Image Based Facial Expression Recognition Using Local Neighborhood Difference Binary Patterns

Sumeet Saurav[1, 2], Sanjay Singh[1, 2], Madhulika Yadav[3] and Ravi Saini[1, 2]

[1] Academy of Scientific & Innovative Research (AcSIR), Chennai, India
[2] CSIR-Central Electronics Engineering Research Institute, Pilani, India
[3] Department of Electronics, Banasthali Vidyapith, Rajasthan, India
sumeetssaurav@gmail.com

Abstract. Recently, automatic facial expression recognition (FER) has gained enormous interest among the computer vision researchers because of their potential deployment in a number of industrial, consumer, automobile and societal applications. There are a number of techniques available in the literature for FER, among them many appearance-based methods such as local binary pattern (LBP), local directional pattern (LDP), local ternary pattern (LTP), gradient local ternary pattern (GLTP) and improved local ternary pattern (IGLTP) have been shown to be very efficient and accurate. In this paper, we have proposed a new descriptor called Local Neighborhood Difference Binary Pattern (LNDBP). This new descriptor is motivated by the recent success of local neighborhood difference pattern (LNDP) which have been proven to be very effective in image retrieval. The basic characteristics of LNDP as compared to the traditional LBP is that it generates binary patterns based on a mutual relationship of all neighboring pixels. However, in the case of LBP the mutual relationship does not exist, here only the neighboring pixels are compared with the central pixel to generate the binary pattern and hence there is a loss of information which is well captured by LNDP. Therefore, in order to utilize the benefit of both LNDP and LBP, we have proposed LNDBP. We have also employed a dimensionality reduction technique to reduce the dimension of the LNDBP features. The reduced features are then classified using kernel extreme learning machine (K-ELM) classifier. In order to validate the performance of the proposed method, experiments have been conducted on two different FER datasets. The performance has been observed using well-known evaluation measures such as accuracy, precision, recall, and F1-Score. The proposed method has been compared with some state-of-the-art works available in literature and found to be very effective and accurate.

Keywords: Facial Expression Recognition (FER), Local Neighborhood Difference Pattern (LNDP), Principal Component Analysis (PCA), Kernel Extreme Learning Machine (K-ELM).

1 Introduction

Recently, automatic facial expression recognition (FER) has gained enormous interest among the computer vision researchers because of their potential deployment in a number of industrial, consumer, automobile and societal applications. For example, such a technology could be embedded inside a robot for providing home services like talking to children and taking care of elderly people. In addition, FER based technology can be deployed in a car to identify the fatigue level of driver and produce warning alarm in case the fatigue level exceeds the threshold limits, which will definitely avoid many accidents and save the life of people. Automatic FER can also be used by companies to determine the worth of their products before their actual launch. Facial expression provides an important cue which reveals the actual intention and state of mind of a person. Therefore, researchers are trying to develop systems for automatic FER which will have the capability of reading the facial expression of a person just like us and take necessary actions in response.

The techniques available in the literature for automatic facial expression recognition can be broadly classified into two main categories: geometric-based methods and appearance-based methods whose details can be found in [1, 2]. As in this work, we have used the appearance-based method for facial feature extraction and more specifically texture based approach, therefore, a brief overview of different works related to our approach has been discussed below.

