Multimedia Tools and Applications (2021) 80:10187–10212
https://doi.org/10.1007/s11042-020-10105-2

An efficient automatic facial expression recognition using local neighborhood feature fusion

P. Shanthi[1] · **S. Nickolas**[1]

Received: 3 July 2019 / Revised: 14 August 2020 / Accepted: 19 October 2020 /
Published online: 17 November 2020

Abstract
In computer vision, several feature extraction methods have been developed to differentiate the variations of facial expressions. But the effect of the relationship among the neighboring pixel is not considered in the existing texture encoding based method. This paper exploits the method to analyze the association among the adjacent pixels using feature fusion technique. For efficient texture representation, the proposed approach combines the Local Binary Pattern (LBP) with the Local Neighborhood Encoded Pattern (LNEP). The LBP feature encodes the relationship of adjacent pixels with respect to the central pixel whereas LNEP represents the relationship among the two closest local neighboring pixels of the current pixel. After concatenating LBP with LNEP, the most relevant features are selected using chi-square statistical analysis and classified using multiclass Support Vector Machine (SVM). Experimental findings show that the proposed hybrid feature performed better than an individual feature and it achieves an average recognition accuracy of 97.86% and 97.11% on CK+ and MMI dataset, respectively. The effectiveness of the reduced hybrid feature is also evaluated under a noisy environment and the results show better performance in such conditions.

Keywords Emotion · Facial expression · Local binary pattern · Local neighborhood encoded pattern · Feature fusion · Feature selection · Multiclass support vector machine

1 Introduction

Facial expression is one of the outward sign that directly reflect the inner emotional state of a person. The effectiveness of visual data in computer vision is the primary motivation for automatic facial expression analysis. As its requirement and value increase in many applications, Facial Expression Recognition (FER) has received significant interest among

✉ P. Shanthi
shanthianu81@gmail.com

S. Nickolas
nickolas@nitt.edu

[1] Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India

researchers from various fields (e.g., psychology, behavioral science, pattern recognition, etc.). Computer vision based FER approaches, use the Facial Action Coding System (FACS) as the base for analyzing the facial muscle movements which represent those movements as 46 Action Units (AUs) [12]. In general, FER systems consist of three essential parts, namely pre-processing, feature extraction, and classification. Researchers proposed various algorithms for those three phases of FER to improve classification accuracy.

The feature extraction is an important stage in the whole process of FER. Features represent the quantifiable property of the raw image to simplify the process of pattern detection, classification, or recognition. Even the best classifier will fail to achieve accurate recognition if the features are insignificant. Most of the existing methods generally use a predefined model as the primary source of feature extraction [17]. In the model-based approach, a statistical facial model such as Active Appearance Model (AAM) [11] and Active Shape Model (ASM) [39], is used to extract required feature for recognition after model fitting. The main issue of this approach is difficulty in establishing the generalized statistical model suitable for all facial configurations, and also manual intervention is required for model fitting.

According to feature-oriented methods, geometric or appearance based features are extracted from the entire face [19] or the individual facial components [14, 24] such as eye, eyebrow, nose, and mouth. The geometric approach considers the location and scale of the facial features, whereas the appearance-based approach considers shape and texture oriented statistical information. Both methods have significant challenge in terms of the face or facial component alignment. The geometric feature-based approach employs the snake model with fine-tuned parameters for feature point detection [27] to reduce the negative impact of the abovementioned issues. A deep learning based hybrid system is proposed in [28], which combines geometric features with Local Binary Pattern to improve the performance. But its computational cost is high since it involves different type of features.

While investigating the texture of the facial region, local neighborhood pixel relationships should be interpreted cautiously for effective classification. Early methods focused on the statistical analysis of texture details such as Gabor wavelet transforms [52]. These methods provide good results when the test samples have a similar orientation with training samples. Another popular texture descriptor called LBP encodes the local neighborhood pixel relationship with the center pixel to detect micro structures of facial expressions quickly, with a firm texture discrimination capability. Application of this feature and its variants provide promising results in FER system. The Gray Level Co-occurrence matrix (GLCM) [20, 36] is an another statistical feature which encode the frequency of specific intensity pattern. In the existing texture descriptors, the local neighborhood intensity relationship is manipulated differently, to represent texture information. But the role of the relationship among the local neighboring pixels is omitted in the earlier study of texture based FER.

The primary and secondary facial component dependencies and the dynamic appearance distortion of the facial region can be easily understood and analyzed by encoding the image feature using local neighboring pixel relationships. Many of the FER methods extract the features based on the center and neighboring pixels relationship. This paper adopts a theoretically and computationally simple binary encoding system called Local Neighborhood Encoded Pattern (LNEP) along with conventional LBP for the better FER. This LNEP operator analyzes the association among closest neighboring pixel instead of the central pixel. Even though LNEP acquires interactive relationships between adjacent pixels themselves, LBP descriptor cannot be ignored entirely due to its robustness against monotonic illumination change. The core idea of the proposed work is to create two different patterns by

handling the 3×3 pixel cell at once in terms of LNEP and LBP. The two histograms of LNEP and LBP are merged to obtain the final feature vector. From the high dimensional feature space, most contributing features are selected based on chi-square statistical analysis. The reduced feature set supports faster machine learning by reducing the system complexity and improves the overall accuracy by avoiding over fitting. Final optimal hybrid feature effectiveness is evaluated using Multiclass SVM. The robustness of the features against noisy data is also analyzed, and experimental findings show the hybrid feature effectiveness in a noisy environment. The following key points outline the main contribution of the proposed work,

- A multiple local feature fusion based method is proposed to improve the recognition accuracy of FER.
- The two features involved in the feature fusion are computationally simple.
- Experimental findings on two different datasets indicate that the hybrid feature significantly improves the recognition rate with feature selection.
- This combination of LBP and LNEP encourages the awareness of local features as a single unit, which complements each other thereby creating a more robust feature vector which is suitable to handle the noisy images.

The rest of this article is divided into four sections. The background theory and related works are briefly discussed in Section 2, and the proposed approach is described in Section 3. Section 4 reports the experimental findings on the CK+ and MMI datasets. Finally, the conclusion and further research scope is given in Section 5.

2 Related work

For consistent performance, a robust, accurate, and stable feature representation of the facial image is a critical element for FER based applications. Recently various texture features are applied to represent the feature for facial expression recognition. LBP is one of the competent texture features in the field of facial image processing. In 1996, Ojala et al., [34] introduced the conventional LBP operator where the center pixel of a 3×3 cell is used as a threshold to the neighboring pixels, and the local texture is represented as an 8-bit binary code. The compact version of the original LBP called uniform LBP takes into account only the binary patterns with maximum of two transitions from 0 to 1 and vice versa [43]. Its rotational invariant pattern was proposed in [35] to minimize the effect of the rotational pattern using uniform LBP.

In the region-based method, features are extracted from the equally sub-divided regions and the resultant histograms are combined over local regions to achieve robustness against translation. In [18], the informative facial region analyzed using Weighted Projection based LBP (WPLBP), is proposed and the weight factor is added based on the importance of the region in FER. Extended LBP based FER introduced in [13] generates uniform LBP code by extending the number of neighboring pixel (P) located at the distance from center pixel (R) and then combine it with the covariance matrix transform of Karhunen- Loeve Transform (KLT). The horizontal and diagonal local gradient coding scheme was introduced in [42], and it utilizes the gray pixel level relationship between the neighboring pixels. But this sub-region processing technique did not precisely represent the expression uniqueness when the size of region is very small.

Instead of using a single feature, the discriminative capability of expressive face muscle movements that are changed both locally and globally can be increased using a combination of more than one feature. In [25], global and local features are extracted using Principal

Component Analysis (PCA) and LBP respectively, and the low recognition rate is realized for anger expression due to the absence of upper face for local feature extraction. Many real-time applications have implemented the fusion of 2D- Gabor, and LBP features [22, 50]. Feature fusion using Gabor and LBP based ensemble classification algorithm proposed in [51], requires additional computational effort to extract two sets of features for training and selecting the classifier. In [5], LBP based signature feature is obtained from the salient regions. These regions are identified using AAM, which takes more time to fit the statistical model over the given face. Its extended version builds the hybrid system which combines texture and distance signature feature for emotion recognition, where distance feature is calculated from the grid formulated with landmarks identified using AAM. These features give better performance when both features are used together. In [48], after efficient pre-processing, the LBP operator is applied for feature extraction, and classification is performed using Kullback-Leibler divergence.

Another study uses the subpattern of Compound LBP [1], by concatenating the 16-bit sign and magnitude code. In [3, 32], the role of a merged binary pattern is investigated using the 16-bit code mean gradiant generated based on LBP. Their study proves that the holistic feature extraction method is better than a division based feature extraction approach. This work is further extended to handle the illumination effect on the real-time data and confirms the importance of texture feature in FER [13]. Some of the recent appearance-based FER methods include the use of Gabor wavelet transform (GWT) [2], Local Directional Patterns (LDP) [16, 29], and Gradient Local Ternary Patterns (GLTP) [15]. Like LBP, GLTP generates three-level encoding based on the threshold. Recently, Gabor wavelet filter extracts multi-scale and multi-orientation appearance feature from the given image, and multiclass classification is performed using Multiple Kernel Learning Decision Tree Weighted Kernel Alignment [2].

