International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Performance Improving of ANN with Preprocessing Stage in Human Face Expression Recognition System

Sumalakshmi C. H., P. Vasuki

Abstract: ***In many face recognition systems, the important part is face detection. The task of detecting face is complex due to its variability present across human faces including color, pose, expression, position, and orientation. So, by using various modeling techniques it is convenient to recognize various facial expressions. The system proposed consists of three phases, the facial expression database, pre-processing and classification. To simulate and assess recognition efficiency based on different variables (network composition, learning patterns and pre-processing), we present both the Japanese Female Facial Expression Database (JAFFE) and the Extended Cohn-Kanade Dataset (CK+). Comparative approaches of data preprocessing include face detection, translation, normalization of global contrast and histogram equalization. Significant results were obtained with 85.52 percent accuracy particularly in comparison with some other pre-processing phases and raw data in single pre-processing phases. The result indicates the ANN classifier representation produces a satisfactory result which reaches more accuracy.***

Index Terms: Artificial neural network, facial expression, database, face detection, emotion

I. INTRODUCTION

Emotions are an inherent part of any communication between people. We can be represented in many different forms that can be seen with the naked eye or not. Thus, any previous or subsequent gestures can be identified and remembered with the right tools. Face recognition appears to be a problematic biometric challenge since no methodology can provide a robust solution for all conditions, such as facial expression variance, invariant pose, and occlusion, among other circumstances. In Figure1 there are displayed sample images of actors whose facial expressions correspond to emotional states such as neutral, joy, surprise, anger, sadness, fear, and disgust. The scale of facial expressions is still inadequate to apply the neural network. Preprocessing is a tool that can be used to progress the effectiveness of the FER framework and can be conducted before the usable extraction method in [2].

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
Sumalakshmi C. H.*, Research Scholar, Sethu Institute of Technology, Kariapatti, Tamilnadu, India.
Dr. P. Vasuki, Professor, Department of ECE, Sethu Institute of Technology, Kariapatti, Tamilnadu, India.

Image preprocessing includes different types of processes such as image clarity and scaling, contrast adjustment, and additional enhancement processes [3] to improve the expression frames [4].

Fig. 1. Facial expressions presented to users [1]

Cropping and scaling are carried out on the face image where the nose of the face is used as a midpoint and the other essential facial features are physically included [5]. For the reduction of the face image size, the Bessel down sample has been used, but somehow it preserves features as well as the perceptive nature of the original image [6]. The Gaussian filter is often used to resize input images to give the images sharpness [7].

Normalization is the pre-processing technique that can be meant to reduce light and change dramatically facial images [8] using the median filter and to change the perception of the face. The standardization approach is also used to derive eye positions that make physical differences for the FER program quite stable and allows input images simpler. Localization is a preprocessing technique that utilizes the process of Viola-Jones [9] to identify the facial expressions of the image data.

Many investigations then employ augmentation approaches well into the preprocessing stage, like cropping, scaling, translating or mirroring, to maximize the variation and thus the size of the data Techniques of image preprocessing such as resizing, facet identification, cropping, noise inclusion, and normalization. We assess and evaluate the performance of each preprocessing step to see the variation in levels of accuracy. In this paper, we propose a framework for the identification of face expression dependent through an ANN. Next, we use hair-like characteristics and histogram equalization to identify the nose. Then we create an ANN framework with three layers.

The main contributions and organization of this paper are summarized as follows: In section 2 we describe background details of work regarding facial expressions by the some authors. Section 3 discusses the proposed work. Section 4 deliberates results and discussions. Finally, in section 5, we concluded the paper.

II. BACKGROUND WORKS

The implementation of a scheme capable of recognizing the various human expressions in color images. Viola and Jones explain the strategy for face detection with the AdaBoost Haar classifier in this paper [10].

After the pre-processing process is carried out, it is understood that the system is simple and reliable. The scheme of the researcher will interpret the facial expression according to the threshold value.