Well, known techniques which come under appearance-based feature extraction methods include the use of local binary patterns (LBP), local ternary pattern (LTP), local derivative pattern (LDP), local directional number pattern (LNDP), local directional texture pattern, local directional ternary pattern and so on. The first successful demonstration of LBP for the purpose of FER has been reported in [3]. In this work, the authors have reported a comprehensive study on the role of LBP for FER. Although LBP is very effective and computationally efficient feature descriptor, it has been found to perform poorly under the presence of non-monotonic illumination variation and random noise as even a small change in grey-level values can easily change the LBP code [4]. To overcome this limitation, different new techniques have been developed as well as different modifications of the original LBP has been done over the time. One such modification is Sobel-LBP [5]. The performance of the operator has been investigated on a facial recognition application and found to outperform the traditional LBP operator in terms of recognition accuracy. However, this operator also fails in uniform and near-uniform regions where it generates inconsistent patterns just like LBP as it also uses only two discrimination levels. To overcome this LDP [6] was developed which employs a different texture coding scheme to that of LBP, where directional edge response values are used instead of grey-level intensity values. While LDP has been shown to outperform LBP, it also tends to produce inconsistent patterns just like Sobel-LBP in uniform and near-uniform regions due to similar reasons. In order to overcome the limitations of LDP and Sobel-LBP, LTP was developed. LTP adds an extra discrimination level and uses the ternary code as opposed to binary codes in LBP. More recently, a technique called gradient local ternary pattern (GLTP) [7] has been developed for the purpose of FER which combines Sobel operator with LTP operator. GLTP uses a three-level discrimination ternary coding scheme like LTP of gradient magnitudes obtained after Sobel operation to encode the texture of an image. As expected, GLTP has proved to be more effective for FER task compared to the earlier discussed operators. Another feature descriptor which was developed to overcome the limitations of the LBP is the Weber local descriptor (WLD) [8]. An important property of WLD is that it is less sensitive to noise and illumination changes and has been adopted for the purpose of FER in [9]. A more recent face descriptor called local directional ternary pattern (LDTP) has been developed for FER [2]. LDTP efficiently encodes information of emotion-related features by using the directional information and ternary pattern in order to take advantages of the robustness of edge patterns in the edge region while overcoming weaknesses of edge-based methods in smooth regions. Another recent method for FER which has been motivated by GLTP is improved gradient local ternary patterns (IGLTP) proposed by the authors in [10]. The improvements over GLTP includes the use of enhanced pre-processing, a more accurate Scharr gradient operator, dimensionality reduction via principal component analysis (PCA) and facial component extraction. A very recent work for FER has been proposed that makes use of multi-gradient features and five level encoding called Elongated Quinary Pattern (EQP) [11] and the operator has been proved to be very effective for the purpose of FER.

The remainder of the paper is organized as follows: In section 2 we have provided a brief description of the proposed methodology used in our work which has been followed by experimental results and discussion in section 3. Finally, section 4 concludes the paper.

2 Proposed Methodology

The algorithmic pipeline used for the implementation of proposed facial expression recognition (FER) system has been shown in Fig 1.

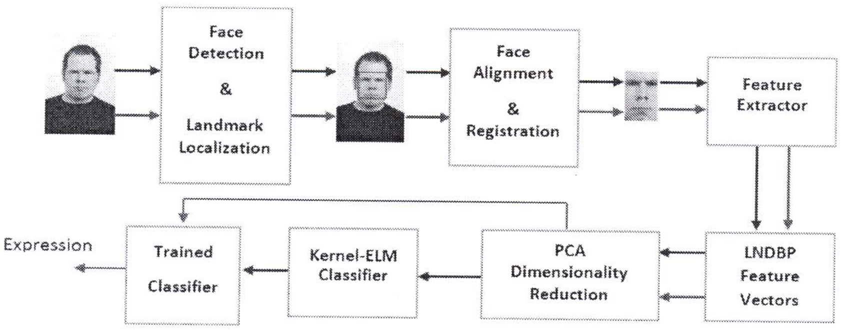

Fig. 1. Algorithmic Pipeline of the Proposed Facial Expression Recognition System

As shown in the figure, the first step detects a human face in the image under investigation and the detected face is then registered. The face registration step is very much essential as it allows the system to get facial features for different expressions at similar facial locations which give discriminative features. The vector of features extracted is then passed to dimensionality reduction to reduce the dimensions. The reduced feature is finally used for classification of the facial emotion using K-ELM classifier. A brief description of the different sequence of steps has been discussed in the further sub-sections.

2.1 Face Detection and Registration

In the first step, human faces are detected using Viola and Jones frontal face detector [12]-[13] which is then passed to the facial landmark detection unit [14] which mark the location of different landmarks on the face. Using coordinates of different landmarks from the left and right eyes, the positions of the eyes center is calculated. Based on the location of the eye's center, the image is rotated and in the subsequent step, the area of interest is cropped and scaled to the specified size of (147 x 108) in order to obtain the registered facial image.

