Batch: A1 Roll No.: 16010120015

Experiment No. 9

Grade: AA / AB / BB / BC / CC / CD /DD

Signature of the Staff In-charge with date

Title: : Case study on anyone Special Neural Networks

Aim : To understand the special neural network and their applications

Expected Outcome of Experiment:

CO3 : Understand perceptron's and counter propagation networks

Books/ Journals/ Websites referred:

- J.S.R.Jang, C.T.Sun and E.Mizutani, "Neuro-Fuzzy and Soft Computing", PHI, 2004, Pearson Education 2004.
- Davis E.Goldberg, "Genetic Algorithms: Search, Optimization and Machine Learning", Addison Wesley, N.Y., 1989.
- S. Rajasekaran and G.A.V.Pai, "Neural Networks, Fuzzy Logic and Genetic Algorithms", PHI, 2003.
- http://library.thinkquest.org/C007395/tqweb/history.html

Discuss the algorithm and application of following special networks

-

TITLE :

Facial emotion recognition using convolutional neural networks (FERC)

Received: 16 July 2019 / Accepted: 12 February 2020 / Published online: 18 February 2020

Link for research Paper : https://rdcu.be/c0J4s

Abstract

Facial expression for emotion detection has always been an easy task for humans, but achieving the same task with a computer algorithm is quite challenging. With the recent advancement in computer vision and machine learning, it is possible to detect emotions from images. In this paper, we propose a novel technique called facial emotion recognition using convolutional neural networks (FERC). The FERC is based on two-part convolutional neural network (CNN): The frstpart removes the background from the picture, and the second part concentrates on the facial feature vector extraction. In FERC model, expressional vector (EV) is used to fnd the fve diferent types of regular facial expression. Supervisory data were obtained from the stored database of 10,000 images (154 persons). It was possible to correctly highlight the emotion with 96% accuracy, using a EV of length 24 values. The two-level CNN works in series, and the last layer of perceptron adjusts the weights and exponent values with each iteration. FERC difers from generally followed strategies with single-level CNN, hence improving the accuracy. Furthermore, a novel background removal procedure applied, before the generation of EV, avoids dealing with multiple problems that may occur (for example distance from the camera). FERC was extensively tested with more than 750K images using extended Cohn–Kanade expression, Caltech faces, CMU and NIST datasets. We expect the FERC emotion detection to be useful in many applications such as predictive learning of students, lie detectors, etc.

INTRODUCTION

Emotion recognition is a process of identifying human emotions most likely from facial expressions as well as from speech. The application of emotion recognition system is that it promotes emotion translation between cultures that can be used in multi-cultural communication systems. Human beings have capability to recognize emotions easily, but it is difficult for the computers to do the same. If computers could recognize these emotional inputs, they could give specific and appropriate help to users in ways that are more in tune with the user’s needs and preferences. Facial expressions help computers in detecting emotions. This paper deals with helping computers to recognize human emotions in real-time.

1.2 PROBLEM STATEMENT

Human computer interaction has been an important field of study. If computer could understand the feelings of humans, it can cater the services based on the feedback received. In this model we give the overview of the work done in the past related to Emotion Recognition using Facial expressions along with our approach towards solving the problem. The approaches used for facial expression include classifiers like Support Vector Machine (SVM), Convolution Neural Network (CNN) are used to classify emotions based on certain regions of interest on the face like lips, lower jaw, eyebrows, cheeks and many more..

1.3 OBJECTIVE AND SCOPE

- Proposing a architecture with features like skip connections,
- Testing the model on different modalities of medical image segmentation.
- Evaluating performance of the model with other existing models and analyze the results.

PROPOSED SYSTEM

Video of a person is recorded using web camera. The video is converted to frames and provided as an input to a classifier to get the desired emotion.

This system has three modules. They are

Pre-processing : Real-time video is captured using the camera at the rate of 30 frames per second (fps). The frames are in BGR (Blue Green Red) format. It is converted into greyscale format which makes computing easy.

