Facial Emotion Recognition – A gift for the visionless

Aneesh Kini U, Hasan Alp Zengin, Smitha N Pai

Abstract— Most of the mundane activities these days are automated. There are others, which do not have a fixed pattern. For instance, recognition of emotions of an individual given the facial expressions. Using the Machine Learning concepts, a model is trained with various facial images having varied expressions, of single and multiple individuals. In the current work, face detection and emotion recognition is carried out at real time even when an individual is on the move. The findings of the paper can be useful in identifying missing individual, helping individuals in emotional distress. It can also help the visionless analyze the mood of the person with whom he is interacting. The name of the individual whose identity is verified is also verbally provided as an assistance to the impaired. Various face recognition algorithms and the relative comparison and analysis is also brought out usic plays an essential role in the well-being of many people. It can be therapeutic, motivational and can even unite people.

Index Terms— Convolution Neural Network, Deep Learning, Eigen Faces, Emotion recognition, Facial recognition, Fisher Face, Local Binary Pattern, Machine learning.

1 INTRODUCTION

Automated Face recognition is gaining more momentum because of the enormous scope it has for various applications. The concept of machine and deep learning with artificial intelligence plays a pivotal role in these cases. Face recognition deals with analyzing patterns based on a person's facial contour. Facial emotional recognition on the other hand, uses the facial expression to identify emotions. The same facial expression can convey varied emotions based on an individual. Deep learning adheres to deep neural network models for different tasks from image processing to speech and language processing. The training for these models requires large processing capabilities with GPU's. Due to the development in technology the deep learning has become supreme in many disciplines such as image recognition and automatic speech recognition. Some of the Deep learning concepts are implemented using Convolution Neural Network, Recurrent Neural Network, Auto Encoders and Deep Belief Networks.

Emotions of a person can be recognized by other persons through feelings towards each other. Machines are not in a position to recognize emotions as they are not associated with any feelings. Emotions of Human changes dynamically and instantaneously with feedback from the surroundings. Even this varies with individuals. Voice if used along with facial expression can help detect the emotions of the human. The current implementation suggests an almost accurate analysis of the emotion of the individual after face recognition.

- *Aneeh Kini U, is currently pursuing masters degree program in computer and networking, MIT, MAHE,Manipal India, E-mail: aneesh.u@learner.manipal.edu*
- *Hasan Alp Zengin is currently a Full Stack Developer at Stuggart Germany. E-mail: hasalp38@gmail.com*
- Smitha N Pai, Dept. of ICT, Manipal Institute of Technology, Manipal Academy of Higher Education, India Email:smitha.pai@manipal.edu

Here, the facial emotion recognition is carried out using trained models which accesses the database, to retrieve images of individuals with different emotions like sad, happy, disgust, surprise, angry, fear and neutral. The image of the person with emotion, under test is compared with existing repository images. This in turn infers the state of the test image. Real time face recognition and emotion recognition using on the go video through convolution neural networks and OpenCV using python environment is addressed in the current work.

2 RELATED WORK

A peripheral vision of a person helps not only in recognition of the persona, but also in detecting the mood of the person. The study [1] involved detecting the moods of various individual who were in various emotional conditions. It involved people with various states like surprise, anger, happy, fear, sad, neutral, disgust. It was found that angle played a role in detecting the mood of the person. In an expressive face the best recognized expressions are Surprise and Happiness. Angry and Sad were poorly recognized and detected. The study also suggested that images conveyed emotions of individual. Trained individuals were able to detect the emotions of the person.

