International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Merged Local Neighborhood Difference Pattern for Facial Expression Recognition

P. Shanthi, S. Nickolas

Abstract: ***Facial expression based emotion recognition is one of the popular research domains in the computer vision field. Many machine vision-based feature extraction methods are available to increase the accuracy of the Facial Expression Recognition (FER). In feature extraction, neighboring pixel values are manipulated in different ways to encode the texture information of muscle movements. However, defining the robust feature descriptor is still a challenging task to handle the external factors. This paper introduces the Merged Local Neighborhood Difference Pattern (MLNDP) to encode and merge the two-level of representation. At the first level, each pixel is encoded with respect to center pixel, and at the second level, encoding is carried out based on the relationship with the closest neighboring pixel. Finally, two levels of encodings are logically merged to retain only the texture that is positively encoded from the two levels. Further, the feature dimension is reduced using chi-square statistical test, and the final classification is carried out using multiclass SVM on two datasets namely, CK+ and MMI. The proposed descriptor compared against other local descriptors such as LDP, LTP, LDN, and LGP. Experimental results show that our proposed feature descriptor is outperformed other descriptors with 97.86% on CK+ dataset and 95.29% on MMI dataset. The classifier comparison confirms the results that the combination of MLNDP with multiclass SVM performs better than other combinations in terms of local descriptor and classifier.***

Keywords: emotion, facial expression, merged local neighborhood difference pattern, support vector machine.

I. INTRODUCTION

Emotion is one of the cognitive processes of the brain and facial expressions are known to be one of the ways of expressing emotion, which plays a significant role in the transmission of emotional state information. Appropriate presentations of internal emotional states may be helpful for the improvement of social relations. The most prominent work on this topic comes from the psychologist Paul Ekman, and he discovered that there are some common facial expressions [1]; even blind people use the same faces to convey the same feelings, including those that signify joy, disappointment, sorrow, shock, and surprise. With this background work, several researchers from different fields are trying to develop an automated facial expression recognition system (FER) through which people can improve their mental-emotional state by investigating their behavior patterns.

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
P. Shanthi*, Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. Email: shanthianu81@gmail.com
S. Nickolas, Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. Email: nickolas@nitt.edu

FER systems have a wide variety of uses such as social marketing, computer interactions, and health-care systems.

An effective expression recognition system plays a significant role in the modeling and parameterization of human faces as avatars and computer animations [2] by defining facial structure and muscle movement. The robust FER is required in vision-based technologies such as augmented reality [3] and virtual reality [4] to introduce a normal, friendly interaction with humans. Nevertheless, studying facial expression is an extremely challenging process due to various positioning transient facial muscle contractions in real-world images/videos captured from the complex noisy environment [5]. The most important stage in automatic FER is the feature extraction stage. The overall system performance purely depends on the quality of the feature. The facial feature with high discriminative capability must re-duce variations in intra-class and significantly increases variations between classes. This discriminative property can assure the establishment of a competent and accurate expression prediction model.

Based on the type of feature involved in facial expression recognition techniques [6-8], it can be differentiated into two streams namely geometric feature-based and appearance feature-based FER system. The geometric type refers to local facial features that are derived based on the geometric relationship among facial components (eye, eye-brows, and mouth) in terms of shape, position, and angle. The feature vector of the geo-metric model contains the absolute and relative geometrical relationship of the facial component using a neutral face as the reference. The latter approach employs different coding techniques to represent texture or appearance information as a feature vector.

The approaches based on appearance are better and more prominent than the methods based on geometric features when the images/videos are taken from uncontrolled environments and real-world applications. The systematic study of facial feature extraction methods from unconstrained or real-world environments clearly explains such issues [9]. Alternatively, appearance features express the facial texture in terms of edges, corners, statistical features, etc., Among the various appearance-based approach, Principal Component Analysis (PCA) [10], Histograms Oriented Gradients (HOG) [11], Scale Invariant Feature Transformation (SIFT) [12], Local Directional Number Pattern (LDN) [13], are the most popular features extraction methods. The quality of the feature is decided based characteristics of the feature such as discrimination power, complexity, dimensionality, sensitivity against noise, low intra-class variations, etc., For a local descriptor, it is hard to achieve all of these properties. For example, the most popular local descriptor called the Local Binary Pattern (LBP) [14] is more prone to random noise, which is computationally simple.

Therefore, defining the robust, complete local descriptor is still an open problem for object representation and classification in many applications such as texture description and face processing.

Currently, deep learning models have been extensively researched for several pattern recognition challenges, such as object recognition [15], face recognition [16], forecasting human pose [17], but it needs an enormous quantity of training data to enhance overall performance. Al-so deep learning-based FER, increase the scope of research in this domain and the achievement of deep features based machine learning models discussed in [18],[19]. Even though the improvement in the recognition rate is high using deep features, the controversy among manually extracted features and deep features is indeed present. Recently, the hand-crafted feature introduced by Benitez-Garcia et al. [20] is capable of attaining a higher recognition level than a deep learning model. This indicates that the handcrafted features and domain-specific familiarity are still successful and favorable in computer vision-based classification. This paper presents a handcrafted local descriptor called Merged Local Neighborhood Difference Pattern (MLNDP) for facial expression recognition to validate the importance of relationship among closest neighboring pixels along with the central pixel relationship.

Emotion classification using a machine learning algorithm is the second critical phase of a FER system. The classification algorithms such as Support Vector Machine (SVM) [21], Random Forest [22], AdaBoost (AB) [23], K-Nearest Neighbor (KNN) [24], and deep neural networks [25] have been used in FER systems to date. A robust classifier is highly desirable in order to obtain very reliable facial expression identification using the proposed local descriptor. Here, we have used a multiclass support vector machine for the final classification. The ten-fold cross-validation method is applied to ensure that all data participate in the training process.