Face detection : A pre-trained classifier called Haar-cascade provided by OpenCV is used for face detection. It also returns face coordinates. These co-ordinates are used to crop the image and obtain only the face.

Classifier : Previously built Convolution Neural Network (CNN) model is used and the license to use this CNN model is provided in [6]. The code available in the link [6] was used to train the model. The trained data is used to predict emotion. A list with probablities of all 7 emotions is obtained as an output. The required output is the maximum of these values and the corresponding emotion is predicted as the final output

LITERATURE SURVEY

The Abstracts explains the Objective of the paper and gives the context of their project.

Facial expression is the common signal for all humans to convey the mood. There are many attempts to make an automatic facial expression analysis tools [7] as it has application in many fields such as robotics, medicine, driving assist systems, and lie detector [8–10]. Since the twentieth century, Ekman et al. [11] defined seven basic emotions, irrespective of culture in which a human grows with the seven expressions (anger, feared, happy, sad, contempt [12], disgust, and surprise). In a recent study on the facial recognition technology (FERET) dataset, Sajid et al. found out the impact of facial asymmetry as a marker of age estimation [13]. Their finding states that right face asymmetry is better compared to the left face asymmetry. Face pose appearance is still a big issue with face detection. Ratyal et al. provided the solution for variability in facial pose appearance. They have used three-dimensional pose invariant approach using subject-specific descriptors [14, 15]. There are many issues like excessive makeup [16] pose and expression [17] which are solved using convolutional networks. Recently, researchers have made extraordinary accomplishment in facial expression detection [18–20], which led to improvements in neuroscience [21] and cognitive science [22] that drive the advancement of research, in the field of facial expression. Also, the development in computer vision [23] and machine learning [24] makes emotion identification much more accurate and accessible to the general population. As a result, facial expression recognition is growing rapidly as a sub-field of image processing. Some of the possible applications are human–computer interaction [25], psychiatric observations [26], drunk driver recognition [27], and the most important is lie detector [28].

3 Methodology

Convolutional neural network (CNN) is the most popular way of analyzing images. CNN is different from a multi-layer perceptron (MLP) as they have hidden layers, called convolutional layers. The proposed method is based on a two-level CNN framework. The first level recommended is background removal [29], used to extract emotions from an image, as shown in Fig. 1. Here, the conventional CNN network module is used to extract primary expressional vector (EV). The expressional vector (EV) is generated by tracking down relevant facial points of importance. EV is directly related to changes in expression. The EV is obtained using a basic perceptron unit applied on a background-removed face image. In the proposed FERC model, we also have a non-convolutional perceptron layer as the last stage. Each of the convolutional layers receives the input data (or image), transforms it, and then outputs it to the next level. This transformation is convolution operation, as shown in Fig. 2. All the convolutional layers used are capable of pattern detection. Within each convolutional layer, four filters were used. The input image fed to the first-part CNN (used for background removal) generally consists of shapes, edges, textures, and objects along with the face. The edge detector, circle detector, and corner detector filters are used at the start of the convolutional layer 1. Once the face has been detected, the second-part CNN filter catches facial features, such as eyes, ears, lips, nose, and cheeks. The edge detection filters used in this layer are shown in Fig. 3a. The second-part CNN consists of layers with 3 × 3 kernel matrix, e.g., [0.25, 0.17, 0.9; 0.89, 0.36, 0.63; 0.7, 0.24, 0.82]. These numbers are selected between 0 and 1 initially. These numbers are optimized for EV detection, based on the ground truth we had, in the supervisory training dataset. Here, we used minimum error decoding to optimize filter values. Once the filter is tuned by supervisory learning, it is then applied to the background-removed face (i.e., on the output image of the first-part CNN), for detection of different facial parts (e.g., eye, lips. nose, ears, etc.) To generate the EV matrix, in all 24 various facial features are extracted. The EV feature vector is nothing but values of normalized Euclidian distance between each face part, as shown in Fig. 3b.