The process of this scheme can be modified in real time and explains quickly how the image is captured and the gestures understood. The implementation of an automated system for face recognition using the neural network Facial recognition offers an essential trait for emotion research. Such two neural network models define the automatic system for the analysis of facial expression in this paper, the radial-based feature network (RBFN) and the multi-layer perceptron (MLP). The researcher implemented the point counter detection method; the scheme may remove the characteristics from the face using this approach.

In [11], the authors suggested a target-based approach for the analysis of facial expression using the Neural Network. Target based in sense, in comparison to the motion, focused method, where face-secular knowledge is derived from a series of images, that the recognition of a facial expression can be achieved with the aid of a single image of the face. The pre-processing on images involves the identification and position of the features, extraction of facial characteristics when only unique traits like mouth and eyes are removed from the image being used as feedback for the recognition grid-the most recognizable traits of the human face [12].

III. PROPOSED SYSTEM

Most methods for identification for facial expression usually track the elementary mechanism as presented in Figure 2. With approaches utilizing conventional machine learning techniques, the identification of facial expression has various key steps as a pattern problem: Data collection and processing, extraction selection evaluation and decision-making are included. Various strategies, particularly in the choice of features and classification tools, can be selected from each point. In this paper, we mainly examine the preprocessing unit in specifics to facilitate classification.

A. Data collection and preprocessing:

For facial expression images, a range of considerations such as light variation, head tilt angle variance, off-plan head movement, and partial occlusion, and so on can also be used to improve the process effectiveness and the precise verification of facial expressions. Much pre-processing work must be done a little more before identification. We conduct a two-stage step in our image preprocessing to eliminate disturbance in the original images, namely, face detection and histogram equalization.

1) Face Detection

Face detection is the first step of image pre-processing. In the face detection section, the hair-like feature is used for detection results. Such models are classified into three sections: edge dimensional, and middle. Depending on this, Viola and Jones's models are used where the design is made up of the black and white portion. Upon positioning it on a certain part of the image, it can be obtained by removing all pixels inside the cover of the white rectangle and that within the cover of the black rectangle. The purpose of these black and white rectangles is, thus, to measure facial features to get dispersed face details and eventually differentiate between the non-face and face parts.

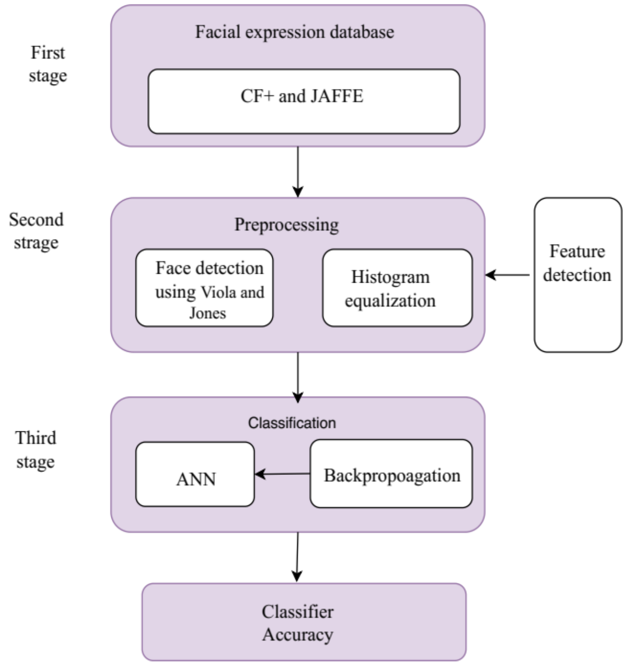

Fig. 2. System Flow Diagram

2) Improved Histogram equalization

Once the quite face section of the image is obtained, certain troubling problems must also be regarded. Despite the different lighting conditions of images, the parts of the human face will also show different luminosity, which will eventually intervene greatly with identification performance. Therefore, before identification, we agree to execute histogram equalization (HE).