The importance of two-level neighboring pixel encoding in automatic facial expression recognition is analyzed against similar local descriptors called Local Gradient Pattern (LGP) [26], Local Directional Pattern (LDP) [27], Local Ternary Pattern (LTP)[28], and Local Directional Number (LDN)[29]. All types of spatial descriptors are extracted from the static images of CK+ and MMI dataset. The comparative results demonstrate that the proposed MLNDP provides 97.86% accuracy on the CK+ dataset and 95.29% accuracy on MMI dataset. To test the MLNDP effectiveness in FER, We compare the quality of the proposed framework with classifiers that are commonly used: K-Nearest Neighbor (KNN), Random Forest (RF), Ada Boost (AB) and Gradient Boosting (GBT).

The rest of the paper is structured as follows. Section 2 summarizes the relevant works, and the suggested approach is discussed in Section 3. Section 4 presents the experimental outcomes of the proposed method using two CK+ and MMI datasets. Section 5 provides the interpretation of the proposed approach and the future scope for further study.

II. PROPOSED METHODOLOGY

The basic FER framework contains the following stages: 1) face detection and pre-processing 2) local features extraction, 3) classification based on feature vectors.

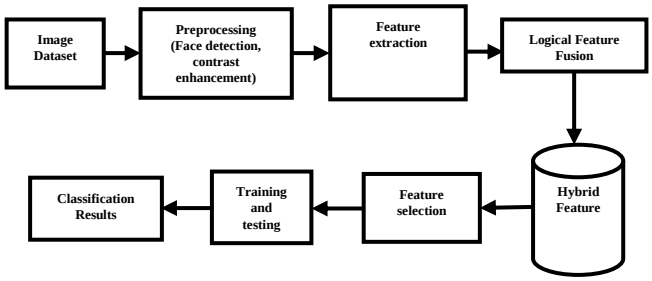


Fig. 1.Illustration of the proposed scheme

The results of the classification depend almost on each of t he above procedures. Many recent works, however, concentr ated on establishing more robust local characteristics [7], [11], [30]. Fig. 1. Show the steps followed in the proposed method. The following subsection explains each step.

A. Face detection and preprocessing

For face detection, Viola-Jones' object detection algorithm is used. In the preprocessing stage, the visual quality is enhanced using CLAHE [31]. In standard histogram equalization, contrast enhancement is performed globally so that important information is not appropriately preserved. However, using CLAHE, the image contrast is increased and also preserves the texture information, and the sample output is shown in Fig. 2. After preprocessing, MLNDP is extracted, and most contributing features are identified using the chi-square feature selection technique. Finally selected features are forwarded to the classifiers for expression classification.

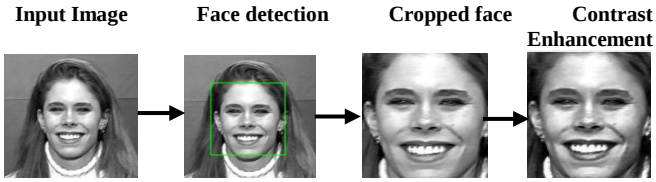


Fig. 2. Image enhancement

B. Construction of MLNDP

The region-based LBP feature extraction technique for the emotion recognition system is proposed in anima [30]. The region-based uniform LBP feature extraction increases the feature dimension when all the sub-region feature vectors are concatenated, but it did not precisely represent the correlation among facial component and expression uniqueness when the number of blocks is very small. To overcome this problem, texture features are extracted from the whole face in the proposed approach.

In the proposed approach, the 8-bit coding scheme is introduced that calculates three bits for each neighborhood pixel. These three neighboring pixels are logically fused to capture micro-changes in the expressive face image. If the pixel intensity is greater than or equal to threshold then it is encoded as 1 otherwise encoded as 0. Here two types of thresholds are used to generate an 8-bit code.

In the first level, the center pixel (I_c) from the 3x3 cell is used as the threshold for the current pixel (I_p). At the second level, the current pixel is used as the threshold for the two adjacent pixels($I_{(p+1)_{mod P}}, I_{(p-1+P)_{mod P}}$). Final feature fusion is performed on the outcome of the above two comparisons. The 8-bit logical feature fusion produces a feature vector of length 256. Fig. 3. shows the general structure used to generate the MLNDP code.

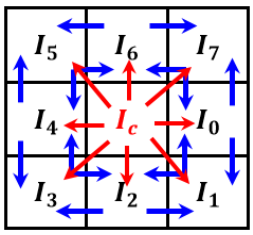


Fig. 3.The encoding structure of MLNDP

The proposed work focuses on the new local neighboring pixel encoding algorithm, which combines the three neighboring pixel relation using the binary operator. The three neighboring pixels with respect to the center pixel (I_c) in the selected window is denoted asI_p ,$I_{(p+1)_{mod P}}, I_{(p-1+P)_{mod P}}$, where p is the current pixel varies from$0,1,\cdots,7$, P is the total number of neighboring pixel

For example, for the current pixelI_5, the neighboring pixels are identified by $I_{(5+1)mod 8} = I_6$and$I_{(8+5-1)mod 8} = I_4$. The binary encoding is carried out based on the equations (1), (2), and (3). Finally, the logical fusion of the outcome of the above three equations will provide the feature vector of the single image.

