IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Facial Landmark Detection for Expression Analysis

Takrim Ul Islam Laskar[1*], Parisimita Sarma[2]

[1,2] Department of Information Technology, Gauhati University, Guwahati, India

[*]*Corresponding Author: takrimulislam@gmail.com, Tel.: +(91)-7896983050*

DOI: https://doi.org/10.26438/ijcse/v7i5.16171622 | Available online at: www.ijcseonline.org

Accepted: 22/May/2019, Published: 31/May/2019

Abstract — In this paper we have developed a system which is able to extract the facial landmarks like jaw, eyebrows, nose, eye and mouth from human face. This is generally done in order to use the extracted data for analysis of the emotions that is depicted in human face. We have used openCV and Dlib library to detect the facial landmarks. There are many feature extraction techniques like Geometry-based Technique, Template-based Technique, Appearance-based Technique, Colour-based Technique, etc[9]. The Pre-trained file that we used to detect the facial landmarks was trained with an Ensemble of Regression Trees. Using the shape predictor of Dlib we passed the file over the input image and the detection was estimated through pixel intensity. The extracted pixel values were stored using pickle C object in python. Any suitable neural network may be farther used to train a model, from the extracted data from dataset/datasets, which is able to analyse the different emotions on human face. Our aim is to proceed further and train a model with neural network for Expression Analysis with special concentration on children.

***Keywords*—**Digital Image Processing, Facial Landmark Detection, Face Detection, Computer Vision.

I. INTRODUCTION

The different facial features are jaw, eyebrow, eye, nose, mouth, etc. These features in the face produce specific gestures to signify specific meanings and thereby the associated emotion is mapped. The project deals with face detection and feature extraction phase.

A. Human face

Face is the frontal side of the human head from fore-head to the chin and between the ears. It is a crucial part of the human body which emits perceptions related to the emotion of an individual. Besides this, speech and gesture can also convey perception of the emotion. However, facial expression is the most accurate for detection.

B. Computer Vision

Human beings are capable of seeing the world and objects with the help of the eyes and acting accordingly. The computers should be availed a system through which the aforementioned task can be accomplished. Therefore, the system through which a computer undergoes a similar task of human's seeing is referred to as computer vision.

C. Face Detection

To be able to locate the facial landmarks, the first and foremost task is to locate the area in the image where the face is available. There are many approaches to detect faces in a digital image. Some of them are face detection with controlled background, colour, motion, unconstrained scenes, etc. We would be using Histogram of Oriented Gradients (HOGs), which falls under the face detection in unconstrained scenes approach to detect the face from still images.

Figure 1 represents the detection of face in an image.

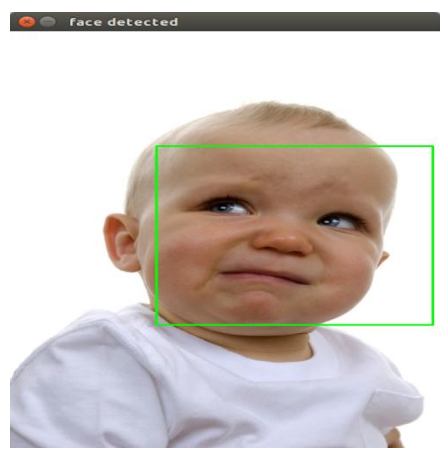

Figure 1. *Image of face Detection.*

D. Facial Landmark Detection

Localization and representation of important regions can be achieved through Facial Landmarks detection. Eyes, eyebrows, nose, mouth, jawline are the regions.

II. Related Work

Rencan Nie et al., in their paper, "Facial Feature Extraction Using Frequency Map Series in PCNN", proposed to extract the facial feature based on Pulse coupled neural network (PCNN). They analysed the limitations of BMS after extraction of one dimensional oscillation time sequence (OTS) and considered the correlation between binary images in BMS and they proposed a method to transform BMS into frequency map series (FMS). Then they convert 2D frequency map to 0D data points. Finally, based on analysis of characteristics they proposed to use the standard Euclidean distance measure as the distance measure of OTS-FMS features. The experimental results showed that the recognition rate of OTS-FMS is significantly higher than PCA and KPCA and also better than method based on BMS extracted OTS-BMS features [1].

