

Edge aware directional coding (EADC) for face expression recognition

P. Surekha *

P. Vidya Sagar †

Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation

Vaddeswaram 522302

Andhra Pradesh

India

G. Ramesh §

Department of Computer Science and Engineering

Gokaraju Rangaraju Institute of Engineering and Technology

Hyderabad 500090

Telangana

India

Abstract

Face Expression Recognition (FER) has immense popularity in research community due to its widespread applicability in different applications. However, the currently available expression descriptors are suffering from several problems like noises, distortions and weak edges. To sort out these issues, this paper proposes an Edge Aware Directional Coding (EADC) that efficiently encodes the prominent expression related features. Initially, EADC finds out the gradients based on Kirsch Compass Mask in eight directions. Then EADC encode each pixel in facial image based on the Gradient Support of its neighbor pixels. Directional Edge Weight (DEW) is measured as support based on the gradient magnitude and gradient direction of edges passing through the neighbor pixels. EADC encode all the texture primitives such as corners and edges such that the recognition system can recognize expressions accurately independent of subjects. For experimental validation, JAFFE and CK+ datasets are used and the performance is measured through recognition accuracy. The obtained results show that the EADC outperforms all the earlier state-of-the-art methods.

Subject Classification: Primary 68, Secondary 97P80.

Keywords: Face expression recognition, Gradients, Directional edge weight, Textures, Directional mask, Accuracy.

* E-mail: prekha.572@gmail.com

† E-mail: pvsagar20@gmail.com

§ E-mail: ramesh680@gmail.com (Corresponding Author)

1. Introduction

From the past few years, the significance of practical requirement of the Human Face Expression Recognition (FER) has been witnessed due to its widespread applicability in different applications including robot control, surveillance, affective computing, driver fatigue surveillance, medical treatments, human machine interaction and many more [1]. Most of these applications need an accurate and robust FER system that provides an adaptive interface for different kinds of users. An FER system can provide important clues regarding the intention of person and the mind state. A robot integrated with the FER technology can perform home services like elderly take care and talking to children [2]. Further, An Advanced Driver Assistance System (ADAS) with the integrated FER technology can determine the fatigue levels of driver. These systems can help the driver from escaping the accidents by warning them when the fatigue levels exceed particular threshold [3]. Next, the industries related to production can use the FER technology to determine the consumer product's worth before sending them to launch in the market [4]. This technology can also help for the company recruiters to determine the hidden emotions of candidates during interviews [5]. The FE technology in hospitals can help in the remote surveillance of patient's health conditions [6]. Further, Ashwin and Guddeti [7] demonstrated the effectiveness of FER technology in the classrooms and online teaching platforms to improve the teaching quality. Hence, from the past few decades, the research has been grown tremendously in the FER to make it suitable for several real world applications.

In general, the FER system composed of three stages; they are face detection, feature extraction and classification. Among these three stages, the feature extraction is more important task which extracts the features of emotion from facial images. Based on the previous studies, the feature extraction methods are categorized into two categories namely geometric features and appearance features [8]. The former features [9], [10] are calculated based on the face geometry to represent the emotions from facial image. But they need an accurate identification of facial components which is tough when the appearance of face changes with time [11]. Even though the detection is done accurately, the emotion description from only such components may lose the huge information belongs to emotion. Thus, the geometry-based features experienced less recognition performance at different environments. Next, the features based on appearance represent the expression through the appearance of face. They

are again classified as global [12] and local features [13] However, a fixed or predefined threshold may not give prominent results under the images with varying contrast and noise levels. Broadly, the drawbacks of the conventional methods are summarized as follows;

1. Less discrimination between flat, noisy and local distorted pixels.
2. More complexity due to 8-bit code representation.