$$f_{hybrid}(x_c, y_c) = f_1(x_c, y_c) \cdot \left(f_2(x_p, y_p) \odot f_3(x_p, y_p)\right) \quad (1)$$

$$f_1(x_c, y_c) = \sum_{p=0}^{7} s(I_c - I_p)2^p \quad (2)$$

$$f_2(x_p, y_p) = \sum_{p=0}^{7} s(I_{(p+1)_{mod P}} - I_p)2^p \quad (3)$$

$$f_3(x_p, y_p) = \sum_{p=0}^{7} s(I_{(p-1+P)_{mod P}} - I_p)2^p \quad (4)$$

$$s(X) = \begin{cases} 1 & if\ X \geq 0 \\ 0 & otherwise \end{cases} \quad (5)$$

After comparing two neighboring pixels with the current pixel using equation. (3) and equation. (4), the XNOR operation is performed by the second part of the equation. (1). In the next step logical fusion is carried out between center pixel relationship and the relationship among the neighboring pixel using the second part of equation. (1). The final binary code is multiplied with the weight assigned to the window, and the sum of the product replaces the center pixel. Fig. 4 and 5 show the output of MLNDP for the given 3X3 cell.

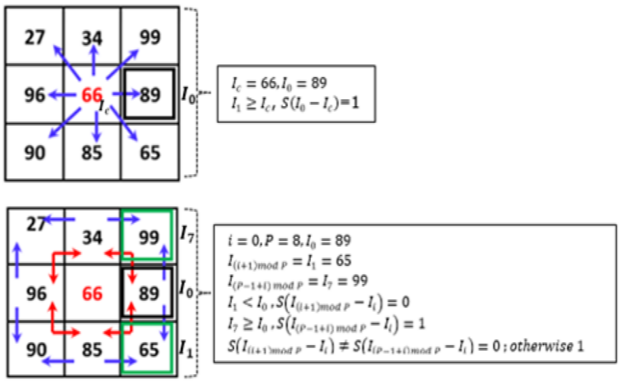


Fig. 4. Sample encoding of MLNDP

Fig. 6 show the sample output of six expressions from CK+ and MMI dataset. The histogram of MLNDP for the given image is given as follows:

$$LNDEP_{P,R}(\tau) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} F(LNDEP_{P,R}(i,j), l); l \in [0, (2^8 - 1)] \quad (6)$$

The MLNDP code yields a more consistent pattern in the existence of noise. For illustration, Fig. 7 shows an original image and its corresponding image with zero mean Gaussian noise. When we compute the LBP and MLNDP code for both images, the MLNDP code remains the same whereas the 6th bit of the LBP code changed from 0 to 1. This example proves that MLNDP code provides an identical pattern in the presence of that noise.

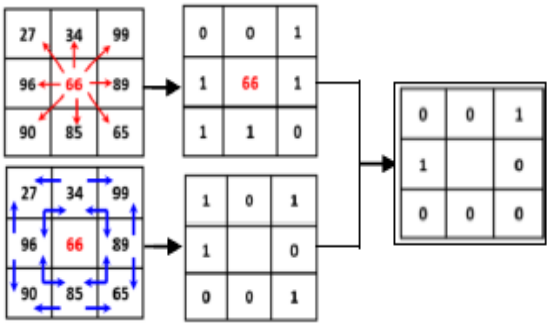


Fig. 5. Neighboring pixel logical fusion using MLNDP

Fig. 6. Sample output of MLNDP (first and the third row is the input image and second and fourth row is its corresponding MLNDP)

Fig. 7. Noise sensitivity analysis of MLNDP

C. Feature selection

This subsection explains the process of selecting the most significant features from the feature space to build a useful predictive model.

Machine learning model performance is usually affected while processing high dimensional data. Feature Selection methods help us to solve these issues by reducing the feature-length without much loss of the overall details. In statistics, the chi-square ($\chi 2$) test is used to check the degree of independence of two events. The feature and class occurrence are the two events that need to be analyzed for feature selection. The chi-square statistics is calculated between every feature and the class variable, and feature selection is performed based on the presence of an association between the features and the class label. The rank is assigned to the feature based on the quantity of ($\chi 2$) which is calculated using equation (7)

$$\chi^2 = \sum_{p_f \in 0,1} \sum_{p_c \in 0,1} \left(N_{p_f p_c} - E_{p_f p_c}\right)^2 / E_{p_f p_c} \quad (7)$$

Where *pf* and *pc* are the feature occurrence and class occurrence, respectively.

N and E are the observed and expected feature values in the given data D. The feature with higher $\chi 2$ value has a higher impact on a dependent feature than the variable with a lower $\chi 2$ value. From the chi-square score based ranked list, the features that are highly correlated to the class labels are identified empirically and classified. Experimental results on the various datasets show the importance of feature selection in the proposed scheme, and it is found to be effective in FER. Algorithm 1 explains the steps involved in proposed approach

Algorithm 1: Steps involved in the MLNDP based FER

Input: Facial image I_{WXH}, No. of neighborhood pixel *P* located at Radius *R*
Output: Selected MLNDP feature based emotion classification

Step 1: For each image in the given database **do**
Perform face detection
Adjust the contrast using CLAHE
Down scale the facial image to the specified size
Step 2: Calculate MLNDP for the given image using equation (1) to (5)
Step 3: Select most contributing features using chi square statistical test
Step 4: Best specified number of features selected based on rank is classified with RBF kernel based multiclass SVM using 10-fold cross validation