Azeem et al., in their paper, "Hexagonal scale invariant feature transform (H-SIFT) for facial feature extraction", proposed a system where hexagonal re-sampling is done to enhance the capacity of low contrast areas on the face image to be highlighted for effective feature extraction [2].

M.H Siddiqi et al., in the paper "Human facial expression recognition using curvelet feature extraction and normalized mutual information feature selection", used curvelet transform for feature extraction. It is able to extract the prominent features keeping the line, curve and edge information from each expression frame.it can be used for image re-construction problems [3].

S.K.A Kamarol, et al., in the paper "Spatiotemporal feature extraction for facial expression recognition", they implemented spatiotemporal texture map (STTM). It generates a 3D texture map given an input video. They established that it outperforms many state-of–the-art appearance based feature extraction techniques on the CK+, CASEME II and AFEW dataset with lower computational cost [4].

Mohammad Rabiei et al., in the paper "System and method for recognizing human emotion state based on analysis of speech and facial feature extraction; Applications to Human-Robot Interaction", developed a hybrid system for emotion detection in where speech and image processing take place simultaneously. For the phase of image processing they used skin texture technique and five facial principal components like: left/right eyes, left/right eyebrows, mouth, lips, nose and head. They state that Eyes' and mouth's localization is of high importance for emotion recognition and influences the localization of eyebrows and nose. The system they extracted facial Action Units(AUs) and they calculated the distance between the input face image and normal face image. In their research, they used a set of 32 facial Action Units, that corresponded to some basic emotional expressions [5].

S. Jadon et al., in the paper "Face Recognition Using Som Neural Network with DDCT Facial Feature Extraction Techniques", dealt with four different techniques for feature extraction of image. They used Self-Organising Map(SOM) Neural Network for training the dataset and simulation of face recognition system. Feature extraction was done based on the Directional discrete cosine transform (DDCT), discrete wavelet transform (DWT), discrete cosine transform (DCT) and Sobel edge detection. The features are then extracted and stored in a feature vector. The whole dataset after feature extraction is stored to feature vectors and all the vectors are stored to a new database. For a target image, they extracted its features and then compared with the vectors to retrieve the similar images. They stated that DCT is a powerful transform to extract proper features for face recognition [6].

H. Liao in the paper "Facial age feature extraction based on deep sparse representation", implemented Sparse Representation Classification (SRC) to three commonly used benchmark datasets namely FG-NET, Morph and IFDB datasets. He faced a problem where it was difficult to separate the faces at adjacent ages due to their similarities. This created overlapping effect of the marginal ages. To solve this problem, he used three layers, first with 2 age group, second with 4 age groups and third with 8 age groups. The best accuracy of classification he achieved with AAM classifier is 97.5%, BIF classifier is 92.1% and Gabor+LBP classifier is 85.5% [7].

Karin Sobottka, et al., in the paper "A novel method for automatic face segmentation, facial feature extraction and tracking", developed a system for segmentation, facial feature extraction and tracking. They did segmentation of face-like structure depending on the skin colour. They first enhanced the input image using the colour information then using SGLD matrices textural features are derived [8].

Ramesha K et al., in their paper "Feature Extraction based Face Recognition, Gender and Age Classification" mentioned a system where the used a small dataset, yet they achieved a good accuracy. The process involved Pre-processing, Feature Extraction and Classification. The features were detected using Canny edge operator and face recognition. The training was based on Posteriori Class Probability and Artificial Neural Network. The accuracy for face recognition was 100%, and gender and age classification was 98% and 94% respectively [11].

V.R. Parihar et al., in their paper "A Novel Approach to Real Time Face Detection and Recognition", used mixture of algorithms and techniques. They used Viola–Jones object detection framework algorithm along with geometric and symmetric information of the landmarks. They solved many existing problems that were faced earlier [12].