3.1 Key frame extraction from input video

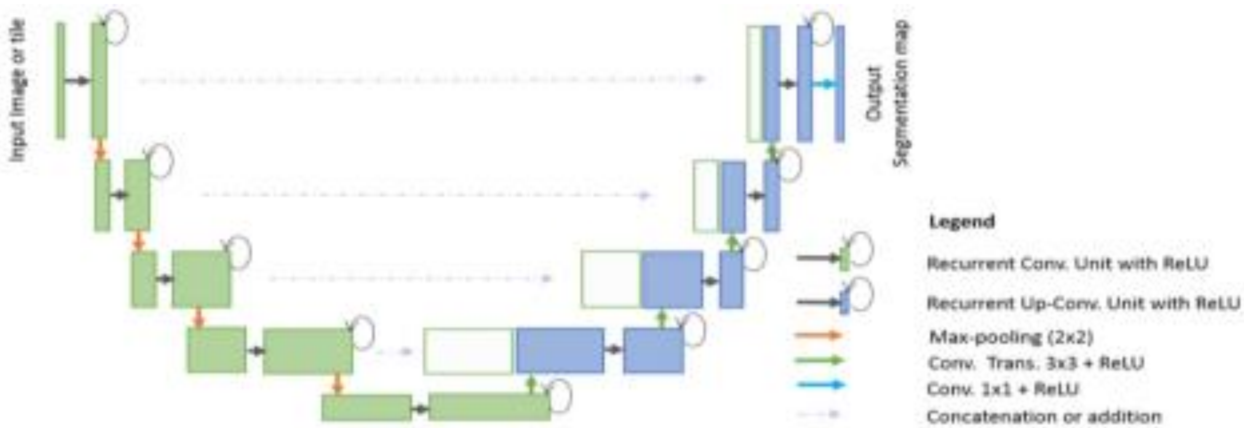
FERC works with an image as well as video input. In case, when the input to the FERC is video, then the diference between respective frames is computed. The maximally stable frames occur whenever the intra-frame diference is zero. Then for all of these stable frames, a Canny edge detector was applied, and then the aggregated sum of white pixels was calculated. After comparing the aggregated sums for all stable frames, the frame with the maximum aggregated sum is selected because this frame has maximum details as per edges (more edges more details). This frame is then selected as an input to FERC. The logic behind choosing this image is that blurry images have minimum edges or no edges.

3.2 Background removal

Once the input image is obtained, skin tone detection algorithm [30] is applied to extract human body parts from the image. This skin tone-detected output image is

Fig. 2: Sample results of emotion

IMPLEMENTATION



The architecture comprises of the following blocks: Convolution Blocks

- This is the first layer and one of the main building blocks of a Convolutional Neural Networks (CNNs).
- They hold the raw pixel values of the training image as input.
- This layer ensures the spatial relationship between pixels by learning image features using small squares of input data.

Recurrent Convolutional Block

- Feature accumulation with recurrent convolutional layers ensures better feature representation for segmentation tasks.
- Recurrent network learns from neighbouring units which helps us to include context information of an image.

Encoding Block

● Takes an input image and generates a high

dimensional feature vector.

Decoding Block

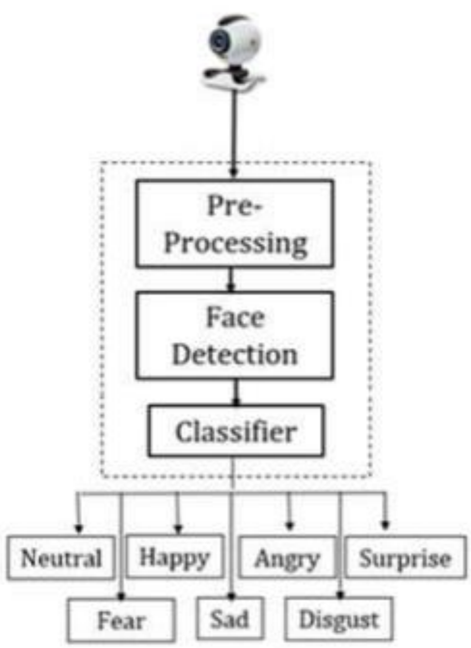
● Takes a high dimensional feature vector and

