Facial Emotion Identification Based on Local Binary Pattern Feature Detector

Sahar Zafar Jumani¹, Fayyaz Ali², Irfan Ali Kandhro¹, Qurban Ali Lakhan¹, Usman Ali³, M. Waqas Haroon³ and Shakeel Ahmed⁴

¹Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan sahar@smiu.edu.pk, irfan@smiu.edu.pk, qurban@smiu.edu.pk
²Department of Software Engineering, Sir Syed University of Engineering & Technology, Karachi, Pakistan; Fayyaz54@gmail.com
³Department of Computing, Abasyn University Islamabad, Karachi, Pakistan; usman007professor@gmail.com, tanoli90@gmail.com
⁴Department of Computer Systems Engineering & Sciences, Baluchistan University of Engineering & Technology Khuzdar; shakeel.sheikh@buetk.edu.pk

Abstract

Objectives: The emotion detection is one of the important fields in computer human interaction and this study plays significant a role for identification facial expression from the images. To identify the single emotion, need a various variability of human shapes such as pose, color, texture, expression, posture and orientation. In this study, we implement Local Binary Pattern (LBP) based filters for identifying the dynamic face textures. And moreover, this approach also provides extension and simplification. **Methods/Statistical Analysis**: We used built-in FER2013 datasets, the database consisting seven classes (Surprise, Fear, Angry, Neutral, Sad, Disgust, Happy). The dataset is divided into three parts testing, validation and training (15% and 70%). The Convolution neural network is trained with feature Descriptor Local Binary Pattern. **Findings**: The experimental results have demonstrated that local LBP representations are effective in spatial dynamic feature extraction, as they encode the information of image texture configuration while providing local structure patterns. The advantages of our approach include local processing, robustness to monotonic grayscale changes and simple computation. The results show that, the performance LBP based Convolution Neural Network (CNN) model is better than conventional CNN. This research study further helps in image classification and image processing fields. **Application/Improvements**: It is recommended that LBP should be used for finding the local regions or pattern from the image. The LBP computation and local processing is quite better with robustness and monotonic changes.

Keywords: Convolution Neural Network (CNN), Facial Emotion, Facial Expression, Face detection, Expressions, Local Binary Pattern (LBP)

1. Introduction

Facial expression is one of most important way for verbal communication, and it demonstrates internal affective intentions and states. Based on various facial expression and representation of human face, the Face is one of the predominant ways for communicating and inferring states with peoples. Over the last decade, programmed based expression detection powers in the field of the human computer interactions. (e.g. emotion analysis, speaking, chatting and image processing)¹. FER recognition system is increasing day by day in field of artificial intelligence prompting for solving the important issue in the scientific community. The issues in FER system become even more

advanced and hard when we detect the expression from the videos and distinct dynamic textures in image. The basic automatic facial expression system is divided into three different phases²: Face preprocessing / acquisition, Face feature detection/extraction and finally the Classification of facial expression. In face recognition system, face acquisition is basic processing stage where we find the important region from the input image. The most real face detector application used most common type cascade classifiers with Harr-wavelet features³. With the help of the eye position and face it detected textures which are quite balanced. After locating and identifying the face regions, the second step is to extract the features from the input original image to show the facial expressions. This task can be done through the two suggested techniques: a) Appearance based methods, and b) Geometric based methods². The Geometric based approach shows the location and shapes of facial expression parts such as noise, mouth, lips and brows. However, the geometric features need the accurate and most reliable facial expression detection which is hard to understand the real time applications. More significantly, the usually geometric features cannot encode into the skin texture such as furrows and face wrinkles that are the serious and life-threatening for facial expression show off. As compare to geometric, the appearance approach shows the changes the skin texture of face, to combine with bulges, wrinkles and furrows. There are some image filters are used to extract the complete face features and extract the important region form the input image and try to find the facial appearance changes with help of Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA) and Gabor Wavelet Analysis (GWA). But, it's quite expensive for utilization of Gabor filters to convolve the face and extracting the multi scale and multi orientation coefficients from the original image. Gabor filter is inefficient with respect of time and memory supply⁴⁻⁷. Recently, the first time Local Binary Pattern (LBP) proposed for the texture analysis, and its nonparametric techniques for extracting the local features and structure of input image⁴. Still good descriptor is not available for finding the appearance of local regions and computes low and high intra-class variance. The main assets of local binary pattern features acceptance brightness updating and simplicity in computation. In facial expression recognition LBP is successfully used for local regions and feature extraction from the image⁵⁻⁷. In final stage of automatic emotion expression is classify

the various emotion-based expressions from extract total features using different classification techniques such as Support Vector Machine (SVM)⁸ and Simple Neural Network (SNN)², Hidden Markov Models (HMM)¹⁰, K-Nearest Neighbor (KNN)¹¹ and rule-based classifiers¹² has been implemented for emotion detection and expression recognition.

