Feed Forward Neural Network - Facial Expression Recognition Using 2D Image Texture

Wisam A. Qader[1] & Saman Mirza Abdullah[2] & Musa M. Ameen[3] & Muhammed S. Anwar[3]

[1]Information Technology Department, Tishk International University, Erbil, Iraq
[2]Software Engineering Department, Koya University, Erbil, Iraq
[3]Computer Engineering Department, Tishk International University, Erbil, Iraq
Correspondence: Wisam A. Qader, Tishk International University, Erbil, Iraq.
Email: wisam.abdulaziz@tiu.edu.iq

Doi: 10.23918/eajse.v8i1p216

Abstract: Facial Expression Recognition (FER) is a very active field of study in a wide range of fields such as computer vision, human emotional analyses، pattern recognition and AI. FER has received extensive awareness because it can be employed in human computer interaction (HCI), human emotional analyses, interactive video, image indexing and retrieval. Human facial expression Recognition is one of the most powerful and difficult responsibilities of social communication. Face expressions are, in general terms, natural and direct methods of communicating emotions and intentions for human beings. GWT is applied as a preprocess stage. For the classification of face expressions, this study employs the well-known Feed Forward Propagating Algorithm to create and train a neural network.

Keywords: Feature Extraction, FER (Face Expression Recognition), Classification, GWT

1. Introduction

Emotion is a state of mind that includes ideas, psychological changes, and expressions that favorably impact intellectual processes such as decision-making, awareness and empathy. Emotions are expressed by humans today in interactions and emotions, portrayed mostly by facial expressions in voice, hand, body gestures (Mehrabian et al., 2017). Facial expressions are a type of nonverbal communication that plays an important role in interpersonal relationships and interactions.

Facial expressions are dynamic components that convey the speaker's attitude, sentiments, intentions, and so on. The major emotional source is the face. In the human facial analysis, facial characteristics play an important role, and features are classed as permanent or transitory. The lines of the facials, brown wrinkles and deep furrows are sample of transitory features. Examples of permanent characteristics include the eyes, lips, brows and cheek (Ekman et al., 1971).

The combination of a facial, eye, lip and nose with various features can represent numerous emotional expressions. Gladness, for example, is a wider separation between left and right corners of the lips and the eyes tend to relax in happiness. Surprise, is usually defined by the broad open mouth, which implies a smaller difference between the upper and lower lips between the left and right corners of the mouth, and eyes tend to be wide open and therefore larger to surprise people (Pantic et al., 2000).

Received: April 5, 2022
Accepted: May 26, 2022
Qader, W.A., Abdullah, S.M., Ameen, M.M., & Anwar, M.S. (2022). Feed Forward Neural Network - Facial Expression Recognition Using 2D Image Texture. *Eurasian Journal of Science and Engineering, 8*(1), 216-224.

Anger is shown by the wide-open eyes, the eye corner height increases, and the nose is elevated and the mouth is pushed, whereas the wrinkle nose is shown as disgust and the upper lip raises for a disgust person. Fear is indicated by the opening of the eyes and the increased breadth of the eye-corners. The eyelids are closed, and the corners of the lip are pushed down to a sorrowful individual. Neutral expression finally does not imply any eye, mouth or nose alterations.

For the search and retrieval process, the design characteristic of an object is highly powerful since its shape is closely related with the functioning and identification of an object. This feature helps differentiate shape from other characteristics that remove visual features, including color or texture. Texture in the surface pattern of materials such as wood, grain, sand, dirt and fabric is seen. The term texture represents the repeating of fundamental texture pieces known as textures including several pixels that can be placed periodically, or randomly.

In numerous mediums such as single-face images, multiple-faceted photographs, films etc., facial expression may be recognized. The abstraction of semanticized information, which is processed by the human brain; a face photo with specific geometrical and colored features remains a common object to separate the target which is susceptible to the processing of emotion ("face"). Face identification is one of the most investigated computer vision problems (Yuille et al., 1992). We are nevertheless interested in face-lifting aspects of the face as psychologists have proven the general facial manifestations of the eyes, eyebrows, lips and nose (Hjelmas et al., 2001). The characterization of facial components rather than the complete face thereby avoids substantial processing times.