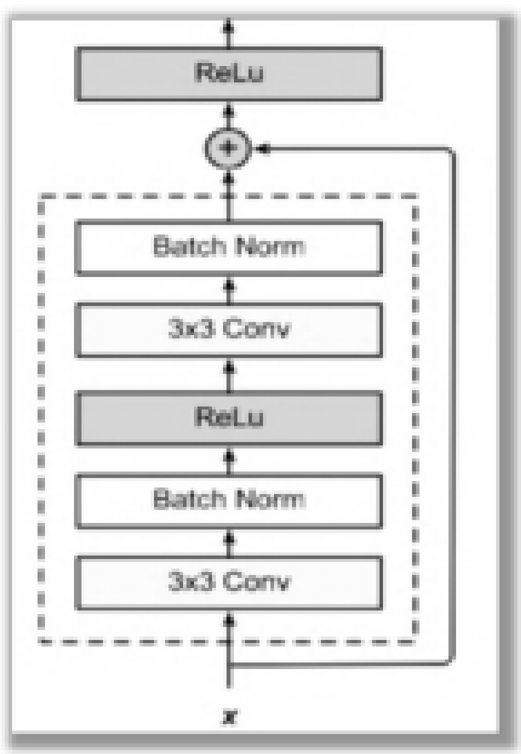
generates a semantic segmentation mask.

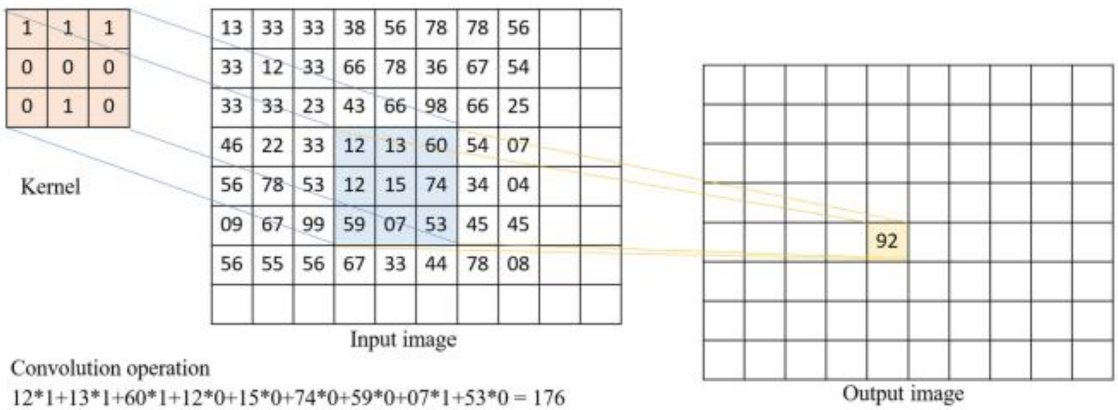
● Decode features aggregated by encoder at multiple levels.