Md. T. Akhtar et al., in the paper "Fast and Real Life Object Detection System Using Simple Webcam", developed a system which is able to detect real life objects by matching with a binary string database which they extracted earlier. Their next aim is to develop a face detection system [13].

A. S. Banu et al., in the paper "SAR Image Classification by Wavelet Transform and Euclidean Distance with Shanon Index Measurement", proposed a system which is Wavelet transform based Euclidean distance with Shanon Indexing measurement to classify the Synthetic Aperture Radar (SAR) image. The steps involved were preprocessing, feature extraction and classification [14].

III. Tools and Environment

A. Tools

- Python
  Python is a general purpose high-level programming language. It is also interpreted and follows object-oriented paradigm. Guido van Rossum was the created of python and he created it in 1990. It is an open source programming language. It has a huge amount of libraries and modules. We used be using python 2.7.

- openCV
  openCV (Open Source Computer Vision Library) is released under BDS License and its free for academic and commercial use. It has C++, java, Python interfaces.
- Dlib
  Dlib is actually a C++ toolkit which contains machine learning algorithms and tools to help creating complex software in C++ to solve real world problems. It can also be wrapped with python.

B. Dependencies

- Shape predictor 68 face landmarks.dat
  It is a pre-trained facial landmark detector available and is used to locate 68(x,y)-coordinates on the face.
- Imutils
  It's a collection of functions to make basic image processing operations such as rotation, resizing, translation and displaying Matplotlib images easier with OpenCV and Python.

C. Environment

- Operating System- Ubuntu 16.04 (Linux based platform)
- RAM- 4GB
- Processor- Intel core i5 7th Gen.

IV. Methodology

The system enables a computer to detect the face and the detect the facial landmarks from the face.

- The first stage would be loading of image.
- Dlib's frontal face detector is used for detection of face in the image. The face would be the region of interest (ROI).
- Facial feature or landmark is then extracted. These features include ears, eyebrows, nose, mouth, jawline.

Figure 2 describes the flow of the system.

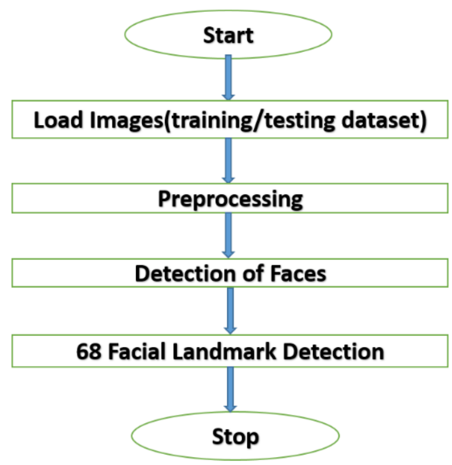

Figure 2. *Flow Chart of the methodology.*

A. Image Loading

The images have to be loaded first from the storage for Facial feature extraction. To do so, openCV function cv2.imread() is used. In this case Colour, grayscale and unchanged image can be loaded.

Colour- cv2.imread('image',1)
Grayscale- cv2.imread('image',0)
Unchanged- cv2.imread('image',-1)

Figure 1 represents an image loaded of bigger size than it should be, to be accommodated by openCV window.

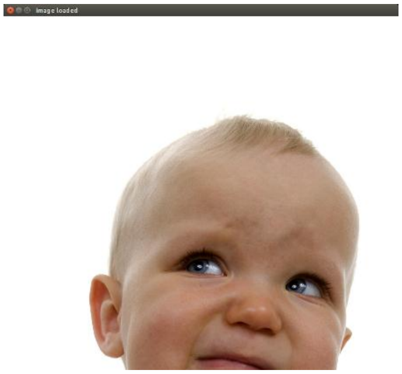

Figure 1. *Image Loaded.*

To read image as CLI argument function cv2.imread(args["image"]) is used.

To read image from directory
directory=os.path.join(os.getcwd(), 'dir')
cv2.imread(directory+'/'+image)

B. Preprocessing

After loading the image, pre-processing is done by resizing the image to 500 pixels and converting the resized image to grayscale.