Facial detection components were investigated intensively; from Yuille et al. (Shin et al., 2014) up to current work as Jongju Shin et al., several methods have been developed (Viola et al., 2004). In this study, we used the well-known and effective hair assessment approach established by viola and Jones (Freund et al., 1997). The AdaBoost algorithm (Saaidia et al., 2007) is used to concentrate just on the primary significant characteristics, just like characteristics by compiling an integral picture, then the vast number of features obtained. The final step is to increase the speed of the election process using a framework of sophisticated classificators that focuses attention on promising image regions. This study aims to construct and train neural network models to classify the facial expressions using a well-known technique known as Gradient Descent Feed Forward Propagation Algorithm.

The difference in form or position of the facial features between the picture and its equivalent image under normal expression is defined by the expression of a face image. As a result, most face recognition systems are based on a series of images or recorded movies comprising facial photographs with various emotions as well as shots with ordinary expressions. The FACE coding system (Friesen et al., 1978) is the most often used method for measuring facial movement.

This research article is organized into five primary sections, the first of which is an introduction to FER. Section II throws Facial expression recognition system. Section III discusses Gabor Wavelet Transform. Section IV discusses the implementation results. Section V contains the conclusion.

2. Facial Expression Recognition System

The facial expression system is divided into two stages: training and testing. During the training phase, known input face data is digitally processed using preprocessing tools and processes. The preprocessed image is further processed to create feature vectors. Feature vectors of known class are utilized to train the network in the supervised neural network. Trained vectors are used for picture classification with unknown expressions. The second part involves testing any unknown expression in a confined setting.

In the test phase, unknown facial patterns are preprocessed, and the face function is extracted from the cropped visual image and then classified into the six fundamental emotional conditions of the individual as seen in Fig. 1.

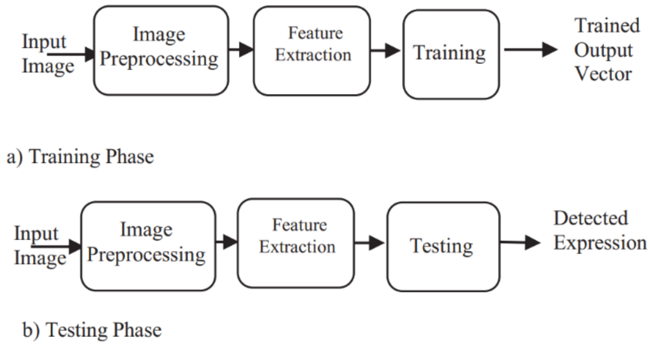

Figure 1: Facial Expression system block diagram.

To minimize noise and lighting fluctuations, the system requires pre-processing of the incoming pictures. The pre-processing module transforms the colored picture into the grayscale and then resizes it for quick processing. The main phases of facial expression detection and classification are:

3. Feature extraction

The FER system's next phase is the feature extraction procedure. The elimination of features identifies and displays good aspects for future image processing. Computer vision extraction is an essential phase in the processing of images which allows the detection of the change from graphical to implicit data. This data representation can then be utilized as a classification input. Feature extraction is divided into five categories; geometric feature method edge-focused, local- and global feature method and the patch-based approach.

The functionality based on texture-based methodology is extracted in the following descriptions. A description of function extraction textures and information on magnitude and phase is provided in the Gabor filter. The Gabor filter magnitude restricts information on the arrangement of the face images. The phase feature includes the full description of the magnitude features (Bashyal et al., 2008).

4. Classification

The FER system's last stage is classification, in which the classifier categorizes expressions including smiling, sad, surprise, rage, fear, disgust, and neutral.

For classification, the Multilayer Feed Forward Neural Network (MFFNN) classifier has three layers: input, hidden, and output layers, as well as a back-propagation method. The weights are initialized and the activation units are estimated during the training stage. The Bayesian neural network classifier is a classification method that contains three layers: input, hidden, and output. For improved accuracy, the traditional back propagation technique is combined with a Bayesian classifier.

Convolutional Neural Network (CNN) is made up of two layers: convolutional layer and subsampling layer, with two-dimensional pictures as input. The feature maps are created in the convolutional layer by intricately combining convolution kernels with two-dimensional pictures, whilst pooling and redeployment are accomplished in the subsampling layer. The CNN also includes two critical perceptions: likely shared weight and sparse connection. The CNN classifier is utilized as many classifiers in FER for distinct facial areas. If CNN is framed for the full-face picture, then frame the CNN first for the mouth area, then for the eye area, and so on for each subsequent location CNNs are framed.