To sort out these problems, this paper proposes a new expression descriptor called as Edge Aware Directional Coding (EADC) based on the local shape structure of facial muscles. For this purpose, EADC finds out the edge response in eight direction through Kirsch Compass Kernel. From the obtained responses, the gradient magnitudes and directions are measured for each pixel. Next, EADC encodes each target pixel based on the support provided by its neighbor pixels. For the assessment of support, EADC introduced a new weight called as Directional Edge Weight (DEW) and it is measured based on the closeness of gradient direction of corresponding pixel with the mean orientation of the local region. The direction of two neighbor pixels those have maximum closeness are encoded as a required code of target pixel.

Remaining paper is stipulated as follows; Section II demonstrates the particulars of literature survey. Section III demonstrates the particulars of developed EADC mechanism for FER. Details of experimental validations are explored in section IV and concluding remarks are explored in section V.

2. Literature Survey

Recently, Cao, Ton-hat, and Choi [14] initially divided the facial image into non-overlapping blocks of equal size. Then each block is subjected to the texture encoding through LBP. After encoding, they employed Support Vector Machine for Emotion classification. For robustness validation, they have simulated their method on three types of images such as large, medium and small. Next, Kauser and Sharma [15] primarily fetch the key features of face such as mouth, nose and eyes. Then they calculated LBP for each element followed by a unique coding template. For classification purpose, they employed "Multi-Layer Perceptron Neural Network (MLP-NN)".

Unlike the LBP which considers the center pixel's intensity as threshold, Goyani and Patel [16] considered the mean pixel intensity of a

local region and encoded the target pixel through a new method called as “Local Mean Binary Pattern (LMBP)”. After the representation of each pixel with LMBP, they employed “Histogram Normalized Absolute Difference (HNAD)” for the classification of expressions. Further, they also employed two more distance based classifiers such as Chi-square and L2-norm to compare with HNAD. However, the major drawback with LBP and LMBP is huge information loss and they can’t provide sufficient discrimination between flat and noise pixels. Moreover, they don’t have a specific method to discriminate edges and their direction of movement.

Similar to the LDN, Lakshmi, and Mohanaiah [17] employed Gabor and Gaussian filters over the facial image to find out the edge directions. They extracted the Gaussian attributes for the noise sensitivity removal and Gabor attributes for view variance removal. They named their method as “Gradient Based Compact Binary Coding (GCBC)”. A very similar method is proposed by Sammaiah and Rao [18] named as “Edge Adaptive Local Direction Binary Pattern (EALDBP)” that encodes the facial image with respect to the maximum top positive and maximum top negative responses. For the determination of edge responses, they applied Gaussian and Robinson Compass masks. These two methods have commonly three drawbacks, they are (1) lengthy 8-bit code, and (2) Prone to noise effect in the local region and (3) no discrimination between flat and noisy pixels.

Liu, Feng, and Wang [19] proposed a deep model which is a three stage FER system. At first phase, a pose guided facial alignment is done to reduce the difference in intra classes.

Compared to the handcrafted features Deep learning features are easy to extract but they have limited scope in describing the expression under different situations like occlusions, noises, distortions etc. They don’t employ specific mechanism to discriminate the flat pixels from noisy pixels. Moreover, compared to the handcrafted features, the deep learning features introduce additional complexity.

3. Methodology

The block schematic of the proposed method is shown in Figure 1, which generates the 7-bit code for each pixel of facial image which represents the local shape. In early research, the authors declare that the facial shape has considerable significance in representing the expression and they are explored by the determination of edges [13]. But the earlier methods cannot guarantee the stability of encoder at several edges. Similar codes may get generated at sometimes even for different edges. Further,