2.2 Facial Feature Extraction and Dimensionality Reduction

The procedure used for feature extraction and dimensionality reduction has been shown in the Fig 2.

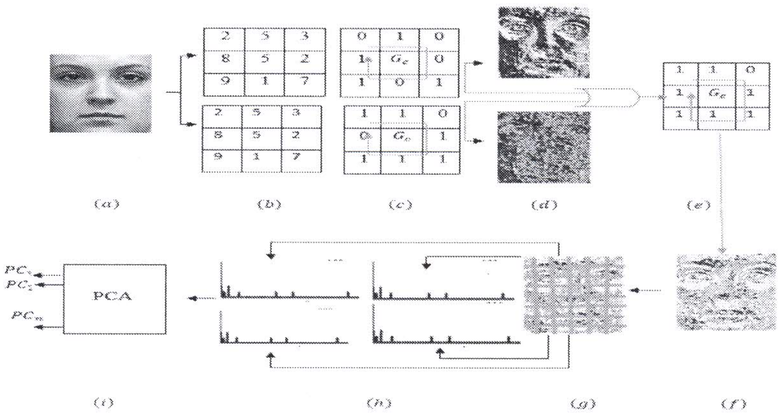

Fig. 2. Representation of Feature Extraction and Dimensionality Reduction Technique

As shown in the figure, the input image is first coded into corresponding LBP and LNDP pattern. An OR operation is performed on these patterns to generate the LNDBP pattern which is then multiplied by weights to generate the LNDBP coded image. Doing so retains the benefits of both the LBP [15] and LNDP [16] operator. The LNDBP coded image is then divided into different cells and an L2-normalized histogram of these cells are concatenated which encodes the complete facial feature information. Since the size of the LNDBP feature is too large, therefore, dimensionality reduction using PCA has been used to reduce the dimension of the features. An important point to note here is that, the algorithm does encode the input image into corresponding LBP and LNDP coded image. These images have been shown in the figure above just to make a clear differentiation between different encoded images. A brief description about LBP and LNDP has been mentioned below.
Local binary pattern (LBP) is a very popular texture descriptor which first computes a corresponding LBP pattern using a 3 x 3 window by comparing the neighboring pix-

els with the center pixel. These patterns are then multiplied by some weights and summed to generate the LBP encoded pixel value for the corresponding center pixel value. Local binary pattern for a center pixel I_c and neighboring pixels I_n (n=1,2,..,8) can be computed as in (1) and (2) respectively.

$$LBP_{P,R}(x,y) = \sum_{n=0}^{P-1} 2^n \times F_1(I_n - I_c) \quad (1)$$

$$F_1(I) = \begin{cases} 1 & I \geq 0 \\ 0 & else \end{cases} \quad (2)$$

where R is the radius of neighboring pixels and P is the number of neighboring pixels. (x, y) are the coordinates of center pixel. The histogram of LBP map is calculated using (3) and (4).

$$His(l)|_{LBP} = \sum_{x=1}^{m} \sum_{y=1}^{n} F_2(LBP(x,y), l); L \in [0, (2^P - 1)] \quad (3)$$

$$F_2(a,b) = \begin{cases} 1 & a = b \\ 0 & else \end{cases} \quad (4)$$

where size of image is $m \times n$ and l is the pattern value.