Remember r to mention the continuous image intensity values, that requires vales in [0, L-1] range, with r=0 signifying black and L-1 designating white. For r to fulfill these requirements, the transforming functionality is specified by:

$$s = T(r) \qquad 0 \le r \le 1 \tag{1}$$

For every pixel value of r corresponding generates intensity level of s in the input image, so that it considers function $T(r)$ that satisfies the following restrictions:

$T(r)$ is single-valued and monotonically growing in the interval, $0 \le r \le 1$ and

$$0 \le T(r) \le 1 \;\; for\; 0 \le r \le 1$$

For a given image all the gray levels are treated to be random variables lies in the range of [0, L-1]. The probability density function can be put in the mathematical form as:

$$p_S(s) = (L-1)\, p_r(r)\left|\frac{dr}{ds}\right| \tag{2}$$

A transformation function of particular importance in image processing has the form

$$s = T(r) = (L-1)\int_{0}^{r} p_r(w)\,dw \quad (3)$$

To get the discrete values in this images, we use summations and probabilities so that the gray level occurrence in the image can be given as:

$$P_r(r_k) = \frac{n_k}{n} \qquad k = 0,1,2,\ldots,L-1 \quad (4)$$

where, n denotes the count of pixels, n_k denotes the corresponding gray levels r_k , and L denotes the possible count for gray intensity levels. Therefore eq. (3) can modified as

$$S_k = T(r_k) = (L-1)\sum_{j=(r_j)0}^{k} p_r \quad (5)$$

$$S_k = (L-1)\sum_{j=0}^{k}\frac{n_j}{n} \quad (6)$$

Then, each pixel is mapped S_k with both the level in the input image to a subsequent pixel with that of the level r_k in the output image using Eq. (6) to generate an output image. The transformation (mapping) given in Eq. (6) is called histogram equalization or histogram linearization.

B. Feature detection:

Local Binary Pattern (LBP) is indeed a texture descriptor that can be used for the extraction of features. These are generated generally with the binary code and can be done by using a threshold between the center pixel and its position pixels. Here feature detection function takes an image as a parameter and returns the LBP histogram for the image from three regions of the face the eye, nose, and mouth. First, it detects the face from the image using the viola Jones function then it finds the LBP histogram for each of the three regions of the detected face using the LBP function and returns the final concatenated histogram of the three regions.

C. Classification:

In this method, the classifier Neural Network accepts grayscale images as input and provides an expression as output. This will be achieved by preprocessing the input image first, training or testing of the image on the neural network which then classifies the image in one of six groups. There are many approaches to a neural network. The most popular one is the backpropagation. For this work, we applied this concept to use it for a face identification application program.

Backpropagation is a common method for training a feed-forward neural network, using mean squared error and gradient descent. If *W* is the set of all network weights, it's easy to find the network gradient. This allows us to update the current *W* by a small step to form the new weight by subtracting the initial weight from the gradient network. One of the more popular activation functions for backpropagation networks is the sigmoid, a real function S_c (0, 1) defined by the expression

$$Sc(x) = \frac{1}{1+e^{-cx}} \quad (7)$$

The weight for the input is used as a multi-layer neural network, and this study utilizes the error back propagation method to know the weight of the neural network. The neural network architecture developed in this study is shown in Figure 3.

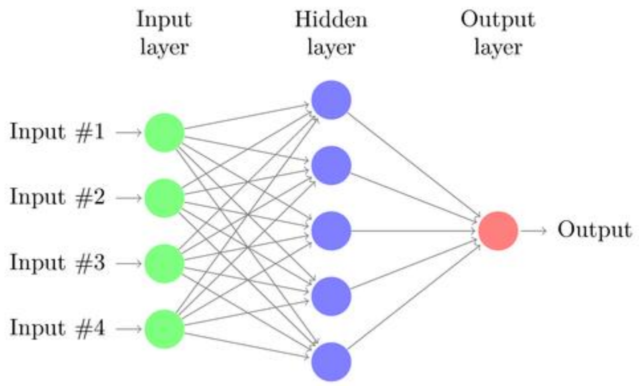

Fig.3. Structure of the Face Recognition neural network

Repeated learning by form (8-9) for the preceding error back propagation technique can provide weight learning between both the input layer as well as the first hidden layer.

