Facial Expression Recognition using Improved Adaptive Local Ternary Pattern

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

[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, there has been a huge demand for assistive technology for industrial, commercial, automobile and societal applications. In such applications, there is a huge requirement of an efficient and accurate system for automatic facial expression recognition (FER). Therefore, FER has gained enormous interest among the computer vision researchers. Although there has been a plethora of work available in the literature, automatic FER system has not yet reached the desired level of robustness and performance. In most of these works, there has been the dominance of 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 Improved Adaptive Local Ternary Pattern (IALTP) for automatic FER. This new descriptor is an improved version of ALTP which have been proved to be effective in face recognition. We have investigated ALTP in more details and have proposed some improvements like the use of uniform patterns and dimensionality reduction via principal component analysis (PCA) are proposed. 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 three different FER datasets. The performance has been observed using well-known evaluation measures such as accuracy, precision, recall, and F1-Score. We have compared our proposed approach with some of the state-of-the-art works in literature and found it to be more accurate and efficient.

Keywords: Facial Expression Recognition (FER), Adaptive Local Ternary Pattern (ALTP), Improved Adaptive Local Ternary Pattern (IALTP), Principal Component Analysis (PCA), and Kernel Extreme Learning Machine (K-ELM).

1 Introduction

Recently, there has been a huge demand for assistive technology for industrial, commercial, automobile and societal applications. An efficient and accurate automatic facial expression recognition (FER) system is often desired in bringing up such a system to reality. This is because facial expression provides an important cue which reveals the actual intention and state of mind of a person.

The techniques available in the literature for automatic FER can be broadly classified into two main categories: geometric-based methods and appearance-based methods as discussed in [1, 2]. Since our proposed scheme comes under the category of appearance-based methods, a brief discussion on appearance-based approaches for FER has been discussed below.

Popular 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 patter and so on. Deployment of LBP for FER for the first time has been reported in [3] wherein 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. To overcome this limitation different new techniques have been developed as well as different modification of the original LBP has been done over the time. One such modification is Sobel-LBP [4]. 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 [5] 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) [6] 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) [7]. 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 [8]. 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. Another recent method for FER which has been motivated by GLTP is improved gradient local ternary patterns (IGLTP) proposed by the authors in [9]. The improvements over GLTP includes the use of enhanced pre-processing, a more accurate Scharr gradient operator, dimensionality reduction via 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 and called Elongated Quinary Pattern (EQP) [10] 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 fol-

lowed 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 automatic FER consists of a sequence of steps which involves: Face detection and registration, feature extraction, feature dimensionality reduction and classification. All of these steps have been discussed briefly below.

2.1 Face Detection and Registration

Face detection and registration step comprise of a face and landmark detection unit. The face detector takes an input image and provides the location of human faces and for this, we have used Viola and Jones frontal face detector [11] with cascade classifier trained using Multi-block local binary pattern (MB-LBP) features. The trained cascade classifier has been made available by [12]. The detected face is passed to the facial landmark detection unit which marks the location of different landmarks on the face. We have used Intraface [13], one of the widely used landmark detector. Finally, 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 in order to obtain the registered facial image.

2.2 Facial Feature Extraction and Dimensionality Reduction

The sequence of steps involved in the proposed IALTP facial feature descriptor has been shown in Fig 2. For feature extraction, we have used a uniform version of Adaptive Local Ternary Pattern (ALTP) proposed in [14] for face recognition. Once the Uniform ALTP coded lower and upper images are obtained they are divided into different cells from which histograms are calculated and concatenated to get the final facial features. Finally, a dimensionality reduction via principal component analysis (PCA) is applied to get the reduced features. All these steps have been further discussed below. The major motivation behind using the uniform version of ALTP (called Uniform Adaptive Local Ternary Patterns) has been taken from the work of [15] wherein an ALTP pattern is called uniform if it involves at most two circular 0-1 and 1-0 transitions. For example, patterns 00111000, 11111111, 00000000, and 11011111 are uniform, and patterns 01010000, 01001110, or 10101100 are not uniform. Using uniform patterns helps in reducing the length of the feature vector (ALTP histogram) from 256-bins in original ALTP to 59-bins in uniform ALTP. This, in turn, facilitates improving the performance of classifiers both in terms of accuracy and computational burden without any loss in recognition accuracy. Furthermore, ALTP has been used in comparison to other similar descriptors like LTP, GLTP, and IGLTP because in all these descriptors the threshold t is set manually according to the pixel value range which is not an optimal choice in real applications when we face different datasets. This has been demonstrated in the figure below. Thus, in order to overcome this issue, ALTP makes use of Weber's law [16] given in (1) which states that the size of a just noticeable difference (i.e.ΔI) is a constant proportion of the original stimulus value.

