Received: June 30, 2021. Revised: September 3, 2021.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Gradient-Based Compact Binary Coding for Facial Expression Recognition

A. Vijaya Lakshmi[1]* **P. Mohanaiah[2]**

[1]*Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu, A.P, India*
[2]*Department of Electronics and Communication Engineering, N.B.K.R Institute of Science and Technology, Vidyanagar, A.P, India*
* Corresponding author's Email: vijayalakshmivardhan@gmail.com

Abstract: Facial Expressions are one of the most essential and general means for human beings to express their feelings and communicate their emotions; thus, Facial Expression Recognition (FER) has gained a significant research interest and an attractive topic in Human-Computer Interactions (HCI). Towards such, this paper proposes a novel face descriptor, Gradient Direction Pattern (GDP), for facial expression recognition. With the help of gradients, GDP encodes the structure of the facial image in a more compact way such that the FER is robust to pose variations. Furthermore, the GDPs are measured for edge feature maps, extracted from Gaussian filtered and Gabor filtered facial images at different orientations. These edge feature maps make the recognition system robust to noise, illumination, scaling, and orientational variations. Initially, the facial image is divided into small regions, and GDPs are extracted from them. Then these patterns are concatenated, and Histograms are measured, called GDP Histograms (GH). Simulation experiments were conducted over two standard datasets, such as CK+ and JAFFE, and the average accuracy is observed as 95% and 91% approximately.

Keywords: Expression recognition, Gradients, Directional pattern, Gabor map, Gaussian map, Local binary pattern.

1. Introduction

In recent years, due to the increased availability of powerful computers and electronic devices, human-centered user interfaces have gained much popularity, responding quickly to naturally occurring human communications [1]. An important responsibility of such interfaces is to understand and analyze the emotions represented by facial expressions. Facial Expressions are the most effective and natural tools which allow humans to interact with each other, expressing their intentions, and communicate their emotions [2]. All these aspects have emphasized the necessity of automatic Facial Expression Recognition (FER), which has been an important research topic since past few years.

In FER, the vital issue is to develop an efficient face descriptor that covers all the details of face appearance [3]. The effectiveness of the face descriptor depends on its representation and the way of extracting it from a facial image. A facial descriptor is useful when it has a high variance among several classes like different persons or different expressions and less variation within classes like the same person or same expression under different environments. Several methods are developed earlier, focusing mainly on the extraction of active features from the facial image. Based on the methodology of feature extraction, the earlier FER methods are classified as Geometric feature-based and Appearance-based methods [4, 5]. In the former case, the features are extracted based on the localization of facial landmarks and facial geometry. These methods used the facial image characteristics such as shape, location of facial components (including nose, eyebrow, eyes, and mouth), and distance between pairs of facial landmark points. Though these methods have been achieved a compelling performance in the FER [6, 7], they are susceptible to misalignments of the face due to the

improper detection and tracking of facial landmark points under various challenges like illumination, low-resolution, and occlusions, etc. The appearance-based methods used image filters either on the whole face or on some portions of the face [8]. The appearance-based methods extract global features when the filters are applied on the whole face image and creates local features when the filters are applied on some portions. Since these methods can discover the appearance changes in the facial image, they have gained an excellent performance compared to the geometric-based methods.

In earlier, there are so many methods that are developed to extract the global feature descriptor from the facial image. Eigenfaces extracted through Principal Component Analysis (PCA) [9], and Fisher faces extracted from linear discriminant analysis (LDA) [10] are the two best examples that have been widely used in facial expression recognition [11]. However, these methods are not robust for pose variations and illumination changes in the facial images. Local Descriptors have gained significant interest due to their capacity to analyze the illumination and pose variations locally. Local Binary Pattern (LBP) [12] is the most effective local descriptor which analyses the texture of facial images, and hence it has achieved better performance in the FER. However, this method suffers from several constraints, such as changes in the pose, age, monotonic illumination variations, and the expression environments. Moreover, the LBP cannot discover the direction of emotion.

To solve these problems, in this paper, we develop a new face descriptor, Gradient Direction Pattern (GDP), for effective emotion recognition from facial expressions. GDP encodes the intensity variations and structural variations at different scales and orientations. Initially, the edge feature maps are extracted with the help of Gaussian Filter and Gabor filter, and then the GDP encodes the structure of a local neighborhood based on the direction information. The edge feature map obtained through Gaussian Filter helps in the discovery of intensity variations due to illuminations and noise. Next, the edge feature map obtained through Gabor Filter helps in the detection of scaling and rotation variations.

The rest of the paper is organized as follows; Section 2 reveals the details of the literature survey. Section 3 reveals the details of the proposed descriptor for face and emotion recognition. Further, the simulated results are discussed in Section 4. Finally, we present concluding remarks in section 5.

2. Literature survey

For an automatic recognition of emotions from facial images, several approaches are developed. The common aspect of all these approaches is to detect the face region and extracting the geometric features or appearance features. In the case of geometric features, the feature descriptor is constructed based on the relationship between various facial landmarks [13, 14]. For example, the facial expression recognition system developed by D. Ghimire, and J. Lee [14] focused on the two geometric features such as position and phase angle of 52 facial landmark points. First, this method measured the Euclidean distance and angle for each pair of landmark points of a current facial image frame in the video sequence. Next, these angles and Euclidean distances are subtracted from the first facial images frame of the video sequence. Moreover, this system accomplished two machine learning algorithms, such as the Adaboost algorithm and Support Vector Machine Algorithm for emotion recognition. However, the considered two geometric features, angle, and Euclidean distance, are sensitive to illumination and brightness variations in the facial image. These methods work well for images that have constant illumination.

On the other hand, the main advantage of appearance-based features is they can alleviate the variations due to illuminations, poses, and captured environments. Generally, the appearance features are extracted from the global face region [15] or different regions of a facial image, carrying various types of information [16, 17]. For example, the facial expression recognition method proposed by S. L. Happy, A. George, A. Routray [15] proposed extracting feature vectors from the global face region using LBP Histogram with different sizes of blocks. Next, the emotion classification is accomplished through PCA. Though this method had achieved a good performance in real-time, the recognition accuracy tends to decrease because the feature vectors are not consistent with the local variations of the facial components. The global features won't contribute local information, which is very much crucial for noise analysis. Hence the local variations need to be considered because different face regions have a different significant level of importance. For instance, the mouth, eyebrows, and eyes convey more information about the emotion than the forehead and cheek. Unlike the global appearance features, the local feature methods compute the features from local regions of the face and then gather information into one feature vector [27]. Some examples for local region-based

appearance analysis methods are Local Features Analysis [18], Gabor Features [19], Elastic Bunch Graph Mining [20], and LBP [12]. Among these methods, LBP has gained much importance due to its effective analysis of the image's texture. In the remaining methods, they can't analyze the edge features of an image with significant muscle movement. Generally, most local region analysis methods employed LBP for feature extraction. But, there are several problems with LBP. The first problem is limited accuracy and the second is colossal information loss. Finally, it makes the recognition method very much sensitive to noise.