Similarly, HOG and HOG-TOP are the other popular texture descriptors utilized in FER [8, 37]. Encoding the edge responses in a different direction is the characteristic element of LDP. But the existence of intensity distortions or noise in the selected region may lead to erroneous LDP patterns. Also, the issue of appearance-based methods of FER is that feature dimension tends to be very large, and some redundant features can degrade both the efficiency and precision of classification algorithm. PCA has been broadly used as a feature extraction technique for face processing [15] and recently it has also been used as a dimension reduction technique [25, 26, 38]. Apart from handcrafted features, deep learning approaches performed better in FER and became more popular in the computer vision community [6, 9, 41, 44]. But it requires extensive training data to achieve optimal accuracy. A deep learning based hybrid system was proposed in [28], to improve the performance by combining geometric features with LBP. Due to the unavailability of sufficient data in the current expression dataset, most of the deep learning methods have employed image augmentation at the cost of high computational power [7, 21, 33]. This kind of artificially created data is not required here to prove the strength of the proposed descriptor, and hence such deep model approaches are considered to be out of the scope in the proposed work.

Most of the appearance-based approaches represent useful features by manipulating neighborhood pixel relationship, especially with a central pixel. Here, a new, dominant local descriptor called LNEP is employed along with LBP, to demonstrate the significance of the correlation among nearest neighboring pixels along with the central pixel relation and this combined feature has already proved its effectiveness in content-based image retrieval [46]. Like LBP, this new feature generates 8-bit binary code by encoding the relationship of two closest

neighboring pixels with the current neighboring pixel. Experimental results on two datasets endorse the effectiveness of the proposed hybrid feature in FER with and without noise.

3 Proposed methodology

The proposed FER framework includes the following stages: preparing images for subsequent stage, feature extraction, feature space reduction using feature selection, and finally classification of identified patterns. The classification accuracy depends on almost every stage of the FER system. Figure 1. illustrate the stages related in the proposed method.

3.1 Preprocessing

Initially, the face is isolated from background using Harr cascade feature based object detection algorithm [47, 49]. The detected facial images are dissimilar in dimensions. As reported in [25], all images are rescaled to 120 x 120 pixel. The images in all data sets contain many variations concerning illumination. It will affect the efficiency of the subsequent steps. So, the dynamic stretch limits GL'_{max} and GL'_{min} are identified to adjust the illumination and pixel intensity variations where GL_{max} and GL_{min} are the standard stretch limits. The given image brightness is regulated using (1).

$$I[m,n] = \begin{cases} \frac{GL'_{max}-GL'_{min}}{GL_{max}-GL_{min}}[I(x,y)-GL_{min}]+GL'_{min} & if\, GL_{min} <= I[m,n] >= GL_{max} \\ GL_{min} & if\, I[m,n] <= GL_{min} \\ GL_{max} & if\, I[m,n] >= GL_{max} \end{cases} \quad (1)$$

This contrast normalization step enhances the appearance and perception of an image. From the enhanced image two types of local appearance features are extracted.

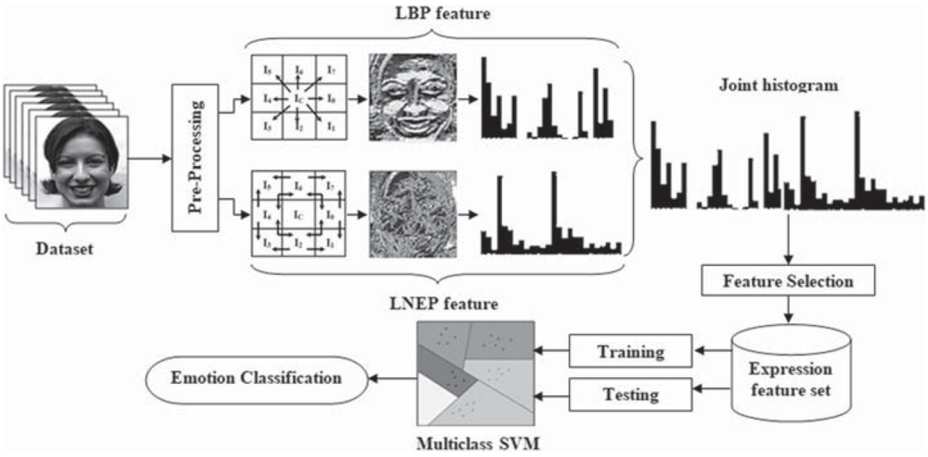


Fig. 1 Design of proposed scheme

3.2 Feature extraction

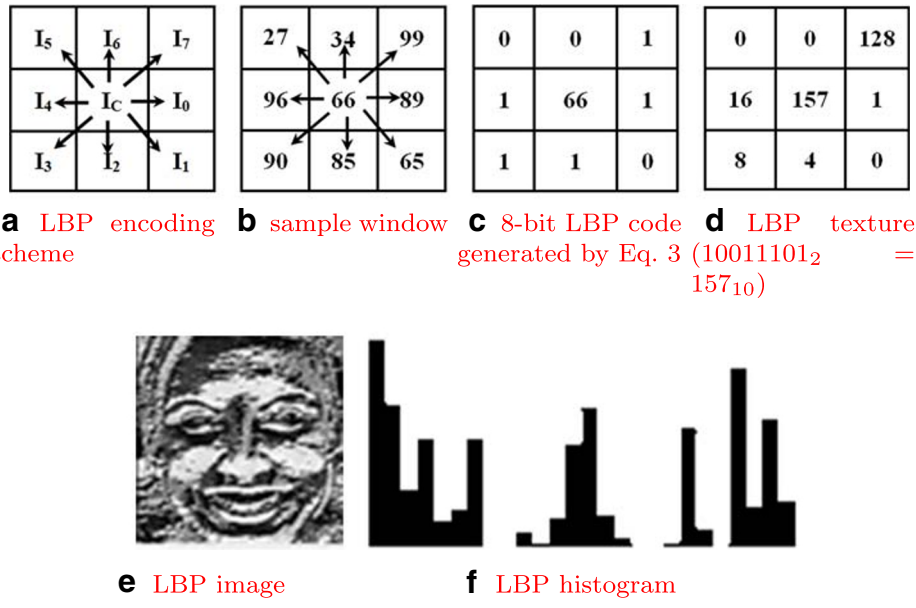
3.2.1 Feature extraction with LBP

The LBP is one of the extensively used texture descriptors to symbolize local spatial information by encoding neighborhood pixel values into binary, using the central pixel as a threshold. For that, the two parameters, namely R and P define a circular symmetric neighborhood structure, where, R is the distance at which neighborhood pixels are located, and P represents the number of neighborhood pixels. Here, 3×3 pixel cell(P = 8 and R = 1) is considered to encode the neighborhood pixels (I_i) relationship with the central pixel (I_c). After thresholding using (3), the resulting sequence of 0s and 1s are multiplied by position weight from I_0 to $I_7(2^0, 2^1, 2^2, ..., 2^7)$. The sum of the product of its pixel weight replaces the central pixel using (2). It is mathematically formulated as follows,

$$LBP_{P,R}(I_c) = \sum_{i=0}^{P-1} s(I_i - I_c)2^i \tag{2}$$

$$s(I_i, I_c) = \begin{cases} 1 & if\, I_i >= I_c \\ 0 & otherwise \end{cases} \tag{3}$$

The LBP code dimension will be increased significantly when the number of neighboring pixel (P) is increased, and it will create a problem in model identification at the classification stage. According to [35], a subset of the 2^p patterns from conventional LBP is adequate to represent given image texture. This subset of pattern is said to be a uniform LBP pattern if it holds almost two transitions, 0 to 1 or 1 to 0 [43]. The uniform LBP encode the presence of micropatterns with the feature size of 59. Therefore, it can successfully decrease the



a LBP encoding scheme **b** sample window **c** 8-bit LBP code generated by Eq. 3 **d** LBP texture ($10011101_2 = 157_{10}$)

e LBP image **f** LBP histogram

Fig. 2 Computation of LBP code

length of conventional LBP code. But, the facial component region based feature description increase the feature length as all regional features are concatenated to formulate the feature vector. This region based feature extraction did not exactly reflect the texture variations when the size of divided regions are very small. Here, shape information of the facial component is preserved by handling entire face for feature extraction and the Fig. 2 show the steps involved in LBP feature extraction.