$$\frac{\Delta I}{I} = k \quad (1)$$

Thus, in ALTP the threshold is automatically set using Weber's law and thus the method is called adaptive uniform local feature descriptor. The parameter k is known as Weber's parameter and is determined experimentally. The threshold is set according to (2)

$$t = G_c \times k \quad (2)$$

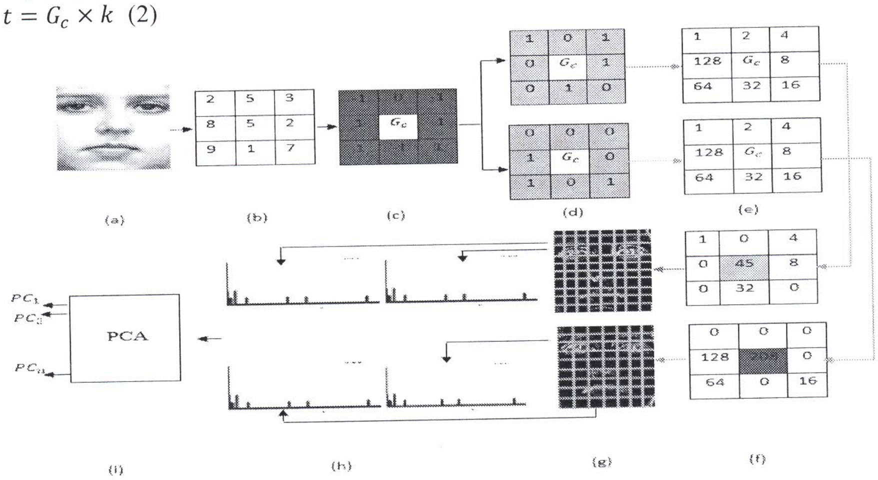

Fig. 1. Systematic representation of sequence of steps involved in the proposed IALTP

Once the Weber's parameter k illustrated in (1) has been determined and threshold is set according to (2), it is applied around a center pixel value G_c of 3×3 pixel neighborhoods throughout the input facial image as shown in Fig 1. (c). Neighbor pixels falling in between $G_c + t$ and $G_c - t$ are quantized to 0, while those below $G_c - t$ to -1 and finally those above G_c + t to 1 using (3). In (3), S_{ALTP} are the quantized value of the surrounding neighbors as shown in Fig. 1 (c).The resulting eight S_{ALTP} values for each results in a much higher number of possible patterns when compared to that of LBP therefore to reduce the dimensionality, each ALTP code is split into its positive and negative parts and treated as individual codes as shown in Fig 1 (d). The formula for converting each binary ALTP code to positive P_{ALTP} and negative N_{ALTP} decimal codes are given in (4) and (5) respectively and is shown in Fig 1. (e) and Fig. 1(f).

$$S_{ALTP}(G_c, G_i) = \begin{cases} -1, G_i < G_c - t, \\ 0, G_c - t \leq G_i \leq G_c + t, \\ +1, G_i > G_c + t. \end{cases} \quad (3)$$