$$W^{k+1} = W^{k} - n\left(\frac{\eth E^{k}}{\eth Wij}\right) \quad (8)$$

$$\frac{\eth E}{\eth Wij} = \frac{\eth E}{\eth hi}\frac{\eth hi}{\eth wi,j}, \frac{\eth E}{\eth hi} = \frac{\eth E}{\eth si}\frac{\eth si}{\eth hi,j} \quad (9)$$

$$E = \frac{1}{2}\sum_{i=1}^{r}(t1-y1)^{2}, si = \frac{1}{1+e^{-\lambda hi}} \quad (10)$$

$$hi = \sum_{i=1}^{q} wilXi + Wiq + 1 \;; y1 = \sum_{i=1}^{r} Vil\,Sl + Vir + 1 \quad (11)$$

k is the number of repetition, *i, j* are the number of the input layer and hidden layer node, respectively, η is the training parameter, *s* is the activation function, *h* is the intermediate sum, *y* is the reference value, *t* is the target value, *m* is the input node number, *r* is the output node number and *V* is the weight of learning between hidden layer and input layer. Formula (12) calls for the weight learning between the hidden layer and the output layer.

$$V^{k+1} = V^{k} - n\frac{\eth E^{k}}{\eth vil}, \frac{\eth E}{\eth vil} = \frac{\eth E}{\eth yi}\frac{\eth yi}{\eth Vi,j} \quad (12)$$

The training is done using two learning weight formulas (8), when the optimal value is achieved by continuous learning using the formula (12), and used as a classifier.

IV. RESULTS AND DISCUSSION

For the simulation, we use two traditional facial expression repositories that are mostly recognized by academics. JAFFE includes 213 images of 10 Japanese women, whereas CK+ includes images of all race groups and a total of 328 images. JAFFE software is used in most studies. JAFFE includes ten expressions of the Japanese girl, with seven facial expressions and 213 images. Each image in the JAFFE database has a resolution of 256 to 256 pixels. Figure 4 shows a few of the sample images of the JAFFE files.

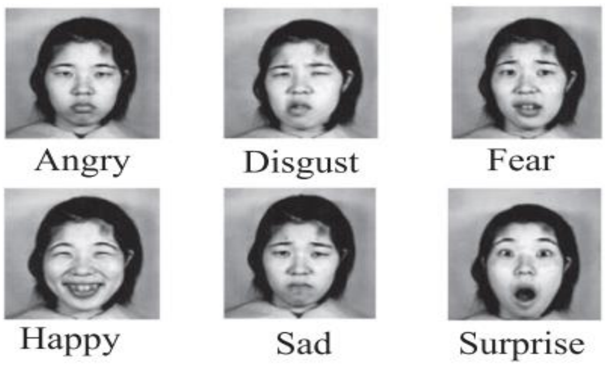

Fig.4. Sample images from JAFFE database

The CK database also includes seven expressions including 132 natural and smile subjects. It includes 486 image samples with a resolution of 640 to 490 pixels of gray images. A few of the CK database sample images are shown in the Fig. 5.

Fig. 5. Sample images from CK database.

After reviewing and outlining the associated works, we propose different rational dimensions for assessing different approaches of recognition of facial expression. The above structure consists of 512 input layer neurons, i.e. one-dimensional vector of 32 per 32 binary images. The output layer has three neurons, one for each type of emotion i.e. amazed, normal and happy, the network was trained by various combinations of hidden layers and neurons per layer. For this methodology network, a hidden layer with 80 neurons was selected on the basis of the least error Parameters such as momentum and learning rate are often selected by hit and check process as shown in Table 1. This work is used to train a back-propagation method and a sigmoid function to activate the network.

Table 1. ANN parameters for evaluation

Layer	No of Neurons
Input layer	512
Hidden layer 1	16
Output layer	6
Bounds: Learning rate 0.1 Momentum 0.3	

Accuracy: It is the process of showing the performance levels of the specific classifier functionality. It is clear that the more accurate system provides better facial expression detection from the dataset.

Table 2 shows the 3 parameters of databases when applied to ANN classifier and its values.

