EXPOSING FACIAL EXPRESSIONS THROUGH NEURAL NETWORKS AND LANDMARK DETECTION

Rameshkumar J[1], Mahendran P[2], Vijayanandh T[3] and Veeramani R[4]

[1]Department of Electronics and Communication Engineering, K.S.R. College of Engineering, Namakkal, Tamil Nadu, India.

[2]Department of Electronics and Communication Engineering, K.S.R. College of Engineering, Namakkal, Tamil Nadu, India.

[3]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India.

[4]Department of Electronics and Communication Engineering, K.S.R.College of Engineering, Namakkal, Tamil Nadu, India.

Abstract: Facial expression recognition is an integral component of facial recognition, and its importance and demand are rapidly increasing. While there are existing methods that utilize machine learning and artificial intelligence techniques to identify expressions, this project aims to employ deep learning and image classification methods for expression recognition and categorization based on the available data. Various forms of data are explored and analyzed to train the expression recognition model. The implementation of Local Binary Patterns Histograms is utilized for accurate expression recognition. The Local Binary Patterns Histograms receptive field, which represents the extent of connectivity, is a hyper parameter.

Keywords: Deep Learning, Facial Recognition, Image Recognition, Convolutional Neural Network (CNN)

I. INTRODUCTION

The face location determination stage, facial landmark detection stage, feature extraction stage, and emotion feelings classification stage are the four general steps of the implementation. It was necessary to gather an acceptable facial database. It functions as both our training and testing data set and simply consists of pictures of people with labels applied to their expressions [1]. The first step relates to the facial recognition techniques that will be applied. These algorithms primarily handle image pre-processing and normalization to remove unnecessary regions from the images. The second stage states that the important aspects of emotion identification are our facial fiducial landmarks, which include the mouth, nose, eyes, and eyebrows. Their feature points are then retrieved to identify theemotion for this purpose, these feature points are mostly taken from standard edge detection and corner point detection processes in the Shi Tomasi corner point detection and Sobel horizontal edge detection, in addition to cutting-edge face landmark identification techniques. After that, the extracted points for the facial features were used to produce the input feature vectors

The connections within the Local Binary Patterns Histograms are local in space, spanning width and height, but encompass the entire depth of the input volume. This architectural design ensures that the learned filters generate the strongest response to a spatially local input pattern. However, in certain cases where the input data, such as centered faces, possess specific structures, the parameter-sharing assumption may not be applicable. For instance, different facial features like eyes or hair may require distinct learning in different parts of the image. In such scenarios, it is common to relax the parameter-sharing scheme and refer to the layer without it.

II. RELATED WORK

Numerous strategies and tactics have been used to deploy Facial Expression Recognition systems. While there are linguistic, paralinguistic and hybrid methods which are used as the main base of one another, there is a big group of the other approaches which use the facial features analysis and examination as their main tool. The appearance traits are often taken from various face areas that contain various kinds of information [4, 5] or the global face region [3]. Happy et al. [3] used salient facial patch traits to identify facial expressions. Following the extraction of facial landmark features, dimensionality reduction, and performance enhancement were achieved through the use of a PCA-LDA hybrid technique.

Some approaches [6] for hybrid methods have integrated geometry and appearance elements to enhance each other's beneficial effects and, in some situations, produce better results. Szwoch et al. [9] used only the depth channel from the Microsoft Kinect sensor to identify mood and facial expression in the absence of a camera. Relationships between particular emotions and local movements in the face region are used to understand facial expressions.

Moreover, Sujono et al. [10] applied face tracking using the Active Appearance Model (AAM) and the Kinect sensor to achieve facial elements like eyes, mouth, etc. Presenting face micro-expression identification in movies taken from a high-speed camera shooting at 200 frames per second (fps) was done by Polikovsky et al. [11].

Using this method, the face areas are divided into specific localized regions. The motion in each local region is then used to build a 3D gradient orientation Histogram.

