Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters

Milad Mohammad Taghi Zadeh
Department of Electrical Engineering
Khatam University
Tehran, Iran
m.mohammadtaghizadeh@khatam.ac.ir

Maryam Imani
Department of Electrical Engineering
Tarbiat Modarres University
Tehran, Iran
maryam.imani@modares.ac.ir

Babak Majidi
Department of Computer Engineering
Khatam University
Tehran, Iran
b.majidi@khatam.ac.ir

***Abstract* - The feelings contain in human face have a great influence on decisions and arguments about various issues. In psychological theory, emotional states of a person can be classified into six main categories: surprise, fear, disgust, anger, happiness and grief. Automatic extraction of these emotions from images of human faces can help in human computer interaction as well as many other applications. Machine learning algorithms and especially deep neural network can learn complex features and classify complex patterns. In this paper, a deep learning based framework for human emotion detection is presented. The proposed framework uses the Gabor filters for feature extraction and then the deep convolutional neural network. The experimental results show that the proposed features increases the speed and accuracy of training the neural network.**

Keywords—Facial emotion recognition, Gabor filter, Convolution neural network

I. INTRODUCTION

Emotions have an important role in our everyday lives, and directly affects decisions, reasoning, attention, prosperity and quality of life of human beings. Establishing communication between people is through emotions and facial expressions. Nowadays, with the influence of computers on human lives and the mechanization of the lives of individuals, the establishment of human and computer interaction (HCI) has played a crucial and very important role [1]. There is now a strong interest in improving the interaction between humans and computers. Many people believe in this theory and there is a positive and useful emotional response to establishing a good and useful cognitive link between computers with users. This interaction between the computer and the human beings can be created through speech [2]. Psychological theory states that human emotions can be classified into six different forms: surprise, fear, hatred, anger, happiness and sadness. By making changes in the facial muscles, human can represent this group of emotions.

In this paper, a deep learning based framework for human emotion detection is presented. The proposed framework uses the Gabor filters for feature extraction and then the deep convolutional neural network. The experimental results show that the proposed features increases the speed and accuracy of training the neural network. The rest of this paper is organized as follows. After describing the related literature in Section II, the proposed fast facial emotion recognition framework is detailed in Section III. The experimental design and the simulation scenarios are discussed in Section IV. Finally Section V concludes the paper.

II. RELATED WORKS

Kwolek et al. [3] performed facial recognition using the Gabor filter to extract features. It was shown in this paper that using the Gabor filter improves the accuracy of convolutional neural network from 79% to 87.5%. Wu et al. [4] suggested that Gabor motion energy filters (GME) could be used to detect facial expressions. In this method, GME filter was compared with Gabor energy (GE) filter and the results showed that the proposed method was 7% better. Li et al. [5] presented a deep fusion convolutional neural network (DF-CNN) method using 2D+3D images. In this method, three-dimensional scan images of the faces are used. These images make up a total of 32 dimensions. This paper explains that prediction is done using two methods. 1- Classification by using the SVM from the 32-dimensional features. 2- The normal prediction of the Softmax function using the six-state probability vector. Experimental results shown DF-CNN achieved good results. Li et al. [6] argued that the existing face databases are not reliable for the real world applications. This article presents a new database called RAF-DB that contains 30,000 facial expressions from thousands of people. In the gathering of this database, it has been seen that real faces often have mixed feelings, in other words, a mixture of several feelings. By reviewing the RAF-DB database, it's the first natural database that is much more diverse than the seven most commonly used datasets. As a result, a DLP-CNN suggested a way of expressing a difference between feelings. These experiments were performed based on 7 basic conditions and 11 combinations. By examining this method on the SFEW and CK + databases, it was shown that the proposed DLP-CNN method is the best method for detecting the state of the face emotion. Tzirakis et al. [7] are suggested that both voice processing and image processing be used to detect human situations. In this method, the sound is separately conveyed to a neural network. In another direction, the image also provides a 50-layer convolutional neural network. This method was performed on the RECOLA database and it is argued that the results of the experiments show better performance than other methods. Lee et al. [8] suggested that the eye can be used to extract emotions. In this method, which was performed using STFT and CNN, the STFT extracts the necessary features in two