Table 2. Database parameters for evaluation

Database	MAE	Training error	Accuracy
JAFFE	0.0389	0.088	57.89%
CK+	0.0369	0.049	85.25%

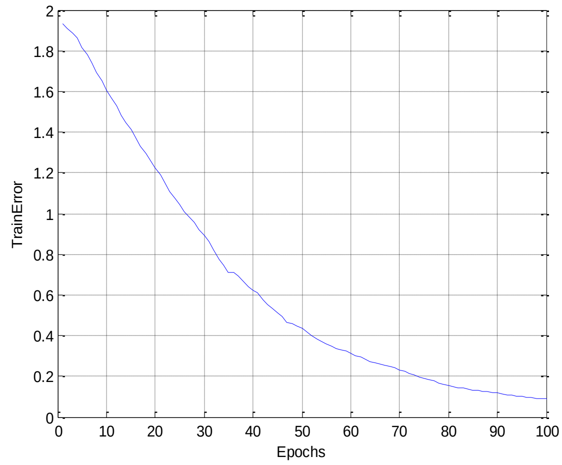

Fig. 6. Performance of the classifier accuracy for 100 epochs of JAFFE database

Fig. 6 where the x-axis shows the number of epochs and the y-axis shows the error of training obtained under ANN. This paper analyzes the training error of ANN in the JAFFE database to calculate the mean error rate. As seen by Fig.6, the training error at 50 epochs is 0.4 in situations where the most computational complexity occurs in the JAFEE database.

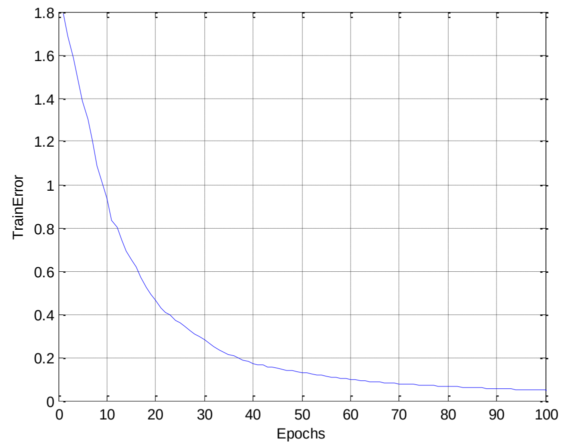

Fig. 7. Performance of the classifier accuracy for 100 epochs of CK database

Fig. 6 where the x-axis shows the number of epochs and the y-axis shows the error of training obtained under ANN. This paper analyzes the training error of ANN in the JAFFE database to calculate the mean error rate. As seen by Fig.6, the training error at 50 epochs is 0.18 in situations where the less computational complexity occurs in the CK database.

Fig. 8. Training error of the ANN classifier for 2 databases

Fig. 8 shows the training error of two widely used datasets JAFEE and CK+ when applied to the ANN classifier. To detect the emotion on facial expressions, mostly CK+ database is having less training error of 0.049 as related to JAFFE database of 0.089.

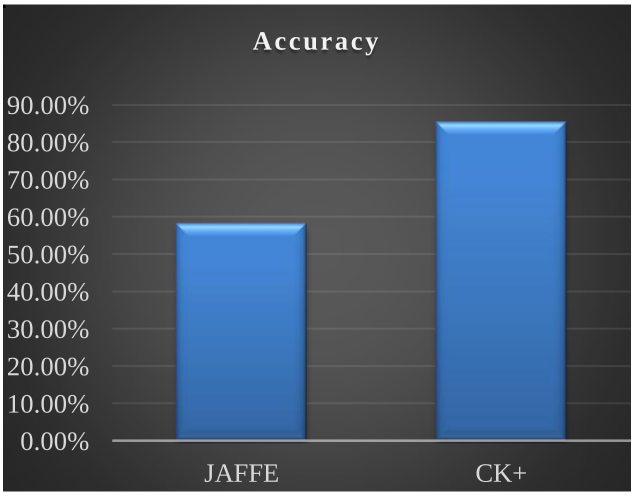

Fig. 9. Accuracy of the ANN classifier for 2 databases

Fig. 9 shows the accuracy of two widely used datasets JAFEE and CK+ when applied to the ANN classifier. To detect the emotion on facial expressions, mostly CK+ database is having more accuracy of 85.25% as related to JAFFE database of 58%.