In the literature [2], an intellectual model for eldercare using robot is addressed. Gabor channel, K-Nearest Neighbor, Support Vector Machine and Local Gabor Binary Pattern techniques helps in perceiving the facial expression of the elderly. The smart home is automated based on the facial expression of the elderly. In the study [3], optical flow concepts are used to detect the motion of the head. Shift Invariant features of the head movement are subtracted from the different frames to obtain the variant emotions of each face. A reference vector is considered from nose tip to midpoint of the eye. Facial muscle movement in terms of its expansion and contraction with respect to the nose helps in obtaining the spatio-temporal features. Images with illumination variations, occlusions are also handled. In the work [4], in addition to basic emotions an additional emotion Contempt is added. A combination of all the seven emotions are used for analysis. Convolution neural network(CNN) helps in feature extraction. Inception v3 concept is used for recognizing the emotions. Project [5], is carried using 167 features of the face. In addition, 37 more features representing the angles between various landmarks are obtained. Angles between the corners, center of eyebrows, it's shape, amount it is raised are some of the features which are addressed. Angle between corners of the nose and its tip measure the nose scrunch. Softmax regression and support vector machine were used to obtain the result. Facial feeling acknowledgment in uncontrolled conditions is an exceptionally difficult assignment because of expansive intra-class varieties caused by components, for example, light and posture changes, impediment, and head development.

Some of the simplest and original of the face recognition techniques are Eigen face, Fisher face and Local Binary Pattern (LBP) process. **Eigen face** uses the Principal Component Analysis (PCA) method for recognizing the face. PCA helps in identifying best few features amongst the large number of available features. This reduces the computational cost with reduced feature set. The eigen values are obtained from the covariance matrix The highest eigen values are chosen to obtain the eigen faces. These faces are also the eigen vectors. This method is resilient to noise. This process also helps in reconstruction of partial faces. The images are linearly projected onto low dimensional subspace.[6,7] An image of face can be estimated to and described by a point in an element space, spread over by various eigen faces. During face verification there is a 1:1 match done to authenticate if it is the same person. During face recognition it is a 1:k match to check who is the person among the faces available in the database.

Fisher face uses a subspace of the features which has large variations in data. Linear Discriminant Analysis finds the subspace from the set of face images. The bias vectors which defines the are known as the Fisherface. A study on how Eigenfaces vary with Fisher is addressed in [8,9,10]. The inter class distance is widened and intra class faces distance is reduced in this case. Hence the effect of light illumination is reduced.

The L**ocal Binary Pattern (LBP)** creates the feature vector which are useful to analyze using support vector machine (SVM). It tackles images which are gray scaled caused by illumination. The neighbours of a pixe, with values above the threshold are combined together to form a binary data. A group of similar values are plotted in the form of a histogram. Any face to be identified is compared with the histogram using the Euclidean distance. The advantage of this process is its simplicity [11].

FaceNet [12] learns neural network that encodes the face image to vector. It uses the Euclidian embedding for each image using the deep neural network. It uses the Inception model which carries out the convolution step, normalization and activates the neural network. Face verification is to threshold the distance between the embedding and recognition is to find the KNN classification to which it belongs. The triplet loss function encodes the image of the same person together and pushing others apart. In DeepFace [13], 32 feature maps are followed into maximum pooling layer of stride value of 2, which is trailed by 16 filter to capture low level features. This is followed by fully connected network. Last layer is fed to softmax regression with k-way to obtain class label. The feature distance between the anchor image to be recognized with the anchor positive (trained) image is kept minimum. The distance between the anchor image and all other anchor neative image is keep as far as possible. DeepFace and FaceNet along with the partial trained model is used to reduce the work of recomputing the weights [12],[13].

In the current work the concepts addressed in the literature are used to recognize an individual and identify the emotions associated with them. Convey the name and the identity of the person.

3 METHODOLOGY

The first stage in emotion recognition is the man-machine interface. Input face are images containing various emotions like happy, sad, angry, fear, disgust, surprise and neutral. The open-cv tool is used to capture image from webcam and then it serves as an input to trained convolution network. This model places a bounding box on the face and recognizes the landmark points. They include the eyes, nose tip and corners of the lip. A directory is created for each individual. The directory name is the name of the person. During the testing phase, a picture is given to the trained model, which verifies for the presence of face features. Face recognition algorithm such as LBPH, FisherFace, EigenFace and FaceNet are used for training and for prediction of the face. The individual after being recognized is sent to the emotion recognition system. The Text to Speech converter pronounces the name of the person along with the emotions. This process can benefit a blind person to hear the name of the person along with his emotions. The flow of events is as shown in Fig. 1.

The dataset uses the Tensor Flow libraries to implement deep learning algorithm. Shallow convolution network does not suffice to capture proper facial emotions. Fig. 1 shows the high level overall process involved in the face and emotion recognition mechanism.