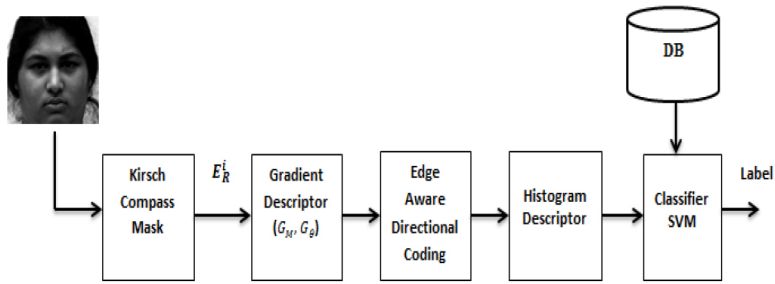


Figure 1

Overall block schematic proposed method

they were not guaranteed about the stability in the case of edges with distortions. More over the earlier methods encode the noise and flat pictures with same codes at sometimes. To sort out these problems, the proposed method measures the consistency of the pixel towards the edge based on the support of its neighbor pixels. The proposed method calculates the support through Directional Edge Weight (DEW) which is computed with the help of direction of edges at neighbor pixels. Towards the calculation of DEW, initially we compute the gradients of each pixel through Kirsch Compass Mask (KCM) compass mask. KCM finds out the gradients in 8 directions by convolving each pixel with eight different kernels. The gradient is calculated as the maximum value of its edge responses of 8 kernels. Then these responses are subjected to Encoding through the DEW. The details of proposed method are explored here.

4. Experimental Analysis

Here in the following section, we explain the performance of developed FER mechanism by validating it over several standard benchmark and publicly available databases. For the experimental validation, we used MATLAB tool. Here in this section, initially, we first present the details of the databases which are used, followed by details of the performance metrics. Finally, we provide a comparative analysis based on these metrics.

A. Databases

We used totally two standard and publicly available and benchmark facial expression databases, they are “Japanese Female Face Expression (JAFPE)” [20] and “Cohn-Kanade (CK+)” [21]. The details of these two

databases, i.e., number of samples for each expression, resolution, number of expression types etc. are discussed here.

JAFPE is a most popular facial expression database and most of the researchers used it to validate their methods. JAFPE is acquired with the help of ten Japanese models and the total number of images present here are 213. All the images are of .TIFF format and they all have uniform resolution as 256×256 . In every expression image, the acquirement takes care about the landmarks at which the major expressions incur. For this purpose, all the models tied their hair back. All models are asked to keep a frontal view such that the full expression is acquired. Some sample images are shown in Figure 2. The emotions present in this dataset are 7, they are namely Neutral, Disgust, Fear, Sad, Happy, Angry and Surprise.

Next, CK+ dataset is one more popular and facial expression dataset which consists of totally 486 sequences. All these images under this dataset are acquired with the help of 90 subjects. It composed of both posed and non-posed expressions and the resolution of each image is observed as 490×640 .

The entire images are in .PNG format and some facial images are shown in Figure 3. Here, every sequence has a neutral expression at starting and it rises to the corresponding expression after some time. The emotions present in this dataset are 8, they are namely Neutral, Contempt, Disgust, Fear, Sad, Happy, Angry and Surprise.

B. Results

Under the JAFPE simulation, we conduct two types of experimental investigations namely; seven-class and six-class expression. Under the 1st simulation case, the extra label Neutral is added and in the second case, the neutral case is removed. Since the JAFPE dataset have limited number of facial expression images, from those images, set up the images both for training and testing. Almost all images are employed for testing and the peak expression image is used for training. For the both 7-class and 6-class simulation, the images employed for testing are Neutral – 31, Surprise – 30, Sadness – 31, Happy – 31, Fear – 32, Angry – 30, and Disgust – 29. Neutral expression is excluded in the 6-class expression simulation while keeping the remaining expression same. The diagonal elements show the correctly recognized samples (True Positives (TPs)) and the end value in each row show the total input samples. Excluding the highlighted value in each row, the sum of remaining values results in False Negatives (FNs). Similarly, after the exclusion of highlighted value in each column, the sum of remaining values results in False Positives.