V. CONCLUSION

In this paper, we address the influence of using ANN classifier on well-known facial expression databases JAFFE and CK+ to detect the emotion on facial expressions. The proposed work mainly concentrated on the preprocessing stage where it helps increase the accuracy of classification with the help of Viola and Jones's models and improved histogram equalization technique. For evaluating the performance of the proposed ANN classifier on 2 databases JAFFE and CK+ in terms of training error and accuracy are computed. Also, it is clear the highest recognition accuracy of 85.25% in the case of the ANN classifier and it identifies the numerous expressions.

REFERENCES

1. Przybyło J., Automatyczne rozpoznawanie elementów mimiki twarzy w obrazie i analiza ich przydatności do sterowania, rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków, 2008.
2. Poursaberi, A., Noubari, H.A., Gavrilova, M., Yanushkevich, S.N., 2012. Gauss – Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. EURASIP J. Image Video Process. 1–13.
3. Bashyal, S., Venayagamoorthy, G.K.V., 2008. Recognition of facial expressions using Gabor wavelets and learning vector quantization. Eng. Appl. Artif. Intell. 21, 1056–1064.
4. Taylor, P., Siddiqi, M.H., Ali, R., Sattar, A., Khan, A.M., Siddiqi, M.H., Ali, R., Sattar, A., Khan, A.M., Lee, S., 2014. Depth camera-based facial expression recognition system using multilayer scheme. IETE Tech. Rev. 31, 277–286.
5. Zhang, L., Member, S., Tjondronegoro, D., 2011. Facial expression recognition using facial movement features. IEEE Trans. Affect. Comput. 2, 219–229.
6. Owusu, E., Zhan, Y., Mao, Q.R., 2014. A neural-Ada boost based facial expression recognition system. Expert Syst. Appl. 41, 3383–3390.
7. Biswas, S., 2015. An Efficient Expression Recognition Method using Contourlet Transform. Int. Conf. Percept. Mach. Intell. pp. 167–174.
8. Ji, Y., Idrissi, K., 2012. Automatic facial expression recognition based on spatiotemporal descriptors. Pattern Recognit. Lett. 33, 1373–1380.
9. Zhang, L., Tjondronegoro, D., Chandran, V., 2014. Random Gabor based templates for facial expression recognition in images with facial occlusion. Neuro computing 145, 451–464.
10. Happy, S.L., Member, S., Routray, A., 2015. Automatic facial expression recognition using features of salient facial patches. IEEE Trans. Affect. Comput. 6, 1–12.
11. Sreevatsa N. A. N, Sathish Kumar. K. G, Rakeshsharma. S and Mohd. Mansoor Roomi. "Emotion Recognition from Facial Expressions: A Target Oriented Approach using Neural Network". ICVGIP, Proceedings of the Fourth Indian Conference on Computer Vision, Graphics & Image Processing, (2004).
12. I.Michael Revina, W.R. Sam Emmanuel, "A Survey on Human Face Expression Recognition Techniques", Journal of King Saud University – Computer and Information Sciences, September 2018.

AUTHORS PROFILE

Sumalakshmi C. H., is currently pursuing PhD from Anna University in the research area of Digital Image Processing. She obtained her Bachelor of Engineering Degree in Computer Science and Engineering under Anna Univercity and her M.Tech in Computer Science and Engineering under Calicut University. She has 8 years and 4 months of teaching experience.

Dr. P. Vasuki, is currently working as Professor, Department of Electronics and Communication Engineering, Sethu Institute of Technology, Kariapatti .She obtained her Bachelor of Engineering Degree in Electronics and Communication Engineering and her M.E Degree in Computer Science and Engineering from Anna University. She obatained her Ph.D. in Digital Image Processing. Her Recognizitions are, Registered Research Guide of Anna University, one scholar completed Ph.D. under her guidance and currently she is guiding 7 Research Scholars in the field of Digital Image processing.