D. Evaluation Metrics

The efficacy of the proposed local descriptor is evaluated using four standard measures, such as precision, recall, F1-measure, and accuracy. Conventionally, multi-class accuracy is defined as the average number of correct predictions:

$$Accuracy = \frac{1}{N} \sum_{k=1}^{|G|} \sum_{x:g(x)=k} I(g(x) = \hat{g}(x)) \quad (8)$$

where the indicator function I, returns 1 if the classes match and 0 otherwise. N is the number of observations and G is the number of class labels. $g(x)$ is the actual class and $\hat{g}(x)$ is the predicted class. To be more sensitive to the performance for individual classes, we can assign a weight w_k to every class such that 1. The higher the value of w_k for an individual class, the higher is the influence of observations from that class on the weighted accuracy. The weighted accuracy is determined by:

$$weighted\ accuracy = \sum_{k=1}^{|G|} w_i \sum_{x:g(x)=k} I(g(x) = \hat{g}(x)) \quad (9)$$

To weight all classes equally, we set $w_i = \frac{1}{|G|}, \forall k \in \{1, \cdots, G\}$. Another widely used measure is the F1-measure, which takes into account the trade-off between precision and recall. Precision reflects the percentage of the facial expressions that are appropriately classified in the total dataset. It is mathematically represented by Eq. (19)-(21).

$$F1 - measure\ (j) = \frac{2 \cdot recall\ (j) \cdot precision\ (j)}{recall\ (j) + precision\ (j)} \quad (10)$$

$$precision\ (j) = \frac{correctly\ classified\ as\ positives}{total\ predicted\ as\ positives} \quad (11)$$

$$recall\ (j) = \frac{correctly\ classified\ positives}{total\ positives} \quad (12)$$

Another commonly used metric for the multiclass problem is log-loss that tends to increase as the predicted possibility deviates from the actual label. Log-loss for multi-class is defined as

$$log - loss = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log(p_{ij}) \quad (13)$$

where N is the number of instances in the test data, M is the number of fault class, y_{ij} is 1 if the observation goes to class j; else 0 and p_{ij} is the prediction probability of the instance assigned class j. The macro average is another way of interpreting the confusion matrix in multi-class settings. Here, we need to compute a confusion matrix for every class $g_i \in G = \{1, \cdots, K\}$ such that the i^{th} confusion matrix considers class g_i as the positive class and all other classes g_j with $j \neq i$ as the negative class. The micro averaged precision, recall and F1-measure is defined as follows,

$$precision_{macro} = \frac{\sum_{k=1}^{|G|} P_i}{|G|} \quad (14)$$

$$recall_{macro} = \frac{\sum_{k=1}^{|G|} R_i}{|G|} \quad (15)$$

$$F1 - measure_{macro} = 2 \frac{precision_{macro} \cdot recall_{macro}}{precision_{macro} + recall_{macro}} \quad (16)$$

III. RESULT AND DISCUSSION

In this section, the classification results are reported on two publically available well-known datasets, namely extended Cohn-Kanade (CK+) [32] and MMI [33] by using the method outlined in the previous section.

- The CK+ data set comprises 593 videos of 123 persons, where 31% are male, and 69% are female with the age group between 18-50 years. The image sequences contain universally common facial expressions, namely angry, disgust, fear, happiness, sad, surprise, and neutral.
- The MMI face dataset contains a frontal and profile view with basic facial expressions. In our experiments, the frontal view images (Part II) are considered, which consists of 238 videos of 28 subjects.

Each of the sequences is grouped into six basic emotions. Fig.8 and 9 show the sample images from two datasets.

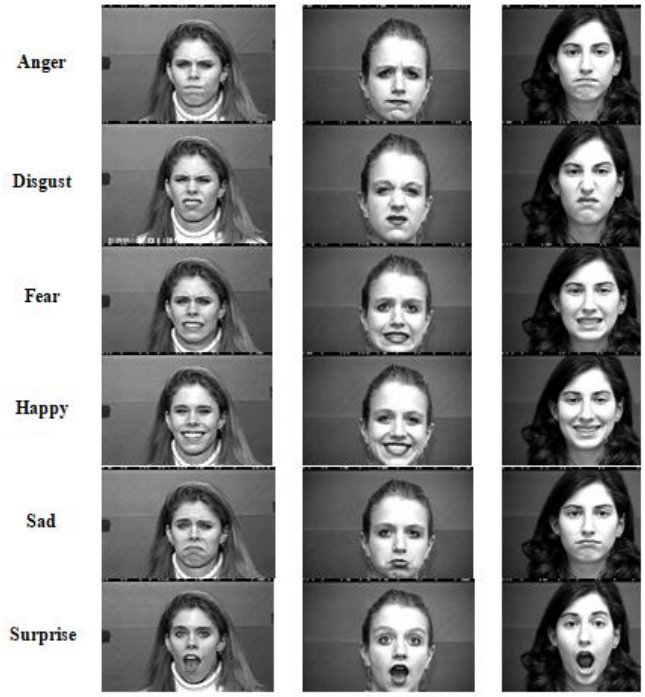


Fig. 8. Sample images from CK+ dataset

After preprocessing and feature extraction, the dataset is divided into tenfold which contains approximately equalized data samples. In this type of cross-validation method, except one fold, remaining folds are used for training. This process continues ten times by keeping one fold for testing. The tenfold cross-validation method is applied in such a way that all data will participate in training at least once. The dataset partitioning is decided based on the number of images taken from each subject, and the results are averaged as the final accuracy. Hyper parameter tuning is performed using a grid search. Three types of experiments are carried out on two datasets. In Experiment I, we compare the proposed MLNDP feature with other related features namely LDP, LTP, LDN, and LGP. In Experiment II, we compare the classifiers for the given proposed feature. In Experiment III, both features and classifiers are compared.