The procedure used for the computation of LNDP feature has been shown in Fig 3. As shown in the figure, in a 3 x 3 block each neighboring pixel is compared with the two most adjacent and appropriate pixels which are either vertical or horizontal pixels as shown in (5)-(7).

$$k_1^n = I_8 - I_n \,, k_2^n = I_{n+1} - I_n \,, \quad for \quad n = 1 \quad (5)$$

$$k_1^n = I_{n-1} - I_n \,, k_2^n = I_{n+1} - I_n \,, \quad for\, n = 2, 3, \ldots \ldots, 7 \quad (6)$$

$$k_1^n = I_{n-1} - I_n \,, k_2^n = I_1 - I_n \,, \quad for\, n = 8 \quad (7)$$

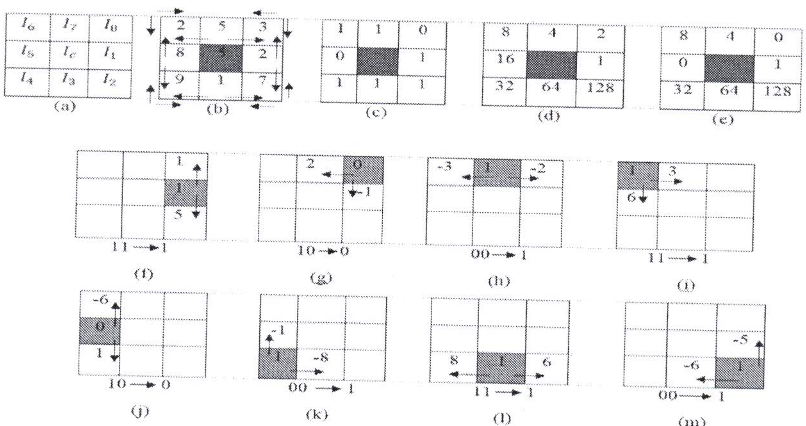

Fig. 3. Sequence of steps used in the computation of LNDP

where I_n (n=1,2,…,8) are the neighboring pixels of a center pixel I_c andk_1^n, k_2^n is the difference of each neighborhood pixel with two other neighborhood pixels. A binary number is assigned to each neighboring pixel with the help of equation (8).

$$F_3(k_1^n, k_2^n) = \begin{cases} 1, if\ k_1^n \geq 0\ and\ \ k_2^n \geq 0 \\ 1, if\ k_1^n < 0\ and\ \ k_2^n < 0 \\ 0, if\ k_1^n \geq 0\ and\ \ k_2^n < 0 \\ 0, if\ k_1^n < 0\ and\ \ k_2^n \geq 0 \end{cases} \quad (8)$$

By using the above binary values, LNDP value for the center pixel can be computed as (9) and then the histogram is calculated as in (10).

$$LNDP(I_c) = \sum_{n=1}^{8} 2^{n-1} \times F_3(k_1^n, k_2^n) \quad (9)$$

$$His(l)|_{LNDP} = \sum_{x=1}^{m} \sum_{y=1}^{n} F_2(LNDP(x,y), l); L \in [0, (2^8 - 1)] \quad (10)$$

Having too few features within a feature vector will most often result in classification failure even when using the best classifiers. On the other hand, having a very large feature vector will make the classification process slow and is not guaranteed to increase classification accuracy. To solve this, in this paper the authors have used PCA.

2.3 Kernel Extreme Learning Machine (K-ELM) Classifier

An Extreme Learning Machine (ELM) [17] classifier is basically a machine learning algorithm used for fast training a Single Layer Feed-Forward Neural Network (SLFN). Unlike the traditional back propagation algorithm which is used for training a neural network, the ELM does not involve any iteration, therefore computation of the output weights β does not involve any iteration and have a direct solution. ELM invariably makes use of a feature mapping function h(x) which helps in learning non-linearity. If the mapping function is not known, kernel technique can be applied into ELM based on Mercers' condition [18]. The output vector f(x) of a kernel ELM can be represented as

$$f(x) = h(x)\beta = h(x)H^T\left(\frac{I}{C} + HH^T\right)^{-1} T \quad = \begin{bmatrix} \phi(x, x_1) \\ \cdot \\ \cdot \\ \phi(x, x_{N_K}) \end{bmatrix} \left(\frac{I}{C} + \Phi\right)^{-1} T \qquad (11)$$

where

$$\Phi = HH^T = \begin{bmatrix} \phi(x_1, x_1) & \cdots & \phi(x, x_{N_k}) \\ \vdots & \ddots & \vdots \\ \phi(x_{N_k}, x_1) & \cdots & \phi(x_{N_k}, x_{N_k}) \end{bmatrix}$$

and N_k denotes the number of training samples used for the kernel ELM. These $N_k$samples are randomly selected from the training set. In this paper, Gaussian function is used as the kernel ϕ which is represented as in (12).

