Facial Expression Recognition using Efficient LBP and CNN

Sonali Sawardekar[1], Prof. Sowmiya Raksha Naik[2]

[1]M.Tech Student, Dept. of Computer Engineering, VJTI College, Maharashtra, India
[2]Assistant Professor, Dept. of C.E.& I.T, VJTI College, Maharashtra, India

---***---

Abstract – *Automated facial expression recognition is challenge in computer vision domain. Many techniques have been applied to gain accurate and efficient results in identifying face expressions. For monitoring security, treating patients in medical field, Human-machine interaction, marketing research and E-learning are some of the application of facial expression recognition. Feature extraction is the first step in facial expression recognition, followed by classifier to classify input face expressions. Local Binary Pattern is a texture description method that describes the local texture feature of an image in a gray-scale range. Convolutional Neural Networks is one of the most representative network structures in deep learning technology, and it has achieved great success in the field of image processing and recognition. In this paper, facial expression recognition using Efficient Local Binary Pattern (LBP) images and convolutional neural network (CNN) for classification is presented. The proposed algorithm is tested using Cohn-Kanade dataset which results 90% accuracy.*

Key Words: **Facial Expression Recognition, deep learning, Convolutional Neural Network, Local Binary Pattern, Feature Map**

1.INTRODUCTION

Much of research interest have been attracted by Artificial intelligence because of it's recognition ability in human emotion. Human vision is easily replicated by computers because of the evolution of Artificial intelligence. Computer learns human vision and performs necessary action to get accurate output. computer plays key role in human computer interaction by recognizing facial expressions. psychological characteristics such as heartbeat and blood Pressure, speech, hand gestures, body movements, Facial expressions identify emotions of person. facial expressions are more effective among all this characteristics. Mehrabian[1] research indicates that facial expressions convey 55 of message in face to face communication. A facial expression is Change in position of muscles beneath the facial skin. These movements on face indicates the emotional state of an individual.

Facial expressions refers to very powerful nonverbal communication that we use to communicate. anger, disgust, fear, happiness, sadness, and surprise are six basic facial expressions identified by Ekman et al. These six expressions are broadly categorized into positive and negative emotions. Surprise and Happy emotions are included under positive emotions and Fear, Sad, and Angry, Disgust expressions are included under negative emotions. By observing Facial features and facial muscle movements , one can identify whether individual is in pain or frustrated or happy and so on. Many applications nowadays, such as Medicine, E-Learning, Marketing research, data-driven animation, interactive games, entertainment make use of Automated facial expression recognition system.

CNN which is a deep learning method is successful for outstanding classification in the area of image classification and recognition. Convolutional Neural Network (ConvNet) has ability of automatic feature extraction and translation invariance which makes it feasible neural network for image classification. Each Layer in CNN extract unique feature from the given input image which makes it more powerful neural network. The objective of the convolutional neural network is to transform a set of inputs into accurate and meaningful outputs. The Local Binary Pattern Convolutional Neural Network is employed in the research to achieve maximum efficiency. Cohn_Kanade Facial Expression Database (CKFED)[2] is standard database used as training and testing data for expression recognition.

2.RELATED WORK

2.1 Status of Face Recognition Research

Face recognition is carried out using one of the three methods - Geometric based method , appearance based method and neural network based methods. face geometry was the first traditional way for face recognition. In Geometrical feature based approaches, face is represented by set of facial landmark points. Angle, Distances between those facial points determine location and shape of facial components. Feature vectors that represents face is given to the classifier to classify input face. The Main difficulty in this is to find out facial point detectors which can locate landmark points on face. In appearance based methods , features are extracted from the pixel intensity values in facial image. Intensity of light emitted from image determines pixel intensity. Local Binary Pattern (LBP) , Local Gabor Binary Patterns (LGBP), principal component analysis (PCA) including Eigenface, linear discriminate analysis (LDA), multi-orientation multi-resolution Gabor wavelets are some of the appearance based methods which does not include an information of facial points.