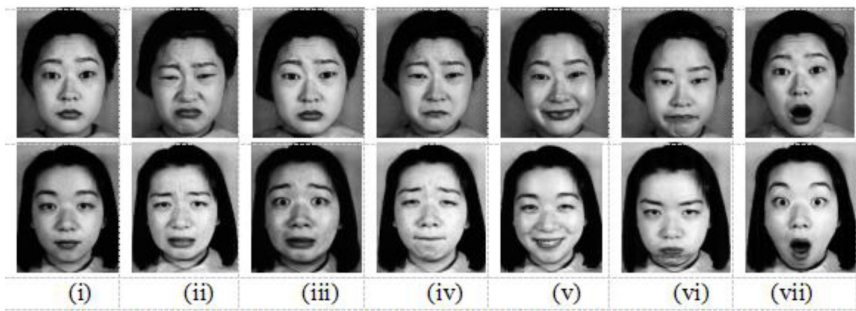


Figure 2

Samples from JAFFE dataset (i) Neutral (ii) Disgust (iii) Fear (iv) Sad (v) Happy (vi) Angry and (vii) Surprise



Figure 3

Samples from JAFFE dataset (i) Neutral (ii) Disgust (iii) Fear (iv) Sad (v) Happy (vi) Angry and (vii) Surprise

Based on the obtained TPs, FPs, FNs, and TNs, the recognition accuracy is calculated as $Accuracy = (TP + TN) / (TP + TN + FP + FN)$. The accuracy observed for both datasets at two simulations studies is shown in the Figure.4. For CK+ dataset, the average accuracy of proposed method for 6-expression and 7-expression s observed as 92.18% and 95.65% respectively. Next, for JAFFE dataset, the average accuracy of proposed method for 6-expression and 7-expression s observed as 92.25% and 89.67% respectively. Since the JAFFE dataset composed of similar face attributes between Japanese models, the system can't discriminate them effectively. Hence the recognition accuracy of JAFFE is less compared to the accuracy of CK+.

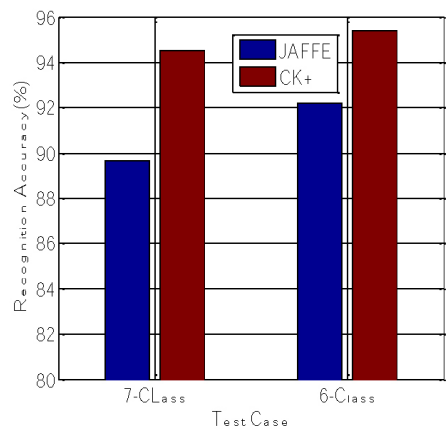


Figure 4
Recognition accuracy for different simulations

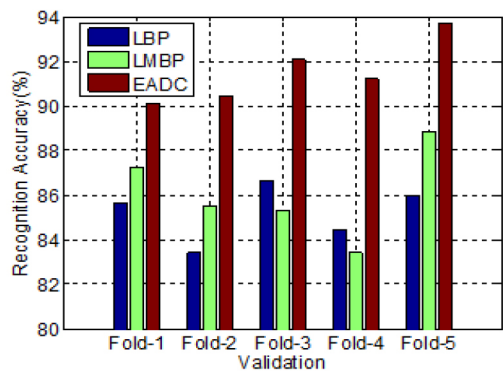


Figure 5
Recognition accuracy at five-fold cross validation on CK+ dataset

Figure 5 explores the recognition accuracy particulars under different cross validations. Towards this simulation, the subjects are changed at every validation; means the subjects used for training and testing are different. Moreover, the combination of subjects employed in one validation is not entertained in the other validation. In this way, we conduct a fivefold cross validation and the recognition is measured for LBP, LMBP and the proposed EADC. From the results, we can see that the improvement in recognition accuracy from LBP to EADC is approximately 6% while from LMBP to EADC, it is observed as 4.5%. Moreover, the average accuracy of EADC is noticed as 91.6% while it is 85.8% and 85%

for LMBP and LBP methods respectively. As the LBP and LMBP considers only the pixel intensities but not the edges for texture encoding, they had shown limited accuracy.

