the java programming language on the NetBeans Integrated Development Environment. The proposed system is developed on a development machine consisting of a standard configuration in which the processing requirements are realized by the Intel Core i5 processor along with 4GB of RAM and 500GB of storage. The MySQL database server realizes the database obligations.

The RMSE or the Root Mean Square Error is executed for the motive of the assessment of the error in the accurate facial image recognition from the sketch input. For the accurate evaluation and realizing the error rate, two evaluation parameters are assessed for this objective. The two continuously correlated entities that are utilized for this purpose are the expected Facial Image Recognition and the achieved Facial Image Recognition.

Equation 3 given below is utilized for the measurement of the Error rate of the proposed methodology.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2}{n}} \quad \_ (3)$$

Where,

$\sum$ - Summation

$(x_1 - x_2)^2$ - Differences Squared for the summation in between the expected Facial Image Recognition and the achieved Facial Image Recognition

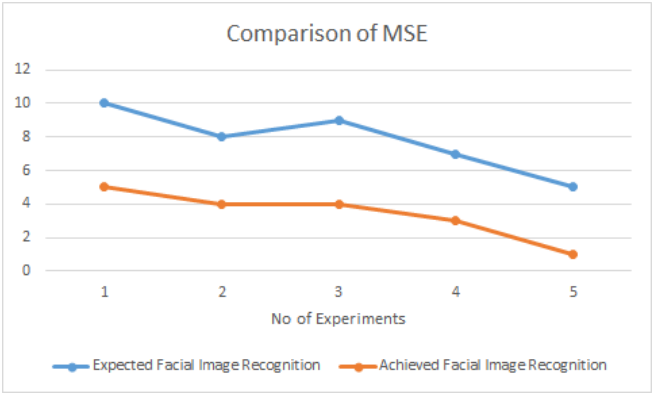
n - Number of samples or Trails

The exhaustive evaluation of the presented methodology is implemented using the RMSE paradigm, and the outcomes of the experimentation are listed in Table 1 given below.

Experiment No	Expected Facial Image Recognition	Achieved Facial Image Recognition	MSE
1	10	5	25
2	8	4	16
3	9	4	25
4	7	3	16
5	5	1	16

Table 1: Mean Square Error measurement

Figure 2 given below plots the experimental data listed in Table 1 above. The plotted graph represents the Mean Square Error rate between the expected Facial Image Recognition and the achieved Facial Image Recognition for a significant number of trials or experimentations that are accomplished by the means of in-depth analysis. There is an exhaustive number of trials that are executed in each experiment. The values of MSE and RMSE measured through the experimental outcomes are 19.6 and 4.42 respectively. The calculated values of RMSE for facial image recognition from the sketch input are



well under the criteria which indicates the increased accuracy offered by the Convolutional Neural Network approach. The RMSE values extracted through the experimentation discloses that the proposed methodology offers reduced execution time and increased reliability and accuracy.

Figure 2: Comparison of MSE in between expected Facial Image Recognition V/s achieved Facial Image Recognition

V CONCLUSION AND FUTURE SCOPE

The process of facial image recognition through a sketch of the face given as input is implemented successfully. The image taken as an input is subjected to various processes to achieve the accurate recognition. The normalized image is then processed using the Sobel Edge formation that utilizes the pixel overturning. The resultant information is then fed to the Convolutional Neural Networks. The CNN utilizes the shape of the edges in the various layers of the network. Subsequently with the application of the block thresholding an accurate facial recognition is performed. The proposed methodology has been tested for its errors using the RMSE paradigm which achieved the value of 4.42. This indicates that the first attempt for facial image recognition has been successful and achieves satisfactory results.

For the Future research prospects, the proposed technique can be executed in real-time. This methodology can also be deployed in the cloud platform with a large database of facial images.

REFERENCES

[1] SeyedehsamanehShojaeilangari, Wei-Yun Yau, KarthikNandakumar, Li Jun, and EamKhwangTeoh, "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning" http://www.ieee.org/publications_standards/publications/rights/index.html 2016

[2] RojaGhasemi, Maryam Ahmady," Facial Expression Recognition Using Facial Effective
Areas And Fuzzy Logic" 978-1-4799-3351-8/14/,IEEE, 2014

[3] Sina Mohseni1, Niloofar Zarei2, Saba Ramazani," Facial Expression Recognition using
Anatomy Based Facial Graph " 2014 IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, USA October 5-8, 2014

[4] Biao Yang, Jinmeng Cao, Rongrong Ni, Yuyu Zhang," Facial Expression Recognition using Weighted Mixture Deep Neural Network Based on Double-channel Facial Images" JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015.

[5] GozdeYolcu, Ismail Oztel, Serap Kazan, Cemil Oz, KannappanPalaniappan, Teresa E. Lever, FilizBunyak," Deep Learning-based Facial Expression Recognition for Monitoring Neurological Disorders " IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2017.

[6] Faisal Ahmed, Padma Polash Paul, Marina L. Gavrilova," Evolutionary Fusion of Local Texture Patterns for Facial Expression Recognition" 978-1-4799-8339-1/15©IEEE 2015.

[7] Nan Song Hongwu Yang," A Gesture-to-Emotional Speech Conversion b Combining Gesture Recognition and Facial Expression Recognition" First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia) 2018.

[8] S. Kim et al, "Deep Explanation Model for Facial Expression Recognition through Facial Action Coding Unit", IEEE International Conference on Big Data and Smart Computing (BigComp), 2019.

[9] T. Shen et al, "Facial Expression Recognition Using Depth Map Estimation of Light Field Camera", IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2016.

[10] S. Gaglio et al, "Human Activity Recognition Process Using 3-D Posture Data", IEEE Transactions on Human-Machine Systems, vol. 45, no. 5, October 2015.

[11] S. Happy et al, "Automatic Facial Expression Recognition Using Features of Salient Facial Patches", IEEE Transactions on Affective Computing, 2015.

[12] K. Zhang et al, "Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks", IEEE Transactions on Image Processing, 2015.

*****