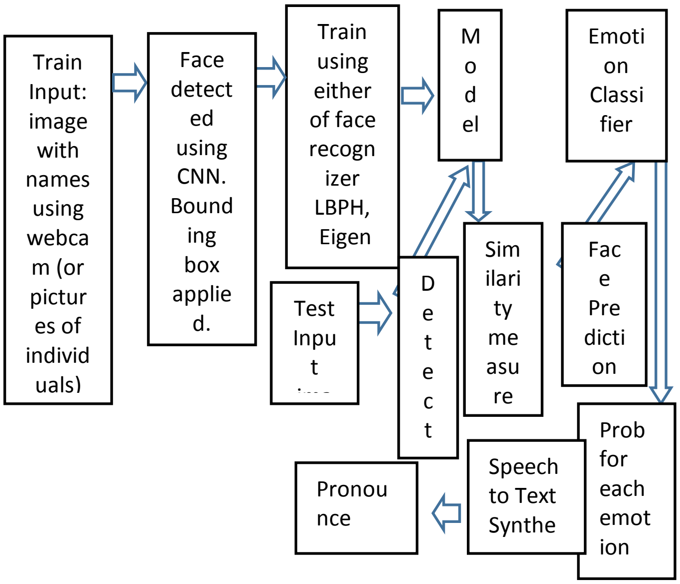


Fig. 1 Flow of events

The face features are captured using both Harr cascade and multi task convolution neural network (MTCNN) [14] . It was observed that MTCNN performs better than harr cascade [15]. Harr cascade has a set of feature vectors and these are moved across the image. This helps in identifying the presence of various features on the face. MTCNN, is a Convolution neural network working in three stages. The first stage finds the overall features, the second helps in refining and the third stage helps in obtaining the landmark points. The last stage of MTCNN is as shown in the figure 2. The various stages include the Convolution and Maximum pooling, which is followed by the fully connected network. The result is checking for presence of face, obtaining the bounding box parameters and the landmark points of the face (eyes, nose, lips corners).

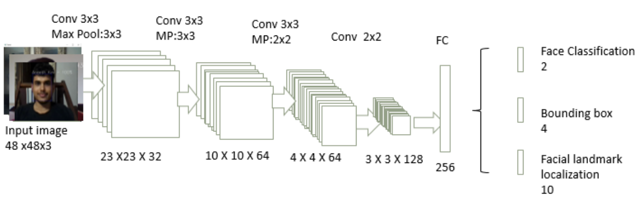
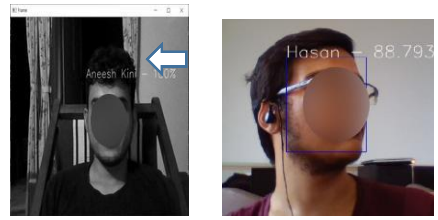


Fig. 2 The last stage of MTCNN [14]

The emotion classifier uses the Deep Convolution Neural Network. A trained model mini-Xception is used for emotion recognition [16]. It comprises of 2D Convolution with Batch Normalization. The same combination followed by max pooling four times. This followed by 2D Convolution and global averaging and 2D pooling. The results are obtained after softmax regression. The averaging of pooling, that is carried out globally, assists feature map size reduction. The fully connected layer is replaced by depth-wise separable convolutions and residual modules.

4 Results and Discussions

The experiments have been carried out using a camera for real time images to be captured and saved in the database with the names of the person as the folder directory. Fig. 3 shows the bounding box on the area of interest which is the face region. The algorithm detects the face at the front with 100% as in Fig. 3a and with reduced accuracy as in Fig. 3b. The faces are blurred to hide the privacy of the person.