2. Proposed Methodology

2.1 LBP

The LBP based image filter operator is developed for texture description and feature extraction. This filtering operator assigned class in each pixel of the image and set threshold (3×3) with neighborhood per pixel with center pixel value and results will generated in binary numbers (0,1). Furthermore, histogram with labels can be used to describe the texture description¹³.

To deal with various texture scales, the LBP filter used to combine different neighborhood hyperparameters^Z demonstrating the local neighborhood needed to be sampling points and each pixel points are evenly adjusted with centered of the circle. Further, labeled with number of sampling points and radius in the operator LBP linked with gray scale and rotation in-variant texture filter. The bilinear interpolation is used whenever the point does not exist in the center. The notation P and R indicates the pixel neighborhood where P shows the sampling points and R shows the radius. Another type of original image filtering operator is uniform pattern the LBP operator also called uniform operator if the binary pattern contains binary transition (bitwise) between 0 - 1^{14} . Vice versa when the binary bits in circular form then 0-transitions (0000000), 2-transitions (01110000) and 2 transition are uniform whereas the pattern 4-transitions (11001001) are uniform 4-transitions (11001001) and 6-transitions (01010011) are not. Figure 1 Simple LBP feature descriptor operator and along with fixed and selected m neighbors the code of LBP operator is given for the center pixel *t* given by:

$$BP_{(m,r)}(tc) = \sum_{n=0}^{m=1} U(t_r - t_s) 2^i$$

	85	99	21	Thursday	1	1	0	
-	54	54	86	I nresnoid	1		1	Binary: 11001011
	57	12	13] ([1	0	0	f

Figure 1. Simple LBP feature descriptor operator.

The Eq (1) configure and setup the LBP condition where values are 0 and 1 and output of the filter LBP shows with binary labels, as denoted with m-bit binary numbers with 2 discrete values shows the smoothness and texture alignment in local region. if $t_r \le t_s U() = 0$ and if $t_r > t_s$ then 1 the local binary pattern computed in clockwise direction for example the sequence of binary numbers of number 83 (11001010). Assuming the coordinate t_s (0,0) and t_r of rsin (2 π), rcos (2 π) with each neighbor and hyperparameter set (m, r). In bilinear interpolation estimation, the neighbors do not exist in the center position. After completing and obtaining the local binary pattern operator code with existence of histogram¹⁵ in Figure 2 shows the results LBP of dynamic texture with histogram orientation of edge, flat and corner.



Figure 2. LBP with histogram (edge, flat and corner).

2.2 Convolution Neural Network (CNN) Based Facial Expression

Figure 3 represents the system architecture of LBP based Convolution Neural Network (CNN). The CNN is combined with two types of layers, respectively called C-layer (Convolutional layer) and S-layer (sub-sampling layer) distinct from conventional deep learning methods. Addition to this, Convolutional neural network takes input as two-dimensional (2D) image. So far, this approach has unique advantages in area of image recognition. As shown in Figure 2, the images data feed as input in the classic