$$\phi(x, x_1) = exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (12)$$

3 Experimental Results and Discussion

3.1 Datasets

We have used two different datasets in our experiments. The first one is the extended Cohn-Kanade (CK+) dataset [19] which is an extended version of the CK dataset. In our experimental setup, we have used both 6 class and 7 class expression images which were obtained from 309 labeled sequences selected from 106 subjects. For 6-class expression recognition, the three most expressive images from each sequence were selected, resulting in 927 images and for 7-class expression, the first image of neutral expression from each of the 309 sequence was selected and added to the 6-class dataset, resulting in a total of 1236 images.
The second dataset used in the experiments is the recently introduced Radbound Faces database (RFD) [20]. The dataset contains images of 67 subjects performing 8 facial expressions (anger, disgust, fear, happiness, contemptuous, sadness, surprise and neutral) with 3 gaze directions. However, in our experiments, we have only used frontal gaze direction images comprising 7 expressions (anger, disgust, fear, happy, neutral, sad and surprise) for a total of 469 images.

3.2 Parameter Selection

There are a number of parameters involved in the design of any automatic FER system. The optimal value of these parameters needs to be determined for an accurate and efficient FER. Our first experiment involved determining the registered facial image size and the size of the cell in which the LNDBP encoded image is divided. From the recent work in [10] and [21], in our work, we have adopted two different facial image resolution of size 65 x 59 and 147 x 108 pixels and a number of cell-sizes. The experiments were performed on CK + 7 expression dataset with 10-fold cross validation strategy which is repeated 10-times. Kernel Extreme Learning Machine (K-ELM) with regularization parameter C and kernel parameter γ value of 100 and 200 respectively has been employed. The results of the experiments in terms of average overall accuracy, precision, recall and f1-score has been tabulated in table 1 and table 2 for image resolutions of 65 x 59 and 147 x 108 respectively for different cell sizes. Based on the tables, the facial image with a resolution of 147 x 108 and cell size of 12 x 11 performed well compared to other combinations of facial image and cell size and therefore was used in all our further experiments.

From the equations of K-ELM, we can find out that there are two parameters involved which needs to be determined to obtain better classification accuracy. In order to obtain the optimal value of these parameters, we performed a grid search along with 10-fold cross-validation which is again repeated 10-times as done in the above experiments. Experiments were performed on both the datasets involved in this work and results have been tabulated in table 3. The range of values for both C and γ have been taken here is [1:10] in logarithmic scale of base 2. In all our further experiments, we used the value of the optimal K-ELM parameters for deriving the accuracy of individual datasets.

Table 1. Determination of cell-size on 65 x 59 image size

[65, 59]	[5, 4]	**[6, 5]**	[7, 6]	[8, 7]	[9, 8]	[10, 9]	[11,10]	[12,11]
Avg. Acc. 10 runs	98.8±0.2	**99.3±0.2**	99.0±0.2	98.7±0.2	98.5±0.2	98.2±0.2	96.9±0.2	97.6±0.3
Feature Dim.	43008	**28160**	20736	14336	12544	9216	6400	6400
Avg. Acc.	99.03	**99.44**	99.27	98.94	98.63	98.46	97.24	98.13
Avg. Prec.	99.04	**99.52**	99.40	98.87	98.81	98.35	96.58	98.02
Avg. Rec.	99.21	**99.53**	99.28	99.37	98.68	98.85	98.03	98.38
Avg. F1-S	99.12	**99.53**	99.34	99.11	98.74	98.59	97.27	98.20