Skip Connections

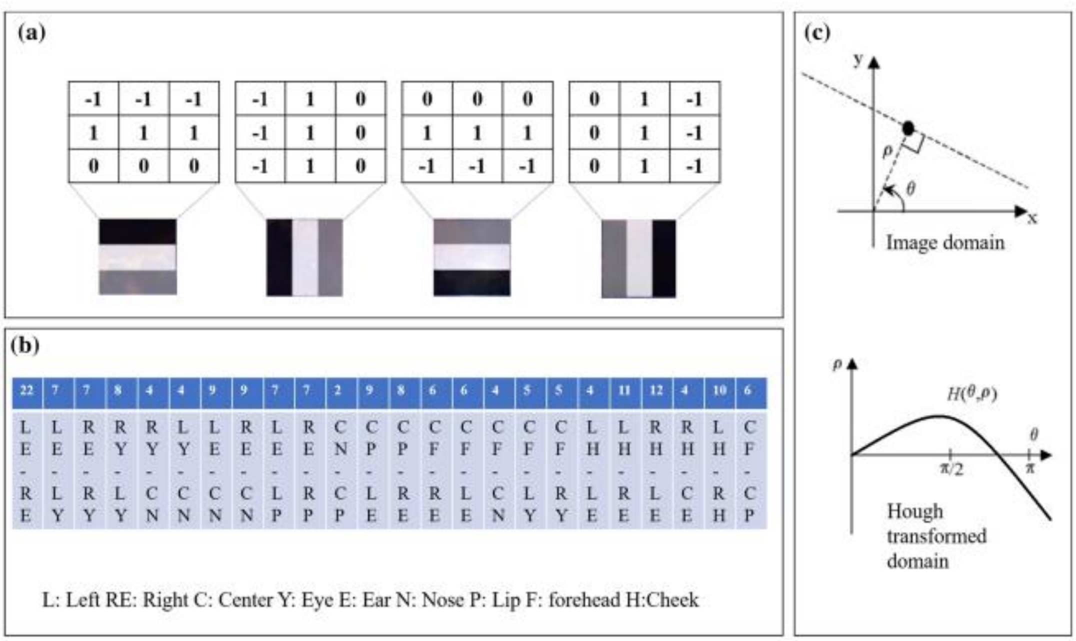
● The information from the initial layers is passed to deeper layers by matrix addition.







a binary image and used as the feature, for the frst layer of background removal CNN (also referred to as the frst-part CNN in this manuscript). This skin tone detection depends on the type of input image. If the image is the colored image, then YCbCr color threshold can be used. For skin tome, the Y-value should be greater than 80, Cb should



range between 85 and 140, Cr value should be between 135 and 200. The set of values mentioned in the above line was chosen by trial-and-error method and worked for almost all of the skin tones available. We found that if the input image is grayscale, then skin tone detection algorithm has very low accuracy. To improve accuracy during background removal, CNN also uses the circles-in-circle flter. This flter operation uses Hough transform values for each circle detection. To maintain uniformity irrespective of the type of input image, Hough transform (Fig. 3c) was always used as the second input feature to background removal CNN. The formula used for Hough transform is as shown in Eq. 1

$$H(\theta, \rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y)\delta(\rho - x\cos\theta - y\sin\theta)\mathrm{d}x\mathrm{d}y \quad (1)$$

3.3 Convolution filter

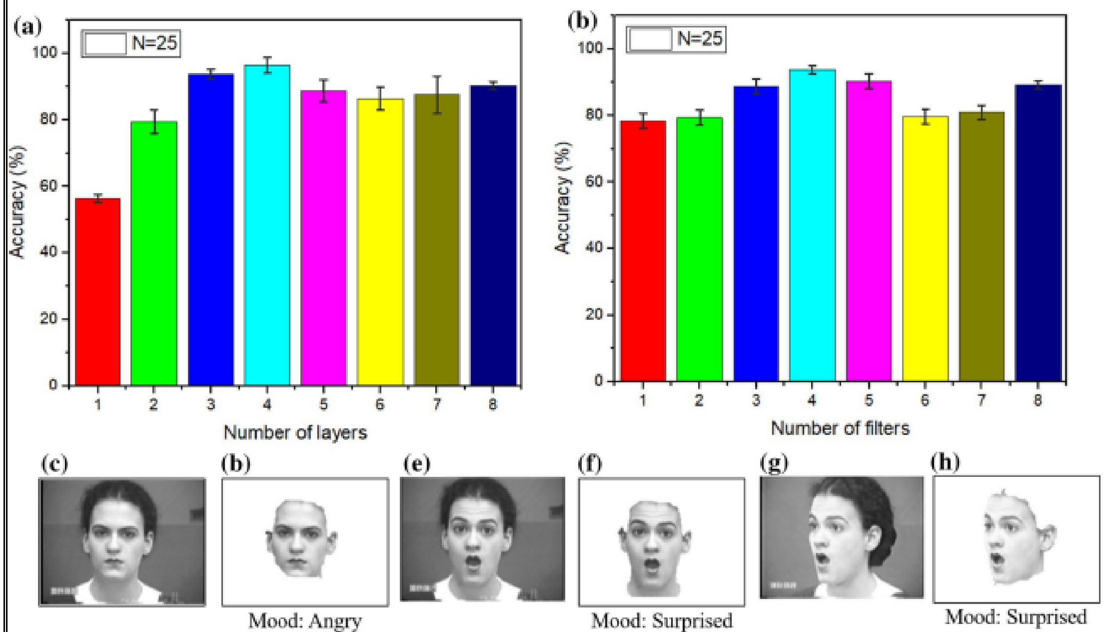
As shown in Fig. 2 for each convolution operation, the entire image is divided into overlapping 3 × 3 matrices, and then the corresponding 3 × 3 flter is convolved over each 3 × 3 matrix obtained from the image. The sliding and taking dot product operation is called ‘convolution’ and H(휃, 휌) = (1) ∫ ∞ −∞ ∫ ∞ −∞ A(x, y)훿(휌 − x cos 휃 − y sin 휃)dxdy hence the name ‘convolutional flter.’ During the convolution, dot product of both 3 × 3 matrix is computed and stored at a corresponding location, e.g., (1,1) at the output, as shown in Fig. 2. Once the entire output matrix is calculated, then this output is passed to the next layer of CNN for another round of convolution. The last layer of face feature extracting CNN is a simple perceptron, which tries to optimize values of scale factor and exponent depending upon deviation from the ground truth.