$$P_{ALTP} = \sum_{i=0}^{7} S_P\left(S_{ALTP}(i)\right) \times 2^i, \quad (4)$$

$$S_P(v) = \begin{cases} 1, if\ v > 0, \\ 0, otherwise \end{cases}$$

$$N_{ALTP} = \sum_{i=0}^{7} S_N\left(S_{ALTP}(i)\right) \times 2^i \quad (5)$$

$$S_N(v) = \begin{cases} 1, if\ v < 0, \\ 0, otherwise \end{cases}$$

Once the positive P_{ALTP} and negative N_{ALTP} decimal coded images are available, these are converted into their respective uniform version using a look-up table. In our case using 8 sampling points, the number of different output labels for mapping for patterns of 8 bits is 59 wherein out of 256 patterns 58 uniform patterns are given output labels from 0-58 and the rest of the non-uniform patterns are grouped with a single output label 59. Finally, the uniform positive and negative ALTP coded image is divided into m x n regions as shown in Fig 1. (g). A positive ($H_{P_{IALTP}}$) and negative ($H_{N_{IALTP}}$) IALTP histogram is computed for each region using (6) and (7).

$$H_{P_{IALTP}}(\tau) = \sum_{r=1}^{M} \sum_{c=1}^{N} f(P_{ALTP}(r,c),\tau), \quad (6)$$

$$H_{N_{IALTP}}(\tau) = \sum_{r=1}^{M} \sum_{c=1}^{N} f(N_{ALTP}(r,c),\tau) \quad (7)$$

$$f(a,\tau) = \begin{cases} 1 & a = \tau \\ 0 & otherwise \end{cases}$$

In (6), M and N are the width and height of the uniform ALTP coded image whereas r and c denotes dimension of the encoded image. The value of τ ranges from 0-58 (as compared to 0-255 in case of ALTP) and for which the frequency of occurrence is calculated using the above listed equations. Finally, the positive and negative histograms for each region are concatenated together to form the feature vector as in Fig 1. (h). Since the size of the feature vector obtained is quite high in dimension, therefore we have proposed another improvement over traditional ALTP wherein we have make the use of dimensionality reduction technique namely PCA as shown in Fig 1. (i) to reduce the dimension of the feature vector.

2.3 Kernel Extreme Learning Machine (K-ELM) Classifier

In order to classify the facial attributes into their corresponding emotion label, we have used kernel extreme learning machine (K-ELM) multi-class classifier in our proposed work. K-ELM is a popular classifier and is usually treated as the kernelized variant of extreme learning machine (ELM) classifier used for fast training a Single Layer Feed-Forward Neural Network (SLFN) [17]. The major benefit of using K-ELM as compared to the traditional backpropagation algorithm based neural network architecture is that the training using K-ELM does not involve any iteration and the output weights are calculated using a direct solution. An ELM classifier invariably makes use of a feature mapping function h(x) which helps in learning non-linearity.

However, if the mapping function is not known, kernel technique can be applied into ELM based on Mercers' condition. For more details about K-ELM and necessary mathematical formulations, we refer to work available in [18]. The output vector f(x) of a kernel ELM can be represented as shown in (11).

$$f(x) = h(x)\beta = h(x)H^T\left(\frac{I}{C} + HH^T\right)^{-1}T \quad = \begin{bmatrix} \phi(x, x_1) \\ \cdot \\ \cdot \\ \phi(x, x_{N_K}) \end{bmatrix}\left(\frac{I}{C} + \Phi\right)^{-1}T \qquad (8)$$

where

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

In this paper, Gaussian function is used as the kernel ϕ which is represented as in (9).

$$\phi(x, x_1) = exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (9)$$

3 Experimental Results and Discussion

In this section, we discuss our experimental setup and various experiments which were performed on different FER datasets. All the experiments have been performed using Matlab 2015a running on a windows platform with 64 GB RAM.

3.1 Datasets

In order to validate the performance of our proposed approach, we have used three FER 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 set-up, 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 expression (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. The third dataset with which is the Japanese female facial Expression (JAFFE) dataset [21]. The dataset contains 7 different prototypic facial expression images of 10 female subject with a total of 213 images.