network CNN. The model is convoluted with various kernels and generates the unique features, and moreover these features mapped into the layer C1. In addition, the layer C1 features will be reduced and sub-sampled into the layer S1 with respect of size and form. Usually, the size of pooling method is 2×2 , and further this process also repeated for the other layers such as C2 and S2. Once the important features are extracted, the 2D pixels are converted into the 1D form and pass to the classic neural network. in applied computer applications, we mostly utilized softmax as final manifold classifier. The model used input 64×64 grayscale images and classification outputs results are rendered in single class from seven different classes, representing happy or other labels. There are four various sizes Convolutional layers are used from C1 to C4 and combine with three MaxPooling layers from P1 to P3 and addition to this, the regularization approach used as dropout layers in network and these layers are fully connected with input and output layers as shown in Figure 4. In C1convolutional layer used filters 64 \times 64 the input image of 32 different automatic learnable kernels with size 3×3 to give results in 32 matrices and dimensional of 62×62 . The convolutional results passed from C1 to P1 MaxPooling layer with (32) size of 2 by 2 learnable kernels. And results received from the P1 will be (32) matrices with dimension 31 by 31 size. The results from C1 to P1 passed from the second Convolutional layer C2 which has (32) self-learnable kernels size of $3 \times$ 3 to receive the 32 matrices with size of 29 by 29. Then, MaxPooling layer P1 used (32) learnable kernels of size (2, 2) implements for continue the previous computed results to get (32) matrices of size 14×14 . Next, the layers C3 to C4 with (32) automatic learnable kernels with size of (3,3) on every layer. The results get C4 are 64 matrices with size of (10,10) which passed to the P3 with auto learnable kernels sizes of (2, 2) to get the results in form of the (64) matrices with size 5×5 . The results are passed from flatten layer with 1600 values. Further process the data, two fully connected layers are used, the layer 1(1024) hidden units and other with (512) and final outputs generates results with two classes "happy" or other classes from FEA2013 dataset.

3. Results and Discussion

In this section, the effectiveness of proposed approach is evaluated through the publicly available Facial



Figure 3. System architecture of LBP based convolution neural network.





Expression Recognition (FER2013) dataset. The database includes 3,589 samples and all samples are grayscale images and have resolution of 48 x 48. Facial expression task categorizes based on each face texture shown in the emotion recognition dataset, furthermore, the system performed facial recognition based on predefined labels (0) Angry, (1) Disgust, (2) Fear, (3) Happy, (4) Sad, (5) Surprise and (6) Neutral, the dataset is divided into Training, Testing and Validation sets with 28709,3589 and 3589 images. The results indicate that usage of modern deep learning models has potential effect on system to improve the performance of Facial expression database. Our proposed LBP based CNN model performed well as compare to other feature extraction operators with CNN. By merging with LBP feature extraction operator with CNN, we achieved 74% testing accuracy which is quite better than HOGCNN¹⁶. Addition to this, the LBP based architecture obtains good results utilizing additional samples of training data or features. We expect that addition of comprehensive data augmentation and face registration techniques with FER specific improve the further results. Table 1 summarized the results of proposed LBPCNN model on FER2013 database. The Figure 4 indicates training and testing accuracy with 98%,74% and similarly, Figure 5 shows the training and testing losses, where 0.05 training and 2.01 testing. In Table 2 summarized the results of proposed FER-LBPCNN system with various hyperparameters. The numerous results have been achieved with combination of epochs (15 to 50) and batch sizes (8 to 1024). By using batch 32 and epochs 25 our model represents state of art performance. In Figure 6 depicts the results of confusion matrix with predefined labels. Further, this evaluation method analysis the performance of model in more detail with "True" and "False" positive and True and False negative. The matrix shows the results labelwise, where label "Happy" correct predication is 541 which is comparatively better than six labels. Figure 7 shows the classification report, where individually



Figure 5. Training and testing loss of FER-LBPCNN.



Figure 6. Confusion matrix of FER-LBPCNN.

	precision	recall	f1-score	support
0	0.78	0.65	0.71	386
1	0.72	0.76	0.74 0.66	38 406
3	0.87 0.60	0.85 0.69	0.86 0.64	640 448
5	0.82 0.76	0.83 0.65	0.82 0.70	325 451
accuracy			0.74	2694
macro avg weighted avg	0.74 0.75	0.74 0.74	0.73 0.74	2694 2694

Figure 7. Classification report of FER-LBPCNN

Table 1.FER-LBPCNN accuracy and loss

Model	Acc	Loss	Val_Acc	Val_Loss
FER-CNN	98%	0.05	72%	2.01
FER-LBPCNN	98%	0.04	73%	2.00

 Table 2.
 FER-LBPCNN results with various hyperparameters

Epochs	Batch Size	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
15	8	95%	90%	0.07	2.03
20	16	97%	93%	0.07	2.03
25	32	98%	97%	0.04	2.02
30	64	98%	97%	0.04	2.01
35	128	96%	93%	0.05	2.02
40	256	96%	91%	0.06	2.03
45	512	96%	92%	0.06	2.04
50	1024	95%	91%	0.06	2.04

precision and recall computed categorize-wise and addition to this, the classification report summarized in Accuracy 74%, macro 73% and weighted averages 74%.