Fig. 9. Sample Images from MMI dataset

Table-I: Performance of MLNDP on CK+ dataset

	LTP			LDP			LGP			LDN			Proposed		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Anger	0.75	0.8	0.77	0.78	0.75	0.76	0.75	0.8	0.77	0.94	0.97	0.95	0.99	0.99	0.99
Disgust	0.66	0.71	0.68	0.69	0.71	0.7	0.66	0.71	0.68	0.93	0.97	0.95	0.96	0.95	0.95
Fear	0.72	0.66	0.69	0.72	0.69	0.7	0.72	0.66	0.69	0.98	0.93	0.96	0.97	0.93	0.95
Happy	0.74	0.7	0.72	0.75	0.81	0.78	0.74	0.7	0.72	0.99	0.97	0.98	0.99	1	0.99
Sad	0.51	0.58	0.55	0.7	0.6	0.65	0.51	0.58	0.55	0.98	0.98	0.98	0.97	0.98	0.98
Surprise	0.84	0.77	0.8	0.78	0.82	0.8	0.84	0.77	0.8	0.99	0.99	0.99	0.99	0.99	0.99
Macro avg	0.7	0.7	0.7	0.74	0.73	0.73	0.7	0.7	0.7	0.97	0.97	0.97	0.98	0.97	0.97
weighted avg	0.72	0.71	0.71	0.74	0.74	0.74	0.72	0.71	0.71	0.97	0.97	0.97	0.98	0.98	0.98
Accuracy	0.71			0.74			0.71			0.97			0.98		

A. Experiment I :Comparison of MLNDP with other descriptors

Here, the SVM is used for feature comparison, which is known to be one of the best multiclass classifiers in the computer vision field. The proposed MLNDP is compared against other local descriptors like LTP, LDP, LGP and LDN and results are reported in Table I and II. The LTP descriptor adds one extra discrimination level relative to LBP. It uses a 3-level coding method with respect to the center pixel with certain threshold (±t).

The

LDP is a binary code of eight bits assigned to each neighbori

Table-II: Performance of MLNDP on MMI dataset

	LTP			LDP			LGP			LDN			Proposed		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Anger	1	0.97	0.99	0.56	0.47	0.51	0.59	0.72	0.65	0.97	0.88	0.92	0.95	0.93	0.94
Disgust	1	0.96	0.98	0.5	0.46	0.48	0.66	0.68	0.67	0.93	0.93	0.93	1	0.93	0.96
Fear	0.98	0.93	0.96	0.55	0.37	0.44	0.82	0.67	0.74	0.9	1	0.95	0.98	0.91	0.94
Happy	0.96	1	0.98	0.48	0.75	0.58	0.9	0.83	0.86	0.94	0.98	0.96	0.96	1	0.98
Sad	1	0.97	0.98	0.55	0.39	0.45	0.86	0.77	0.81	0.97	0.97	0.97	0.97	0.94	0.95
Surprise	0.93	1	0.96	0.49	0.51	0.5	0.55	0.62	0.58	1	0.92	0.96	0.88	1	0.94
Macro avg	0.98	0.97	0.98	0.52	0.49	0.5	0.73	0.72	0.72	0.95	0.95	0.95	0.96	0.95	0.95
weighted avg	0.98	0.97	0.97	0.52	0.51	0.5	0.74	0.72	0.73	0.95	0.95	0.95	0.96	0.95	0.95
Accuracy	0.97			0.51			0.72			0.95			0.95		

ng pixel in the selected cell. This pattern is determined by comparing it in multiple directions,

and the relative edge response value is encoded as the eig ht-bit code. The LGP is calculated as the differences among the specified pixel and its adjacent pixels. Then, the average gradient value of the neighbors is used as a margin for encoding. The LDN represents the local structure by examining maximum positive and negative directional information to produce a meaningful descriptor.

From Tables I. and II. it is observed that the average recognition rate of MLNDP is 97.86%, and the second-highest recognition rate is 97.09% using LDN, which is 0.77% lower than MLNDP on CK+ dataset. This difference indicates that MLNDP is much more capable of encoding the micro-texture. The main problem with other descriptor is that it will generate the same code for two different local structures. This inefficient nature of other descriptors will increase the false detection rate. For the unique feature representation, three local neighborhood pixels are encoded separately and then logically fused. The resultant feature space is further reduced by identifying the most contributing features using a chi-square statistical test.

In our proposed work, only the best 100 features are identified using the chi-square test. After feature selection, the proposed local descriptor provides the highest recognition rate of 99.26% for surprise and the lowest recognition rate of 93.03% for fear on CK+ dataset. Like CK+, MMI also achieves an accuracy of 95.29%, which is slightly better than

other descriptors. The comparison clearly shows that the proposed MLNDP with feature selection performs well

compared to other approaches. The highest misclassification occurs between fear, disgust, anger, and sadness. Because these expressions involve major muscle movements around eyebrow and mouth regions. Besides, eyebrow and mouth muscles are deformed when surprised, whereas some people only raise their eyebrows. This will also create confusion among the expressions. Hence, to improve the detection rate in such situations, temporal information may be incorporated.

In addition, poor representation or pose given by the subject is also one of the reasons for misclassification. In all cases, increasing the number of illustrations per expression in the training set can further improve the recognition rate. Henceforth a secondary feature like forehead wrinkle, nose side wrinkle may also be incorporated to improve the detection rate. Here, the proposed approach handles the images taken from a controlled environment with good recognition accuracy. But, this has to be extended to process real-time data also. Since the proposed non-overlapping local neighborhood feature fusion based MLNDP captures most of the local texture information, it can be more significant to solve the real-time problem. In addition, multiclass SVM prediction probability has less deviated from the actual class, so that log loss is very low when compared with other classifiers. The second best recognition rate is achieved LTP with 97% on the MMI dataset, but its noise handling capability is low.