Shen et al. [12] extracted horizontal and vertical temperature variations from several face subregions using infra-thermal videography. Expression Recognition uses the Adaboost algorithm with the k-nearest Neighbor weak classifiers. Because visible light images depend far more on illuminationthan infraredimages do, some researchers [9, 10, 12, 13, and 14] have attempted to identify facial expressions using infrared photos instead of visible light images. Near-infrared (NIR) video sequences and LBP-TOP feature descriptors were used by Zhao et al. [13].

This study combines geometry and appearance information about the face using component- based facial traits. Conventional field-effect relay systems are investigated for application in mobile phones due to their high degree of precision, dependability, and low processing power. Generally speaking, these systems use a great deal less memory and processing power than their deep learning-based counterparts.

III. DATA COLLECTION

To evaluated the methods using the Karolinska Directed Emotional Faces dataset [1] after conducting a thorough search for relevant datasets.It has been designed and produced in the Psychology Section of the Department of Clinical Neuroscience at the Karolinska Institute, Stockholm, Sweden.The set of 4900 photos includes 70 persons (35 men and 35 women), each of whom displays one neutral emotion and six fundamental expressions. Every expression is captured twice, from five distinct perspectives. 562 x 762 pixels make up the original image size. Figure 1 displays a selection of photos displaying each of the seven expressions. The training and testing set of data are split 90:10 each.

IV. IMAGE PRE-PROCESSING

The initial stage after receiving the picture is image pre-processing, which is done to improve the contrast and remove unwanted noise from the image. After using a Low Pass 3x3 Gaussian filter, the image was smoothed and the gradient intensity values were normalized. For illumination corrections, contrast adaptive histogram equalization was then applied. Pre-processing is usually done to make sure the input photographs are of the same size and shape. This wouldn't apply to the photographs in the KDEF database because each picture's face location and orientation are consistent, negating the requirement for any pre-processing unique to the database.

V. FACIAL DETECTION

Figure 1 A selection of photos featuring six fundamental emotions from the KFED facial database

The objective of the facial detection stage is to find the face. As such, it also sets up our region of interest, so that all the subsequent operations in the following step will happen within that particular region of interest. This facial detection step was finished with the existing very reliable Haar classifier technology.

VI. DETECTION OF LANDMARK

A. *Detection of eyes*

The aim is to extract feature points that are linearly proportional to the eyes, and to this end the ROIs of the left and right eyes will be taken. The specific method contained the action of the Circular Hough Transform with the input gray level plot of the suitable region of the face [16, 17].

The characteristic equation of a circle with radius r and center (m, n) is as follows:

$$(c-m)+(d-n)^2=r^2 \tag{1}$$

The objective of the Hough transform is to determine the parameters of triplet (a, b, r) which yield the circles in the output. Circular Hough transform makes the problem going from three dimensions of space into two dimensions of space easier by initially solving only the center of the circle (m, n) on two-dimensional parameter space (a, b) with unknown radius.

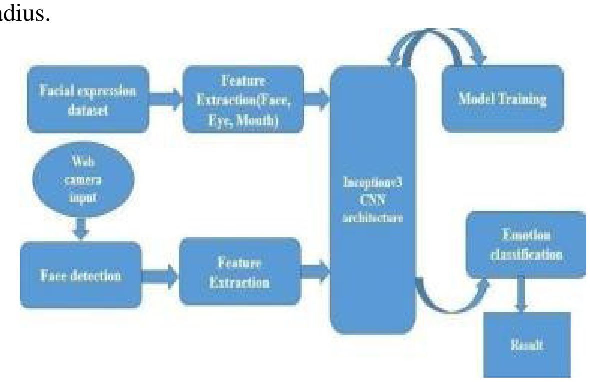


Figure 2 Proposed FER approach flowchart

The next step is to determine the radius, which is denoted by r, of each circle, associated with the local maximum. The degree of this spatial gradient around each circle wherever an eye is centered to is computed to obtain the resultant electric field. The eyes are chosen to meet the following requirements to reduce false positives caused by the Hough change:

(1) The interpupillary line's actual slant is less than 20 degrees from the horizontal.
(2) The pixels that make up the interpupillary distance range from 80 to 240.