3.2 Parameter Selection

A number of parameters are usually involved in the design of an efficient FER system. Therefore, an optimal value of these parameters is often desired to achieve good recognition accuracy. Thus, to determine the optimal values of different parameters, we have conducted a series of experiments. Firstly, we tried to determine the optimal facial image size and cell-size and for this, we experimented with two different facial image resolution of size 65 x 59 as in [9] and 147 x 108 pixels [22]. We experimented with 8 different cell sizes in which the facial image is divided. The experiments were performed on CK + 7 expression dataset with 10-fold cross validation strategy which is repeated 10-times using K-ELM classifier with regularization parameter C and kernel parameter γ value of 100 and 200 respectively. The value of Weber's parameter k fixed in the experiment is 0.12. Table 1 and table 2 depicts the results of the experiment for 65 x 59 and 147 x 108 resolution facial images respectively. Based on the experimental results, we found that the facial image with a resolution of 147 x 108 and cell size of 9 x 8 performed well compared to other possible combinations of image and cell size and therefore this value of facial image size and cell size has been used in all our further experiments.

Table 1. Performance of different cell-size on 65 x 59 pixels facial image

[65, 59]	[5, 4]	[6, 5]	[7, 6]	[8, 7]	[9, 8]	[10, 9]	[11, 10]	[12, 11]
Avg. Acc. 10 runs	99.1±0.2	99.2±0.2	98.9±0.2	99.0±0.2	98.5±0.3	97.6±0.4	96.5±0.5	96.7±0.1
Feature Dim.	19824	12980	9558	6608	5782	4248	2950	2950
Avg. Acc.	99.35	99.35	99.35	99.19	99.03	98.30	97.17	97.01
Avg. Prec.	99.29	99.27	99.33	99.12	98.91	98.66	97.88	97.57
Avg. Rec.	99.34	99.34	99.33	99.23	99.08	97.86	96.10	96.09
Avg. F1-S	99.32	99.30	99.33	99.17	98.99	98.24	96.94	96.76

Table 2. Performance of different cell-size on 147 x 108 pixels facial image

[147, 108]	[7, 6]	[8, 7]	[9, 8]	[10, 9]	[11, 10]	[12, 11]	[13, 12]	[14, 13]
Avg. Acc. 10 runs	99.1±0.2	99.1±0.2	**99.3±0.2**	98.9±0.2	99.2±0.1	99.0±0.2	98.6±0.2	98.7±0.2
Feature Dim.	40120	31860	**24544**	18172	15340	12744	10384	9440
Avg. Acc.	99.35	99.35	**99.51**	99.11	99.27	99.27	98.95	98.95
Avg. Prec.	99.30	99.33	**99.46**	99.04	99.26	99.15	99.09	98.75
Avg. Rec.	99.46	99.45	**99.70**	99.31	99.28	99.38	98.98	99.05
Avg. F1-S	99.38	99.39	**99.58**	99.17	99.27	99.26	99.03	98.89

In our second experiment, we tried to determine the optimal value of the Weber's parameter k for every dataset. The results of the experiments have been tabulated in table 3- table 5. Our third experiment dealt with the determination of the K-ELM parameters i.e. the optimal value of regularization coefficient C and kernel parameter γ. For this experiment, we used the CK+ 7 expression dataset with fixed image size, cell size and k obtained in our earlier experiment. The range of both C and γ have been taken here is [1:10] in logarithmic scale of base 2. The result of the experiment has been tabulated in table 6.

Table 3. Performance of different k on CK+ 7 Expression Database

K	0.01	0.03	0.06	0.09	0.12	0.15	0.18	0.21
Avg. Acc.	**99.35**	99.09	99.20	99.21	99.31	98.97	98.77	98.59

Table 4. Performance of different k on JAFFE 7 Expression Database

K	0.01	0.03	0.06	0.09	0.12	0.15	0.18	0.21
Avg. Acc.	**95.53**	94.97	95.44	94.52	92.86	93.32	92.71	91.45

Table 5. Performance of different k on RFD 7 Expression Database

K	0.01	0.03	0.06	0.09	0.12	0.15	0.18	0.21
Avg. Acc.	96.43	**97.42**	96.47	95.43	95.15	94.36	93.46	93.39

In our next experiment, we tried to determine the optimal number of principal components for all the three FER datasets. To determine this, we fixed the value of all other parameters involved to the optimal value determined earlier. The experimental result for CK+ 7 expression dataset has been tabulated in table 7. Similar experiments were also carried out for JAFFE and RFD dataset but we have not shown the results due to limited paper length. From the experiments, we found that the optimal number of principal components is 224 for CK+, 96 for JAFFE and 256 for RFD dataset.