Features extracted from above methods are given to neural network which then identifies input image. Classifiers such as Support Vector Machines, (SVMs),Artificial Neural Networks (ANNs), Hidden Markov Models (HMMs), K-Nearest Neighbors (KNNs) outputs input image into one of the expression based on the descriptive features obtained from geometric and appearance based methods. However recently the methods as called as deep learning have appeared as a promising one to perform facial expression recognition. Deep learning methods extract both low-level and high level features without a method. Deep learning methods has shown great results in signal processing and image processing.

Yavuz Kahraman et al.[3] implemented geometry-based features for face expression recognition. This Technique searched for 153 possible distances among 18 critical candidate points/landmarks. correlation-based feature subset selection (CFS) method was applied to select 16 of these distances having significant contribution to accuracy. This CFS+ANN method has 91.2% correct classification rate.

Zhou Ji-liu et al.[4] used automatic fiducial point location algorithm locating 58 fiducial points and calculated the Euclidean distances between the center of gravity coordinate and the fiducial points coordinates of the face. person's neural expression and the other seven basic expressions are used to extract geometric deformation difference features. This feature vector acts as input to multiclass SVM classifier which classifies data input seven basic expressions. . Archana Shirsat et al.[5] proposed FER using Efficient Local Binary Pattern (LBP) for feature extraction and artificial neural network (ANN) for classification. Local Binary pattern is illumination variant and detects the features using very simple calculation. Extracted features from LBP were given to ANN for classification. This algorithm improves recognition rate for 64x64 window size.

Reza Azmi[6] used three feature extraction methods, Gabor filters and the local binary pattern operator (LBP) and local Gabor binary pattern (LGBP). The K-NN classifier with sum of absolute differences vector distance measure is used as classifier. LGBP because of its high robustness and effectiveness shows high accuracy under a variety of occlusion conditions on FER. Banu, Simona et al.[7] designed a model for detection of face, eyes and mouth which uses Haar functions and applied Bezier curves to extract the distances between facial parts. Two layered feed-forward neural network is used along with Kmeans algorithm for Pre-classification resulting in accuracy of 85%. JunWang et al.[8] proposed facial expression recognition method using Hidden Markov Model. The relative displacement of the feature points between the current frame and the neutral frame are extracted as the facial features. Classification entropy threshold and model parameters are found out using iterative algorithm during training process. During testing, an image sequence is assigned an expression category when the entropy of the expression likelihood obtained from early HMMs is below the threshold by gradually increasing sequence length. The overall recognition rates achieve about 82% using about 45% sequence length on CK+ database, and about 52% with about 23% sequence length on MMI database.

Jun Wang et al.[9] proposed a deep convolutional neural network (CNN) for facial expression recognition system which is used for deeper feature representation of facial expression to achieve automatic recognition. The proposed system results 76.7442% and 80.303% accuracy in the JAFFE and CK+, respectively.

2.2 The main work of this paper

The main work of this paper is to study extraction of facial expression based on efficient Local Binary Pattern (LBP) and recognition of facial expression based on deep neural network i.e convolution neural networks (CNN).

In this paper, we provide LBP feature map as the input of CNN to improve the understanding and learning of CNN which will provide guidance for the selection of CNN learning data.

2.2.1 LBP feature Map

LBP (Local Binary Pattern) describes local texture features of images. Rotation invariance and gray invariance is the main advantage of Local binary pattern. [10]. Local Binary pattern is a simple tool for the detection of the features and is robust to the illumination variations in an image. Because of its simplicity and robustness, LBP is widely used method for the feature extraction in many of the object recognition methods as well as the facial expression detection. Local Binary Pattern was first introduced in 1996 by Ojala et al as a basic binary operator. Local Binary Pattern works as a powerful texture classifier.