B. Experiment II: Classifiers performance analysis using MLNDP

In this experiment, four classifiers are considered in addition to the Multiclass SVM. Parameters are set based on grid search method. Tables III and IV show the confusion matrices of SVM classification method on CK+ and MMI dataset respectively. From Table V, average recognition rate using Adaboost classifier is 63.15%, with highest prediction rate of 84% for the happy and lowest prediction rate of 47% for sad. The weighted average accuracy of gradient boosting is 85.74%, with highest prediction rate of 98% for happiness and lowest prediction rate of 70% for fear. The recognition accuracy of the proposed method using KNN and Random Forest (RF) is 94.01% and 81.4%, respectively. The highest recognition rate of those classifiers is achieved on happy.

Table-III: Confusion matrix of CK+ dataset using MLNDP

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**0.99**	0	0	0	0.01	0
Disgust	0.01	**0.95**	0.02	0	0.01	0.01
Fear	0	0.03	**0.93**	0.02	0.01	0
Happy	0	0	0	**1**	0	0
Sad	0	0	0.01	0	**0.98**	0.01

Surprise	0	0	0	0.01	0	**0.99**
Average Recognition = 97.86%						

Similarly in MMI dataset also, MLNDP with SVM achieves an average recognition rate of 95.29% with 100% as the highest individual recognition rate for happiness and surprise and lowest recognition rate for 93% for anger and disgust. The next highest recognition accuracy is realized with KNN and MLNDP combination as 93.59%. The lowest recognition is achieved with Adaboosting classifier with 43.59%. Using multiclass SVM with RBF kernel, the maximum average recognition rate is obtained for the selected feature from the whole face. We have noted that the proposed model with MLNDP+SVM combination provides better accuracy on both datasets than any other classifiers suggesting that the characteristics derived and defined are ro bust to changes in lighting, noise, and other variances in environment. Because the proposed method uses a minimum number of features, this suggests that this work is computationally less expensive. The summary of classifier performance is presented in Table V.

Table-IV: Confusion matrix of MMI dataset using MLNDP

	Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Anger	**0.93**	0	0.03	0.03	0	0.03
Disgust	0	**0.93**	0	0.04	0.04	0
Fear	0	0	**0.91**	0	0	0.09
Happy	0	0	0	**1**	0	0
Sad	0.06	0	0	0	**0.94**	0
Surprise	0	0	0	0	0	**1**
Average Recognition = 95.3%						

Table-V: Classifier performance analysis using MLNDP

Classifier	**CK+**		**MMI**	
	Accuracy %	**Log - loss**	**Accuracy %**	**Log - loss**
KNN	94.01283	0.197993	93.58974	0.262081
RF	81.39701	0.841996	74.78633	1.196589
ABC	63.15039	1.762611	43.58974	1.660928
GBT	85.74483	0.546563	65.38462	0.964417
Multiclass SVM	97.86173	0.085466	95.29915	0.137276

C. Experiment III: Performance analysis of features vs. classifiers

In this experiment, all features considered for comparisons are classified with all five classifiers and the results on CK+ and MMI dataset are shown in Table VI and Table VII and its graphical representation is shown in Fig. 10 and 11 respectively. Due to the various computational parameters such as experimental setup, preprocessing techniques, cross-validation strategies, amount of data used for testing and so forth, experimental results cannot be compared directly. However, the comparison gives insight into the discrimination power of different approaches. As for the features concern, the proposed MLNDP achieves a higher recognition rate using multiclass SVM in both datasets. This combination is accurate in learning local facial representations with feature selection because it includes two types of neighboring pixel relationship. Integration with global representation adds more strength to the proposed method. Even the presence of noise also will not affect the proposed MLNDP code. Compared with other combinations of feature with classifiers, MLNDP+SVM can discover local exclusionary information.

This type of feature fusion capture most of the texture details from the local neighboring pixel relationship, and this non-overlapping feature fusion eliminates redundant information. Hence, the combination of three neighboring pixel-based MLNDP descriptor is an effective method for FER. Nevertheless, there is no guarantee that this combination performs well on discriminative facial regions when expression intensity decreases. Here, the proposed approach handles the images taken from a controlled environment with good recognition accuracy. However, this has to be extended to process real-time data also. Since the proposed non-overlapping local neighborhood feature fusion based hybrid feature captures most of the local texture information, it can be more significant to solve the real-time problem.

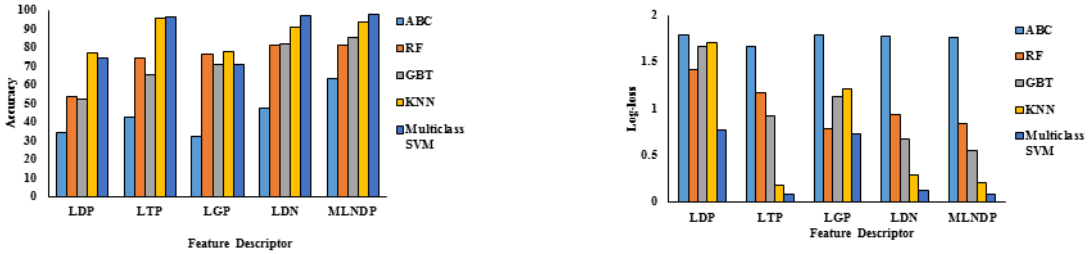
Table-VI: Local descriptor performance on various classifiers on CK+ dataset