Next, a bounding box is created around the eyeballs that have been detected, with the Hough Circle serving as the box's centroid

B. *Nose detection*

The nose is made of nine points. These points are directly taken from the 49-point landmark detector proposed for the human face by the Interfaces and heuristic techniques ensure that they place well on the nose and so appear in the image overall.

C. Detection of lips and eyebrows

Impressively, Interface's 49-point landmark recognition can localize the points for the lips (18 points) and eyebrows. (10 points). An edge detector can identify the distinctive edge produced by the top lip and eyebrows. These algorithms would accurately detect edges; these include popular ones like Sobel, Canny, Prewitt, and Roberts. Utilizing the gradient-based Sobel edge detector, the lip and eyebrow detection features were the most precise and thus should be chosen for this purpose.

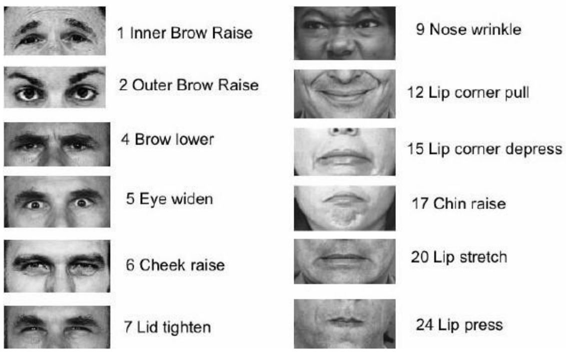


Figure 3 A selections of photos from the KFED facial database demonstrating distinct variations in facial landmarks for a range of expressions.

As finding the derivative at various points along the X-axis mentality is the main mission of edge detection, we apply the horizontal operator of the Sobel and Sobel horizontal operator for lip and eyebrow edge detection. A threshold amount which the operator may apply in vain to obtain only the strong edges from the horizontal peaks from the various operations will be used.Surplus edges persisted even after the edge detection process was completed. It was necessary to go through a trial-and-error procedure to see if binary thresholding techniques worked. This efficacy stemmed from the degree to which the superfluous edges were eliminated, allowing us to concentrate on the key feature points. When it came to this, Otsu's thresholding outperformed other binary thresholding methods as the optimal method. Its bi-modal histogram design makes the goal easier to achieve. These fictitious components that had an area below a predetermined cutoff were also eliminated.

VII. FEATURE EXTRACTION OF CORNER POINTS

The corner point detector is built by using the region of interest windows that were produced for the eyebrows and lips to obtain the corner point features. By identifying which windows, when moved in both A and B directions, result in very substantial fluctuations in intensity, the Shi Tomasi technique [20] calculates the A and B gradients. Once such a window has been located, a score R is calculated. There is a window function = w(m,n) associated with each corresponding window.

$$\Sigma(x,y) I_y\, I_x \rightarrow \Sigma(x,y) I_y^{2} \qquad (2)$$

Where,

Derivatives of the image in the x and y directions are represented by the variables Ia and Ib, respectively.

The selected feature corner points are:

$$-\Sigma(x,y) I_x^{2} \rightarrow \quad \Sigma(x,y) I_x I_y \qquad (3)$$

$$\Sigma(x,y) I_y\, I_x \rightarrow \Sigma(x,y) I_y^{2} \qquad (4)$$

The value R for each window:

$$R = min\,(\lambda 1, \lambda 2) \qquad (5)$$

Where,

λ1 and λ2 are the eigenvalues of M

Notable corners are located and labeled when they are reached at a certain point in the R processing process. Besides the adjustments like corners of the image's minimum quality, the maximum Euclid distance between both corners and a window size of computing the matrix derivative of the co variation for the return of the corner coordinates of the fiducial landmark is done.

- The selected feature's points of intersection are:
- Ten eyebrow feature points: five on the left & five on the right.
- Twelve feature points total—six on the left and six on the right—for the eyes.
- Nine feature points go toward the nose. Eighteen feature points are for the lips.