(a) (b)

Fig. 3 Face contour detection with bounding box across it

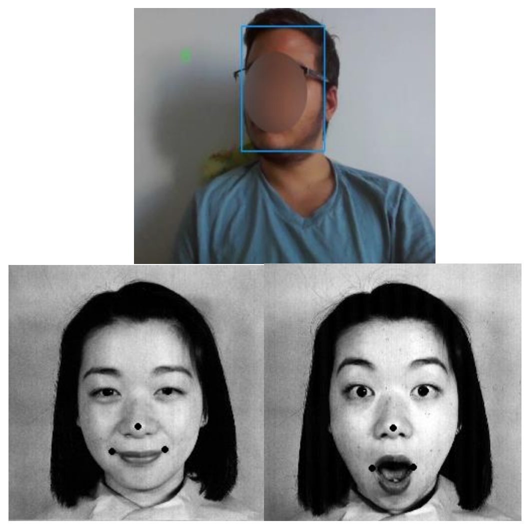


Fig. 4 Landmarks identification of eyes, nose and corners of lips for neutral, happy and surprise expressions.

Fig. 4 shows the landmarks points which are used to recognize the emotions of an individual. These points are also displayed for the static images from the JAFFE database [18].

A stepwise implementation of Eigen face recognizer is shown with Fig. 6a as the training input, Fig. 6b the testing input. Fig. 7a shows the average of all the faces, Fig. 7b the normalized face. On computing the eigen values and the eigen vector, only ten significant faces (eigen vectors with top ten eigen values) are chosen. Fig. 8 shows the matched and unmatched result of computation.

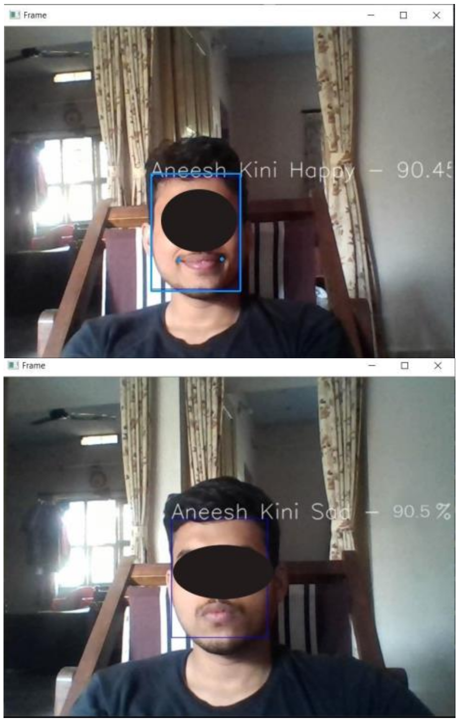
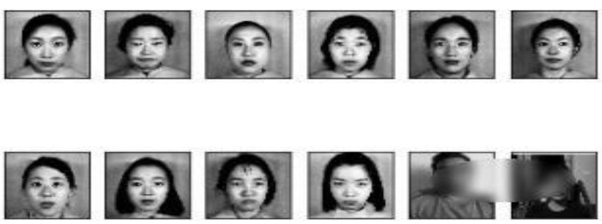


Fig. 5 Face recognition with happy and sad emotions.



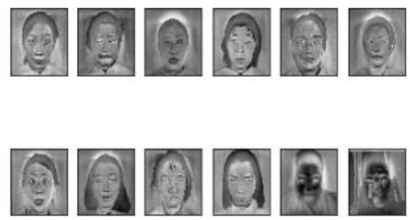
(a)

(b)

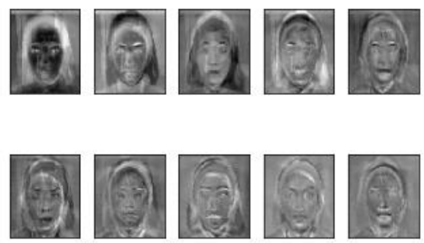
Fig 6a) Training input b) Testing input

Fig. 5 shows the emotional expression "happy" along with the face recognition with 90% accuracy. "Sad" is also detected with similar accuracy. Large size of the database for each individual improves the result tremendously. Results are obtained for different individual and the accuracy varies based on the angle of the image capture. Fig. 6a shows the variation of accuracy levels for different individuals. The x-axis signifies the various test inputs and y-axis implies the accuracy. Other representation of accuracy is depicted in Fig. 6b. The x-axis contains false positive rate which states, in how many cases the facial expression was falsely detected. The y-axis holds the values for true positive rate which defines how many face expressions were truly detected. The area under the curve is 0.845 of the total space, i.e. with 84.5% accuracy for face recognition. True Positive Rate signifies the number of results which are positive among all the samples which are positive. False Positive Rate states all the results which are positive but incorrect in the samples which are negative.