Some variants are proposed to overcome these problems with LBP, for example, Local Ternary Pattern (LTP) [21], Local Directional Pattern (LD_iP) [22, 24], and Local Derivative Patterns (LD_eP) [23]. Instead of pixel intensity information, the LD_iP encodes the directional information in the neighborhood, and LD_eP uses higher-order derivatives. Both methods use other information to overcome illumination and noise problems. Although these methods use more information to encode the pixel, which produces a stabilized binary code for each pixel, they still encode the information similarly to LBP. Despite this simple coding strategy, these methods also discard most of the information from the neighborhood. For example, these methods do not focus on the sign of a directional gradient, which can make the code flip, resulting in code with different characteristics. Moreover, these methods are very much sensitive to noise, illuminations, scaling, and orientations.

Some more feature extraction methods are developed based on LBP and combining it with several handcrafted methods. M. Z. Uddin et al. [25] proposed a new feature extraction method called as Local Directional Position Pattern (LDPP) for FER. The texture features are extracted using LDPP, PCA and Generalized Discriminant Analysis (GDA). The expressions are characterized using Deep Belief Network (DBN). However, the PCA transforms the pixels into principal components which are not readable and interpretable as original pixels. Moreover, for PCA, data standardization is mandatory, which diminishes the range of facial image pixels.

M. Sajjad et al. [26] combined the Uniform-Local Ternary Operator (U-LTP) with a Histogram of Oriented Gradients (HOG) to describe the texture and shape features of the entire face of an image. The two features are concatenated into a single vector and classified using SVM. However, the HOG features are susceptible to rotation because they do not determine pixel movements from the local region to region. Moreover, this method is not concentrated on the relation between neighbor pixels.

Combining the covariance matrix with K-L transform with Extended LBP (ELBP), M. Guo X. Hou, Y. Ma [28] proposed a FER recognition system. First, ELBP is used to extract the feature matrix of facial expression images, and then the covariance matrix transform is accomplished for dimensionality reduction. And finally, recognition is done using an SVM classifier. The covariance matrix determines the linearity between pixels, and upon the occurrence of sudden muscle movement, the linearity does not exist, and the corresponding pixel is considered as a redundant pixel.

M. Abdul and R. S. Holambe [29] combined Directional wavelet transform (DIWT) with LBP to extract the facial features. Initially, the facial image is decomposed into directional sub-bands, and an adaptive direction selection method is accomplished based on quadtree partitioning to obtain top-level DIWT sub-bands. Next, the LBP histograms are extracted from the selected top bands to get a local descriptive feature set. Once the image is transformed into sub-bands through DIWT, the correlation between image pixels brokes and the accomplishment of LBP over sub-bands does not show much significance in the expression recognition.

The method proposed by M. L. Seyed, and Z. M. Hussain [30] combined LBP with Local Phase Quantization (LPQ) and Log-Gabor filter. Initially, the facial image is adopted for feature map extraction through the Gabor filter [39] at five scales and eight orientations. Then the Gabor maps are encoded with LBP and LPQ and then processed for classification through the SVM algorithm. However, the directional information is not considered during the encoding process through LBP. Similarly, the FER system proposed by I. M. Revina and W. R. S. Emmanuel [31] proposed an extension to the local Directional Pattern, called Dominant Gradient Local Ternary Pattern (DGLTP), to extract the local dominant texture features of a facial image. In pre-processing, to remove the noise, a new filter called Enhanced Modified Decision Based Unsymmetric Trimmed Median Filter (EMDBUTMF) is used. Finally, the histogram features are extracted through DGLTP and then fed to SVM classifier for emotion classification. Though this approach employed directional information, the method is not robust for scaling and rotational variations.

S. Sammaiah and K. V. Rao [40] proposed a salient binary coding scheme that encodes the directional information after extracting edge features

through two edge filters, namely the Gaussian filter and the Robinson filter. This method is an extended version of LBP, which encodes the directional information. However, this method is sensitive to image rotations.

Recently, some more methods are developed with the aim of robustness in FER by applying a deep learning-based classifier. M. M. Thiruthuvanathan, and B. Krishnan [41, 42] adopted Convolutional Neural Networks (CNNs) and enabled them with residual components to enhance the learning rate of the network. They employed this model to classify classroom engagement through facial expression recognition. The residual network consists of a shallow depth of 50 layers and is applied to Diasee dataset. However, these methods do not employ either geometrical or appearance-based feature extraction methods. Due to this reason, they suffer from so many problems like a vast computational burden, sensitivity to noise and edge cuttings etc.

K. A. El Dahshan et al. [43] proposed FER based on Deep Belief Networks (DBN) and Quantum Particle Swarm Optimization (QPSO). Initially, the facial image is pre-processed for Region of Interest (ROI) extraction, and it was divided into several blocks. Then the size of the image is reduced by downsampling and then applied DBN for classification. The parameters of DBN are optimized through QPSO. However, this method didn't focus on the local feature analysis, making the recognition system less efficient and not robust for noise.

3. Proposed approach

3.1 Overview

This section describes the details of the proposed emotion recognition mechanism. This mechanism is composed of three principal stages, 1) Face Feature Edge map Construction, 2) Face Descriptor through Gradient Directional Pattern (GDP) and 3) Emotion Recognition. In the first stage, given the facial image's expression, this work constructs two different feature edge maps based on the Intensities and Orientations of pixels. The Intensities-based feature edge map is constructed with the help of the Gaussian filter, and the Orientations-based feature edge map is constructed with the help of the Gabor filter. In the second stage, the two feature edge maps are processed for GDP evaluation followed by face descriptors through Histograms. The final step executes emotion recognition through the Support Vector Machine (SVM) Classifier. Figure.1 shows the block diagram of the proposed Face Expression recognition mechanism.

3.2 Edge feature map construction

Given a facial image containing an expression/emotion, we would like to describe it with informative features extracted from it. The descriptor is expected to capture the intensity and orientation information, the primary cues of an expression/emotion. Under this phase, two filters are applied over the input facial image to construct two different feature maps. The main focus of this feature map construction is to highlight the edge features of an image. Since the edges are more prominent features by which emotion can be distinguished. The two filters are namely Gaussian Filter and Gabor filter. The details are discussed in the following subsections.