Table 2. Determination of cell-size on 147 x 108 image size

[147, 108]	[7, 6]	[8, 7]	[9, 8]	[10, 9]	[11, 10]	**[12, 11]**	[13, 12]	[14, 13]
Avg. Acc. 10 runs	98.9±0.3	99.1±0.2	99.3±0.1	99.3±0.1	99.1±0.2	**99.4±0.2**	99.0±0.1	99.1±0.3
Feature Dim.	87040	69120	53248	39424	33280	**27648**	22528	20480
Avg. Acc.	99.27	99.43	99.51	99.51	99.35	**99.51**	99.27	99.43
Avg. Prec.	99.55	99.64	99.70	99.64	99.54	**99.70**	99.52	99.64
Avg. Rec.	99.21	99.42	99.46	99.52	99.37	**99.52**	99.27	99.42
Avg. F1-S	99.38	99.53	99.58	99.58	99.45	**99.60**	99.39	99.53

Table 3. Determination of K-ELM Parameter

Performance Measure/ Datasets	**CK+ 7 Expressions**	**RFD 7 Expression**
Avg. Accuracy 10- runs	99.47+0.14	97.58+0.23
Kernel Parameter (γ)	64	1024
Regularization Parameter (C)	32	64

Hereafter, we fixed the value of the regularization parameter C and kernel parameter γ for the rest of the experiments.

In our next experiment, we tried to determine the optimal number of principal components. To determine this, we fixed the value of all other parameters. For K-ELM, we used the value of kernel parameter γ and regularization coefficient C obtained after grid-search. The experimental result of CK+ dataset has been tabulated in table 4 and that of RFD in table 5.

3.3 Results on CK+ Dataset

In order to determine the performance of the proposed FER pipeline on CK+ dataset, we performed 10-fold cross validation which was repeated 10 times using K-ELM with the value of C and γ determined using grid-search. On CK+ 6 expression the accuracy achieved using both LNDBP and LNDBP+PCA is 100 and on CK+ 7 expression dataset the FER pipeline achieved an accuracy of 99.5±0.1 using both LNDBP and LNDBP+PCA. The performance in terms of different measures has been shown in table 6 and table 7 corresponding to CK+ 6 and CK+ 7 expressions respectively.

Table 4. Determination of no. of principal component using CK+ 7 expression dataset

No. of PCA	32	64	96	128	160	192	224	256
Avg. Acc. 10 runs	98.7±0.2	99.2±0.2	99.3±0.2	99.4±0.1	99.3±0.2	99.3±0.1	**99.5±0.1**	99.4±0.2
Avg. Acc.	98.95	99.51	99.51	99.60	99.51	99.43	**99.60**	99.60
Avg. Prec.	99.25	99.71	99.71	99.76	99.71	99.65	**99.76**	99.75
Avg. Rec.	98.75	99.41	99.41	99.51	99.41	99.34	**99.51**	99.57
Avg. F1-S	98.99	99.56	99.56	99.63	99.56	99.50	**99.63**	99.65
Comp. Time (Sec)	0.150	0.153	0.164	0.184	0.198	0.182	**0.273**	0.275

Table 5. Determination of no. of principal components using RFD 7 expression dataset

No. of PCA	32	64	128	160	224	256	288	320
Avg. Acc. 10 runs	93.3±0.7	94.7±0.5	96.2±0.3	97.2±0.3	97.3±0.2	97.4±0.3	97.4±0.3	**97.7±0.3**
Avg. Acc.	94.24	95.31	97.01	97.65	97.44	97.87	97.87	**98.08**
Avg. Prec	94.24	95.31	97.01	97.65	97.44	97.87	97.87	**98.08**
Avg. Recall	94.37	95.40	97.13	97.70	97.47	97.90	97.91	**98.11**
Avg. F1-S	94.24	95.28	96.99	97.65	97.43	97.86	97.86	**98.07**
Comp. Time (Sec)	0.150	0.153	0.184	0.198	0.273	0.275	0.297	**0.328**

Table 6. Performance of LNDBP+PCA (CK+ 6 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Sa	Su	Recall
An	135	0	0	0	0	0	100
Di	0	177	0	0	0	0	100
Fe	0	0	75	0	0	0	100
Ha	0	0	0	207	0	0	100
Sa	0	0	0	0	84	0	100
Su	0	0	0	0	0	249	100
Precision	100	100	100	100	100	100	
F1-Score	100	100	100	100	100	100	