3.4 Hardware and software details

All the programs were executed on Lenovo Yoga 530 model laptop with Intel i5 8th generation CPU and 8 GB RAM with 512 GB SSD hard disk. Software used to run the experiment were Python (Using Thonny IDE), MATLAB 2018a, and ImageJ.

4 Results and discussions

To analyze the performance of the algorithm, extended Cohn–Kanade expression dataset [31] was used initially. Dataset had only 486 sequences with 97 posers, causing accuracy to reach up to 45% maximum. To overcome the problem of low efciency, multiple datasets were downloaded from the Internet [32, 33], and also author’s own pictures at diferent expressions were included. As the number of images in dataset increases, the accuracy also increased. We kept 70% of 10K dataset images as training and 30% dataset images as testing images. In all 25 iterations were carried out, with the diferent sets of 70% training data each time. Finally, the error bar was computed as the standard deviation. Figure 4a shows the optimization of the number of layers for CNN. For simplicity, we kept the number of layers and the number of flters, for background removal CNN (frst-part CNN) as well as face feature extraction CNN (the second-part CNN) to be the same. In this study, we varied the number of layers from 1 to 8. We found out that maximum accuracy was obtained around 4. It was not very intuitive, as we assume the number of layers is directly proportional to accuracy and inversely proportional to execution time. Hence due to maximum accuracy obtained with 4 layers, we selected the number of layers to be 4. The execution time was increasing with the number of layers, and it was not adding signifcant value to our study, hence not reported in the current manuscript. Figure 4b shows the number of flters optimization for both layers. Again, 1–8 flters were tried for each of the four-layer CNN networks. We found that four flters were giving good accuracy. Hence, FERC was designed with four layers and four flters. As a future scope of this study, researchers can try varying the number of layers for both CNN independently. Also, the vast amount of work can be done if each layer is fed with a diferent number of flters. This could be automated using servers. Due to computational power limitation of the author, we did not carry out this study, but it will be highly appreciated if other researchers come out with a better number than 4 (layers), 4 (flters) and increase the accuracy beyond 96%, which we could achieve. Figure 4c and e shows regular front-facing cases with angry and surprise emotions, and the algorithm could easily detect them (Fig. 4d, f). The only challenging part in these images was skin tone detection, because of the grayscale nature of these images. With color images, background removal with the help of skin tone detection was straightforward, but with grayscale images, we observed false face detection in many cases. Image, as shown in Fig. 4g, was challenging because of the orientation. Fortunately, with 24 dimensions EV feature vector, we could correctly classify 30° oriented faces using FERC. We do accept the method has some limitations such as high computing power during CNN tuning, and also, facial hair causes a lot of issues. But other than these problems, the accuracy of our algorithm is very high (i.e., 96%), which