5. Gabor Wavelet Transform

Gabor functions were developed by Dennis Gabor as a technique for recognizing signals in noise. Gabor (Gabor et al., 1946) shown that there is a "quantum principle" for information; for no signal to conquer less than a certain minimal region, the conjoint time-frequency domain for 1D signals must be quantized. A directional microscope is used to assess Gabor decomposition's orientation and scaling sensitivity. Curves generate a low-level feature map of picture intensity because they incorporate certain low-level prominent properties in an image.

A wavelet filter from Gabor is a Gaussian kernel, tempered by a sinusoidal plane wave, as in (1).

$$\psi_g(\mathrm{x},\mathrm{y}) = \frac{f2}{\eta\gamma\pi}\exp(\beta^2 y'^2 - a^2 x'^2)\exp(2\pi\, j\, f\, x')$$

$$x' = x\cos\theta + y\sin\theta, \qquad [1]$$

$$y' = y\cos\theta - x\sin\theta,$$

where f defines central frequency of the sinusoidal plane wave, the anticlockwise rotation of the Gaussian defined by θ and the envelope wave defined by α, which is the sharpness of the Gaussian along the major axis parallel to the wave and the sharpness of the Gaussian minor axis perpendicular to the wave defined by β. To keep the ratio between frequency and sharpness constant $\gamma = f/\alpha$ and $\eta = f/\beta$ is defined (Eleyan et al., 2009). (2) defines the 2D Gabor wavelet which has Fourier transform.

Where f defines the main frequency of the sinusoidal wave plane, the Gaussian rotation, defined in the antilock direction by the axis of α, and the wave envelope defined by α, that is the Gaussian sharpness, along the main axis, parallel with the wave, and the Gaussian minor axis, perpendicular to the β wave. The ratio of frequency to sharpness is described as $\gamma = f/\alpha$ and $\eta = f/\beta$ (Eleyan et al., 2009). (2) defines the Fourier-transforming 2D Gabor wavelet.

$$\psi_g\,(u,v) = \exp\left(-\pi^2\left(\frac{(u'-f)^2}{\alpha^2} + \frac{v'^2}{\beta^2}\right)\right)$$

$$u' = u\,cos\theta + v\,sin\theta, \qquad [2]$$

$$v' = v\cos\theta - u\sin\theta.$$

Since GWT is mainly developed for vision applications and systems, one of the most significant applications it can be used for is face recognition. GWT usage in vision area and applications was first utilized by Daugman in the 1980s. Lately, a face recognition system based on Gabor wavelet was developed by B. S. Manjunath (Chellappa et al., 1992).

In expressing both the spatial relationship and the structure of the spatial frequency, Gabor wavelet transformation can be employed. A single input image with Gabor filters and eight orientations captures the whole frequency spectrum and the reaction is always complicated. Fig. 2 illustrates the Gabor filter's amplitude and phase responses on a single input image.

(a)

(b)

(c)

Figure 2: (a) The original image, (b) the magnitude and (c) the single-scale, eight-orientation Gabor kernels phase

6. Experimental Results

We utilized our "Facial Expression" dataset FER-2013 (Goodfellow et al., 2013) consisting of 32 x 32 gray-scale pixels of the face of 6 category facial expressions: disgusted, angry, fear, sadness, happy, and neutrality as appear in Fig. 3 shows. FER-2013 is a collection of face expression recognition algorithms. There are 1,5724 training images, 3,589 test images, all of which are 1,8507. The faces were automatically recorded from the web such that each face is centred around the same region inside each image. The training data may also include noise on a label that does not show the correct facial expression on the labels of several faces.

Figure 3: Examples of six facial expressions from database

The algorithm's effectiveness is then tested using the various pictures for the training and testing for the four block sizes. The findings for the picture database are compared with the accuracy. The neural network type utilized in the test to effectively classify features after applying GWT. The technique presented to recognize facial emotions is to assess the impact of the partition of image blocks on the classification of the neural network.