1. Resizing

Scaling down the images allows faster processing. To resize the images imutil function is used. imutils.resize(img_file, width=500).

Figure 2 represents the resizing of the loaded image.

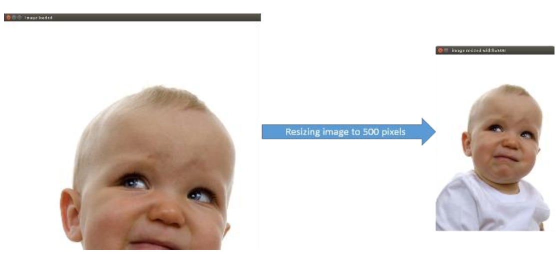

Figure 2. *Resizing image to 500 pixels.*

2. Grayscale Conversion

We converted the image from BGR colour model to grayscale colour model. The openCV supports BGR colour model and dlib supports RGB colour model. And grayscale colour model is supported in both. Grayscale has two channels black and white and is used for 2D image processing. Luminance is a very important factor in distinguishing visual features. The colour images distract the edge detection due to presence of three different colour channels. Edges are better depicted in grayscale images.

To convert the image to grayscale openCV function cv2.cvtColor(img_file, cv2.COLOR_BGR2GRAY) is used where 'img_file' is the name of the image file and 'cv2.COLOR_BGR2GRAY' is the parameter passed which converts the image from BGR to grayscale colour model.
Figure 3 represents the colour model conversion of the resized image.

Figure 3. *Grayscale Conversion of the image.*

The grayscale image is then scanned with the dlib frontal frontal face detector. The detector detects the faces in the image and forms dlib rectangle around them. Each of the face rectangles are then converted to openCV-based face bounding box where (x , y) coordinate of top-left position of the box and the width(w) and height(h) is obtained. These values are used to draw the rectangles around the detected faces.

C. Face Detection

The first step of the system is to detect the face from the dataset images for training and testing and after implementation the task is to detect the face from the camera in real-time. We used frontal face detector which comes with dlib library. It is based on Histogram of Oriented Gradients (HOG) feature descriptor and Support Vector Machine (SVM). The model was built with 5 HOG filters which are front looking, left looking, right looking, front looking but rotated left and front looking but rotated right. The dataset used for training consists of 2825 images, obtained from LFW dataset and manually annotated by Davis King, the author of dlib[10].
Figure 4 represents the detection of face in the image and its representation with bounding box.

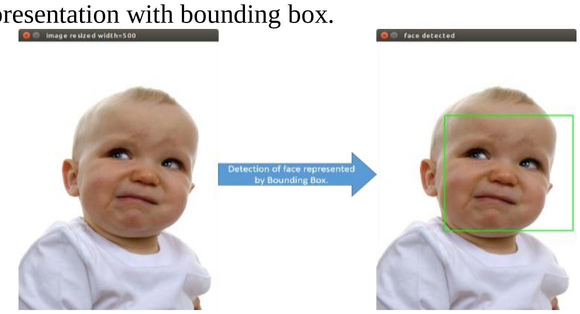

Figure 4. *Detection of face represented with bounding box.*

D. Facial landmark extraction

We have used the dlib's shape predictor and analysed the input images with a pre-trained file, 'shape_predictor_68_face_landmarks.dat' to get the facial landmarks detected. The trained file is an implementation of the 'One Millisecond Face Alignment with an Ensemble of Regression Trees' paper written by Kazemi and Sullivan(2014). To train the detector they used a manually labelled facial landmarks specifying specific (x , y)-coordinates of regions surrounding each of the facial structure. An ensemble of regression trees is used to train the file for estimation of the facial landmark positions through pixel intensities themselves.
The facial landmark detector is used to estimate the location of 68 (x , y)-coordinates that map to facial structure on the face.

Figure 5 represents the annotations are based on 68 point iBUG 300-W dataset which the dlib facial landmark predictor was trained on.