VIII. FINDING FEATURE VECTORS

The next step after deriving the feature points from the pictures would be to select the feature vectors that will be inputted and taken into consideration for the training of our artificial neural network. The 49 landmark locations that were identified were inserted into their intra-landmark Euclidean distance vectors. It was further shown that these intra-landmark vectors contain distance parameters unique to a combined landmark and normalized face. Additionally there is a need for the inclusion of a distance parameter to frame of reference. Consequently, the neural network will learn from these input feature vectors to categorize the mood of the final output. Table 1 displays a sampling of the feature vector inputs.

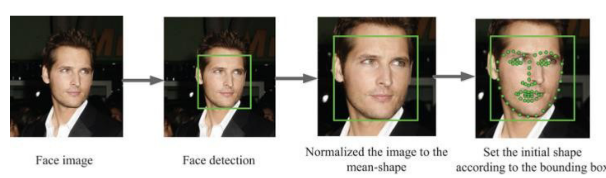


Figure 4 A few example points are displayed together with different phases of face landmark detection and feature extraction

Table 1

Insert the feature vectors and the summary

Definition	Formula
Height of Left eye	A1=f1-f2
Width of Left eye	A2=f4-f3
Height of Right eye	A3=f5-f6
Width of Right eye	A4=f8-f7
Width of Left eye brow	A5=f11-f10
Width of Right eye brow	A6=f14-f13
Width of width	A7=f17-f16
Left eye upper corner and left eye brow center dist.	A8=f12-f1
Right eye upper corner and right eye brow center dist.	A9=f15-f5
Nose center and lips center dist.	A10=f9-f18
Left eye lower corner and lips left corner dist.	A11=f2-f16
Right eye lower corner and lips right corner dist.	A12=f6-f17

IX. NEURAL NETWORK TRAINING

A. *Neural networks and back propagation*

In recent years, neural networks have demonstrated their ability to provide straightforward and effective solutions in the domains of artificial intelligence, computer vision, and signal processing. The back-propagation neural network is the most common implementation of AI because of its flexibility and ease of implementation. The approach of a multi-layer perceptron neural networks equisition of synaptic weights is process by the back-propagation concept. Supervised learning is built on a foundation composed of a descent method which minimizes a global error on output layer. The learning algorithm consists of the actions of forward moving and the actions of back propagation i.e., the feed forwarding phase and the backward propagation phase. As the input pattern represents, the inputs coming from the first layer of processing units are diffused step by step through the layers of processing units and trigger a solution pattern in the output. In the second stage, the synapses will be modified to lower the error after the errors detected in the output layer are transferred back to the hidden layers by the part of back propagation algorithm responsible for reweighing the synapses. This is a cycle that is replicated until the error value created during the first batch reaches a threshold value.

B. *The neural network configuration*

The lightweight MLP neural network consists of an input layer, two hidden layers, and an output layer. Compared to 100 neurons in the first hidden layer, there are 500 neurons in the second hidden layer, which has a SoftMax output layer of seven labels.

Non-linear activation functions are applied to the neurons in an MLP network concerning the activation functions of the nodes. To ascertain the activation function, a, the objective and output of the neural network, where i ϵ [0,6] denotes each of the seven output sensations, must be analyzed. The function of sigmoid.

$$\text{Emotionclassified}=\max(P(Y_i/x))i\epsilon[0,6] \qquad (6)$$

The output of a given k-class classification is a k-dimensional vector, where each of the k values ρ(0, 1) adds up to 1. The SoftMax Sigmoid activation function is suitable for a multi-class classificationWith the image x, the final emotion categorized is therefore = max(P(Yi/x)) ∀ I ϵ [0,6], or the maximum of the probability of all 6 emotionsresults, while Table 3 displays the false positive detection rates for each emotion.