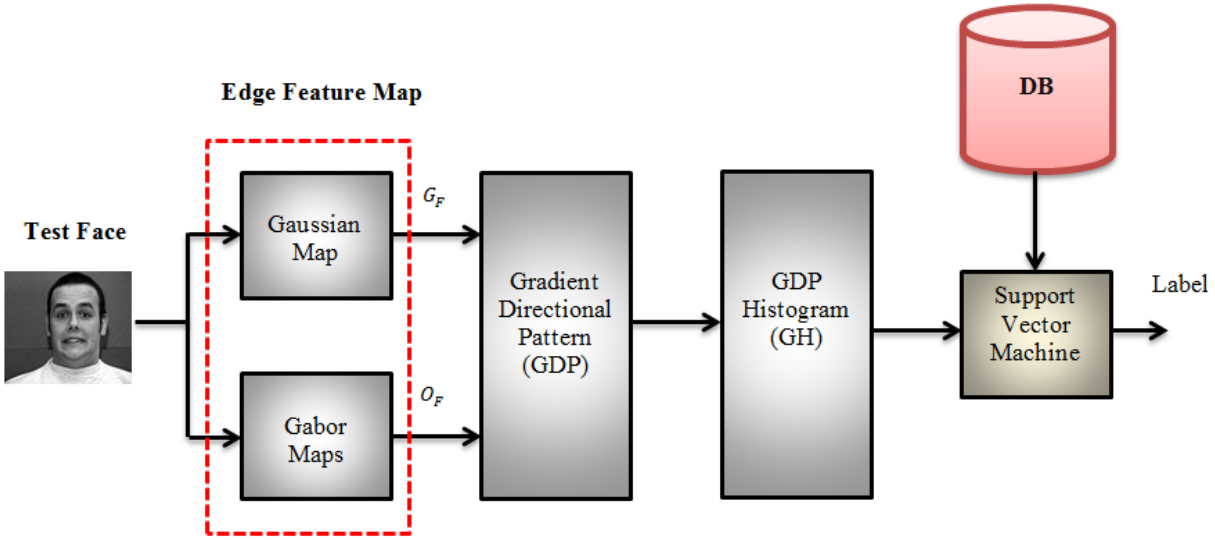


Figure. 1 Block diagram of proposed FER system

3.2.1. Gaussian edge feature map

The main intention of the Gaussian feature map is to represent the facial emotion features robust to illuminations. Further, the Gaussian smoothing results in a stabilized pattern in the presence of noise using the Gaussian mask. Here the proposed Gaussian Feature map construction is inspired by Center Surround (CS) [32] field, which has been identified long ago in the human visual system. The main advantage of CS theory is observed at the enhancement of edges that ensures the detection, location, and tracking of small objects. After the CS accomplishment, the features with different scales, like boundaries and edges, are enhanced. CS operation is successfully used to acquire the intensity information for emotion classification [33]. Based on the inspiration of CS theory, we proposed to construct a seven-level Gaussian Pyramid on the input facial image. The seven-level Gaussian pyramid is constructed by convolving the several copies of input facial image with Gaussian Filter ($\sigma = 2$). With an increase in the pyramid's level, the image's size is reduced through downsampling. At the first level of the Gaussian pyramid, the original input facial image is convolved with the Gaussian Filter. At the second level, the same Gaussian Filter is processed for convolution with the downsampled version of the original input image. In this manner, we construct a seven-level Gaussian Pyramid. Fig.2 shows the process of Gaussian Feature edge map construction. The Gaussian Pyramid construction is done as follows;

$$G_l(x,y) = \sum_i \sum_j W(i,j) \times G_{l-1}(2x+i, 2y+j) \quad (1)$$

Where

$$W(i,j) = \frac{1}{2\pi\sigma^2} e^{\frac{-i^2+j^2}{2\sigma^2}} \quad (2)$$

Where l denotes the level of Gaussian Pyramid, and (x, y) denotes the pixel position in the facial image. Next, the Gaussian CS feature map is measured by subtracting the pixel-by-pixel between various initial and final levels. Here the initial levels are considered as 2 and 3, and the final level is obtained as $Final\ Level = maximum\ level - initial\ level$. Based on this formula, in this proposed method, extracted two feature maps, one is from the second and fifth levels, and another is from the third and fourth levels. The final Gaussian Feature edge map is obtained through max pooling. Max pooling is applied on pixel-by-pixel between two feature maps, and finally, one feature map is constructed, considered as Final Gaussian Feature Edge map, mathematically represented as:

$$G_F(x,y) = \max(FM_{2-5}(x,y), FM_{3-4}(x,y) \quad (3)$$

Where G_F is the final Gaussian Feature Edge map, FM_{2-5} is the feature map obtained by the subtraction of pixel-by-pixel from Gaussian feature map at the second level and fifth level, FM_{3-4} is the feature map obtained by the subtraction of pixel-by-pixel from Gaussian feature map at third level and fourth level, and (x, y) denotes the pixel position in every feature map.

For a subtraction process, the sizes of the two matrices must be the same. Still, the sizes of initial and final frames are not the same because, as the Gaussian pyramid increases, the size of the frame decreases gradually due to downsampling. Hence to obtain the same size as the final level, it is interpolated into the size of the frame at the initial level and then subtracted.

3.2.2. Gabor edge feature map

The main intention of the Gabor filter is to extract the orientation features that play an essential role in recognizing emotions from facial images under different orientations. These features are more effective in providing proper discrimination between various emotions under multiple orientations [34].

Towards the extraction of orientations features, Gabor filter is used due to its property of orientation selection. Gabor filter is a widely used concept in several domains to obtain orientation information. Here the Gabor filter is accomplished at various scales such as 5×5, 7×7, 9×9, and 11×11, and eight orientations such as 0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, and 315^0. So totally for an input facial image, we will get $4 \times 8 = 32$ feature maps. Fig.3 shows the process of Gabor Feature edge map construction. The necessary mathematical representation for Gabor filter is stipulated as:

$$G(x,y) = exp\left(\frac{X^2+\gamma Y^2}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda}x\right) \quad (4)$$

$$X = xcos\theta - ysin\theta, \quad Y = xsin\theta + ycos\theta \quad (5)$$

Where (x, y) is the position relative to the center of the filter. According to the mathematical expressions (4) and (5), the θ value varies from 0^0 to 315^0 with an angular deviation of 45^0. For instance, consider the scale 7×7; initially, the action frame is scaled, and then the Gabor filer is applied over it for eight orientations. Similarly, the facial image is

processed for the remaining scales and obtained a total of 32 feature maps. Further to achieve the main Gabor feature maps, applied max-pooling operation among different scales. For every orientation, we obtained four feature maps at four different scales. The max-pooling mechanism evaluates only one feature map with maximum dominant features in that particular orientation. Let's $S = \{5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11\}$ and $\theta = \{0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, 315^0\}$, the max-pooling at different scales is formulated as:

$$O_{F,i} = \max_{\substack{i=1\ to\ length(\theta) \\ j=1\ to\ length\ (s)}} (x,y)\{f_{S_i}(x,y,\theta_j)\} \quad (6)$$

Where f_{S_i} represents the feature map at i_{th} orientation and θ_j represents the j_{th} scale. For $i = 1$, the orientation $\theta = 0^0$ is picked up and the feature maps obtained at four scales are chosen, and the expression (3) picks up the final feature map with all maximum values. For a given co-ordinate (x, y), the expression (3) searches for maximum value in the total four feature maps obtained at i_{th} orientation. Finally, eight feature maps are obtained, covering almost all scale and rotation invariant features for a given facial image. These feature maps are the Gabor filter responses that have dominant features at respective orientations.