4.1 Comparison with other methods

As shown in Table 2, FERC method is a unique method developed with two 4-layer networks with an accuracy of 96%, where others have just gone for a combined approach of solving background removal and face ing both issues separately reduces complexity and also the tuning time. Although we only have considered fve moods to classify, the sixth and seventh mood cases were misclassified, adding to the error. Zao et al. [37] have achieved maximum accuracy up to 99.3% but at the cost of 22 layers neural network. Training such a large network is a time-consuming job. Compared to existing methods, only FERC has keyframe extraction method, whereas others have only gone for the last frame. Jung et al. [38] tried to work with fxed frames which make the system not so efcient with video input. The number of folds of training in most of the other cases was ten only, whereas we could go up to 25-fold training because of small network size. As shown in Table 3, FERC has similar complexity as that of Alexnet. FERC is much faster, compared to VGG, GoogleNet, and Resnet. In terms of accuracy, FERC outperforms existing standard networks. However, in some cases we found GoogleNet out-performs FERC, especially when the iteration of GoogleNet reaches in the range of 5000 and above.

CONCLUSION

FERC is a novel way of facial emotion detection that uses the advantages of CNN and supervised learning (feasible due to big data). The main advantage of the FERC algorithm is that it works with diferent orientations (less than 30°) due to the unique 24 digit long EV feature matrix. The background removal added a great advantage in accurately determining the emotions. FERC could be the starting step, for many of the emotion-based applications such as lie detector and also mood-based learning for students, etc.

VI FUTURE WORK Several emotions such as happiness, anger, surprise and fear are better classified using facial recognition. Some other are classified better using speech input such as sadness and fear. So in order to achieve better accuracy, we need to integrate both audio and facial based systems into a single system

References

1. Mehrabian A (2017) Nonverbal communication. Routledge,

London

2. Bartlett M, Littlewort G, Vural E, Lee K, Cetin M, Ercil A, Movellan

J (2008) Data mining spontaneous facial behavior with automatic expression coding. In: Esposito A, Bourbakis NG, Avouris

N, Hatzilygeroudis I (eds) Verbal and nonverbal features of

human–human and human–machine interaction. Springer,

Berlin, pp 1–20

3. Russell JA (1994) Is there universal recognition of emotion from

facial expression? A review of the cross-cultural studies. Psychol

Bull 115(1):102

4. Gizatdinova Y, Surakka V (2007) Automatic detection of facial

landmarks from AU-coded expressive facial images. In: 14th

International conference on image analysis and processing

(ICIAP). IEEE, pp 419–424

5. Liu Y, Li Y, Ma X, Song R (2017) Facial expression recognition

with fusion features extracted from salient facial areas. Sensors

17(4):712

6. Ekman R (1997) What the face reveals: basic and applied studies of spontaneous expression using the facial action coding

system (FACS). Oxford University Press, New York

7. Zafar B, Ashraf R, Ali N, Iqbal M, Sajid M, Dar S, Ratyal N (2018)

A novel discriminating and relative global spatial image representation with applications in CBIR. Appl Sci 8(11):2242