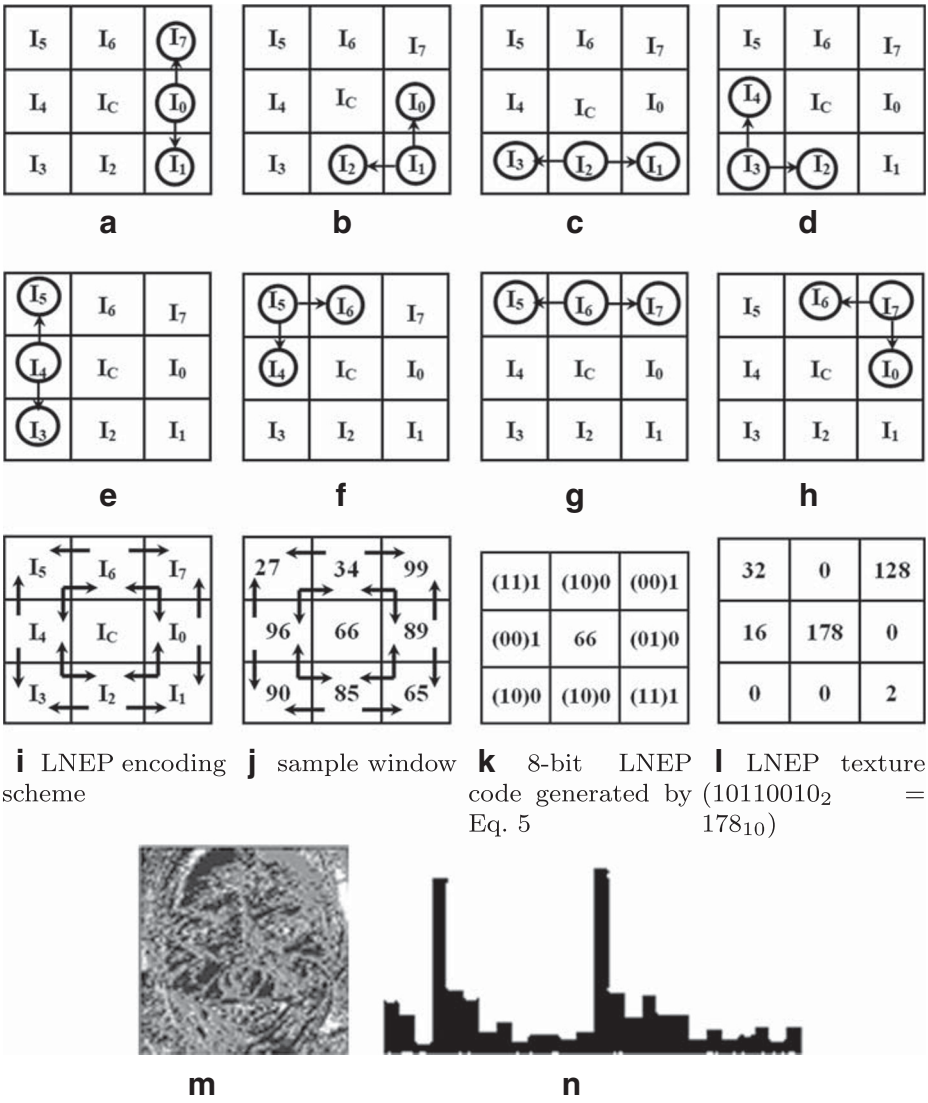


Fig. 3 Computation of LNEP (a-h)Two closest neighboring pixel for the current pixel $I_i (i = 0, 1, .., 7)$ (i-l)LNEP code calculation (m)LNEP image (n)histogram of LNEP image

3.2.2 Feature extraction with LNEP

In uniform LBP, 20% of the texture details are lost, which leads to higher false expression recognition. Generally, the conventional LBP operator considers only the relationship of adjacent pixels with the central pixel, and its derived local patterns encode the neighboring pixel relationship in a different scale and orientation. In many of the present approaches, the adjacent pixel association with the central pixel is encoded in many ways. The current work focuses on the new encoding algorithm for adjacent pixels along with the conventional LBP to take advantage of both operators in FER. The main motivation of the proposed feature descriptor is to represent the facial local sub-structure texture information using a different conjoint relationship of neighboring pixels. Interactions among adjacent pixels in the selected window are as significant as the association between the center pixel and adjacent pixels. For a given pixel cell, the nearest neighboring pixels of current pixel excluding the center pixel is considered as one of the local patterns in the proposed hybrid feature which provides more interrelated information.

Each pixel value is equated against the two nearby pixels which are located either vertical and/or horizontal to the current pixel. These two-pixel positions are represented as $I_{(i+1)modP}$ and $I_{(P+i-1)modP}$ where i is the location of the current pixel. For example, $I_{(5+1)mod8} = I_6$ and $I_{(8+5-1)mod8} = I_4$ are the two adjacent pixels for I_5. After equating the two adjacent pixels with the current pixel using (4) and (5), a logical operation is performed by (6), where $I_i = 1$ iff the value of $I_{(i+1)modP}$ and $I_{(P+i-1)modP}$ are same after comparison otherwise$I_i = 0$. After binarization, its equivalent decimal is obtained by the sum of the product of the weight assigned to the position. Finally, the statistical histogram represents the LNEP feature for the given image in the frequency domain. Computation of LNEP for the selected window is mathematically expressed as follows,

$$LNEP_{P,R}(I_c, I_i) = \sum_{i=0}^{P-1} (s(I_{(i+1)modP}, I_i) \odot s(I_{(P+i-1)modP}, I_i))2^i \tag{4}$$

$$s(I_{(i+1)modP}, I_i) \odot s(I_{(P+i-1)modP}, I_i) = \begin{cases} 1 & if\, s(I_{(i+1)modP}, I_i) == s(I_{(P+i-1)modP}, I_i) \\ 0 & otherwise \end{cases} \tag{5}$$

$$s(I_{(i+1)modP}, I_i) = \begin{cases} 1 & if\, I_{(i+1)modP} >= I_i \\ 0 & otherwise \end{cases} \tag{6}$$

$$s(I_{(P+i-1)modP}, I_i) = \begin{cases} 1 & if\, I_{(P+i-1)modP} >= I_i \\ 0 & otherwise \end{cases} \tag{7}$$

Where $s(I_{(i+1)modP}, I_i)$ and $s(I_{(P+i-1)modP}, I_i)$ denotes the gradient relationships among the two neighboring pixels with the current neighboring pixel. The time taken to extract LNEP feature is 0.0229 seconds whereas for LBP it takes 0.0289 seconds for a single image. Figure 3 show the procedure of LNEP calculation. Algorithm 1 describe hybrid feature extraction steps.

Algorithm 1 Hybrid feature extraction.

Input: Facial image I_{WxH} where W is the width and H is the height of an image, number of adjacent pixels $P = 8$ located at Radius $R = 1$
Output: Hybrid feature vector
for every facial expression image **do**
    Isolate face from the background
    Contrast modification with (1)
    Resize the facial image
    for $w = 2$ to $W - 1$ **do**
        for $h = 2$ to $H - 1$ **do**
            $I_c = I(w, h)$
            $i = 0$
            while $i <= P - 1$ **do**
                Indentify local neighborhood $I_i = I(w \pm 1, h \pm 1)$
                if $I_i >= I_c$ **then**
                    Compute LBP code $F_{LBP} = F_{LBP} + 2^i$
                end if
                if $(I_{(i+1)modP} >= I_i \odot I_{(P+i-1)modP} >= I_i) \oplus (I_{(i+1)modP} <= I_i \odot I_{(P+i-1)modP} <= I_i)$ **then**
                    Compute LNEP code $F_{LNEP} = F_{LNEP} + 2^i$
                end if
                $i = i + 1$
            end while
        end for
        Compute hybrid feature $F_{LBP+LNEP}(I) = F_{LBP}(I) + F_{LNEP}(I)$
    end for
    Return $F_{LBP+LNEP}(I)$
end for

3.2.3 Feature vector formation

After the encoding process of LNEP and LBP using the respective pixels of an image, histogram of both features are merged using (8) and (9).

$$F_{LBP+LNEP}(I) = F_{LBP}(I) + F_{LNEP}(I) \tag{8}$$

$$F_{LBP+LNEP}(I) = \sum_{i=0}^{P-1} s(I_i - I_c)2^i + \sum_{i=0}^{P-1} (d_1 \odot d_2)2^i \tag{9}$$

The first part represents the LBP code and the second part describes the LNEP code. Figure 4. Illustrates the computation of hybrid feature using LBP and LNEP which represents texture feature in terms of two different neighborhood pixel relationships.