The pixels of the image are labeled by the binary operator by comparing the center pixel value with the 3x3 neighborhood of each pixel values to form a binary number (8 bit) which is then converted to the decimal value. The vertical and horizontal projection is obtained which is a one dimensional feature vector for the 2D face image. For a given pixel at (x_c, y_c), LBP code is obtained using the following equation.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{7} S(g_p - g_c)2^p, \quad S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where:

g_c = gray value of center pixel
g_p = gray value of neighbouring pixel of g_c
P = 8 maximum of 8 neighbours of center pixel
R = 1 for selected box of 3*3

Thus, there can be total $2^8 = 256$ different values that can be assigned to a pixel.

In Fig. 1, of 3x3 size image block is considered. Central pixel value is 7 and center pixel is surrounded by the 8 pixels in all eight directions. The LBP converts all 9 pixel values in to a single value. This will be done by comparing the pixel value of every neighboring pixel with the central pixel value (that is the intensity value at that point). The pixel values are usually considered in the grayscale. The pixel value greater than or equal to the central pixel is assigned 1 and lower value is assigned a 0 since only binary values of 0 and 1 can be assigned to the pixels. Now byte is formed by these 8 pixels surrounding center one.

12	15	18
5	7	3
8	1	2

⇨

G_7	G_6	G_5
G_0	G_c	G_4
G_1	G_2	G_3

Fig. 1. input block of size 3*3.

B_7	B_6	B_5
B_0		B_4
B_1	B_2	B_3

⇨

1	1	0
0		0
1	0	0

Fig. 2. input block coded in LBP

LBP code is obtained by circularly tracing the bins in clockwise direction.
Binary LBP code = 1 1 0 0 0 0 1 0
Decimal code = 194
This resultant byte then is turned into a decimal number. As long as we are consistent with the binary values, any pixel block can be encoded into a byte as seen above. Converting this byte into a decimal value yields a single decimal value for the block. Thus using the values obtained for each of the block in the complete image feature vectors obtained. These feature vectors acts as the input for the classification process. Even if there are changes in lightening conditions, the relative pixel difference between the central pixel and the neighboring ones remain remains the same in LBP and it is main advantage of efficient LBP. The binary pattern remains the same irrespective of the illumination and rotation conditions. Overall values of the pixels increase or decrease if the brightness of an image changes which makes relative difference same. This property of the LBP makes it very suitable for the real time applications. There are various variants of the LBP that can be incorporated for better results as: Transition Local Binary Pattern, Direction Coded Local Binary Pattern, Modified Local Binary Pattern, Multi-block LBP, Volume LBP and RGBLBP.

Our experiment proposes efficient LBP as the feature extraction technique which yields accurate and efficient classification results.

2.2.2 Introduction to convolutional neural network (CNN)

Convolution neural network[11],[12] is one of the representative network structures in depth learning, and has become a hotspot in the field of speech analysis and image recognition. Data in the form multiple arrays are processed by ConVnets. CNN can take raw image as an input,thus avoiding feature extraction and data reconstruction procedure in the standard learning algorithms. Its weight-sharing network structure makes it more similar to the biological neural network, which reduces the complexity of the network model and reduces the number of weights. Convolutional Neural Networks are invariant to translational, scale, tilt, or other forms of deformation. Local area perception is the important idea frame of convolution neural; CNN architecture mainly constitutes of three layers viz. Convolutional Layers, Pooling Layers and Fully Connected Layers. Output of convolutional layer is given to pooling layer and so on. This convolution layer extract the features, and then combined to form a more abstract feature, finally, forming the description of the image object characteristics. The CNN structure diagram is as shown in Fig.3.