3.3 Gradient directional pattern

The GDP is a six-bit binary code assigned to each pixel of an input facial image to represent the texture and structure and its intensity variations. According to some previous studies [35, 36], edge magnitudes are very insensitive to lighting changes.

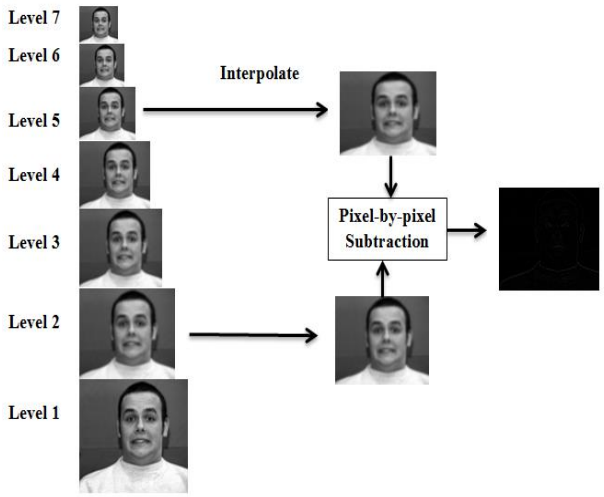


Figure. 2 Gaussian feature edge map construction

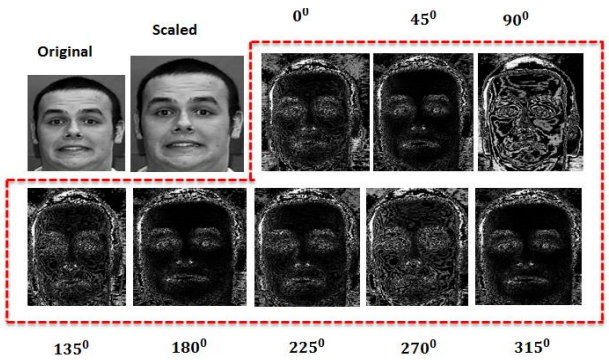


Figure. 3 Gabor feature edge map construction

Hence we compute our GDPs over the edge responses of input facial images. The edge responses are obtained through Gaussian Filter and Gabor filter.

The Gaussian filter produces the edge feature maps that are robust to illuminations and noise. Next, the edge feature maps produced through Gabor filer are robust to scaling and rotational variations. The proposed GDP is based on directional gradients observed in the edge feature maps. Here the GDP is evaluated for the Gaussian edge feature map based on the pixel intensities of local neighborhood pixels. In contrast, the GDP for the Gabor edge feature map is assessed based on the gradients of the same pixels in all eight feature maps. The GDP of a pixel is simply related to the maximum and minimum values of gradients in all eight feature maps. The positive and negative gradients provide valuable information about the structure of the neighborhood, as they reveal the direction of the gradient of dark and bright regions in the neighborhood. Hence this distinction between dark and bright responses allows GDP to differentiate between blocks with positive and negative direction swapped by generating a different code for every pixel. It helps in recognition of emotions more accurately. For example, the eyebrows and mouth have different intensity variations when they move up and down in some emotions. For example, in the surprise emotion, the eyebrows raise, and the mouth will open, whereas, in the sad emotion, the eyebrows and mouth move in the opposite direction. These variations are effectively covered by GDP, which helps in the recognition of emotions that have even minor variations. Under this phase, the GDP is evaluated separately from Gaussian edge feature maps and Gabor edge feature maps.

3.3.1. GDP for Gaussian maps

GDP over a pixel (x, y), in the Gaussian Edge feature map is a six-bit binary code, obtained by the concatenation of maximum gradient position and

minimum gradient position. Simply a pixel is represented with maximum and minimum variations in its neighborhood. The detailed process of GDP evaluation for a Gaussian edge feature map is shown in Fig. 4.

3.3.2. GDP for Gabor maps

GDP over a pixel (x, y) in Gabor edge feature maps is a six-bit binary code obtained by concatenating maximum and minimum response positions. To obtain a scale and rotation insensitive GDP, this process considered a total of eight edge feature maps obtained through the Gabor filter. The GDP obtained a pixel (x, y) are scale and rotation invariant, i.e., it can capture the scaled and rotated version of pixel (x, y). The detailed process of GDP evaluation for a Gabor edge feature map is shown in the Fig. 5.

3.3.3. Final GDP

Once the GDPs are measured from both the Gaussian edge feature map and the Gabor edge feature map, the final GDP is obtained by the horizontal concatenation, resulting in a 12-bit binary code. Compared to the normal LBP, GDP is more effective in preserving information as the LBPs are constructed based on the pixel intensities of neighborhood pixels. Furthermore, the LBP is not robust to scaling and rotational variations in the image. GDP provides a more compact representation through the gradients of a facial image, which helps in recognize emotions even with scaling, rotation, and illumination variations.

3.4 Face descriptor

To alleviate the fine to coarse information of a facial image and its local textures, spots, edges, and corners, this work focused on representing it through GDP Histogram (GH). Generally, the histogram represents an image by discovering the occurrences of certain micro-patterns without local information. Hence to aggregate the local information to the face descriptor, we divide the facial image into several blocks, $\{B_1, B_2, \ldots, B_N\}$ and measures histogram H_i from each block B_i. Here, each code is considered as a bin, and the occurrences are aggregated to create a histogram H_i, as:

$$H_i(c) = \sum_{\substack{GDP(x,y)=c \\ (x,y)\in B_i}} C \tag{7}$$

Where (x, y) denotes the pixel position in the block B_i, c is a binary GDP code and $GDP(x, y)$ is a