3.3 Feature analysis

This section analyzes the discriminative capability of the various descriptors under consideration. For this purpose, existing descriptors, such as LBP, LDP, LDN, and LNEP, are considered. As shown in Fig. 5, LBP generates the same code for the given edge and corner patch, whereas the LNEP generate a different code. Similarly, while considering edge

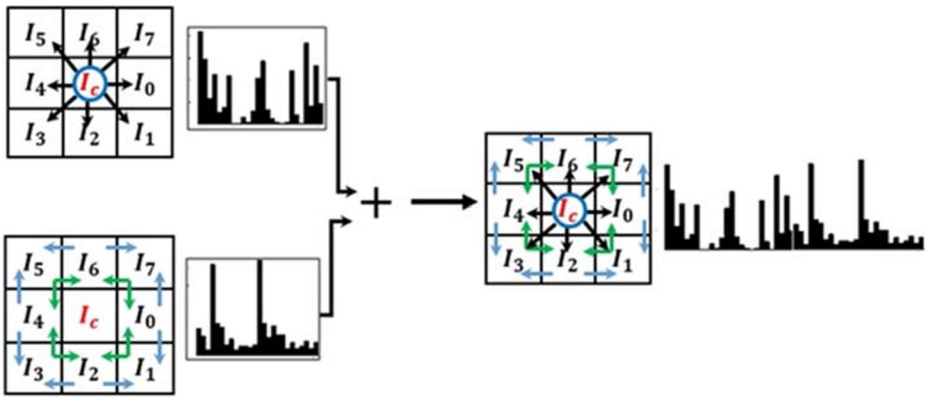


Fig. 4 Hybrid feature

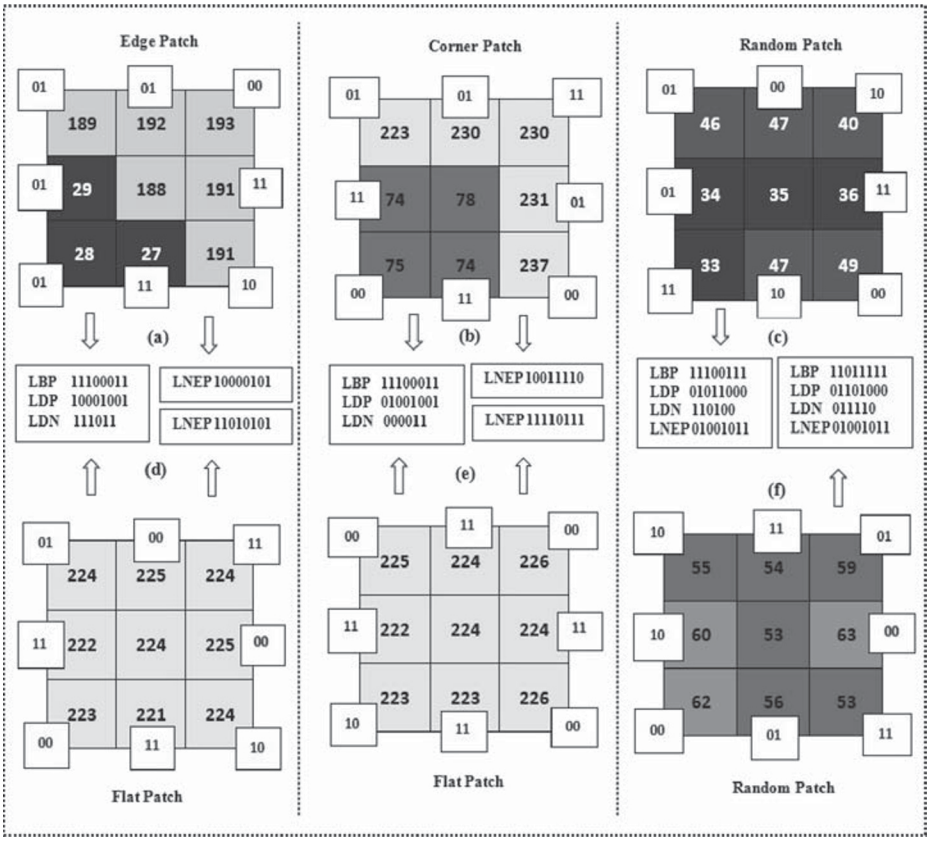


Fig. 5 Comparison of different local descriptor encoding: **a,d** edge vs flat patch; **b,e** corner vs flat patch; **c,f** random patch

and smooth or flat texture, except LNEP, other three descriptors generate the identical code and the same happens for corner and flat texture. Generating exactly similar binary pattern for completely different texture patches creates issues in the interpretation of the features which in turn reduces the classification result. However, in both cases, the LNEP generates different codes so that LNEP takes advantage of the closest neighborhood relationship to distinguish the different textures. In particular, the consideration of positioning relationship of the closest neighboring pixels to define the edges and corners, produce distinctive codes.

Also, for the given two different random patches shown in Fig. 5, LBP, LDP, and LDN generate corresponding binary code. The binary pattern generated from the random patches may disturb other histogram bins and therefore affect the feature descriptor of original expression. Nevertheless, LNEP encodes the relationship among the two closest adjacent pixel of the current pixel and similar code is generated for the specified random patches to discriminate it from edge and corner texture patches. This example results clearly reveal the comparative efficacy of LNEP in the generation of distinguishable codes for distinct textures against the existing descriptors.

Further, the local descriptors robustness and stability is analyzed in an uncertain, noisy environment. For that, zero-mean Gaussian noise with different noise variance (from 1 to 10) is added to the 800 expressive facial images randomly collected from working dataset. At each noise level, local texture features are extracted for various descriptors, such as LTP, LDN, LDP, LBP, and LNEP. The histogram difference between the featured image without noise and its corresponding featured image with noise is computed using chi-square distance for every noise level. The average of histogram dissimilarity at each noise variance level is presented in Fig. 6. From the result, it is observed that dissimilarities for LNEP are comparatively smaller at each noise level than other descriptors, demonstrating its reliability under noise. The primary factor for these consistent performance is that LNEP specifically dismisses the ineffective feature patterns including random noisy patterns and flat textures. These textures can alter the local structure under a noise environment and thus create confusion in feature representation. Therefore, the unpredictable effects of these textures are very low in the LNEP than other descriptors, thus leading to stable under noise.

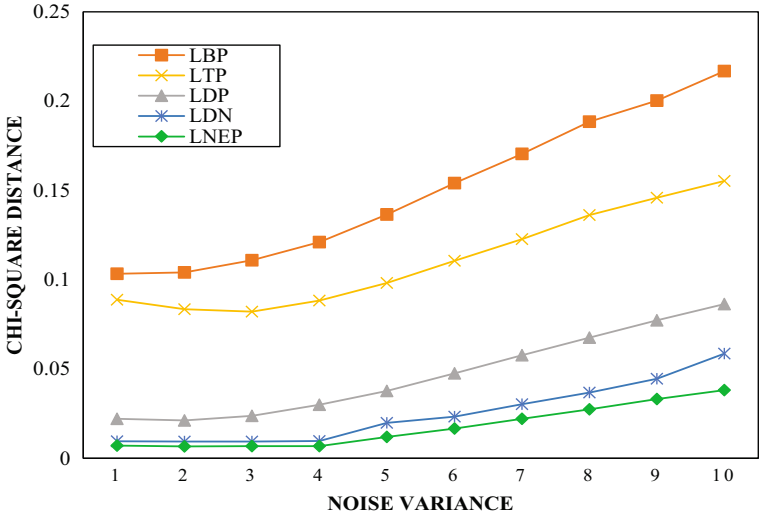


Fig. 6 Histogram dissimilarity of features with and with noise

3.4 Feature selection

Feature selection is the process of selecting the most relevant features to the prediction variable, and it used to build an effective predictive model. Machine learning models are usually congested when they handle very high dimensional data. Because, too many features increase the training time exponentially and the high risk of the model over fitting. Feature Selection techniques can be used to overcome these problems by eliminating irrelevant feature without losing discriminating capability. It is also helps us to understand the importance of feature in the given problem. The chi-square (χ^2) is the one of the statistical method to find the level of independence between two incidents. Here, the degree of independence is evaluated between feature and class variable by calculating chi-square score using (10) and the features are prioritized accordingly. The classification results after feature selection directly reflects the degree of correlation exist between feature and class label and the results also shows the importance of feature selection.

$$\chi^2 = \sum_{p_f \in 0,1} \sum_{p_c \in 0,1} \frac{(N_{p_f p_c} - E_{p_f p_c})^2}{E_{p_f p_c}} \tag{10}$$

where p_f denotes the feature occurrences and p_c represents the class occurrences. The actual and estimated feature values in the dataset D is denoted by N and E. The high χ^2 value indicates that the feature is more dependent on class variable. With the help of chi-square score, the features that are strongly dependent of class variables are identified as relevant feature and then classified. The results obtained through implementing the proposed method on different datasets shows the significance of feature selection in the proposed method.

4 Results and discussion

The effectiveness of the hybrid feature is evaluated on two datasets, namely Extended Cohn–Kanade (CK+) [23] dataset and MMI dataset [45]. The number of samples considered for evaluation on the two benchmark datasets are given in Table 1. The Leave-one-subject-out (LOSO) cross-validation [29] method is applied to ensure the person independent FER system so that there is no coincidence among the training and testing data. Nested cross validation is generally prefered to finetune the hyperparameters during the training process. The given dataset divided into number of folds and each fold size is based on number images taken from each subject. An inner loop is used to select the model via tenfold cross validation on the training fold. After model selection, the test fold is then used to evaluate the model performance. After feature extraction, chi-square test based feature selection is performed to discover the low dimensional merged patterns, and the classification results of the reduced feature set are presented here as a hybrid feature with feature selection. This experimental setup is adopted for both dataset.

Table 1 The number of sample consider from CK+ and MMI dataset

Dataset	Emotion class						Total
	Anger	Disgust	Fear	Happy	Sad	Surprise	
CK+	649	586	610	691	616	961	4113
MMI	130	110	130	190	170	130	860

Fig. 7 Sample images from CK+ dataset

4.1 Results on extended Cohn-Kanade database (CK+)

The Extended Cohn-Kanade Database (CK+) [23] consists of 593 video sequences that show 123 persons performing different expressions. From the entire dataset, 4113 images out of 593 image sequences are labeled and grouped into six basic emotion classes, namely anger, disgust, fear, happy, sad, and surprise. Every facial expression image sequence contains the expressive images of a person that changes from the beginning state (the starting point) to the expression peak state. The proposed scheme effectiveness is evaluated using frontal pose images, ranging from onset to apex state for feature extraction and the Fig. 7 show the sample images taken from CK+ dataset.