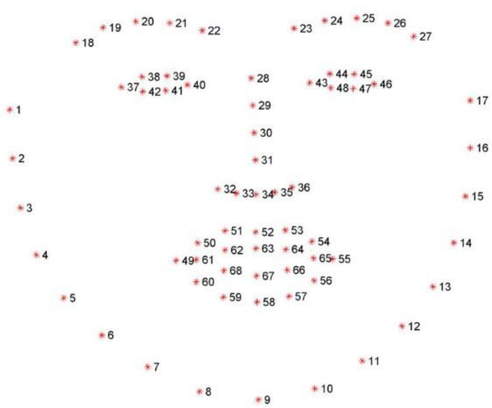

Figure 5. *Annotations based on 68 point iBUG 300-W dataset.*

Table 1 represents the Facial Landmarks and their mapping with the points.

Table 1. Annotation of different landmark points.

Facial Landmarks	**Points**
Jaw	1-17
Right Eyebrow	18-22
Left Eyebrow	23-27
Nose	28-36
Right Eye	37-42
Left Eye	43-48
Mouth	49-68

The facial landmarks are thereby detected. Figure 6 represents landmark detection.

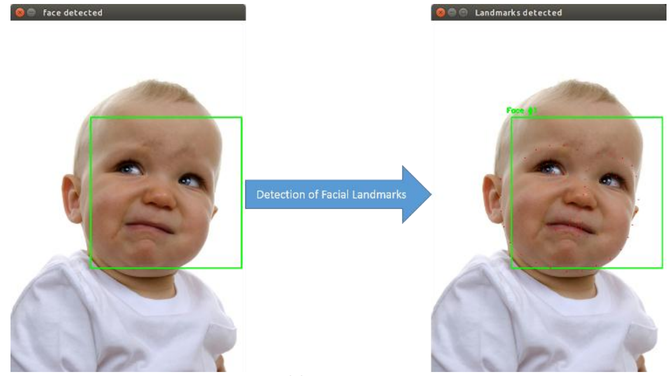

Figure 6. *Detection of facial landmarks within the ROI.*

V. RESULTS AND DISCUSSION

In the system, face detection and facial landmark detection have been completed successfully. Some images of landmark detection in different emotional faces namely happiness, sadness, anger, disgust, surprise, fear and neutral is displayed in Figure 7.

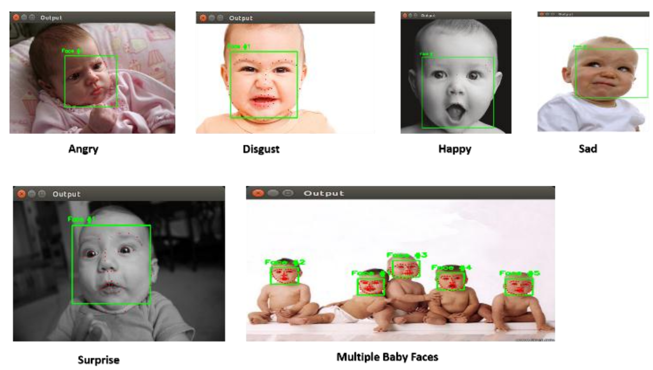

Figure 7. *Face and landmark Detection on Image with Multiple Faces.*

VI. CONCLUSION AND FUTURE SCOPE

In this project we have developed a system for facial landmark detection. For that, we first detected the faces in the image using dlib's frontal face detector. Then, we detected the 68 facial landmarks' coordinates like eyes, eyebrow, nose, mouth, jawline using dilb's shape predictor with the help of 'shape_predictor_68_face_landmarks.dat' file. In a condition of low illumination, the system fails to show a proper result.
In future, the system would be further incremented to build a system for facial emotion detection with special concentration on babies' emotional face.