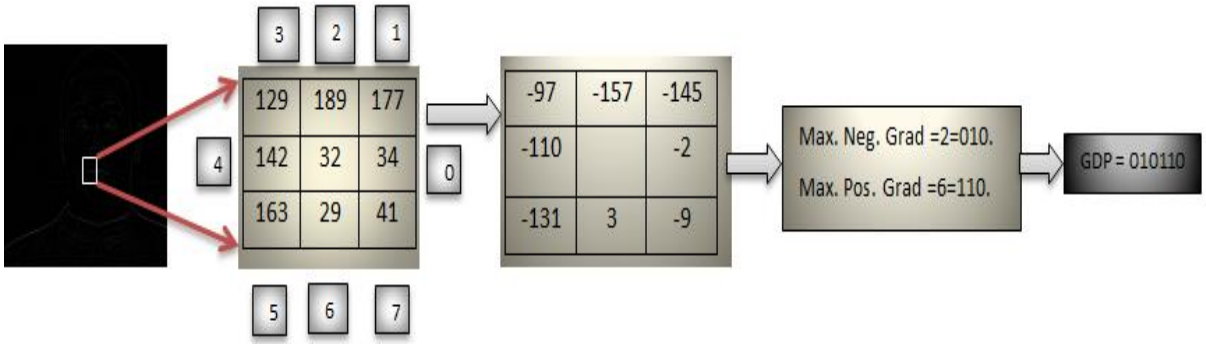


Figure. 4 GDP evaluation from Gaussian edge feature map

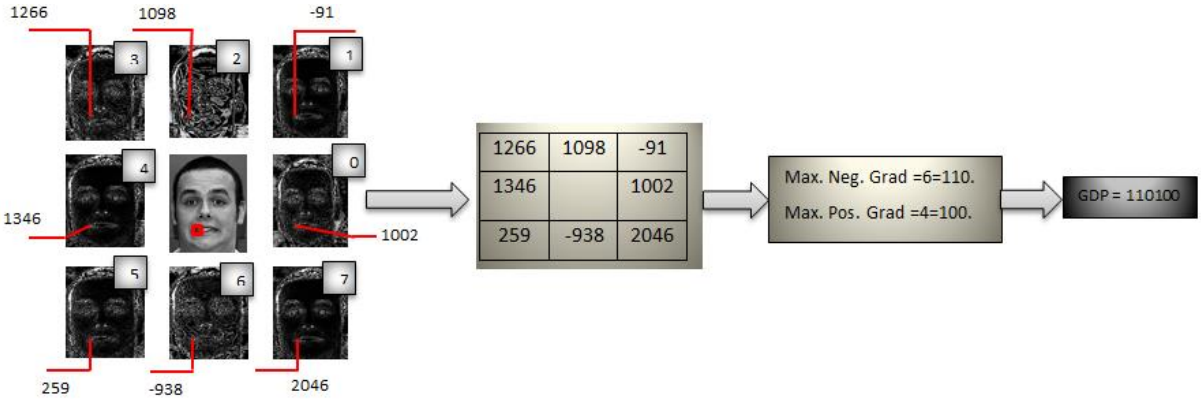


Figure. 5 GDP evaluation from Gabor edge feature map

binary GDP code of a pixel located at position (x, y) and C is an accumulation value. Next, the final GH is calculated by concatenating the Histograms of all blocks as:

$$GH = \prod_{i=1}^{N} H_i \quad (8)$$

Where N is the total number of blocks into which the facial image is divided, and Π denotes the concatenation operation. Here the concatenation is accomplished spatially, and the obtained final GH plays a vital role in global face features for a given input facial image.

4. Simulation experiments

This section describes the details of simulation experiments conducted over the developed FER system. The situation experiments are conducted with the help of the MATLAB tool. Under this simulation study, initially, the details of databases considered for simulation are explored. Next, the details of simulation metrics are explored through which the performance is measured. Finally, a detailed comparative analysis is outlined to highlight the performance effectiveness of the proposed FER method.

4.1 Database details

Totally two different facial expression databases are considered here for simulation. The first one is (Extended) Cohn-Kanade (CK+) [37] and the next one is Japanese Female Face Expression (JAFFE) [38].

CK+ is an extended version of the CK database, which consists of 486 sequences acquired from 97 subjects. This dataset consists of both posed and non-posed (spontaneous) expressions. Compared to the images present in the CK dataset, the sequences of CK+ are increased by 22% and captured with an additional 27% of subjects. In this dataset, every sequence begins with a neutral expression and ends with a peak expression. The peak expression is fully coded with the Facial Action Coding System (FACS) and given an emotion label. The original size of every image is noticed as 490×640, and at the simulation, every test image is cropped according to the requirements, and on average the size of cropped image is noticed as 310×260. All the images are in PNG format. Some samples of the CK+ data set are shown in Fig. 6.

Next, the JAFFE dataset is a facial expression dataset captured through the facial expressions of only Japanese female subjects. JAFFE contains 213 images of 7 different facial expressions as Neutral, Angry, Disgust, Fear, Happy, Sad and Surprised. All these images are captured with the help of 10 Japanese models. In every image, the face is posed in frontal view and the candidate's hair is tied back to represent all the expressive landmarks of face. The original size of each image in this dataset is noticed as 256×256, and at the simulation, every test image is cropped, and on average, the size of cropped is noticed as 170×200. All the images are in TIFF format. Some samples of the JAFFE data set are shown in Fig. 7.

4.2 Performance metrics

Under this phase, the performance is measured through various performance Metris such as Recall or Detection Rate or True Positive Rate (TPR), True Negative Rate (TNR), precision or positive predictive value (PPV), False Positive Rate (FPR), False Negative Rate (FNR), and Accuracy. These performance metrics are obtained based on the following mathematical formulations as:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

$$False\ Negative\ Rate\ (FNR) = \frac{FN}{FN+TP} \quad (11)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

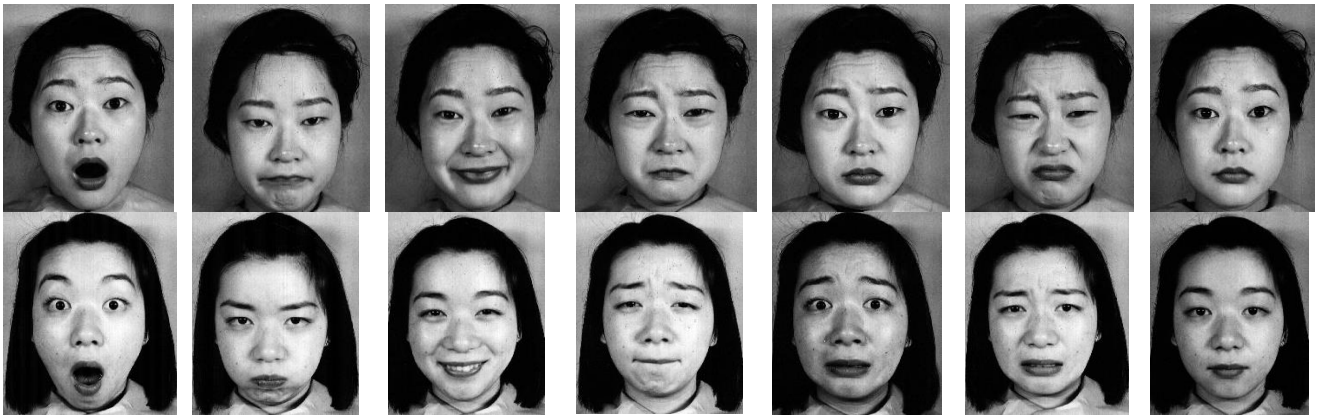
$$F - Score = \frac{2*Recall*Precision}{Recall + Precision} \quad (13)$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TP} \quad (14)$$