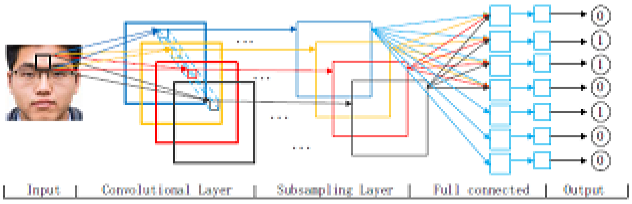


Fig.3 The model of CNN

2.2.2.1 Convolutional layer

The primary purpose of Convolution in case of a Convolutional Neural Network is to extract features from the input image. Convolution preserves the spatial relationship between pixels by learning image features i.e. filter using small squares of input data In CNN terminology, ’filter’ or ’kernel’ or ’feature detector’ is used The matrix formed by sliding the filter over the image and computing the dot product is called the ’Convolved Feature’ or ’Activation Map’ or the ’Feature Map’. In this layer, Filter size and stride (after how many pixel filter should be moved) is to be chosen. Output of convolutional layer is given by
(W-F+2P) /S+1=O
If stride is not chosen properly, neurons do not "fit" neatly and symmetrically across the input. So in order to fit neurons neatly , zero padding is used across image.

2.2.2.2 Relu layer

In this layer, every negative value from filtered image is removed and replaced it with zero This is done to avoid values summing up to zero Rectified Linear Unit transform function only activates a node if the input is above certain quantity, if the input is below zero, the output is zero. But the input rises above threshold, it has linear relationship with dependent variable ReLU has been proven as an effective solution to resolve vanishing gradient problem in training a convolutional neural network
Output of RELU layer is given by

$$f_r = ReLU(x_i) = max(0, x_i)$$

2.2.2.3 Pooling layer

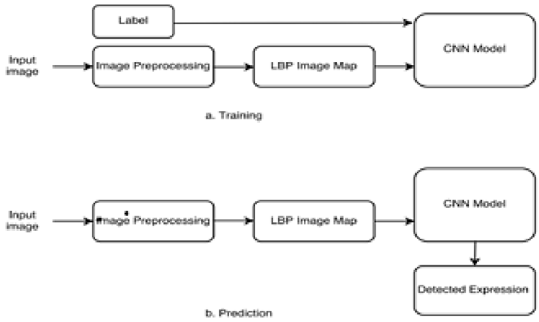
The objective of a pooling layer is to subsample the rectified feature map for reducing its spatial dimensionality thus produces a more compact feature representation. The output of this pooling layer is a pooled featured map. There are two widely used pooling techniques-max pooling and mean pooling.

2.2.2.4 Fully connected layer

The function of Fully Connected Layer is to remap the pooled feature map from a two-dimensional structure into a one-dimensional vector i.e. feature vector. The output is a "flatten" pooled feature map. This feature vector acts as regular Fully connected layer for classification.

3. THE METHOD BASED ON LBP FEATURE FOR CNN

LBP image depicts local texture feature of an image. CNN will perform better if image in texture format is given to it. In general original image is provided as an input to CNN; But in this CNN learns from the pixel level. In the learning process of CNN, the feature extraction starts from the lowest level feature. There is more difficult learning in feature extraction. In view of the above reasons, we propose a method to improve training CNN. Local texture feature of the original image which is provided by LBP is higher than the pixel level, and CNN learns the characteristics of the image better than the pixel level in learning.
This paper presents a method of facial expression recognition based on LBP combination of features and CNN. The CNN is trained by using the LBP feature map of the face image as the input of the CNN. The following block diagram describes in detail the implementation of face recognition based on LBP features for CNN.



In this model, we first process the LBP image to generate the LBP feature map of the image, and then use the LBP feature map of the image as CNN input to train the CNN. Similarly, in the test identification images, it is first LBP feature map is extracted and then LBP image is fed into the CNN classifier for identification

4. EXPERIMENTS

In this section, the effectiveness of proposed CNN model is evaluated on Cohn-Kanade database. The database includes of 2,105 images having the resolution pixel of 640 x 480 or 640 x 490. All image sequences includes a neutral and apex expressions. Especially the last frames cover the most discriminative image. In the experiments, the image size was resized into 28x28. The proposed model was trained for 60 epochs. The learning rate was selected as 0.01 for first twenty epochs and 0.0001 for next 35 epochs and 0.00001 for the end epochs. Batch size is selected as 100. Original images are first pre-processed using viola jones face detection and different augmentation techniques are applied to increase data-size. For