8. Ali N, Zafar B, Riaz F, Dar SH, Ratyal NI, Bajwa KB, Iqbal MK, Sajid

M (2018) A hybrid geometric spatial image representation for

scene classifcation. PLoS ONE 13(9):e0203339

9. Ali N, Zafar B, Iqbal MK, Sajid M, Younis MY, Dar SH, Mahmood MT,

Lee IH (2019) Modeling global geometric spatial information for

rotation invariant classifcation of satellite images. PLoS ONE 14:7

10. Ali N, Bajwa KB, Sablatnig R, Chatzichristofs SA, Iqbal Z, Rashid

M, Habib HA (2016) A novel image retrieval based on visual

words integration of SIFT and SURF. PLoS ONE 11(6):e0157428

11. Ekman P, Friesen WV (1971) Constants across cultures in the face

and emotion. J Personal Soc Psychol 17(2):124

12. Matsumoto D (1992) More evidence for the universality of a

contempt expression. Motiv Emot 16(4):363

13. Sajid M, Iqbal Ratyal N, Ali N, Zafar B, Dar SH, Mahmood MT,

Joo YB (2019) The impact of asymmetric left and asymmetric

right face images on accurate age estimation. Math Probl Eng

2019:1–10

14. Ratyal NI, Taj IA, Sajid M, Ali N, Mahmood A, Razzaq S (2019)

Three-dimensional face recognition using variance-based registration and subject-specifc descriptors. Int J Adv Robot Syst

16(3):1729881419851716

15. Ratyal N, Taj IA, Sajid M, Mahmood A, Razzaq S, Dar SH, Ali N,

Usman M, Baig MJA, Mussadiq U (2019) Deeply learned pose

invariant image analysis with applications in 3D face recognition. Math Probl Eng 2019:1–21

16. Sajid M, Ali N, Dar SH, Iqbal Ratyal N, Butt AR, Zafar B, Shafique T, Baig MJA, Riaz I, Baig S (2018) Data augmentation-assisted makeup-invariant face recognition. Math Probl Eng 2018:1–10

17. Ratyal N, Taj I, Bajwa U, Sajid M (2018) Pose and expression invariant alignment based multi-view 3D face recognition. KSII Trans Internet Inf Syst 12:10

18. Xie S, Hu H (2018) Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans Multimedia 21(1):211

19. Danisman T, Bilasco M, Ihaddadene N, Djeraba C (2010) Automatic facial feature detection for facial expression recognition. In: Proceedings of the International conference on computer vision theory and applications, pp 407–412. https://doi.org/10.5220/0002838404070412

20. Mal HP, Swarnalatha P (2017) Facial expression detection using facial expression model. In: 2017 International conference on energy, communication, data analytics and soft computing (ICECDS). IEEE, pp 1259–1262

21. Parr LA, Waller BM (2006) Understanding chimpanzee facial expression: insights into the evolution of communication. Soc Cogn Affect Neurosci 1(3):221

22. Dols JMF, Russell JA (2017) The science of facial expression. Oxford University Press, Oxford

23. Kong SG, Heo J, Abidi BR, Paik J, Abidi MA (2005) Recent advances in visual and infrared face recognition—a review.

Comput Vis Image Underst 97(1):103

24. Xue Yl, Mao X, Zhang F (2006) Beihang university facial expression database and multiple facial expression recognition.

In: 2006 International conference on machine learning and

cybernetics. IEEE, pp 3282–3287

25. Kim DH, An KH, Ryu YG, Chung MJ (2007) A facial expression imitation system for the primitive of intuitive humanrobot interaction. In: Sarkar N (ed) Human robot interaction.

IntechOpen, London

26. Ernst H (1934) Evolution of facial musculature and facial

expression. J Nerv Ment Dis 79(1):109

27. Kumar KC (2012) Morphology based facial feature extraction

and facial expression recognition for driver vigilance. Int J

Comput Appl 51:2

28. Hernández-Travieso JG, Travieso CM, Pozo-Baños D, Alonso

JB et al (2013) Expression detector system based on facial

images. In: BIOSIGNALS 2013-proceedings of the international

conference on bio-inspired systems and signal processing

29. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S,

Fellenz W, Taylor JG (2001) Emotion recognition in human–

computer interaction. IEEE Signal Process Mag 18(1):32

30. Hsu RL, Abdel-Mottaleb M, Jain AK (2002) Face detection in

color images. IEEE Trans Pattern Anal Mach Intell 24(5):696

[31] Kaggle, URL:

https://www.kaggle.com/c/challengesin-representation-learning-facialexpression recognition-challenge/data[Last accessed: Dec 2018]

[32] License Link,URL:https://github.com/oarriaga/face_classification/blob/master/LICENSE

Date: __30/11/2022__ **Signature of faculty in-charge**

/