In order to find six fundamental emotions, Happiness, Disgust, Happiness, Fear, Angry and Neutral is the suggested Facial Expression Recognition Algorithm. There were several sets of pictures used for training and testing. The program iteratively collects the pictures from the base and calculates the feature vector for the specified block size using the GWT technique. In accordance to the training picture set, the neural network classifies the image into one of the emotions.

The proposed method first applies preprocessing on the images using GWT then the result of the images is sent NN to differentiate the expression based on the result of the network. The network is trained against the train dataset of the FER2013 database.

The tests are conducted using MATLAB 2018b software. The results are satisfactory in terms of recognizing the category of images based on expression of the face. For various expressive epochs, Fig. 4 demonstrates the rate of accuracy of the approach presented. The total of 20 epochs tested for expression recognition.

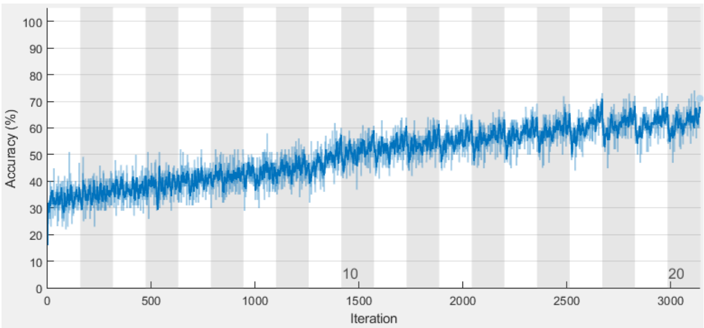

Figure 4: The rate of accuracy based on iteration for 20 epochs

The loss rate of the proposed technique is shown in Fig. 5. The nearest result will be shown as an output to the user as shown in Fig. 5.

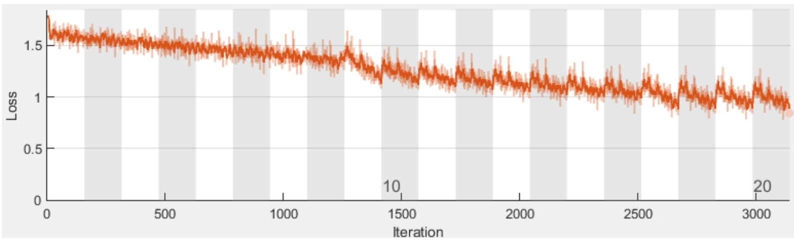

Figure 5: The rate of accuracy based on iteration for 20 epochs

Fig. 6 shows the chart of the accuracy rate per epochs. As it can be analyzed that the rate is increasing proportionally with increasing the number of epochs and iterations on dataset.

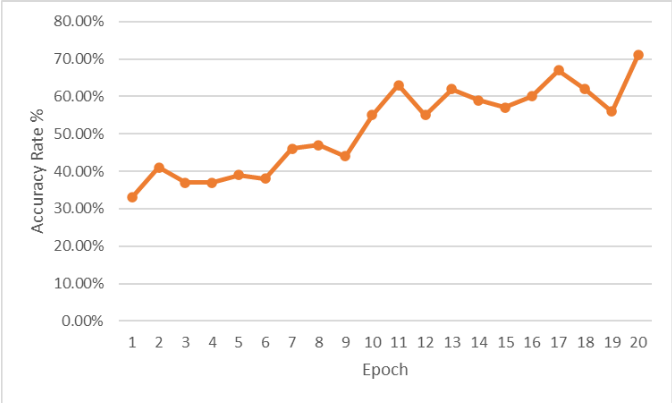

Figure 6: The rate of accuracy based on iteration for 20 epochs

Table I. shows test results of first 10 epochs of the face expression test technique. The results of the proposed technique on epoch 20 is like that: Mini-batch Accuracy is 71.00% and Mini-batch Loss is 0.8442.

The system is tested against different parameters and configuration of the network system and results were analyzed and compared with different systems. The result may get better in terms of performance by using different preprocessed techniques and algorithms which are currently available.