(a)



(b)

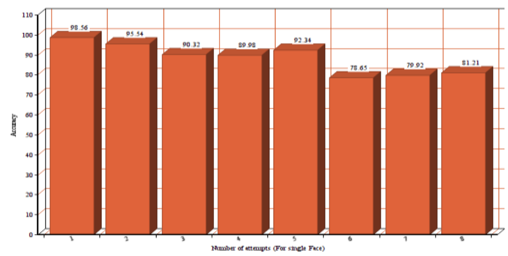


(c)

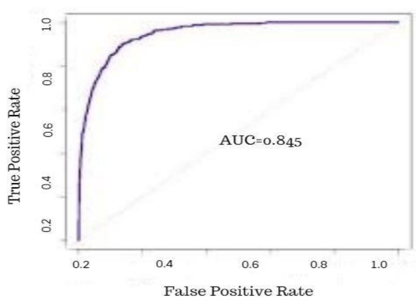
Fig. 7a) Average of all input faces 7b) Normalized face 7c) Ten significant face

Fig 8 Result of comparison of trained and tested input which are recognized with match or wrong match

Table 1 shows the training time taken by various algorithms. It is found that the Eigen Face Recognizer takes longer to train and highly efficient as compared to Fisher and LBPH Face Recognizer. The training time for Fisher though relatively high compared to LBPH, is faster when compared to predicting the result in real time. The database consisting of the facial expression of the Japanese female (JAFFE) [16] has 213 images There are 7 facial expressions with 6 corresponding to happy, sad, surprise, angry, disgust, fear and an addition feature of neutral. This database includes faces from 10 female model. The images are in gray scale of size 256x256 pixel. FaceNet a deep Convolution algorithm has a good accuracy but the training time is equally high as that of Eigen Face, but prediction time takes longer though it is an accurate predictor. This signifies that it is good a face recognizer for large dataset and once it is trained then it is a good predictor. The presence of face in the image is carried out using Multi task Cascaded Convolution Network is used



(a)



(b)

Fig, 8 a) Accuracy for different test inputs b) Accuracy using true positive and false positive

TABLE 1
Comparison of various face recognizing algorithms

OPENCV Face Detection Algorithms Test without Manipulated Test Images						
Face recognition Algorithm	Training size	Training time in ms	Test size	Failure count	Success count	Prediction Time in ms
Eigen Face Recognizer	203	35313	27	0	27	5858
Fisher Face Recognizer	203	25045	27	1	26	1996
LBPH Face Recognizer	203	6525	27	1	26	3862
FaceNet	203	33567	27	0	27	26723
Dataset	: http://www.kasrl.org/jaffedb_info.html					

The algorithm was also tested for image which had noise introduced for the same dataset. The result of comparison is as shown in the Table 2 for the same set of algorithms. There was no much variation in the relative training time in both the cases. The number of testing samples were reduced. The prediction time is relatively the same as the dataset without manipulations. The results show that though Eigen is a good predictor, the accuracy in the case of Facenet algorithm is better than that of Eigen Face recognizer. LBPH performed badly as compared to all of them. LBPH uses the binary pattern around any pixel. This pattern gives how many pixel alternatively varies largely around the pixel. Hence, with mutilated images the histogram obtained deviates largely from correct values. In the case of Eigen Face, faces of the same persons are clustered together in the feature vector space, hence the mutilated pictures do not affect much to the results. Fishers face tries to reduce the intra cluster distance of the feature vector of the same person and increase the inter cluster distance between feature vector of different person. Hence the effect of illumination and mutilated face has not much affect. The FaceNet model extracts features of the face obtained from MTCNN model using the face embedding process. Support Vector method is used to predict the face from the trained model.