Under the experimental evaluation of CK+dataset, totally, we have tested 732 images acquired from 123 subjects. For 7-class prototypical expression recognition, the three most expressive image frames were taken from each sequence that resulted in 732 expression images (Angry - 93, Contempt - 60, Disgust - 108, Fear - 105, Happy - 126, Sadness - 105, and Surprise - 135). Next, to construct the neural expression set, the first frame from each sequence is selected, resulting in an 8-class expression dataset with 855 images. After testing the above test set, the obtained confusion matrix is shown in table.1, and the respective

(a) (b) (c) (d) (e) (f) (g)

Figure. 6 Samples from CK+ dataset: (a) Surprise, (b) Angry, (c) Happy, (d) Sad, (e) Fear, (f) Disgust, and (g) Neutral



(a) (b) (c) (d) (e) (f) (g)

Figure.7 Samples from JAFFE dataset: (a) Surprise (b) Angry (c) Happy (d) Sad (e) Fear (f) Disgust and (g) Neutral

performance metrics are represented in Table 2. Next, Under the experimental evaluation of the JAFFE dataset, tested 183 images were acquired from 10 subjects. For 6-class prototypical expression recognition, all versions of images are considered resulting in 183 expression images (Angry - 30, Disgust - 29, Fear – 32, Happy – 31, Sad – 31 and Surprise - 30). After testing the above test set, the obtained confusion matrix is shown in Table 3, and the respective performance metrics are represented in Table 4.

To achieve a generalized performance to novel subjects, in both datasets, a 10-fold cross-validation testing process is used in simulation experiments. Significantly, the entire test data was divided into ten partitions randomly. Each partition has an equal number of subjects. Nine are used to train the classifier at every validation among the ten partitions, and the last one is used for testing. This process is repeated ten times, and at every turn, the testing group is changed.

After the simulation of the proposed emotion recognition system over the CK dataset, the obtained results are shown in table.1, in the form of the confusion matrix. From table.1, we notice that the maximum True positives are obtained for Surprise emotion (132 for 135 inputs) and minimum TPs are for Angry emotion (70 for 93 inputs), followed by Contempt (46 for 60 inputs) and Sadness (86 for 105 inputs). Due to the similar characteristics in facial expression, the angry emotion has more confusion with disgust emotion. In some disgusted facial expression images, the subjects open their mouth so that the teeth are visible, which looks like an angry emotion. The content emotion almost looks like a neutral emotion due to fewer variations in the facial parts, and the sadness emotion has more confusion with Fear emotion. Based on the confusion matrix shown in the table.1, the performance metrics are measured and depicted in table.2. Form table.2, we noticed that the maximum recall (97.7885%), precision (97.7885%) is observed for Surprise emotion and minimum recall (75.2712%), precision (78.6534%) is for Angry Emotion. Further, the maximum FPR (21.3556%) and FNR (24.7378%) are observed for Angry and minimum FPR (2.2202%), and FNR (2.2202%) is observed for Surprise emotion. Finally, the higher F-Score

is achieved for surprise emotion, and lower is for angry emotion.

After the proposed emotion recognition system simulation over the JAFFE dataset, the obtained results are shown in table.3, in the form of the confusion matrix. From table.3, we notice that the maximum True positives are obtained for Angry emotion (29 for 30 inputs), Disgust (28 for 29 Inputs) and Happy (30 for 31 Inputs) and minimum TPs are for fear (30 for 32 inputs), sad (29 for 31 inputs) and surprise (28 for 30 inputs). Even though the JAFFE dataset has achieved higher TPs for every emotion, the overall accuracy is less because some expressions in JAFFE are very similar to other expressions. Based on the confusion matrix shown in the table.3, the performance metrics are measured and depicted in table.4. Form table.4, we noticed that the maximum recall (96.9712%) is obtained for Angry emotion and minimum recall (93.3309%) for surprise emotion. Next, the maximum precision (100%) is obtained for Disgust, and the minimum (88.2444%) is obtained for fear. Similarly, the highest F-Score (98.2452%) is observed at Disgust, and minimum (90.9128%) is at fear. Next, the highest FPR (11.7685%) is observed at fear, and minimum FPR (0%) is at disgust emotion. Finally, the highest FNR (6.6789%) is observed at surprise, and the lowest FNR (3.2345%) is happy emotion.

Table 1. Confusion matrix of 7-class expression recognition in the CK+ dataset

	Angry	**Contempt**	**Disgust**	**Fear**	**Happy**	**Sadness**	**Surprise**	**Total**
Angry	**70**	02	07	02	09	02	01	**93**
Contempt	02	**46**	01	03	01	07	0	**60**
Disgust	05	0	**101**	02	0	0	0	**108**
Fear	02	01	0	**96**	0	06	0	**105**
Happy	04	0	0	0	**120**	0	02	**126**
Sadness	05	03	01	09	01	**86**	0	**105**
Surprise	01	0	02	0	0	0	**132**	**135**
Total	**89**	**52**	**112**	**112**	**131**	**101**	**135**	**732**

Table 2. Performance metrics of 7-class expression recognition in the CK+dataset

Emotion/Metric	**Recall(%)**	**Precision(%)**	**FPR(%)**	**FNR(%)**	**F-Score(%)**
Angry	75.2712	78.6534	21.3556	24.7378	76.9290
Contempt	76.6790	88.4678	11.5456	23.3334	82.1412
Disgust	93.5212	90.1834	9.8256	6.4878	91.8290
Fear	91.4309	85.7168	14.2957	8.5724	88.4813
Happy	95.2431	91.6042	8.4087	4.7698	93.3800
Sadness	81.9020	85.1513	14.8555	18.1072	83.4929
Surprise	97.7885	97.7885	2.2202	2.2202	97.7885