According to LOSO method, all expression images from each subject is omitted from tarining in each fold and then tested in the nested loop manner. The number of folds are decided based on number of subjects. An inner loop is used to select the model via tenfold cross-validation on the training fold. After model selection, the test fold is then used to evaluate the model performance (outer loop). This process is repeated for all subject and the average performance based on the test fold is reported. The training folds are used to select the optimal parameters. The classification is performed using the RBF kernel based Multiclass SVM, and its regularization parameter C and gamma values were identified using grid

Table 2 Confusion matrix of LBP on CK+ dataset with feature selection

Predicted Label							
True Label		Anger	Disgust	Fear	Happy	Sad	Surprise
	Anger	82%	4.5%		5%		8.5%
	Disgust	5.5%	89%			5.5%	
	Fear			100%			
	Happy	8%			92%		
	Sad	11%				84%	5%
	Surprise	4.5%				4.5%	91%
	Average	89.53%					

Table 3 Confusion matrix of LNEP on CK+ dataset with feature selection

Predicted Label							
True Label		Anger	Disgust	Fear	Happy	Sad	Surprise
	Anger	82%	4.5%	8.5%	5%		
	Disgust		95%				5%
	Fear			100%			
	Happy				92%	4%	4%
	Sad				5%	95%	
	Surprise	5%	4%				91%
	Average	91.94%					

search based parameter tuning. The training folds are used to select the optimal parameters. For the CK+ dataset, highest accuracy is achieved when the C value is 100 and gamma is 50.

Tables 2, 3 and 4 presents the recognition rate of LBP, LNEP and the hybrid features on CK+ dataset and Fig. 8 shows the effect of feature selection on the proposed approach. From Tables 2 and 3, the average recognition rate of LBP is 89.53% whereas for LNEP it is 91.94% which is 2.41% higher than LBP. This result indicates the effectiveness of LNEP in representing facial texture feature using the relationship among neighboring pixels. key problem with LBP is that two different local structures can produce the same LBP code [4]. This inefficient nature of LBP can be handled by adding more features. So that, local neighborhood feature fusion based hybrid feature is introduced by combining LNEP and LBP. The resultant high dimensional feature space is minimized with the help of chi-square score based feature selection. After feature selection, a hybrid feature achieves 97.86% average accuracy on CK+ data set with maximum of 99.26% and the minimum of 93.03% for surprise and fear respectively.

The proposed hybrid feature without feature selection achieved 95.64% average accuracy, which is 2.22% lower than average accuracy rate of the reduced feature set. Also, the feature selection technique reduces the dimension of the hybrid feature by analyzing the correlation between dependent and independent variables and improves the overall recognition accuracy. Results in Fig. 8 show that selected features give better results for all expressions other than happy. Person independent experiments are used to illustrate the strength of the

Table 4 Confusion matrix of hybrid feature on CK+ dataset with feature selection

Predicted Label							
True Label		Anger	Disgust	Fear	Happy	Sad	Surprise
	Anger	99.01%	0.5%			0.5%	
	Disgust	1.2%	95.06%	1.85%		0.6%	1.2%
	Fear		0.3%	93.03%	2.5%	1.5%	
	Happy			0.3%	99.7%		
	Sad			0.6%		98.32%	1.1%
	Surprise	0.4%				0.4%	99.26%
	Average	97.86%					

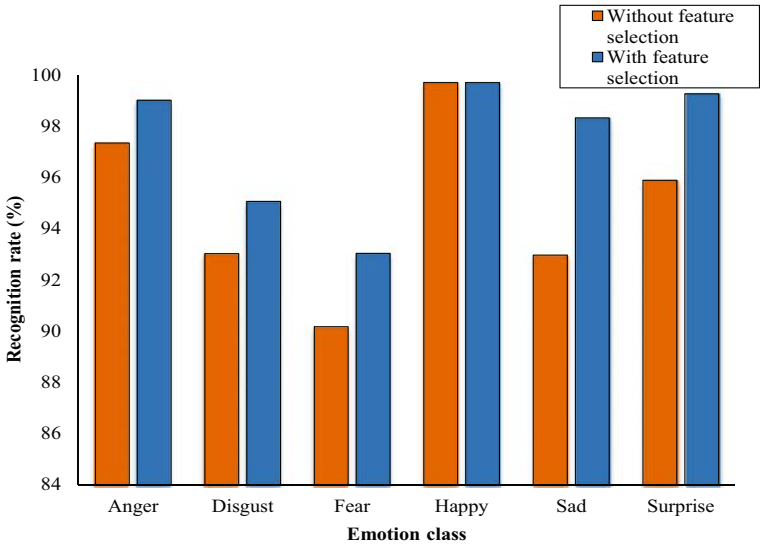


Fig. 8 The effect of feature selection on CK+ dataset

hybrid feature by testing on unseen subjects. Although high accuracy is attained for all six basic emotions, it is observed that most of the misclassification occurs only among anger, disgust, and fear due to the involvement of muscle movements in both upper and lower face, whereas for happy and surprise muscle movements around the mouth region are enough to distinguish them.

In Table 5, the proposed method results on the CK+ dataset are compared against other state-of-the-art techniques that adopted similar protocols. Among all the comparative meth-

Table 5 Performance comparison of different State-of-the-Art approaches on CK+ Database

Feature Descriptor	Avg.Accuracy %
LBP	89.52
LDP	79.54
LTP	88.35
LGP	73.8
LDN	96.89
LNEP	91.94
ELBP+KLT (2013) [25]	89.8
Gabor (2018) [2]	88
GLTP (2017) [15]	86.5
TPOEM (2018)[10]	97.37
HOG with GSP (2019) [30]	97.61
LDSP (2019) [29]	94.49
Geometric and LBP (2018)[28]	98.95
hvnLBP (2017) [31]	90.6
LDPv (2010) [16]	96.7
D-T signature feature (2019) [6]	98.6
Proposed hybrid feature	97.86

ods, the two topmost accuracies are achieved by Anima Majumder [28] and Asit Barman [6] using a hybrid feature in which geometric feature is combined with texture feature. In [28], feature fusion is carried out using three autoencoders to balance the different nature of features and the resultant high dimensional feature space increase the computational complexity. Because, when the dimension of data increases, the time taken to prepare data for visualization using 2D lattice is also increased and the number of distances calculated to find the better similarity map in KSOM will also increase exponentially which makes classification more expensive. In [6], D-T signature feature gets 98.7% average accuracy on CK+, but the experimental results seem to be person dependent.

Other local texture feature-based approaches like Improved GLTP [15], LDPv [16], LDSP [29] and HOG with GSP [37] extract histogram oriented features from the specific region or divide the whole face into a different region. This histogram-based feature description is simple, but, it needs adequate illustrations, and excludes the spatial information inside each region. However, a region-based texture descriptor retains the spatial information but the number of sample code required to describe the micro texture information decreases.In general, our proposed approach perform better than most of the existing approach but not by wide margin. Some facial expression (fear and disgust) are more difficult to classify due to the similar and micro muscle movement. In addition, the less number sample per class is also the one of the reason for this performance.

Also, block or region based feature extraction omits the useful correlation among different features and increases the feature space so that computational cost is high with the risk of over fitting. In our proposed work, the entire face is considered for two type of feature extraction, so that feature space size after feature fusion will double the size of the single feature. the highly co-related features identified with the help of chi-square test, reduce the feature space which in turn reduce the system overhead. The proposed hybrid feature with feature selection increases the recognition rate to 97.86%, which is 8.34% and 5.92% higher than LBP and LNEP, respectively. From the experimental results, it is witnessed that the relationship among the neighboring pixel is as important as the relationship of neighboring pixels with a central pixel. When these complementary features are combined to take advantage of different contribution, it collectively improves the recognition rate for all emotions, which in turn improves the overall accuracy of the proposed system.

Fig. 9 Sample images from MMI dataset

Table 6 Confusion matrix of LBP on the MMI dataset with feature selection

Predicted Label							
True Label		Anger	Disgust	Fear	Happy	Sad	Surprise
	Anger	81.4%	3.9%	6.2%	5.4%	2.3%	0.8%
	Disgust	3.6%	78.2%	3.6%	10.9%	2.7%	0.9%
	Fear	3.9%	3.1%	77.7%	6.9%	1.5%	6.9%
	Happy	2.6%	0.53%	3.7%	91.6%		1.6%
	Sad	1.1%	2.2%	4.4%	2.2%	85.6%	4.4%
	Surprise	5.4%		2.3%	6.9%		85.4%
	Average	83.95%					

4.2 Results on MMI

The MMI face dataset [45] comprises more than 1500 samples of videos and static images of facial expression. For our experiments, Part II of this dataset is considered which contains of frontal facial images taken from 19 male and female subjects, showing basic expressions. Figure 9, show the sample images from MMI dataset. A subject-independent experiment on the MMI dataset with feature selection is conducted, and the hybrid feature results are reported in Table 8. As implemented in the CK+ database, expression samples of the single subject is randomly chosen for testing, and the rest are used for training. Tables 6 and 7 present the individual feature contribution in recognizing basic emotions. The average recognition accuracy of LBP is 83.95% whereas for LNEP it is 92.69%.