Descriptor Classifier	**LDP**		**LTP**		**LGP**		**LDN**		**MLNDP**	
	Accuracy	**Log-loss**	**Accuracy**	**Log-loss**	**Accuracy**	**Log-loss**	**Accuracy**	**Log-loss**	**Accuracy**	**Log-loss**
ABC	34.3	1.79	42.3	1.66	32.4	1.78	47.7	1.77	63.2	1.76
RF	53.8	1.42	74.4	1.17	76.5	0.79	81.4	0.94	81.4	0.84
GBT	52.6	1.67	65.4	0.92	70.6	1.13	81.9	0.68	85.7	0.55
KNN	76.9	1.71	95.7	0.18	77.6	1.21	91.3	0.29	94	0.2
Multiclass SVM	74.4	0.76	96.6	0.09	71.3	0.73	97.1	0.13	97.9	0.09

Table-VII: Local descriptor performance on various classifiers on MMI dataset

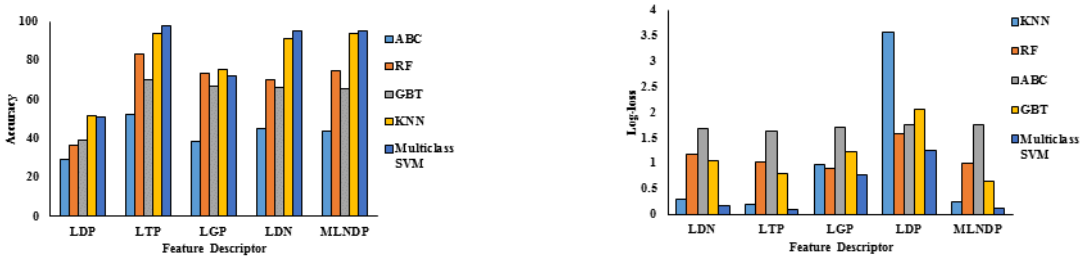
Descriptor Classifier	**LDP**		**LTP**		**LGP**		**LDN**		**MLNDP**	
	Accuracy	**Log-loss**	**Accuracy**	**Log-loss**	**Accuracy**	**Log-loss**	**Accuracy**	**Log-loss**	**Accuracy**	**Log-loss**
ABC	29.06	1.75	52.56	1.62	38.46	1.7	45.3	1.68	43.59	1.66

RF	36.75	1.59	82.91	1.03	73.08	0.9	70.09	1.17	74.79	1.2
GBT	39.32	2.06	70.09	0.8	66.67	1.22	66.24	1.04	65.38	0.96
KNN	51.71	3.56	94.02	0.19	75.21	0.98	91.03	0.3	93.59	0.26
Multiclass SVM	50.85	1.26	97.44	0.09	72.22	0.78	94.87	0.17	95.3	0.14



(a) (b)

Fig. 10. Accuracy (a) and Log-loss (b) of features vs. classifiers on CK+ dataset



(a) (b)

Fig. 11. Accuracy (a) and Log-loss (b) of features vs. classifiers on MMI dataset

IV. CONCLUSIONS

This work analyzes the potential of local neighboring pixel relationships in pattern recognition. Usually, local descriptors are derived by an encoding relationship of the neighboring pixel in different way. This paper presents the new texture descriptor called Merged Local Neighborhood Difference Pattern (MLNDP) for facial expression recognition. This work encourages the use of an intensity amplification scheme to tackle variance in gradient discrepancy and makes this reliable against pose variations, slight occlusion and moderate local lighting issues left in the preprocessing stage. The feature selection is accomplished using chi-square statistical test to increase the performance of the classifier. From the findings, it is evident that the proposed methodology outperforms all of the methods such as LDP, LTP, LDN, and LGP on both datasets with 97.86% on CK+ dataset and 95.29% on MMI dataset where the images are taken under the controlled environment. The combination of MLNDP and multiclass SVM performs better than other local descriptors and classifiers. In addition, holistic approach achieves considerably higher performance by retaining information at the pre-processing stage which captures accurate information for identification of facial expression. In addition to poor lighting, these images can contain different levels of noise. In such situation also the proposed MLNDP is stable and retains the same code. Our future research target is to develop the best computational model to handle occlusion and head pose efficiently in order to achieve more promising results.

ACKNOWLEDGMENT

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

REFERENCES

1. P. Ekman, and K. Dacher, “Universal facial expressions of emotion,” *Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture* (1997): 27-46.
2. M. Liu, S. Li, S. Shan, R. Wang, and X. Chen, “Deeply learning deformable facial action parts model for dynamic expression analysis.” *In Proceedings of the Asian Conference on Computer Vision, Singapore*, 1–5 November 2014; pp. 143–157.
3. C. H. Chen, I. J. Lee, and L. Y. Lin, “Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders”, *Res. Dev. Disabil*. 2015, 36, 396–403.
4. S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa, “ Eyemotion: Classifying facial expressions in VR using eye-tracking cameras”, In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1626-1635). IEEE.
5. Y. I. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis”, *IEEE Trans. Pattern Anal. Mach. Intell*. 2001, 23, 97–115.
6. H. Yan, J. Lu, X. Zhou, “Prototype-based discriminative feature learning for kinship verification”, *IEEE Trans. Cybern*. 2015, 45, 2535–2545.
7. H. Yan, “Transfer subspace learning for cross-dataset facial expression recognition”, *Neurocomputing* 2016, 208, 165–173.
8. H. Yan, “Biased subspace learning for misalignment-robust facial expression recognition”, *Neurocomputing* 2016, 208, 202–209.
9. N. Wang, X. Gao, D. Tao, H. Yang, and X. Li, “Facial feature point detection: A comprehensive survey”, *Neurocomputing* 2018, 275, 50–65.