cases of eye size and motion and serves as an input to the CNN network with two layers. This review shows that the CNN network is also effective in eye-based diagnosis, in addition to being effective in recognizing emotions in different ways. Pons et al. [9] proposed to train a CNN network for the final face emotion classification, which can be used to evaluate the combination of CNN networks. In this method, 72 CNN networks were trained with four layers and with different parameters, and 64 CNN networks were trained with VGG-16. In the final analysis, the output of different networks was classified using the proposed classification and the result was presented. The results of this study show that this method can increase accuracy by 5% in different classification methods. Tang et al. [10] presented three different methods called DGFN, DFSN and DFSN-I. DGFN is actually based on the critical points of psychology and physiology rules and uses the traditional machine learning network. DFSN is designed on the basis of CNN. In the end, DFSN-I combines both DGFN and DFSN, which has both advantages for better performance. Experimental results show that the performance in the CK + database and Oulu CASIA improved.

III. METHODOLOGY

A. Gabor filter

Gabor filters are generally used in texture analysis, edge detection, feature extraction. In (1), the Gabor filter is described. When a Gabor filter is applied to an image, it gives the highest response at edges and at points where texture changes. The following images show a test image and its transformation after the filter is applied.

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma)=\exp(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2})\exp(i(2\pi\frac{x'}{\lambda}+\psi)) \quad (1)$$

in which the real part is:

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma)=\exp(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2})\cos(2\pi\frac{x'}{\lambda}+\psi) \quad (2)$$

and the imaginary part is:

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma)=\exp(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2})\sin(2\pi\frac{x'}{\lambda}+\psi) \quad (3)$$

where:

$$x'=x\cos\theta+y\sin\theta \quad (4)$$

and

$$y'=-x\sin\theta+y\cos\theta \quad (5)$$

The proposed filter is shown in the Figure 1. As shown in Figure 1, the original image of the first Gabor filter is displayed and then we pass the filtered image again from the second Gabor filter. In Figure 2, four samples of the original images are shown.

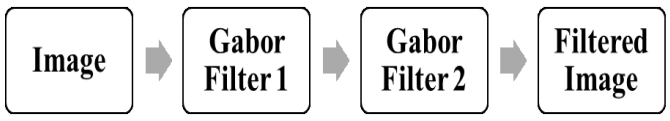

Figure 1 – The proposed filter

Figure 2-Orginal Images

In Figure 3, four sample images are shown after applying the Gabor filter on Orginal images with the following parameters:

$(x,y)=(18,18),\ \sigma=1.5,\ \theta=\pi/4,\ \lambda=5,\ \gamma=1.5$
$\psi=0$

Figure 3-Sample image after applying first Gabor Filter

In Figure 4, four sample images are shown after applying the Gabor filter on first gabor filtered images with the following parameters:

$(x,y)=(18,18),\ \sigma=1.5,\ \theta=3\pi/4,\ \lambda=5,\ \gamma=1.5$
$\psi=0$

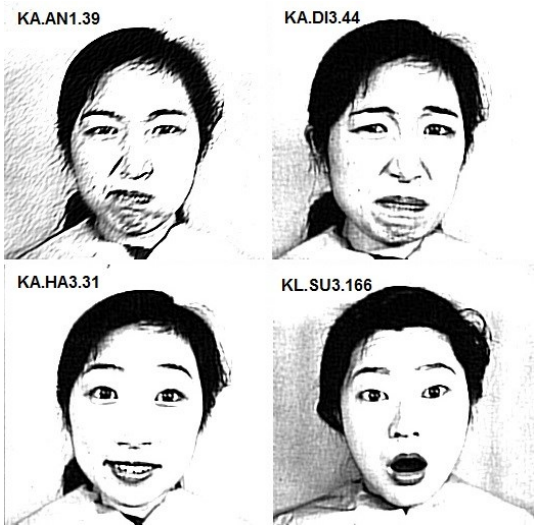

Figure 4 –Sample image after applying second Gabor Filter

B. CNN Architecture

The structure of the convolutional neural network is shown in the Figure 4 . As shown in the table 1, first, the image is received and after passing through different layers and the learning process returns a vector with seven modes as output. In fact, these seven modes are: Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise.