Table -1: Confusion matrix for CK+ dataset with image size 28*28

	No of Images	Angry	Disgust	Fear	Happy	Neutral	Surprise	Unhappy	Accuracy%
Angry	40	31	3	0	0	4	1	1	77.5
Disgust	47	0	38	0	3	5	0	1	80.8
Fear	20	0	2	12	0	3	1	2	60
Happy	73	1	0	2	69	1	0	0	94,5
Neutral	105	0	0	0	0	102	3	0	97.1
Surprise	96	0	0	1	1	1	93	0	96.8
Unhappy	19	0	1	0	0	3	0	15	78.9

cohn kanade dataset, the efficiency of recognition(90.00%) is seen for 28*28 image size.

Table -2: Performance comparison of different state-of-art approaches with our experimental results

Evaluation	Index Accuracy
KNN	77.27%
CNN	80.303%
LBP+CNN	90%

5. CONCLUSIONS

This paper proposes a facial expression recognition model based on LBP feature map and CNN. Firstly, the image is transformed into LBP feature map, then the LBP feature map is used as the input of CNN to train CNN. In face recognition And then into the CNN to identify. By comparing the accuracy of classification, the effect of the proposed classification method is better than the effect of CNN and ANN. The knowledge of CNN learning is a pixel-level feature that is low-level, untreated knowledge When the CNN input is an image. When the LBP image is used as input to the CNN, the knowledge of CNN learning is knowledge of the edge of the processed image. By comparison of the experimental data, we can see that compared with the original knowledge, the processed knowledge is easier to be learned and understood by CNN, and the face recognition is better. So, processed knowledge is preferable as input to CNN training data to get desire output.

REFERENCES

[1] Mehrabian, A. "Silent Messages". Wadsworth, Belmont, California, 1971.

[2] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar,Z., Matthews, The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression I. 2010.

[3] Alptekin D.,Yavuz Kahraman "Facial Expression Recognition Using Geometric Features" The 23rd International Conference on Systems,Signals and Image Processing. 23-25 May 2016, Bratislava,Slovakia.

[4] Lei Gang, Li Xiao-hua* ,Zhou Ji-liu, Gong Xiao-gang "Geometric feature based facial expression recognition using multiclass support vector machines" IEEE International Conference on Granular Computing,2009

[5] Kiran Talele, Archana Shirsat, Tejal Uplenchwar, Kushal Tuckley "Facial Expression Recognition Using General Regression Neural Network" IEEE Bombay Section Symposium (IBSS),2016

[6] Reza Azmi, Sam ira Yegane "Facial Expression Recognition in the Presence of Occlusion Using Local Gabor Binary Patterns" 20th Iranian Conference on Electrical Engineering, (ICEE2012), May 15-17, Tehran, Iran

[7] Simona Maria Banu, Gabriel Mihail Danciu, Gabriel Boboc, Moga and Cristiana "A Novel Approach for Face Expressions Recognition" IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics,September 20-22, 2012,

[8] Subotica, Serbia Alptekin D.,Yavuz Kahraman "Early Facial Expression Recognition Using Hidden Markov Models" International Conference on Pattern Recognition, 22nd 2014

[9] Ke Shan, Junqi Guo, Wenwan You, Di Lu, Rongfang Bie "Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural- Network Structure " IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA),2017

[10] Di Huang, Caifeng Shan, Yunhong Wang and Liming Chen "Binary Patterns and Its Application to Facial Image Analysis: A Survey" IEEE Transactions on Systems, Man and Cybernatic Part- C,: Applications and Reviews, Vol 41, No. 6, November 2011.

[11] L. Deng, D. Yu "Deep learning: methods and applications," Foundations and Trends in Signal Processing, Vol.7. pp. 3â˘A ¸ S4, 2014.

[12] S. Christian, A. Toshev, D. Erhan "Deep neural networks for object detection," Advances in Neural Information Processing Systems, 2013.