Experimental results show that the LNEP feature performs well over the LBP feature with 2.6% higher recognition rate. After feature fusion, the features with the highest chi-square score are selected for classification. The selected optimal features improve the recognition rate of all expressions. From Table 8, it is observed that the overall recognition accuracy of the hybrid feature with feature selection is 97.11%. Like CK+ dataset, the highest recognition is achieved for happy with 98.31%, and the lowest rate is achieved for fear with 95.51%. Due to the similar muscle movements around eyebrow and mouth for fear and surprise, highest misclassification is realized between them. In addition, some people just

Table 7 Confusion matrix of LNEP on the MMI dataset with feature selection

Predicted Label							
True Label		Anger	Disgust	Fear	Happy	Sad	Surprise
	Anger	83.69%	2.1%	5%	5%	0.7%	3.5%
	Disgust		100%				
	Fear	4%		92%			4%
	Happy	1.57%	2.09%		95.81%	0.52%	
	Sad	4.26%	3.19%			92.55%	
	Surprise			6.2%		0.78%	93.02%
	Average	92.69%					

Table 8 Confusion matrix of hybrid feature on the MMI dataset with feature selection

Predicted Label							
True Label		Anger	Disgust	Fear	Happy	Sad	Surprise
	Anger	98.1%	0.8%	1.1%		0	
	Disgust	0.2%	97.22%	0.2%	1.4%	1%	
	Fear			95.51%			4.5%
	Happy		1.5%	0.1%	98.31%	0.1%	
	Sad	1.6%		1.6%		96.81%	
	Surprise	3.2%					96.77%
	Average	97.11%					

lift their eyebrows for surprise expression. It will create uncertainty among the expressions. In such situation, temporal feature and secondary feature like forehead wrinkle and nose side wrinkle, can be used to reduce misclassification rate. The wrong pose or poor representation is also the another reason for misclassification. In all cases, the recognition rate can be further improved by increasing training set size.

Figure 10 shows the effect of feature selection in individual emotion recognition on the MMI dataset. From results, it is observed that the recognition rate of all expression except surprise is significantly enhanced. Figure 11 shows the average recognition rate of CK+ and MMI dataset with feature selection. Here, feature selection improves the recognition accuracy by 2.22% in CK+ dataset and by 18.39% in MMI dataset. Moreover, the subjects

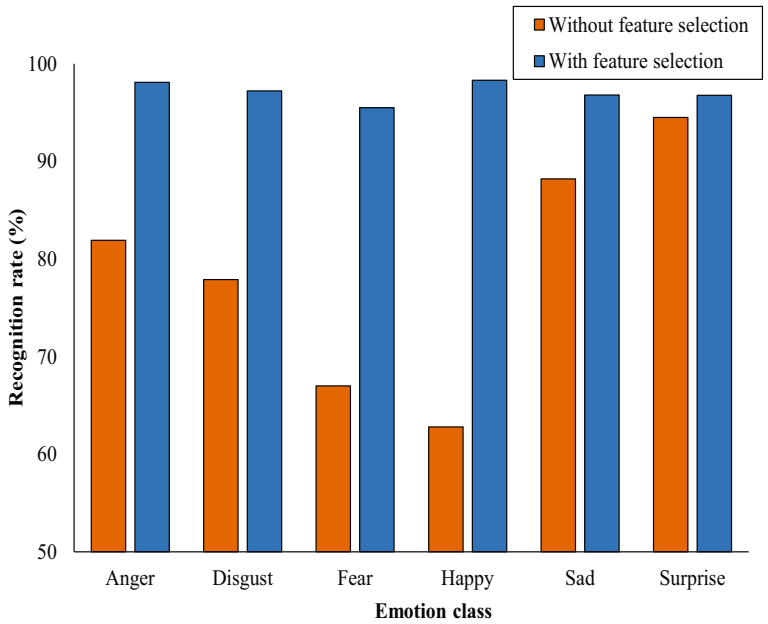


Fig. 10 The effect of feature selection on the MMI dataset

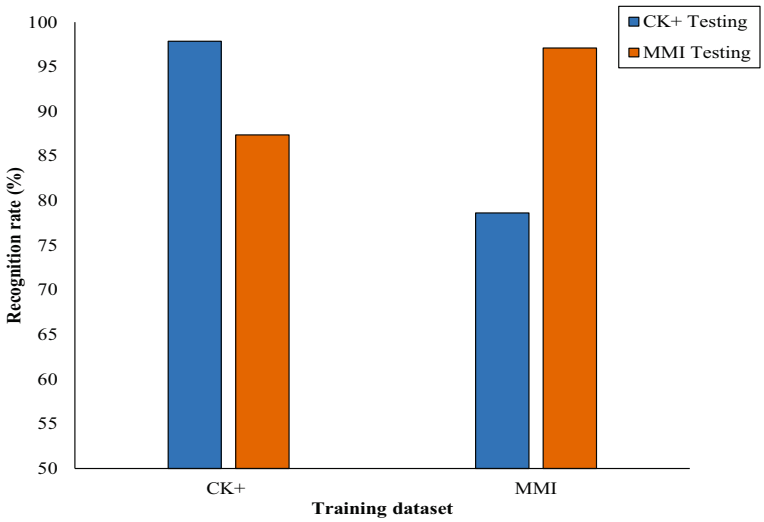


Fig. 11 The average recognition rate of proposed feature with and without feature selection on the CK+ and MMI dataset

wearing eyeglasses are the primary cause for the presence of irrelevant information in feature vector; therefore, this misguiding information may generate uncertainty in the feature vector. However, evaluating the performance of the proposed hybrid feature under these conditions can be another important research issue.

Experimental findings can not be compared explicitly due to the various factors such as experimental setup, volume of data used for evaluation, evaluation method and so on. But still the comparison provides way to understand the strength of various methods. Table 9 shows the comparative results of the proposed method with the state-of-art methods from literature. The performance of the related methods is cited directly from the original references and the rest of the results are obtained by our own implementation. Like CK+, in MMI

Table 9 Performance comparison of different State-of-the-Art approaches on MMI Database

Feature Descriptor	Avg.Accuracy %
LBP	83.95
LDP	60.58
LTP	94.5
LGP	76.3
LDN	95.5
LNEP	92.7
Geometric and LBP (2018) [28]	97.55
LDTP+PCA (2015) [38]	93.7
SWLDA (2015) [40]	96.83
TPOEM (2018) [10]	93.66
LDSP (2019) [29]	69.05
D-T signature feature (2019) [6]	94.3
Proposed hybrid feature	97.11

also [28] achieve the accuracy of 97.55%, which is slightly better than our proposed method. In [28], different dimensional features are fused with the help of three autoencoders and the feature-length used for classification after feature fusion is 230. But, the facial component based two type of feature extraction and feature fusion using three autoencoders increases the system computation cost. In our proposed scheme, the relevant features selected via the chi-square analysis are classified using multiclass SVM. The evaluation clearly indicate that the proposed local texture feature fusion with feature selection perform well in contrast with other methods in basic emotion classification. Here, the proposed scheme achieves good average recognition rate by handling the images taken from controlled environment. But this also needs to be generalized in order to process data in real time.

4.3 Analysis of other possible fusions with LNEP

The feature fusion of LNEP with other local features are evaluated on two data set using LOSO cross-validation method. The minimum, maximum, standard deviation and the average recognition rate are reported in Table 10. The proposed method with the feature fusion of LBP and LNEP is better from all the other possible fusion with LNEP. The average recognition rate of LTP+ LNEP is very close to the proposed feature fusion, but the LTP feature length is very high when compared with the other feature. Facial expression recognition is the challenging task in the real world, and the methodology used to solve such problem need to be computationally simple. The LDP, LDN, and LGP encode the texture information with the feature size of 56, 56, and 7, respectively. The average recognition rate of these features are increased significantly after feature fusion.

4.4 Performance analysis under noisy environment

Noise is the arbitrary discrepancy of pixel intensity in an images produced by the digital camera. Noise generates unwanted effects such as artifacts, unrealistic edges, unseen lines, corners, blurred objects and disturbs background scenes. There are many forms of noise capable of influencing images. The Gaussian noise is the most common noise which naturally affects the image quality. In digital images, Gaussian noise typically disturbs the gray values. Mathematical model of the Gaussian noise typically reflects the closest approximation of real world situations. In face processing, noise may be induced due to wrong image acquisition practices. Sometimes, it is even difficult for the human to identify the person from the noisy facial image. Various noise removal methods are available to denoise the image. But the denoising process sometimes removes some useful visual information which

Table 10 Fusion of LNEP with other descriptors

Descriptors	Feature Length	CK+			MMI		
		Min	Max	Avg.Accuracy	Min	Max	Avg.Accuracy
LDP+LNEP	56+256	91.75	94.19	93.39(±1.03)	50	100	89.21(±10.35)
LTP+LNEP	512+256	96.85	99.02	97.9(±0.69)	76.9	100	97.11(±5.58)
LDN+LNEP	56+256	96.11	98.79	97.37(±0.66)	76.9	100	96.01(±6.14)
LGP+LNEP	7+256	95.39	97.81	96.38(±0.75)	70	100	94.49(±7.38)
LBP+LNEP	256+256	96.84	98.55	97.86(±0.1)	80	100	97.11(±4.98)

Fig. 12 Sample images with different noise variations (σ)

affect the subsequent stages. Here, the various local descriptors robustness and stability are analyzed in an uncertain, noisy environment for direct recognition.