TABLE 2
Comparison face recognizing algorithms with noise introduced

OPENCV Face Detection Algorithms Test with Manipulated Test Images						
Face Recognition Algorithm	Training Count	Training Time in ms	Test Sample Count	Failure Count	Success Count	Prediction Time in ms
Eigen Face Recognizer	183	30757	30	1	29	11253
Fisher Face Recognizer	183	22187	30	0	30	5625
LBPH Face Recognizer	183	5902	30	13	17	8855
Facenet	183	31047	30	0	30	28235
Dataset:	http://www.kasrl.org/jaffedb_info.html + Test Images Manipulated					

The face recognition system helps the visionless to recognize the person along with the emotion he possesses pronouncing the name of the person and his facial expression. The Text to Speech (TTS) [19] package from python helps in reading the label (directory name) associated with each person. Hidden Markov model is used to synthesize the speech.

5 CONCLUSION

Face recognition plays an important role for identification and authentication to enhance security. By spelling out the name of the person, it can inform the impaired about the person accompanying him. The emotional expression on the face helps in identifying if a conversation has to be continued or not. Increase in the training size has shown improvement in the result, The work can be enhanced by finding if the person is stressed based on combination of mood. This can give alerts to the driver of the vehicle that he might be sleepy and hence take time off from driving.

REFERENCES

[1] Smith FW, Rossit S (2018) Identifying and detecting facial expressions of emotion in peripheral vision. PLoS ONE 13(5): e0197160.pp 1-15

[2] HAN Jing, XIE Lun, LI Dan, HE Zhijie, WANG at el. Cognitive Emotion Model for Eldercare Robot in Smart Home China Communications, Vol.12, Issue 4, pp 32-41,April 2015

[3] Seyedehsamaneh Shojaeilangari, et al. Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning IEEE Transactions on Image Processing, Vol. 24, Issue 7, pp. 2140-2152, July 2015

[4] Jianzhu Guo et al., Dominant and Complementary Emotion Recognition from still images of Faces IEEE Access, Vol. 6, pp 26391-26403, Apr. 2018

[5] Wang, Matthew et al.: "Determining Mood from Facial Expressions"; CS 229 Project; Fall 2014; pp. 1-5.

[6] Mathew Turk, Alex Pentland, " Eigen faces for Recognition", Journal of Cognitive Neruro science, Vol. 3, No. 1, 1991

[7] Müge Çarıkçı a, Figen Özen "Face recogntion System based on Eigen faces method", Procedia Technology 1 (2012) 118 – 123

[8] Ziqiang Wang, Xia Sun, "Manifold Adaptive Kernel Local Fisher Discriminant Analysis for Face Recognition", Journal of Multimedia, vol. 7, Dec. 2012.

[9] Peter N Belhumer, Joan P. Hespanha and David J. Kriegman,"Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 19, No.7, July 1997

[10] Mustamin Anggo and La Arapu, Face Recognition using Fisherface Method", Journal of Physics, pp1-9, 2018

[11] Di Huang, caifeng Shan, et. Al, " Local Binary Patterns and its application to Facial Image analysis: A Survey", IEEE Transaction on Systems, Man and Cybernetics- Part C: Appliations and Reviews., Vol. 41, No. 6, Nov. 2011

[12] Florian Schroff, Dmitry Kalenichenko, James Philbin , " FaceNet: A Unified Embedding for Face Recogntion and Clustering", Article, (2015).pp 1-10

[13] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf , " DeepFAce: Closing the gap to human-level performance in face verification",(2014).

[14] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li Yu Qiao, Joint Face Detection and Alignment using Multi-task Cascaded Convolution Networks, IEEE Signal Processing Letters, Vol. 23, no.10, pp. 1499-1503, 2016

[15] R. Padilla, C. F. F. Costa Filho and M. G. F. Costa, "Evaluation of Haar Cascade Classifiers Designed for Face Detection", World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:6, No:4, 2012

[16] Octavio Arriaga, Paul G. Ploger, Matias Valdenegro, "Real-time Convolutional Neural Networks for Emotion and Gender Classification"pp 1-5, Article (2017)

[17] Enrique Correa, Arnoud Jonker, Michacl Ozo, Rob Stolk,"Emotion Recognition using Deep Convolution Network", TU Delft, 2016

[18] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200 -205 (1998) The Japanese Female Facial expression (JAFFE) Dataset http://doi.org/10.5281/zenodo.3451524

[19] Natesh M Bhat, "pyttsx3 Documentation Release 2.6", 2019