Table 2

Matrix of confusion for the classification of emotions

I/O	Happiness	Anger	Disgust	Surprise	Fear	Sadness	Neutral
Happy	97.2%	1.1%	0%	1%	1.3%	0.3%	0.2%
Anger	1%	88.6%	8.3%	3.6%	1.5%	0.7%	0.1%
Disgust	1%	4.4%	86.9%	1.1%	1.6%	8.0%	1.3%
Surprise	0.5%	0.2%	1.2%	92.8%	1.9%	2.4%	0%
Fear	1.4%	1.6%	4.3%	1.8%	82.5%	3.4%	2.4%
Sad	0.4%	1.2%	0.6%	0.6%	7.6%	76.6%	1.5%
Normal	0.7%	0.4%	2.7%	0%	1.4%	3.3%	80.1%

Tables 2 and 3 demonstrate how well the emotions of surprise and happiness are recognized, with over 95% of true positive success rates. This suggests that some emotions lead to more constant facial expressions than others, increasing the chances of success. Anger, sadness, disgust, and fear all have success rates between 84% and 86% and overlap with nearly every other emotion, suggesting that different people may react differently to the same emotion on their faces. In contrast, the recognition rate of the neutral feeling is 90.1%, which is quite good.

Table 3

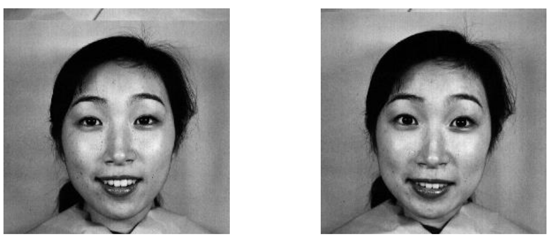
Rate of false positives for each feeling

Emotion	False rate
Happy	1.6%
Anger	13.4%
Disgust	17.1%
Surprise	2.2%
Fear	14.5%
Sad	16.4%
Normal	8.9%

It is common knowledge that people react differently to the same feelings and that expression is necessarily subjective. Person A's expression of dread, for instance,canPerson B's exhibit of revulsion. By focusing on feature vectors associated with the lips, we will only be able tofeature vectors are A7, A10, A11, and A12. This idea operates under the implicit presumption that, for example, a person is more expressive and their feature vector will be larger when they smile. the integration of the InceptionV3 algorithm into a convolutional neural network framework for facial expression recognition emotion detection from facial cues. Through the proposedsystem, we have demonstrated the efficacy of InceptionV3. With the image x, the final emotion categorized is therefore = max(P(Yi/x)) ∀ I ϵ [0,6], or the maximum of the probability of all 6 emotions

X. RESULTS AND CONCLUSION

After training the neural network as previously stated, the performance of the proposed FER system was assessed using our testing set of images from the KDEF database. Table 2 displays the confusion matrix derived from the test in capturing intricate facial details, enabling the discernment of subtle emotional nuances such as sadness, happiness, neutrality, and anger. The diversity of the training dataset has contributed to the adaptability of the model across different individuals and cultural contexts. To comprehensive evaluation and comparative analysis have showcased the superiority of the InceptionV3-based model over traditional CNN architectures, emphasizing the importance of leveraging advanced our concept of subjective expressions in the data are shown in Figure.



(a) Greater (b) lesser

Figure 5 Different degrees of expression are used to describe the emotion of happiness.

Better outcomes across the two groups were obtained from experimenting with these two sets of data, particularly when it came to the emotions of fear and rage. As anticipated, there is a larger degree of association between the "more expressive" and "less expressive" groups in how they communicate their feelings. Future research on the uniqueness and intimate quality of facial expressions can improve our ability to identify emotions. Because each user has unique mannerisms and behaviors, and because these personal assistants will be learning from them exclusively, their capacity to accurately determine the moods and emotions of their specific users will be substantially strengthened. Another limiting restriction is the number of photographs in the face expression databases, as more could produce better results given future developments.

REFERENCES

XI. [1] D. Lundqvist, A. Flykt, A. hman, (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630- 7164-9

XII. [2] D. Ghimire, J. Lee, Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. Sensors 2013, 13, 77147734. [CrossRef] [PubMed]

XIII. [3] S.L Happy, A. Routray, Automatic Facial Expression Recognition Using Features of Salient Facial Patches in IEEE Transactions on Affective Computing - May 2015, DOI 10.1109.