4. Conclusion

This work proposed a novel and automated effective Facial Expression Recognition (FER) system. The FER system is dividing a facial image into small regions and using local binary pattern to compute the features and description of each region. In this study, we conducted experiments on FER2013 dataset which is publicly available and most common image database in FER based CNN. Furthermore, the dataset is divided into three parts: training, testing and validation with 35089, 3589 and 3589. Evaluation metrics are

used to analysis the systems performance's with different way namely Accuracy and Confusion Matrix and Classification report. The FER-LBPCNN training and testing accuracy were of 98%, 73%, respectively. Similarly, Losses were 0.04, 2.00. It is found that the Accuracy of FER-LBPCNN is better than Histogram of oriented Gradients and traditional CNN models. We have identified the existing issues and bottlenecks. Our FER system is computationally simple and robust in terms of rotation variations and grayscale images making as it is very promising in real applications.

5. Future Work

In future, the more advance features descriptor filters are to be implemented for finding the weights for facial images by dividing into the local regions and further, explore temporal information.

6. References

- Cowie R. Emotion recognition in human-computer interaction, IEEE Signal processing magazine. 2001; 18(1):32-80. https://doi.org/10.1109/79.911197.
- Tian YL, Kanade T, Cohn JF. Facial Expression Analysis. Handbook of Face Recognition; 2005. p. 247–75. https:// doi.org/10.1007/0-387-27257-7_12.
- 3. ViolaP, Jones MJ. Robustreal-time face detection, International Journal of Computer Vision. 2004; 57(2):137–54. https://doi. org/10.1023/B:VISI.0000013087.49260.fb.
- Turk MA, Pentland AP. Face Recognition using Eigenfaces. Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 1991. p. 586–91.
- Ojala T, Pietikäinen M, Mäenpää T. Multi resolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 24(7):971–87. https://doi.org/10.1109/TPAMI.2002.1017623.
- Shan C, Gong S, McOwan PW. Robust Facial Expression Recognition using Local Binary Patterns. IEEE International Conference on Image Processing; 2005. p. 1–6.
- Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: A comprehensive study, Image and Vision Computing. 2009; 27(6):803-16. https://doi.org/10.1016/j. imavis.2008.08.005.
- Tian YI, Kanade T, Cohn JF. Recognizing action units for facial expression analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; 23(2):97–115.

https://doi.org/10.1109/34.908962. PMid: 25210210, PMCid: PMC4157835.

- Moore S, Bowden R. Local binary patterns for multiview facial expression recognition, Computer Vision and Image Understanding. 2011; 115(4):541–58. https://doi. org/10.1016/j.cviu.2010.12.001
- Cohen I. Facial expression recognition from video sequences: Temporal and static modeling, Computer Vision and Image Understanding. 2003; 91(1-2):160–87. https://doi.org/10.1016/S1077-3142(03)00081-X
- Kotsia I, Pitas I. Facial expression recognition in image sequences using geometric deformation features and support vector machines, IEEE Transactions on Image Processing. 2006; 16(1):172–87. https://doi.org/10.1109/ TIP.2006.884954. PMid: 17283776.
- 12. Pantic M, Patras I. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics).

2006; 36(2):433-49. https://doi.org/10.1109/TSMCB .2005.859075.

- Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007; 6:915–28. https://doi. org/10.1109/TPAMI.2007.1110. PMid: 17431293.
- Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; 12:2037–41. https://doi.org/10.1109/ TPAMI.2006.244. PMid: 17108377.
- Li W. Local binary patterns and extreme learning machine for hyperspectral imagery classification, IEEE Transactions on Geoscience and Remote Sensing. 2015; 53(7):3681–93. https://doi.org/10.1109/TGRS.2014.2381602.
- Jumani SZ. Facial expression recognition with histogram of oriented gradients using CNN, Indian Journal of Science and Technology. 2019; 12:24–30.