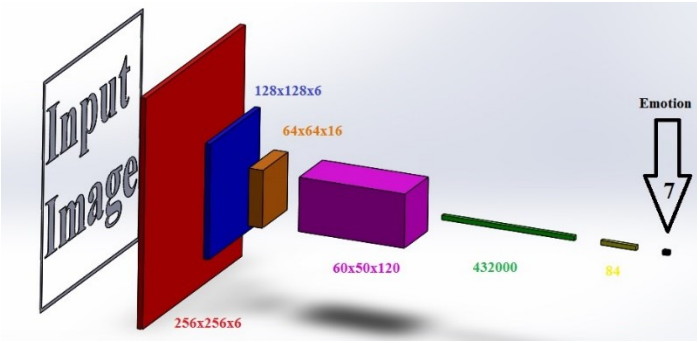

Figure 5 – The CNN architecture

As shown in Table 1, in the first stage, the deep neural network applies a convolution of a 6 × 6 filter on the image. In the next step, using MaxPooling, the dimensions are reduced to 128 × 128 × 6. Next, another convolutional network will be applied to the 16 × 16 filter size, and in the next step, using the MaxPooling function, we will reduce the size to 64 × 64 × 16. The next convolution is applied to the data with a 120 × 120 filter size. In the next step, using the Flatten function, all data is converted to a vector of the size 432000. Then, the vector is converted to a vector of length 84, and at the end it is reduced to seven, which is the 7 categories of emotional states.

Table 1- Proposed CNN architecture method details

Layer type	Details	Output Shape
Conv	Conv (6x6)	256, 256, 6
Activation	Relu	256, 256, 6
MaxPooling	Pool size (2,2)	128, 128, 6
Conv	Conv (16x16)	128, 128, 16
Activation	Relu	128, 128, 16
MaxPooling	Pool size (2,2)	64, 64, 16
Conv	Conv (120x120)	60, 60, 120
Activation	Relu	60, 60, 120
Dropout	----------	60, 60, 120
Flatten	Flatten to a vector	432000
Dense	Input →84	84
Activation	Relu	84
Dropout	----------	84
Dense	Input → Classese Num =7	7
activation	softmax	7

Rectified Linear Unit (Relu) function :

$$f(x) = x^{+} = \max(0, x) \tag{6}$$

Softmax function:

$$\sum_{i} X_i = 1 \qquad x_i = [0,1] \tag{7}$$

C. Proposed facial emotion recognition algoritm:

The proposed method is to first apply a Gabor filter to the images and then convey the output results as inputs to the neural network. The output of the Gabor filter is given to the convolutional neural network.

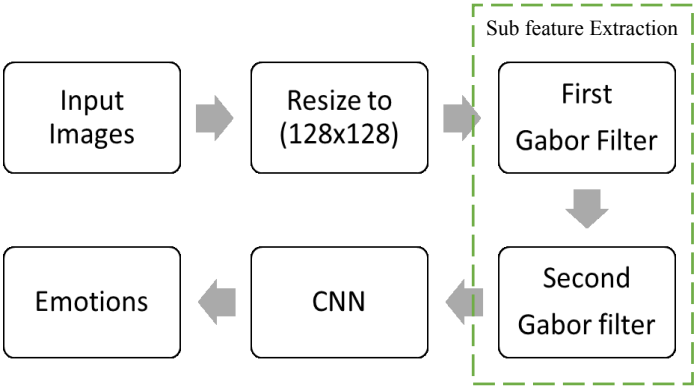

Figure 6 - The proposed method

IV. EXPERIMRNTAL RESULTS

For the experimental results the JAFFE database [11] is used. The database contains 213 Japanese female model images that include seven emotional states of the face, in which six modes, natural states of the face and normal face. Table 2 shows

the specification of the simulation hardware. By comparing Figure 8 and Figure 9 and Table 3, the results show that after 10 epochs, the proposed system reaches 58% accuracy, but non-Gabor filter systems reach 46% accuracy. After 15 epochs, the proposed system reaches 76% accuracy but non-Gabor filter systems reach 59% accuracy. after 25 epochs, the proposed system reach 84% accuracy but non-Gabor filter systems reach 77% accuracy. At the end, the proposed system reach 91% accuracy and non-Gabor filter systems reach 82% accuracy.