XIV. [4] M.H. Siddiqi, R. Ali, A.M Khan, Human facial expression recognition using stepwise linear discriminant analysis and hidden conditionalrandom fields. IEEE Trans. Image Proc. 2015, 24, 13861398.

XV. [5] R.A Khan, A. Meyer, H. Konik, S. Bouakaz, Framework for reliable, real-time facial expression recognition for low-resolution images. Pattern Recognit. Lett. 2013, 34, 11591168.

XVI. [6] D. Ghimire; S. Jeong; J. Lee; S.H Park, Facial expression recognition based on local region specific features and support vector machines. Multimed. Tools Appl. 2017, 76, 78037821.

XVII. [7] M. Suk, B. Prabhakaran, Real-time mobile facial expression recognition systemA case study. In Proceedings of the IEEE Conference on Com- puter Vision and Pattern Recognition Workshops, Columbus, OH, USA, 2427 June 2014; pp. 132137

XVIII. [8] F.D Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, J. Cohn, IntraFace. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia, 48 May 2015; pp. 18.

XIX. [9] M. Szwoch, M. Pieni, Facial emotion recognition using depth data. In Proceedings of the 8th International Conference on Human System Interactions, Warsaw, Poland, 2527 June 2015; pp. 271277.

XX. [10] A.A.S Gunawan, Face expression detection on Kinect using active appearance model and fuzzy logic. Procedia Comput. Sci. 2015, 59, 268274.

XXI. [11] S. Polikovsky, Y. Kameda, Y. Ohta, “Facial micro-expressionsrecognition using a high-speed camera and 3D-gradient descriptor” in Proceedings of the 3rd International Conference on Crime Detection and Prevention, London, UK, 3 December 2009; pp. 16

XXII. [12] P. Shen, S. Wang, Z. Liu, “Facial expression recognition from infrared thermal videos”. Intell. Auton. Syst. 2013, 12, 323333.

XXIII. [13] G. Zhao, X. Huang, M. Taini, S.Z Li, M. Pietikinen, “Facial expression recognition from near-infrared videos”. Image Vis. Comput. 2011, 29, 607619. [CrossRef]

XXIV. [14] W. Wei, Q. Jia, G. Chen, Real-time facial expression recognition for affective computing based on Kinect. In Proceedings of the IEEE 11th Conference on Industrial Electronics and Applications, Hefei, China, 57 June 2016; pp. 161165.

XXV. [15] Paul Viola, Micheal J. Jones, “Robust Real-Time Face Detection” in International Journal of Computer Vision archive Volume 57 Issue 2, May 2004 Pages 137-154.

XXVI. [16] Y. Ito, W. Ohyama, T. Wakabayashi, F. Kimura, “Detection of Eyes by Circular Hough Transform and Histogram of Gradient”, 21st International Conference on Pattern Recognition (ICPR 2012) November 11-15, 2012. Tsukuba, Japan

XXVII. [17] T. Kawaguchi, D. Hidaka, and M. Rizon, “Detection of eyes from human faces by Hough transform and separability filter,” Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), Vancouver, BC, Canada, 2000, pp. 49- 52 vol.1.

XXVIII. [18] Wenshuo Gao, Xiaoguang Zhang, Lei Yang, and Huizhong Liu, “An improved Sobel edge detection,” 2010 3rd International Conference on Computer Science and Information Technology, Chengdu, 2010, pp. 67- 71.

XXIX. [19] Zhang Jin-Yu, Chen Yan and Huang Xian-Xiang, ”Edge detection of images based on improved Sobel operator and genetic algorithms,” 2009 International Conference on Image Analysis and Signal Processing, Taizhou, 2009, pp. 31-35.

XXX. [20] J. Shi, C. Tomasi, ’Good features to track’, IEEE Conference on Computer vision and pattern recognition, June 1994

XXXI. [21] L. Yin, X. Wei, Y. Sun, J. Wang, M.J Rosato, A 3D facial Expression database for facial behavior research. In Proceedings of the International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 1012 April 2006; pp. 211216