REFERENCES

[1] Rencan Nie, Dongming Zhou, Min He, Xin Jin, and Jiefu Yu, "*Facial Feature Extraction Using Frequency Map Series in*

PCNN", Journal of Sensors, Vol. **2016**, Article ID **5491341**, pp. **1- 9, 2016.**

[2] A. Azeem, M. Sharif, J.H. Shah, M. Raza, "*Hexagonal scale invariant feature transform (H-SIFT) for facial feature extraction*", Journal of Applied Research and Technology, Vol. **13**, pp. **3, 2015.**

[3] M.H Siddiqi, R Ali, M Idris, A.M Khan, E.S Kim, et al., "*Human facial expression recognition using curvelet feature extraction and normalized mutual information feature selection*", Multimedia Tools and Applications, Vol. **75**, pp. **935-959, 2016.**

[4] S.K.A Kamarol, M.H Jaward, J Parkkinen, Rajendran Parthiban, "*Spatiotemporal feature extraction for facial expression recognition*", IET Image Processing, Vol. 10(7), pp. **534-541, 2016.**

[5] M. Rabiei, A. Gasparetto, "*System and method for recognizing human emotion state based on analysis of speech and facial feature extraction; Applications to Human-Robot Interaction*", In the Proceedings of the 4th International Conference on Robotics and Mechatronics(ICROM 2016), Tehran, Iran, pp. **266-271, 2016.**

[6] Swati jadon, Mahendra Kumar, Y ogesh Rathi, "*Face Recognition Using Som Neural Network With Ddct Facial Feature Extraction Techniques*", IEEE International Conference on Communications and Signal Pocessing (ICCSP 2015) , India, UAE, pp. **1070-1074**, **2015**.

[7] Haibin Liao, "*Facial age feature extraction based on deep sparse representation*", Multimedia Tools and Applications, Vol. **78**, pp. **2181–2197**, **2018**.

[8] Karin Sobottka, Ioannis Pitas, "*A novel method for automatic face segmentation, facial feature extraction and tracking*", Signal Processing: Image Communication, Vol. **12(3)**, pp. **263-281**, **1998**.

[9] Sanjeev Dhawan, Himanshu Dogra, "*Feature Extraction Techniques for Face Recognition*", International Journal of Engineering, Business and EnterpriseApplications (IJEBEA), Vol. **2(1)**, pp. **1-4**, **2012**.

[10] V Kazemi, J Sullivan, "*One Millisecond Face Alignment with an Ensemble of Regression Trees*", IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. **1867-1874**, **2014**.

[11] Ramesha K, K B Raja , Venugopal K R, L M Patnaik, "*Feature Extraction based Face Recognition, Gender and Age Classification*", (IJCSE) International Journal on Computer Science and Engineering, Vol. **02**, No.**01S**, **14-23,** , **2010.**

[12] Vikramsingh R. Parihar1, Anagha P. Dhote, "*A Novel Approach to Real Time Face Detection and Recognition*", (IJCSE) International Journal on Computer Science and Engineering, Vol. **5**, Issue. **01S**, **62-67,** , **2017.**

[13] Md. T. Akhtar, S. T. Razi, K. N. Jaman ,A. Azimusshan ,Md. A. Sohel, "*Fast and Real Life Object Detection System Using Simple Webcam*", (IJSRCSE) International Journal of Scientific Research in Computer Science and Engineering, Vol. **6**, Issue. **4**, pp. **18-23**, **2018.**

[14] A. Shakin Banu ,P.Vasuki , S. Mohamed Mansoor Roomi, A. Yusuf Khan, "*SAR Image Classification by Wavelet Transform and Euclidean Distance with Shanon Index Measurement*", (IJSRNSC) International Journal of Scientific Research in Network Security and Communication, Vol. **6**, Issue. **3**, pp. **13-17**, **2018.**

Authors Profile

Takrim Ul Islam Laskar is a student of MTech of Gauhati University for the session 2017-19. He is pursuing MTech degree in Information technology and his area of specialization is Digital Image Processing and Deep Learning. He completed B.Tech degree in Computer Science and Engineering from Department of Information Technology,Gauhati University, Guwahati, India.

Parismita sarma is an assistant professor in the department of Information Technology, Gauhati University. She has completed her PhD from Gauhati University in the department of Information Technology. She has graduated BE in computer science and engineering from Jorhat Engineering College and MTech from Tezpur University, Assam. She has almost twenty publications in both conferences and international Journals.