Table 2 – The system specifications

Model	Dell XPS L502
Processor	Intel Core i7 CPU – 2.00GHz
RAM	8 GB
System type	64-bit Operating System

Figure 7 – sample of correct emotional recognition

Table 3 – Comparison of proposed method and simple CNN

Epoch	CNN Methode Accuracy	2Gabor + CNN Methode
1	0.1492	0.1713
10	0.4696	0.5856
15	0.5996	0.7624
20	0.6519	0.8453
25	0.7790	0.8453
30	0.8232	0.9116

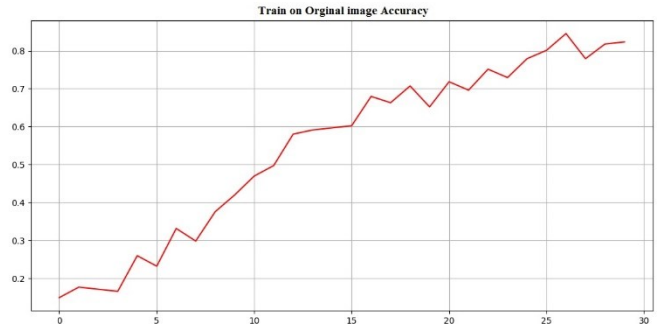

Figure 8 - Accuracy of simple CNN method

Figure 9 – Accuracy of 2Gabor filter+CNN method

V. CONCLUSION

After the Gabor filter applied, the system learning became faster and the accuracy has improved. As seen in Figures 7 and 8. The learning speed of the convolutional neural network has increased profoundly. This is because the Gabor filter actually extracts the image subfeature and gives the neural network. By doing this, the convolutional neural network receives a number of subfeature and takes one step further in extracting the emotions from the faces.

REFERENCES

[1] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Transactions on Affective Computing,* vol. 6, no. 1, pp. 1-12, 2015.

[2] A. Nicolai and A. Choi, "Facial Emotion Recognition Using Fuzzy Systems," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 2216-2221.

[3] B. Kwolek, "Face Detection Using Convolutional Neural Networks and Gabor Filters," in *Artificial Neural Networks: Biological Inspirations – ICANN 2005*, Berlin, Heidelberg, 2005, pp. 551-556: Springer Berlin Heidelberg.

[4] T. Wu, M. S. Bartlett, and J. R. Movellan, "Facial expression recognition using Gabor motion energy filters," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 42-47.

[5] H. Li, J. Sun, Z. Xu, and L. Chen, "Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network," *IEEE Transactions on Multimedia,* vol. 19, no. 12, pp. 2816-2831, 2017.

[6] S. Li, W. Deng, and J. Du, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2584-2593.

[7] P. Tzirakis, G. Trigeorgis, M. Nicolaou, B. Schuller, and S. Zafeiriou, *End-to-End Multimodal Emotion Recognition Using Deep Neural Networks*. 2017.

[8] H. Lee and S. Lee, "Arousal-valence recognition using CNN with STFT feature-combined image," *Electronics Letters,* vol. 54, no. 3, pp. 134-136, 2018.

[9] G. Pons and D. Masip, "Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis," *IEEE Transactions on Affective Computing,* vol. 9, no. 3, pp. 343-350, 2018.

[10] Y. Tang, X. M. Zhang, and H. Wang, "Geometric-Convolutional Feature Fusion Based on Learning Propagation for Facial Expression Recognition," *IEEE Access,* vol. 6, pp. 42532-42540, 2018.

[11] B. Majidi and A. Bab-Hadiashar, "Real time aerial natural image interpretation for autonomous ranger drone navigation," in *Digital Image Computing: Techniques and Applications, 2005. DICTA'05. Proceedings 2005*, 2005, pp. 65-65: IEEE.