In Section 3.3, the robustness against noise is analyzed using chi-square histogram dissimilarity map. However, in this section, performance of the proposed descriptor in noisy environment is explicitly evaluated against other descriptors. For that, zero-mean Gaussian noise is added to mimic the effect of many random noises that occurs in nature with a variance that varies from 1 to 10 to each image of the MMI data set. In our approach, noise intervals are randomly distributed to ensure the ordinary imperfect condition. Figure 12 shows the example of noisy images.

Consequently, we conduct subject-independent recognition for various descriptors and the results are reported in Table 11. From the results it is found that the proposed hybrid feature performs well in a noisy environment than that of other descriptors. As mentioned before, the discriminating capability of the hybrid feature contributes the most to the consistent performance by excluding unclear noisy patterns.

The combination of these two features can effectively integrate the gains mutually by preserving local features of the facial image. These two methods also make up for their deficiencies. The LBP operator possesses the advantage of its invariant nature against monotonic gray-level changes and computational simplicity. The main issue of LBP is sensibility in the presence of noise. In such a situation, the LNEP operator can reduces the impact of the noise on LBP effectively while the LBP enhances the representation of local texture characteristics. Also, the non-overlapping nature of this feature fusion represent all texture without ambiguity. Therefore, the proposed hybrid feature descriptor is an efficient tool for FER in different noisy environments.

Table 11 Recognition rate (%) of the MMI dataset with varying noise

Descriptors	Without Noise	With Noise	
		$\sigma = 2.24$	$\sigma = 3.16$
LBP	85.95	44.935	29.95
LDP	60.58	60.58	20.9
LDN	95.5	63.82	33.72
LTP	94.5	64.32	34.25
LNEP	92.69	57.66	34.89
Hybrid	97.11	64.72	37.21

Table 12 Performance analysis on cross dataset validation using CK+ and MMI

Training vs Testing data set	Anger	Disgust	Fear	Happy	Sad	Surprise	Avg.Accuracy
CK+ vs CK+	99.01	95.06	93.03	99.7	98.32	99.26	97.86
CK+ vs MMI	79.87	79.41	81.85	98.78	76.35	94.62	87.37
MMI vs MMI	98.1	97.22	95.51	98.31	96.81	96.77	97.11
MMI vs CK+	78.81	63.91	66.92	93.03	63.88	86.23	78.63

4.5 Results on cross dataset validation

In real-time applications, test samples are not the same as the training samples, and its acquisition condition is often different, such as variation in illumination, etc. But the generalization is essential for a local descriptor to achieve better performance with real-time data. For that, cross-database validation is carried out by considering the samples from one dataset as the training data, and the remaining dataset as the testing data. This type of evaluation in FER is a challenging task. Table 12 shows the results on the cross-database validation. The results show that the recognition rates of all emotions decrease significantly when the training and test samples are taken from a different dataset.

The maximum recognition rate is achieved when the model is trained using CK+ while tested on MMI in cross-database validation. Because, the total number of samples taken from the CK+ database is more than MMI. In another case, the number of training samples from MMI is less than CK+, which leads to over-fitting, and thus the recognition performance on the test samples is low [7]. In addition, some of the subject's expression in MMI is taken with glasses. This type of external accessories may create uncertainty in the feature description. However, experimental results show the reliability of the proposed scheme, and it can be further improved by increasing the training set size. Figures 13a, b, and c show the performance comparison of the cross-database validation using CK+ and MMI dataset.

4.6 Computation time

The proposed hybrid feature length is double the size of a single feature. The high dimensional feature generally increases the system complexity. Hence, the feature selection method is adopted to reduce the feature dimension. The feature extraction and the classification time with and without feature selection on the MMI dataset is reported in Table 13 and it show that LNEP consume less time in feature extraction. When considering the classification time, the execution time after feature selection is reduced when compared with all other descriptors. Even though the proposed method performs better than the existing state-of-the-art methods in terms of accuracy, the LOSO based person independent cross-validation method increases the time complexity. The feature extraction time per image (MATLAB code) and the classification time without and with feature selection (python code) is calculated using a desktop machine with octa-core CPU running at 3.5 GHz and it is reported in Table 13.

Although the feature extraction and classification time of LBP, LDN, LNEP are faster than the hybrid feature, the combined feature provides a noteworthy gain in the facial expression recognition due to the inclusion of both features mutual relationship using non-overlapping feature fusion.

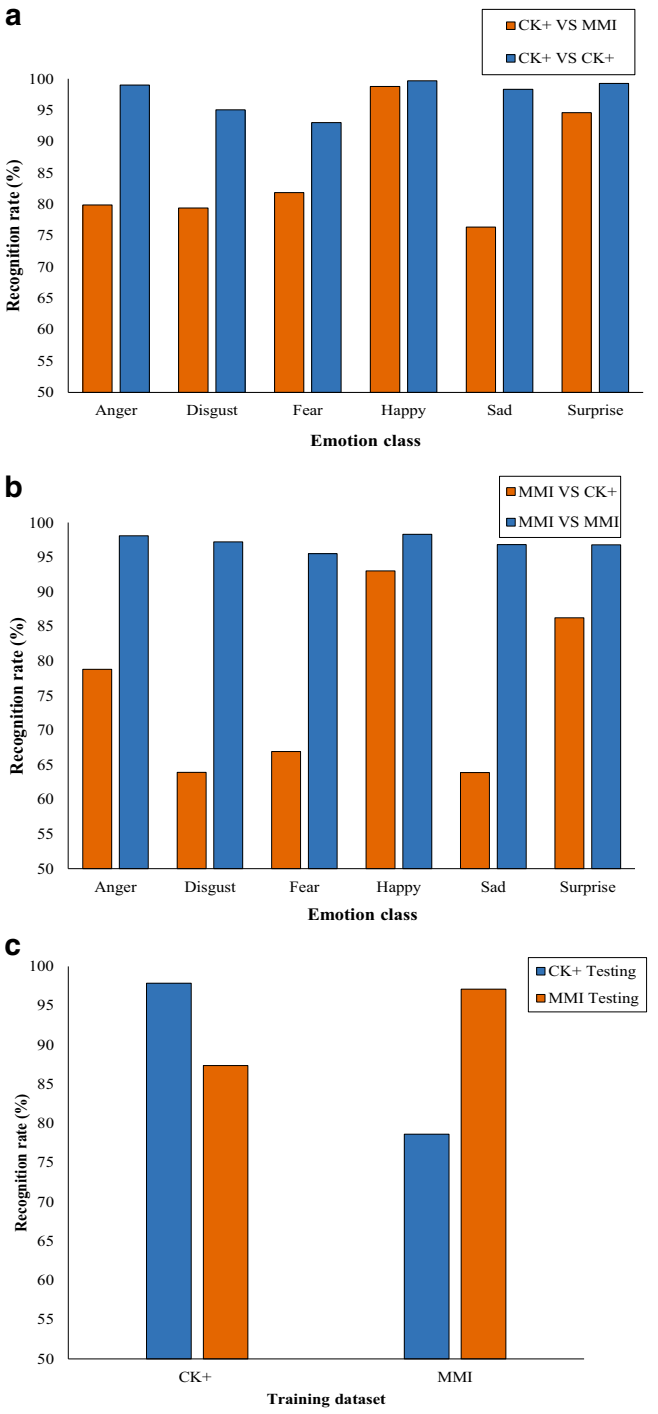


Fig. 13 Performance comparison of cross database (**a**) CK+ vs. MMI (**b**) MMI vs. CK+ (**c**)Average recognition rate on cross-dataset validation

Table 13 The computation time in seconds for feature extraction and the classification time using MMI database

Descriptor	Feature Extraction time (per image)	Classification time (without feature selection)	Classification time (with feature selection)
LBP	0.0289	66	29.69
LDP	0.1069	70	32.01
LTP	0.0514	126	31.4
LDN	0.1329	63	29.08
LNEP	0.0229	66	29.75
Hybrid Feature	0.05	123	29.94

5 Conclusion

An appropriate facial feature representation will significantly influence the effectiveness of a successful expression recognition system. In the proposed system, two complementary features, LBP and LNEP are extracted by considering local neighboring pixel relationship. The LNEP represent the mutual relationship among the closest neighborhood pixel whereas the LBP encode the neighboring pixel relationship with central pixel. The proposed method combines both features in an non-overlapping manner to deal with the facial expression recognition. From the high dimensional combined feature, the most relevant features are selected, using chi-square statistical analysis. The efficiency of the selected feature is analyzed individually and collectively on CK+ and MMI dataset using LOSO cross-validation with Multiclass SVM. The selected hybrid features improve the recognition rate of all expressions with an average rate of 97.86% on CK+ dataset and 97.11% on MMI dataset. Experimental outcomes exhibit that the hybrid feature performs better than other state-of-art methods under lab-controlled environment. Also, the effectiveness of the proposed system is validated over the facial image with noise. Although the noise can severely affect the recognition accuracy, it has been shown by the experiments that the hybrid feature performs better than other descriptors in a noisy environment. In the future, the other issues that arise in a real-time environment such as head pose variation, occlusion, illumination effect, etc., which directly affect the appearance of the face, need to be addressed by combining both geometric, and appearance features. Also, the recent achievement of deep learning methods in expression recognition may drive the way to incorporate the proposed hybrid feature within the deep learning models to improve recognition accuracy.