Table 6. Determination of K-ELM Parameter

Performance Measure/ Datasets	CK+ 7 Expressions
Avg. Accuracy 10- runs	99.47±0.08
Kernel Parameter (γ)	1024
Regularization Parameter (C)	256

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. On CK+ 6 expression the accuracy achieved using uniform ALTP is 99.90±0.14 and that using IALTP is 99.97±0.05. On CK+ 7 expression dataset the FER pipeline achieved accuracy of 99.43±0.2 and 99.49±0.14 using uniform ALTP and IALTP respectively. Table 8 and table 9 depicts the result of our experiment using different performance measure corresponding to the best 10-fold cross-validation for CK+ 6 and CK+ 7.

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

No. of PCA	32	64	96	128	160	192	224	256

Avg. Acc. 10 runs	98.1±0.2	98.6±0.2	99.2±0.2	99.4±0.1	99.4±0.2	99.5±0.1	**99.5±0.1**	99.4±0.1
Avg. Acc.	98.30	98.87	99.43	99.60	99.60	99.60	**99.68**	99.51
Avg. Prec.	98.50	99.04	99.42	99.53	99.59	99.59	**99.63**	99.46
Avg. Rec.	98.30	99.03	99.64	99.75	99.74	99.71	**99.79**	99.70
Avg. F1-S	98.40	99.03	99.53	99.64	99.66	99.65	**99.71**	99.58

Table 8. Performance of IALTP (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

Table 9. Performance of IALTP (CK+ 7 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Ne	Sa	Su	Recall
An	135	0	0	0	0	0	0	100
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	100
Ne	1	0	0	0	307	1	0	99.35
Sa	0	0	0	0	0	84	0	100
Su	0	0	0	0	2	0	247	99.20
Precision	99.26	100	100	100	99.35	98.82	100	
F1-Score	99.63	100	100	100	99.35	99.41	99.60	

Avg. Performance: recall =99.79, precision =99.63, accuracy =99.68, F1-Score = 99.71

3.4 Results on Jaffe Dataset

In order to determine the performance of the proposed FER pipeline on JAFFE we again performed the experiment using the similar setup as above. On JAFFE 6 expression the accuracy achieved using IALTP is 95.77±0.95 and that using IALTP+PCA is 95.93±0.80. On JAFFE 7 expression dataset the FER pipeline achieved accuracy of 95.53±0.68 and 95.92±0.98 using IALTP and IALTP+PCA respectively. The performance of the best 10-fold cross-validation in terms of different measures has been shown in table 10 and table 11.

Table 10. Performance of IALTP (JAFFE 6 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Sa	Su	Recall
An	30	0	0	0	0	0	100

Di	0	28	0	0	1	0	96.55
Fe	0	0	31	0	0	1	96.88
Ha	0	0	0	31	0	0	100
Sa	0	0	1	1	29	0	93.55
Su	0	0	0	1	0	29	96.67
Precision	100	100	96.88	93.94	96.67	96.67	
F1-Score	100	98.25	96.88	96.88	95.08	96.67	

Avg. Performance: recall =97.27, precision =97.36, accuracy =97.27, F1-Score = 97.29

Table 11. Performance of IALTP (JAFFE 7 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Ne	Sa	Su	Recall
An	30	0	0	0	0	0	0	100
Di	0	28	0	0	0	1	0	96.55
Fe	0	0	31	0	0	0	1	96.88
Ha	0	0	0	31	0	0	0	100
Ne	0	0	0	0	30	0	0	100
Sa	0	0	1	1	0	29	0	93.55
Su	0	0	0	1	0	0	29	96.67
Precision	100	100	96.88	93.94	100	96.67	96.67	
F1-Score	100	98.25	96.88	96.88	100	95.08	96.67	

Avg. Performance: recall =97.66, precision =97.74, accuracy =97.65, F1-Score = 97.68

3.5 Results on RFD Dataset

Performance of the proposed FER pipeline on RFD dataset again in terms of different performance measures corresponding to best 10-fold cross-validation run has been tabulated in table 12. The overall average accuracy of the 10 runs of the 10-Cross validation using uniform ALTP and IALTP is 97.4±0.4 and 97.8±0.4 respectively.