Table 3. Confusion matrix of 6-class expression recognition in the JAFFE dataset

	Angry	**Disgust**	**Fear**	**Happy**	**Sadness**	**Surprise**	**Total**
Angry	**29**	0	0	0	01	0	**30**
Disgust	0	**28**	01	0	0	0	**29**
Fear	0	0	**30**	01	0	01	**32**
Happy	0	0	01	**30**	0	0	**31**
Sadness	0	0	02	0	**29**	0	**31**
Surprise	01	0	0	01	0	**28**	**30**
Total	**30**	**28**	**34**	**32**	**30**	**29**	**183**

Table 4. Performance metrics of 7-class expression recognition in the JAFFE dataset

Emotion/Metric	**Recall(%)**	**Precision(%)**	**FPR(%)**	**FNR(%)**	**F-Score(%)**
Angry	96.9712	96.6712	3.3333	3.3333	96.6712
Disgust	96.5543	100.00	0000	3.4523	98.2452
Fear	93.7565	88.2444	11.7685	6.2596	90.9128
Happy	96.7787	93.7520	6.2532	3.2345	95.2466
Sadness	93.5598	96.6745	3.3312	6.4575	95.0803
Surprise	93.3309	96.5574	3.4569	6.6789	94.9125

4.3 Comparative analysis

To further alleviate the proposed method, the performance is compared with various conventional methods such as LBP [8], LDP [22, 24], Extended LBP with K-L Transform [28], Active Shape Model (ASM) [13] and LDN with DGLP [31]. In this comparison, we compare our method with several appearance-based methods. C. Shan [8] reported two methods, such as LBPs and Boosted LBPs, when they use these methods over CK facial expression dataset and JAFFE dataset. However, this method has gained significantly less accuracy at low resolutions, and this less accuracy is due to the information loss at the neighborhood during LBP evaluation. The main disadvantage of LBP is that the encoding process ignores the connection between adjacent pixels and the encoding direction.

Unlike the LBP, the LDP [22, 24] encodes the edge responses around every pixel which is robust to noise and uneven illuminations. However, the LDP considered all eight directions at each pixel position and generated code from relative strength magnitude. Though this method considered more information to get a stable binary code, encoding is similar to LBP, resulting in loss of information at the neighborhood. This method has observed 85.2302% recognition accuracy at the JAFFE dataset, and this is much less than the recognition accuracy obtained at the CK dataset (92.6900%). One of the main reasons is that some expressions of the JAFFE dataset are very much similar to other expressions. Compared to LDP, the proposed approach is much better in the preservation of information loss. Moreover, the proposed approach also considered the directional movements of facial muscles, which makes the system robust for several issues.

Extending the LBP, a new version is proposed by M. Guo, X. Hou, Y. Ma [28] called ELBP to extract the varying texture properties of facial expression. Next, a covariance matrix is applied to reduce the dimensions of ELBP. The ELBP, on the other hand, ignores posture invariant characteristics, making it more vulnerable in pictures with variable poses. This method is observed to have a considerable information loss due to the accomplishment of the covariance matrix at the dimensionality reduction phase.

Next, focusing on the directionality of pixel intensities, I. Michael Revina, W.R. Sam Emmanuel [31] proposed LDN followed by DGLP to extract the directionality information along with the facial landmark movement. For example, in the contempt emotion, only the right or left side of the lips (in the closed format) is moved up, which means the gradient is upward. To recognize such types of emotions, directionality information also needs to be included in the LBP. LDN discovers the directionality and integrates it into the feature vector. However, the directionality discovery at a single orientation is not efficient in recognizing emotion in multiple orientations. On average, the accuracy obtained is noticed as 88.6300% for both CK and JAFFE datasets. Next, the ASM proposed in [13] finds the active shape of facial expression, which is insufficient to describe an expression because some expressions like happy and surprise involve the movements of the mouth and cheeks. These two parts are not extracted through ASM. The method proposed in [30, 39] applied mainly Gabor filter for feature extraction. At expression description, they

Table 5. Comparative analysis

Method	Database	Accuracy (%)
LBP [8]	CK	86.1441
	JAFFE	83.2230
LDP [22]	CK	91.4572
	JAFFE	85.2302
LDP [24]	CK	92.6900
ELBP + KLT [28]	CK	92.7552
	JAFFE	89.4574
ASM + SVM [13]	CK+	85.8002
LDN + DGLP [31]	CK	88.6300
	JAFFE	88.6300
Gabor + LBP [30]	CK	81.7000
Log-Gabor + LBP [39]	CK	82.3000
EALDBP [40]	CK	93.5922
	JAFFE	90.3898
Proposed	CK	**94.6658**
	JAFFE	**91.0856**

employed LBP. Even though Gabor and log-Gabor filters are effective face descriptors in multiple orientations, the LBP-based encoding has many drawbacks. Compared to all these methods, the proposed approach has gained more accuracy for both datasets because it concentrated on facial feature extraction and expression encoding. Even though the recently proposed EALDBP has obtained a better recognition accuracy, it has a significant drawback of scale and rotation invariance, which means the recognition accuracy is less for the scaled and tilted facial images. Table 5 shows that our method outperforms compared to all the conventional approaches at both CK and JAFFE datasets.

5. Conclusion

In this paper, we have developed a novel face expression recognition framework that works based on compact code representation. A novel image coding scheme called GDP is introduced here, which encodes the facial expression through their textures. These textures are analyzed through the gradients of facial expressions at different scales and different orientations. Since this work considered the gradients, instead of pixel intensities, the edges are enhanced through which the expressions are represented more compactly. Moreover, GDP encodes the gradient directions; the proposed FER is robust for pose variations. We also discovered that the proposed Gaussian edge features map is more stable against noise and illuminations. Furthermore, we also found that the proposed Gabor edge feature map is stable against the scaling and rotation of the image. Simulation experiments conducted over two standard datasets such as CK and JAFFE had proven the outstanding performance of the proposed expression recognition system. On average, the obtained improvement in the accuracy through the proposed approach is observed as 8% from LBP, 5% from LDP and 4% from LDN based methods.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing-original draft preparation, writing-review editing and visualization, have been done by 1st author. The supervision and project administration has been done by 2nd author.

References

[1] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The First Expression Recognition and Analysis Challenge", In: *Proc. of International Conf. on Automatic Face and Gesture Recognition*, Santa Barbara, CA, USA, pp. 921-926, 2011.

[2] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: a comprehensive survey", *Image Vision Computing*, Vol. 30, No. 10, pp. 683-697, 2012.

[3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey", *ACM Computer Survey*, Vol. 35, No. 4, pp. 399-458, 2003.

[4] Z. Zhang, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, pp. 39-58, 2009.

[5] Y. L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis", *Handbook of Face Recognition*, Springer, pp. 247-276, 2005.

[6] M. Valstar, I. Patras, and M. Pantic, "Facial Action Unit Detection Using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point Data", In: *Proc. of International Conf. on Computer Vision and Pattern Recognition Workshop*, San Diego, CA, USA, p. 76, 2005.