5. Conclusion

In this article, we deal with the FER and developed a new face expression representator termed as “EADC”. Mainly, EADC is proposed to encode each pixel of facial image with a seven-bit code based on the edge support of its neighbor pixels. For the support calculation, EADC computes the Directional Edge Weight based on the passage of edge through neighbor pixel. Moreover, the EADC encode all types of texture primitives in a discriminative way such that the system can recognize the expressions more accurately irrespective of subject. For classification, SVM with three different kernels is employed. We validate the EADC on different expression datasets with different cross validations and it had shown an outstanding performance in recognizing different kind of expressions.

References

- [1] P. Weinberg and W. Gottwald, “Impulsive consumer buying as a result of emotions,” *Journal of Business Research*, vol. 10, no. 1, pp. 43–57 (1982).
- [2] T.-H.-S. Li, P.-H. Kuo, T.-N. Tsai, and P. C. Luan, “CNN and LSTM based facial expression analysis model for a humanoid robot,” *IEEE Access*, vol. 7, pp. 93998–94011 (2019).
- [3] M. R. Jeong and B. C. Ko, “Driver’s facial expression recognition in real-time for safe driving,” *Sensors*, vol. 18, no. 12, p. 4270, Dec. (2018).
- [4] M. S. Bouzakraoui, A. Sadiq, and N. Enneya, “A customer emotion recognition through facial expression using POEM descriptor and SVM classifier,” in *Proc. 2nd Int. Conf. Big Data, Cloud Appl*, pp. 1–6., Mar. (2017).
- [5] Y. S. Su, H. Y. Suen, and K. E. Hung, “Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews,” *J. Real-Time Image Process.*, vol. 19, pp. 1011–1021, Jan. (2021).

- [6] M. Alhussein, "Automatic facial emotion recognition using Weber local descriptor for e-Healthcare system," *Cluster Comput.*, vol. 19, no. 1, pp. 99–108, Mar. (2016).
- [7] T. Ashwin and R. M. R. Guddeti, "Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks," *Educ. Inf. Technol.*, vol. 25, pp. 1–29, Mar. (2019).
- [8] J. Cai, Z. Meng, A. S. Khan, Z. Li, J. O'Reilly, and Y. Tong, "Island loss for learning discriminative features in facial expression recognition," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Xi'an, China, pp. 302–309, May (2018).
- [9] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, Jan. (2007).
- [10] N. Bourbakis, A. Esposito, and D. Kavraki, "Extracting and associating meta-features for understanding people's emotional behavior: Face and speech," *Cogn. Comput.*, vol. 3, no. 3, pp. 436–448 (2011).
- [11] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, Jun. (2009).
- [12] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86 (1991).
- [13] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI J.*, vol. 32, no. 5, pp. 784–794, Oct. (2010).
- [14] N. T. Cao, A. H. Ton-hat, and H. I. Choi, "Facial expression recognition based on local binary pattern features and support vector machine," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 28, no. 6, 1456012 (2014).
- [15] N. Kauser and J. Sharma, "Facial expression recognition using LBP template of facial parts and multilayer neural network," in *Proc. Int. Conf. I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Palladam, India, pp. 445–449 (2017).
- [16] M. M. Goyani and N. Patel, "Recognition of facial expressions using local mean binary pattern," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 16, no. 1, pp. 54–67 (2017).

- [17] A. Vijaya Lakshmi and P. Mohanaiah, "Gradient-based compact binary coding for facial expression recognition," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 6, pp. 377–390 (2021).
- [18] S. Sammaiah and K. V. Rao, "A salient binary coding scheme for face and features of pixel and geometry," *IEEE Access*, vol. 9, pp. 18876–18889 (2021).
- [19] J. Liu, Y. Feng, and H. Wang, "Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning," *IEEE Access*, vol. 9, pp. 69267–69277 (2021).
- [20] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, pp. 200–205, Apr. (1998).
- [21] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, San.

Received October, 2024