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Ahmed F, Bari H, Hossain E (2014) Person-independent facial expression recognition based on compound local binary pattern (clbp). Int Arab J Inf Technol 11(2):195–203

2. Akputu OK, Seng KP, Lee Y, Ang L-M (2018) Emotion recognition using multiple kernel learning toward e-learning applications. ACM Trans Multimed Comput Commun Appl (TOMM) 14(1):1
3. Arshid S, Hussain A, Munir A, Nawaz A, Aziz S (2017) Multi-stage binary patterns for facial expression recognition in real world. Clust Comput pp 1–9
4. Banerjee P, Bhunia AK, Bhattacharyya A, Roy PP, Murala S (2018) Local neighborhood intensity pattern–a new texture feature descriptor for image retrieval. Expert Syst Appl 113:100–115
5. Barman A, Dutta P (2017) Facial expression recognition using distance and shape signature features. Pattern Recognit Lett
6. Barman A, Dutta P (2019) Facial expression recognition using distance and texture signature relevant features. Appl Soft Comput 77:88–105
7. Chang T, Li H, Wen G, Hu Y, Ma J (2019) Facial expression recognition sensing the complexity of testing samples. Appl Intell
8. Chen J, Chen Z, Chi Z, Fu H (2018) Facial expression recognition in video with multiple feature fusion. IEEE Trans Affect Comput 9(1):38–50
9. Chen L, Zhou M, Su W, Wu M, She J, Hirota K (2018) Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction. Inform Sci 428:49–61
10. Cruz EAS, Jung CR, Franco CHE (2018) Facial expression recognition using temporal poem features. Pattern Recogn Lett 114:13–21
11. Da Silva FAM, Pedrini H (2016) Geometrical features and active appearance model applied to facial expression recognition. Int J Image Graph 16(04):1650019
12. Friesen E, Ekman P (1978) Facial action coding system: a technique for the measurement of facial movement. Palo Alto p 3
13. Guo M, Hou X, Ma Y, Wu X (2017) Facial expression recognition using elbp based on covariance matrix transform in klt. Multimed Tools Appl 76(2):2995–3010
14. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. IEEE Trans Affect Comput 6(1):1–12
15. Holder RP, Tapamo JR (2017) Improved gradient local ternary patterns for facial expression recognition. EURASIP J Image Video Process 2017(1):42
16. Jabid T, Kabir MH, Chae O (2010) Robust facial expression recognition based on local directional pattern. ETRI J 32(5):784–794
17. Ko B (2018) A brief review of facial emotion recognition based on visual information. Sensors 18(2):401
18. Kumar S, Bhuyan MK, Chakraborty BK (2016) Extraction of informative regions of a face for facial expression recognition. IET Comput Vis 10(6):567–576
19. Lai C-C, Ko C-H (2014) Facial expression recognition based on two-stage features extraction. Optik-Int J Light Elect Opt 125(22):6678–6680
20. Li R, Liu P, Jia K, Wu Q (2015) Facial expression recognition under partial occlusion based on gabor filter and gray-level cooccurrence matrix. In: 2015 international conference on computational intelligence and communication networks (CICN), IEEE, pp 347–351
21. Liang D, Liang H, Yu Z, Zhang Y (2019) Deep convolutional bilstm fusion network for facial expression recognition. Vis Comput pp 1–10
22. Liu Z, Wu M, Cao W, Chen L, Xu J, Zhangm R, Zhou M, Mao J (2017) A facial expression emotion recognition based human-robot interaction system
23. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE computer society conference on computer vision and pattern recognition-workshops, IEEE, pp 94–101
24. Luo RC, Huang CY, Hsiao CC (2011) Recognition of facial expressions using component-based active appearance models for human-robot interactions. In: IECON 2011-37th annual conference of the IEEE industrial electronics society, IEEE, pp 4244–4249
25. Luo Y, Wu C-M, Yi Z (2013) Facial expression recognition based on fusion feature of pca and lbp with svm. Optik-Int J Light Elect Opt 124(17):2767–2770
26. Luo Y, Zhang T, Yi Z (2016) A novel fusion method of pca and ldp for facial expression feature extraction. Optik-Int J Light Elect Opt 127(2):718–721
27. Majumder A, Behera L, Subramanian VK (2014) Emotion recognition from geometric facial features using self-organizing map. Pattern Recogn 47(3):1282–1293
28. Majumder A, Behera L, Subramanian VK (2018) Automatic facial expression recognition system using deep network-based data fusion. IEEE Trans Cybern 48(1):103–114
29. Makhmudkhujaev F, Iqbal MTB, Ryu B, Chae O (2019) Local directional-structural pattern for person-independent facial expression recognition. Turkish J Elect Eng Comput Sci 27(1):516–531
30. Meena HK, Joshi SD, Sharma KK (2019) Facial expression recognition using graph signal processing on hog. IETE J Res pp 1–7

31. Mistry K, Li Z, Neoh SC, Lim CP, Fielding B (2017) A micro-ga embedded pso feature selection approach to intelligent facial emotion recognition. IEEE Trans Cybern 47(6):1496–1509
32. Munir A, Hussain A, Khan SA, Nadeem M, Arshid S (2018) Illumination invariant facial expression recognition using selected merged binary patterns for real world images. Optik 158:1016–1025
33. Nguyen H-D, Yeom S, Lee G-S, Yang H-J, Na I-S, Kim S-H (2018) Facial emotion recognition using an ensemble of multi-level convolutional neural networks. Int J Pattern Recognit Artif Intell
34. Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern Recognit 29(1):51–59
35. Ojala T, Pietikäinen M, Topi M (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell, (7):971–987
36. Punitha A, Kalaiselvi Geetha M (2013) Texture based emotion recognition from facial expressions using support vector machine. Int J Comput Appl 80(5):1–5
37. Rathee N, Ganotra D (2017) Modelling facial features for emotion recognition and synthesis. Iete J Res 63(6):845–852
38. Rivera AR, Castillo JR, Chae O (2015) Local directional texture pattern image descriptor. Pattern Recogn Lett 51:94–100
39. Shbib R, Zhou S (2015) Facial expression analysis using active shape model. Int J Signal Process Image Process Pattern Recognit 8(1):9–22
40. Siddiqi MH, Ali R, Khan AM, Park Y-T, Lee S (2015) Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans Image Process 24(4):1386–1398
41. Sun W, Zhao H, Jin Z (2018) A visual attention based roi detection method for facial expression recognition. Neurocomputing 296:12–22
42. Tong Y, Chen R, Cheng Y (2014) Facial expression recognition algorithm using lgc based on horizontal and diagonal prior principle. Optik-Int J Light Elect Opt 125(16):4186–4189
43. Topi M, Timo O, Matti P, Maricor S (2000) Robust texture classification by subsets of local binary patterns. In: Proceedings 15th international conference on pattern recognition. ICPR-2000, IEEE, vol 3, pp 935–938
44. Turan C, Lam K-M (2018) Histogram-based local descriptors for facial expression recognition (fer): a comprehensive study. J Vis Commun Image Represent 55:331–341
45. Valstar M, Pantic M (2010) Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In: Proceedings 3rd intern workshop on EMOTION (satellite of LREC): corpora for research on emotion and affect, p 65
46. Verma M, Raman B (2018) Local neighborhood difference pattern: a new feature descriptor for natural and texture image retrieval. Multimed Tools Appl 77(10):11843–11866
47. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vision 57(2):137–154
48. Vupputuri A, Meher S (2015) Facial expression recognition using local binary patterns and kullback leibler divergence. In: 2015 international conference on communications and signal processing (ICCSP), IEEE, pp 0349–0353
49. Wang H, Hu J, Deng W (2018) Face feature extraction: a complete review. IEEE Access 6:6001–6039
50. Xie S, Shan S, Chen X, Chen J (2010) Fusing local patterns of gabor magnitude and phase for face recognition. IEEE Trans Image Process 19(5):1349–1361
51. Zavaschi THH, Britto AS Jr, Oliveira LES, Koerich AL (2013) Fusion of feature sets and classifiers for facial expression recognition. Expert Syst Appl 40(2):646–655
52. Zhang Z (1999) Feature-based facial expression recognition: sensitivity analysis and experiments with a multilayer perceptron. Int J Pattern Recognit Artif Intell 13(06):893–911