[7] P. Kakumanu and N. Bourbakis, "A local-global graph approach for facial expression recognition", In: *Proc. of International Conf. On Tools Artif. Intell.*, Arlington, VA, USA, pp. 685-692, 2006.

[8] C. Shan, S. Gong, and P. W. Mcowan, "Facial expression recognition based on local binary patterns: A comprehensive study", *Image Vis. Comput.*, Vol. 27, No. 6, pp. 803-816, 2009.

[9] M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.

[10] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images", *Journal of Optical Society Americe A*, Vol. 14, No. 8, pp. 1724-1733, 1997.

[11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisher faces: Recognition using class specific linear

projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, 1997.

[12] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 28, No. 12, pp. 2037-2041, 2006.

[13] M. Suk and B. Prabhakaran, "Real-time mobile facial expression recognition system—A case study", In: *Proc. of International Conf. on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, USA, pp. 132-137, 2014.

[14] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines", *Sensors*, Vol. 13, No. 6, pp. 7714-7734, 2013.

[15] S. L. Happy, A. George, A. Routray, "A real-time facial expression classification system using local binary patterns", In: *Proc. of the 4th International Conf. on Intelligent Human-Computer Interaction*, Kharagpur, India, pp. 1-5, 2012.

[16] M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park, and S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields", *IEEE Trans. on Image Proc.*, Vol. 24, No. 4, pp. 1386-1398, 2015.

[17] R. A. Khan, A. Meyer, H. Konik, S. Bouakaz, "Framework for reliable, real-time facial expression recognition for low-resolution images", *Pattern Recognit. Lett.*, Vol. 34, No. 10, pp. 1159-1168, 2013.

[18] P. Penev and J. Atick, "Local feature analysis: A general statistical theory for object representation", *Netw., Comput. Neural Syst.*, Vol. 7, No. 3, pp. 477-500, 1996.

[19] D. Gabor, "Theory of communication", *J. Inst. Electr. Eng. III, Radio Commun. Eng.*, Vol. 93, No. 26, pp. 429-457, 1946.

[20] L. Wiskott, J. M. Fellous, N. Kuiger, and C. V. D. Malsburg, "Face recognition by elastic bunch graph matching", *IEEE Trans. on Pattern Anal. Mach. Intell.*, Vol. 19, No. 7, pp. 775-779, 1997.

[21] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions", *IEEE Trans. Image Process.*, Vol. 19, No. 6, pp. 1635-1650, 2010.

[22] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (LDP) for face recognition", In: *Proc. of IEEE International Conf. on Consum. Electron.*, Las Vegas, NV, USA, pp. 329-330, 2010.

[23] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor", *IEEE Trans. Image Process.*, Vol. 19, No. 2, pp. 533-544, 2010.

[24] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern", *ETRI J.*, Vol. 32, No. 5, pp. 784-794, 2010.

[25] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network", *IEEE Access.*, Vol. 5, pp. 4525-4536, 2017.

[26] M. Sajjad, A. Shah, Z. Jan, S. I. Shah, S. W. Baik, and I. Mehmood, "Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery", *Cluster Comput.*, Vol. 21, pp. 549-567, 2018.

[27] D. Ghimire, S. Jeong, J. Lee, S. H. Park, "Facial expression recognition based on local region specific features and support vector machines", *Multimed. Tools Appl.*, Vol. 76, pp. 7803-7821, 2017.

[28] M. Guo, X. Hou, and Y. Ma, "Facial expression recognition using ELBP based on covariance matrix transform in KLT", *Multimedia Tools and Applications*, Vol. 76, pp. 2995-3010, 2017.

[29] M. Abdul and R. S. Holambe, "Local binary patterns based on directional wavelet transform for expression and pose invariant face recognition", *Appl. Comput. And Informatics*, Vol. 15, No. 2, pp. 163-171, 2019.

[30] M. L. Seyed and Z. M. Hussain, "Facial Expression Recognition Using Log-Gabor filters and Local Binary Pattern Operators", In: *Proc. of International Conf. on Computer and Communications*, China, pp. 349-353, 2016.

[31] I. M. Revina and W. R. S. Emmanuel, "Face expression recognition using LDN and Dominant Gradient Local Ternary Pattern descriptors", *Journal of King Saud University -Computer and Information Sciences*, Vol. 33, No. 4, pp. 392-398, 2018.

[32] C. Yang, M. Schmalz, W. Hu, and G. Ritter, "Center-surround filters for the detection of small targets in cluttered multispectral

imagery: Background and filter design", *SPIE*, Vol. 2496, pp. 637-648, 1995.

[33] D. Song and D. Tao, "Biologically inspired feature manifold for scene classification", *IEEE Trans. Image Process.*, Vol. 19, No. 1, pp. 174-184, 2010.

[34] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex", *Nat. Neurosci.*, Vol. 2, pp. 1019-1025, 1999.

[35] H. Chen, P. Belhumeur, and D. Jacobs, "In search of illumination invariants", In: *Proc. of IEEE Conf. on Comput. Vis. Pattern Recognition*, Hilton Head, SC, USA, pp. 254-261, 2000.

[36] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs, "A study of face recognition as people age", In: *Proc. of IEEE 11th Int. Conf. on Comput. Vis.*, Rio de Janeiro, Brazil, 2007, pp. 1-8.

[37] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression", In: *Proc. IEEE Comput. Soc. Conf. on Comput. Vis. Pattern Recognition Workshops*, San Francisco, CA, USA, pp. 94-101, 2010.

[38] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets", In: *Proc. of 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Nara, Japan, pp. 200-205, 1998.

[39] Q. Zhao, B. Chang, P. J. Jian, and T. Y. Yuan, "Facial Expression Recognition Based on Fusion of Gabor and LBP Features", In: *Proc. of International Conf. on Wavelet Analysis and Pattern Recognition*, Hong Kong, China, pp. 362-367, 2008.

[40] S. Sammaiah and K. V. Rao, "A Salient Binary Coding Scheme for Face and Expression Recognition from Facial Images, *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 2, pp. 67-80, 2021.

[41] M. M. Thiruthuvanathan and B. Krishnan, "EMONET: A Cross-Database Progressive Deep Network for Facial Expression Recognition", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 6, 2020.

[42] M. M. Thiruthuvanathan, B. Krishnan, and M. Rangaswamy, "Engagement Detection through Facial Emotional Recognition Using a Shallow Residual Convolutional Neural Networks", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 2, 2021.

[43] K. A. E. Dahshan, E. K. Elsayed, and A. A. E. A. Ebeid, "Recognition of Facial Emotions Relying on Deep Belief Networks and Quantum Particle Swarm Optimization", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, 2020.