Table 12. Performance of IALTP (RFD 7 Expressions)

Actual/Predicted	An	Di	Fe	Ha	Ne	Sa	Su	Recall
An	66	0	0	0	0	1	0	98.51
Di	0	67	0	0	0	0	0	100
Fe	0	0	67	0	0	0	0	100
Ha	0	0	0	67	0	0	0	100
Ne	0	0	0	0	67	0	0	100
Sa	1	0	1	0	4	61	0	91.04
Su	0	0	0	0	0	0	67	100
Precision	98.51	100	98.53	100	94.37	98.39	100	
F1-Score	98.51	100	99.26	100	97.10	94.57	100	

Avg. Performance: recall =98.51, precision =98.54, accuracy =98.51, F1-Score = 98.49

Comparison result of the proposed FER framework with other state-of-the-art work available in literature has been shown in table 13. From the table we find that our proposed approach has achieved superior performance and is more accurate and effective.

Table 13. Comparison of recognition accuracy (%) on CK+ and RFD Dataset

Method	CK+ 6 Expression	CK+ 7 Expression	RFD 7 Expression	JAFFE 6 Expression	JAFFE 7 Expression
LBP [23]	90.1	83.3	---	---	NA
LDP[23]	93.7	88.4	---	---	NA
LTP[23]	93.6	88.9	---	---	NA
GLTP[6]	97.2	91.7	---	77.0	74.4
Improved GLTP[9]	99.3	97.6	---	83.3	81.7
HOG[22]	95.8	94.3	94.9	---	---
Uniform ALTP	**99.9**	**99.4**	**97.4**	**95.8**	**95.5**
IALTP Proposed	**100**	**99.5**	**97.8**	**95.9**	**95.9**

4 Conclusion

In the presented paper, we have proposed a new facial feature descriptor named improved adaptive local ternary pattern (IALTP). The proposed descriptor is a modified version of ALTP wherein we have investigated ALTP from FER perspective and proposed some improvements like the use of uniform ALTP patterns and dimensionality reduction via principal component analysis. We have used K-ELM classifier for classifying the facial expressions. The performance of the proposed approach has been validated using 10-fold cross validation which has been repeated 10 times for a fair comparison with the existing works. The experiments have been performed on three FER datasets viz. CK+, JAFFE, and RFD and performance have been observed using precision, recall, accuracy, and f1-score measures. 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 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. Zhao, S., Gao, Y., Zhang, B.: Sobel-lbp. In 15th IEEE International Conference on Image Processing, pp. 2144-2147, (2008).

5. 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).
6. Ahmed, F., & Hossain, E.: Automated facial expression recognition using gradient-based ternary texture patterns. Chinese Journal of Engineering, 2013.
7. 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.
8. Alhussein, M.: Automatic facial emotion recognition using weber local descriptor for e-Healthcare system. Cluster Computing, 19(1), 99-108, (2016).
9. 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).
10. 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).
11. Viola, P., Jones, M. J.: Robust real-time face detection. International journal of computer vision, 57(2), 137-154, (2004).
12. Martin, K.: "Efficient Metric Learning for Real-World Face Recognition", http://lrs.icg.tugraz.at/pubs/koestinger_phd_13.pdf
13. 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).
14. Yang, W., Wang, Z., & Zhang, B.: Face recognition using adaptive local ternary patterns method. Neurocomputing, 213, 183-190, (2016).
15. Lahdenoja, O., Poikonen, J., Laiho, M.: Towards understanding the formation of uniform local binary patterns. ISRN Machine Vision, (2013).
16. Jain, A. K.: Fundamentals of digital signal processing, (1989).
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. Lyons, M.J., Akamatsu, S., Kamachi, M., Gyoba, J. and Budynek, J.: The Japanese female facial expression (JAFFE) database. In Proceedings of third international conference on automatic face and gesture recognition, pp. 14-16, 1998.
22. Carcagnì, P., Coco, M., Leo, M., & Distante, C.: Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus, 4(1), 645, (2015).
23. 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.