10. M. Turk, and A. Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, *3*(1), 71-86.
11. M. Nazir, J. Zahoor, and M. Sajjad. "Facial expression recognition using histogram of oriented gradients based transformed features." *Cluster Computing* 21.1 (2018): 539-548.
12. T. Kalsum, S. M. Anwar, M. Majid, B. Khan, and S. M. Ali, "Emotion recognition from facial expressions using hybrid feature descriptors." *IET Image Processing* 12.6 (2018): 1004-1012.
13. Rivera, Adín Ramírez, J. R. Castillo, and O. Chae. "Local directional texture pattern image descriptor." *Pattern Recognition Letters* 51 (2015): 94-100.
14. Ojala, M. Pietikainen, and D. Harwood , "A comparative study of texture measures with classification based on featured distribution," *Pattern Recog*., 29(1) 51–59,1996.
15. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *in: Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
16. F. Schroff, D. Kalenichenko, and J. Philbin, Facenet, "A unified embedding for face recognition and clustering", *in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823
17. A. Toshev and C. Szegedy, and Deeppose, "Human pose estimation via deep neural networks", *in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1653–1660.
18. X. Zhao et al., "Peak-piloted deep network for facial expression recognition", *European Conference on Computer Vision, Springer*, 2016, pp. 425–442.
19. H. Ding, S.K. Zhou, and R. Chellappa, "Facenet2expnet: Regularizing a deep face recognition net for expression recognition", in: Automatic Face & Gesture Recognition (FG 2017), 2017 *12th IEEE International Conference on, IEEE*, 2017, pp. 118–126.
20. G. Benitez-Garcia, T. Nakamura, and M. Kaneko, "Facial expression recognition based on local fourier coefficients and facial fourier descriptors", *J. Sign. Inform. Process*. 8 (2017) 132.
21. S. Naik, and R.P.K. Jagannath, "GCV-Based Regularized Extreme Learning Machine for Facial Expression Recognition", *In Advances in Machine Learning and Data Science; Springer: Singapore*, 2018; pp. 129–138.
22. S. Benini, K. Khan, R. Leonardi, M. Mauro, P. Migliorati, "Face analysis through semantic face segmentation", *Signal Process. Image Commun*. 2019, 74, 21–31.
23. V. K. Verma, S. Srivastava, T. Jain, A. Jain, "Local Invariant Feature-Based Gender Recognition from Facial Images", *In Soft Computing for Problem Solving; Springer: Berlin/Heidelberg, Germany*, 2019; pp. 869–878.
24. J. García-Ramírez, J.A. Olvera-López, I. Olmos-Pineda, M. Martín-Ortíz, "Mouth and eyebrow segmentation for emotional expression recognition using interpolated polynomials", *J. Intell. Fuzzy Syst.* 2018, 34, 1–13.
25. N. Zeng et al., "Facial expression recognition via learning deep sparse autoencoders", *Neurocomputing* 2018, 273, 643–649.
26. M.S. Islam, "Local gradient pattern-A novel feature representation for facial expression recognition", *J. AI Data Min.* 2 (2014) 33–38.
27. T. Jabid, M.H. Kabir, and O. Chae, "Local directional pattern (LDP)–A robust image descriptor for object recognition, Advanced Video and Signal Based Surveillance (AVSS)", *in: 2010 Seventh IEEE International Conference on, IEEE*, 2010, pp. 482–487.
28. F. Bashar, A. Khan, F. Ahmed, and M.H. Kabir, "Robust facial expression recognition based on median ternary pattern (MTP)", *in: Electrical Information and Communication Technology (EICT), 2013 International Conference on, IEEE*, 2014, pp. 1–5.
29. A.R. Rivera, J.R. Castillo, and O.O. Chae, "Local directional number pattern for face analysis: Face and expression recognition", *IEEE Trans. Image Process*. 22 (2013) 1740–1752.
30. A. Majumder, L. Behera, and V. K. Subramanian, "Automatic facial expression recognition system using deep network-based data fusion" *IEEE transactions on cybernetics*, 2016, *48*(1), 103-114.
31. S.M. Pizer, E.P. Amburn, J.D. Austin, "Adaptive histogram equalization and its variations", *Comput. Vision Graphics Image Process*. 39 (1987) 355–368
32. P. Lucey et al., "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression", (2010, June), *In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops* (pp. 94-101). IEEE.
33. M. Valstar, and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database", (2010, May), *In Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect* (p. 65).

AUTHORS PROFILE

P. Shanthi received her B.E. degree in computer science and engineering from Baradhidasan University, Tamilnadu, India in 2002, and M.E (CSE) in 2012 from Anna University, Tamilnadu, India. She is currently pursuing Ph.D. degree at the Department of Computer Applications, NIT, Trichy, India. Her current research interests include object detection and tracking, pattern recognition, facial analysis, and computer vision.

S.Nickolas is a Professor in the Department of Computer Applications, National Institute of Technology, Trichy, Tamilnadu, India. He received his M.E. Computer Science from REC, Trichy in 1992 and Ph.D. in the year 2007 from NIT, Trichy. He is the Professor In-Charge of the Massively Parallel Programming Laboratory, NVIDIA CUDA Teaching Centre, NIT, Trichy. His research interest includes Evolutionary Algorithms, Data Mining, Big Data Analytics, Distributed Computing, Cloud Computing, and Software Metrics.