Avg. Performance: recall =100, precision =100, accuracy =100, F1-Score = 100

3.4 Results on RFD Dataset

Performance of the proposed FER pipeline on RFD dataset in terms of avg. precision, avg. accuracy, avg. recall and avg. f1s corresponding to best 10-fold cross-validation run out of 10 has been mentioned in table 8. The overall average accuracy of the 10 runs of the 10-fold cross-validation using LNDBP and LNDBP+PCA is 97.7±0.2 and 97.7±0.3 respectively. Here also we used the optimal value of the parameters determined in our earlier experiments.

Table 7. Performance of LNDBP+PCA (CK+ 7 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Ne	Sa	Su	Recall
An	135	0	0	0	0	0	0	98.54
Di	0	177	0	0	0	0	0	100
Fe	0	0	75	0	0	0	0	100
Ha	0	0	0	207	0	0	0	99.52
Ne	2	0	0	1	305	1	0	99.67
Sa	0	0	0	0	0	84	0	99.82
Su	0	0	0	0	1	0	248	100
Precision	100	100	100	100	98.71	100	99.60	
F1-Score	99.26	100	100	99.76	99.19	99.41	99.80	

Avg. Performance: recall =99.51, precision =99.76, accuracy =99.60, F1-S = 99.63

Table 8. Performance of LNDBP+PCA (RFD 7 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Ne	Sa	Su	Recall
An	67	0	0	0	0	0	0	100
Di	0	67	0	0	0	0	0	100
Fe	0	0	62	0	1	1	3	92.54
Ha	0	0	0	67	0	0	0	100
Ne	0	0	0	0	67	0	0	100
Sa	1	0	1	0	1	64	0	95.52
Su	0	0	0	0	1	0	66	98.51
Precision	98.53	100	98.41	100	95.71	98.46	95.65	
F1-Score	99.26	100	95.38	100	97.81	96.97	97.06	

Avg. Performance: recall = 99.11, precision =98.08, accuracy =98.08, F1-S = 98.07

In order to test the performance of the proposed algorithm, we also performed cross-dataset performance evaluation using LNDBP+PCA wherein, we kept one of the datasets as training data and the other one as a test. The results of the experiment have been tabulated in table 9 and seem to be very effective.

Table 9. Performance of LNDBP+PCA (Cross-Dataset)

Database/Perf. Measure	Avg. Precision	Avg. Recall	Avg. F1-s	Testing Time (Sec)	Testing Accuracy	C	γ
Train: CK+ 7 Expressions Test : RFD 7 Expressions	86.8	81.0	78.71	0.02	81.0	32	64
Train : RFD 7 Expressions Test : CK+ 7 Expressions	81.7	79.9	79.3	0.02	81.1	64	1024

Comparison result of the proposed FER framework with other state-of-the-art approaches has been shown in table 10. As shown in the table, the proposed approach has superior performance compared to the other approaches available in the literature.

Table 10. Comparison of recognition accuracy (%) on CK+ and RFD Dataset

Method	CK+ 6 Expressions	CK+ 7 Expressions	RFD 7 Expressions
LBP [22]	90.1	83.3	----
LDP[22]	93.7	88.4	----
LTP[22]	93.6	88.9	----
GLTP[7]	97.2	91.7	----
Improved GLTP[10]	99.3	97.6	----
HOG[21]	95.8	94.1	94.9
LNDBP Proposed	100	99.5	97.7
LNDBP+PCA Proposed	100	99.5	97.7