Table 1: Sample output of test results for 10 epochs

Epoch	Iteration	Time Elapsed	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:02	16.00%	1.7918	0.01
1	50	00:00:13	34.00%	1.5748	0.01
1	100	00:00:27	34.00%	1.5629	0.01
1	150	00:00:40	33.00%	1.5826	0.01
2	200	00:00:56	36.00%	1.5355	0.01
2	250	00:01:10	38.00%	1.5695	0.01
2	300	00:01:23	41.00%	1.53	0.01
3	350	00:01:36	40.00%	1.4793	0.01
3	400	00:01:47	36.00%	1.5471	0.01
3	450	00:01:59	37.00%	1.4858	0.01
4	500	00:02:10	47.00%	1.3863	0.01
4	550	00:02:22	39.00%	1.5572	0.01

4	600	00:02:35	37.00%	1.4665	0.01
5	650	00:02:46	38.00%	1.4753	0.01
5	700	00:02:58	31.00%	1.495	0.01
5	750	00:03:08	39.00%	1.5018	0.01
6	800	00:03:21	43.00%	1.3813	0.01
6	850	00:03:34	43.00%	1.4078	0.01
6	900	00:03:46	38.00%	1.4832	0.01
7	950	00:03:58	47.00%	1.4238	0.01
7	1000	00:04:11	33.00%	1.4375	0.01
7	1050	00:04:25	46.00%	1.3886	0.01
8	1100	00:04:36	47.00%	1.3728	0.01
8	1150	00:04:49	47.00%	1.3237	0.01
8	1200	00:05:01	52.00%	1.192	0.01
8	1250	00:05:12	47.00%	1.3622	0.01
9	1300	00:05:25	51.00%	1.2801	0.001
9	1350	00:05:37	52.00%	1.2304	0.001
9	1400	00:05:49	44.00%	1.284	0.001
10	1450	00:06:03	47.00%	1.3917	0.001
10	1500	00:06:15	48.00%	1.258	0.001
10	1550	00:06:27	55.00%	1.1864	0.001

The results are calculated for all images in the test set. In the suggested approach to trains and test sets, the recognition rates are somewhat higher than the one achieved without the deployment of the preprocess technique.

7. Conclusion

One of the most common neural networks is the ML Neural Network based on back propagation algorithm. The proper number of neurons hided before training is one of the problems of using the MLP network. The constructive and pruning algorithms are two techniques proposed in the literature to tackle this challenge. Recent publications address significant future improvements such as FER for side view faces, which combine subjective information from facial sub-regions and different factors to describe the posture of the face in real-time applications. In real time, FER are used for monitoring the safety of drivers, medical imaging, the interface of robots, forensics and detecting deception. A facial expression recognition system was examined using GWT. The comparison and analysis of preprocessing pictures using GWT to further increase recognition efficiency is a primary emphasis of our study. It is shown that the system performs better with the GWT than with simply the ML neural network. The additional scope might encompass the utilization of the many other classification methods for diverse datasets.

References

Bashyal, S., & Venayagamoorthy, G. K. (2008). Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence*, 21, 1056-1064.

Chellappa, R., Manjunath, B. S., & Malsburg, C. V. D. (1992). A feature based approach to face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 373-378.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124.

Eleyan, A., Özkaramanli, H., & Demirel, H. (2009). Complex wavelet transform-based face recognition. *EURASIP Journal on Advances in Signal Processing*, 2008, 1-13.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119-139.

Friesen, E., & Ekman, P. (1978). Facial action coding system: a technique for the measurement of facial movement. *Palo Alto*, 3, 5.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-part III: Radio and Communication Engineering*, 93, 429-441.

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. *International Conference on Neural Information Processing*, Springer, Berlin, Heidelberg, 117-124.

Hjelmås, E., & Low, B. K. (2001). Face detection: A survey. *Computer vision and image understanding*.

Mehrabian, A. (2017). Communication without words. *In Communication theory*. Routledge, 193-200.

Pantic, M., & Rothkrantz, L. J. (2000). Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 18, 881-905.

Saaidia, M., Chaari, A., Lelandais, S., Vigneron, V., & Bedda, M. (2007). Face localization by neural networks trained with Zernike moments and Eigenfaces feature vectors. A comparison. *IEEE Conference on Advanced Video and Signal Based Surveillance,* 377-382.

Shin, J., & Kim, D. (2014). Hybrid approach for facial feature detection and tracking under occlusion. *IEEE Signal Processing Letters*, 21, 1486-1490.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57, 137-154.

Yuille, A. L., Hallinan, P. W., & Cohen, D. S. (1992). Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8, 99-111.