4 Conclusion

In the presented paper, a new facial feature descriptor has been proposed and named as local neighborhood difference binary patterns (LNDBP). The proposed feature extractor incorporates the benefits of LBP which computes the relationship of neighboring pixels with center pixel and LNDP which extracts the relationship among neighboring pixels by comparing them mutually. LNDBP is computed by using an OR operation on the local binary pattern and local neighborhood difference pattern which has been proved here to be very effective in extracting discriminate facial attributes. Dimensionality reduction technique using PCA has been used to reduce the dimension of the LNDBP features. K-ELM classifier has been used for classifying the facial expressions. We have tested the performance of the proposed approach using 10-fold cross-validation which has been repeated 10 times and cross-dataset wherein one dataset is used as a training set and the other as a test set. The experiments have been performed on two FER datasets viz. CK+ and RFD and performance has been observed using precision, recall, accuracy, and f1-score. We have also compared the performance of our approach with some of the state-of-the-art works available in literature and our performance measure clearly indicate that the proposed method outperforms other methods in terms of recognition accuracy and efficiency.

References

1. Rivera, A. R., Castillo, J. R., Chae, O. O.: Local directional number pattern for face analysis: Face and expression recognition. *IEEE transactions on image processing*, *22*(5), 1740-1752 (2013).
2. Ryu, B., Rivera, A. R., Kim, J., Chae, O.: Local directional ternary pattern for facial expression recognition. *IEEE Transactions on Image Processing*, *26*(12), 6006-6018 (2017).

3. Shan, C., Gong, S., McOwan, P. W.: Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, *27*(6), 803-816 (2009).
4. Zhou, H., Wang, R., Wang, C.: A novel extended local-binary-pattern operator for texture analysis. *Information Sciences*, *178*(22), 4314-4325 (2008).
5. Zhao, S., Gao, Y., Zhang, B.: Sobel-lbp. In 15th IEEE International Conference on Image Processing, *pp.* 2144-2147, (2008).
6. Jabid, T., Kabir, M. H., Chae, O.: Facial expression recognition using local directional pattern (LDP). In 17th IEEE International Conference on Image Processing, pp. 1605-1608 (2010).
7. Ahmed, F., & Hossain, E.: Automated facial expression recognition using gradient-based ternary texture patterns. *Chinese Journal of Engineering*, *2013*.
8. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: WLD: A robust local image descriptor. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), 1705-1720, 2010.
9. Alhussein, M.: Automatic facial emotion recognition using weber local descriptor for e-Healthcare system. *Cluster Computing*, *19*(1), 99-108, (2016).
10. Holder, R. P., Tapamo, J. R.: Improved gradient local ternary patterns for facial expression recognition. EURASIP Journal on Image and Video Processing, 2017(1), 42, (2017).
11. Al-Sumaidaee, S. A. M., Abdullah, M. A. M., Al-Nima, R. R. O., Dlay, S. S., Chambers, J. A.: Multi-gradient features and elongated quinary pattern encoding for image-based facial expression recognition. Pattern Recognition, 71, 249-263, (2017).
12. Viola, P., Jones, M. J.: Robust real-time face detection. International journal of computer vision, 57(2), 137-154, (2004).
13. Martin, K.: "Efficient Metric Learning for Real-World Face Recognition", http://lrs.icg.tugraz.at/pubs/koestinger_phd_13.pdf
14. Xiong, X., De la Torre, F.: Supervised descent method and its applications to face alignment. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 532-539, (2013).
15. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1), 51-59, 1996.
16. Verma, M., Raman, B.: Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. Multimedia Tools and Applications, 1-24, 2017.
17. Huang, G. B., Zhou, H., Ding, X., Zhang, R.: Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2), 513-529, (2012).
18. Huang, Z., Yu, Y., Gu, J., & Liu, H.: An efficient method for traffic sign recognition based on extreme learning machine. IEEE transactions on cybernetics, 47(4), 920-933, (2017).
19. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94-101, (2010).
20. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D.: Presentation and validation of the Radboud Faces Database. Cognition and emotion, 24(8), 1377-1388, (2010).
21. Carcagnì, P., Coco, M., Leo, M., & Distante, C.: Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus, 4(1), 645, (2015).
22. Ahmed, F., & Kabir, M. H.: Directional ternary pattern (dtp) for facial expression recognition. In IEEE International Conference on Consumer Electronics (ICCE), pp. 265-266, 2012.