

Automatic facial expression recognition combining texture and shape features from prominent facial regions

Naveen Kumar H N¹ | A Suresh Kumar² | Guru Prasad M S² | Mohd Asif Shah³

¹Department of Electronics and Communication Engineering, Vidyavardhaka College of Engineering, Gokulum III Stage, Mysuru, Karnataka, India

²Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun, India

³Department of Economics, College of Business and Economics, Kebri Dehar University, Jigjiga, Ethiopia

Correspondence

Mohd Asif Shah, Department of Economics, College of Business and Economics, Kebri Dehar, University, Po Box 250, Ethiopia.
Email: drmohdasifshah@kdu.edu.et

Abstract

Facial expression is one form of communication which being non-verbal in nature precedes verbal communication in both origin and conception. Most of the existing methods for Automatic Facial Expression Recognition (AFER) are mainly focused on global feature extraction assuming that all facial regions contribute equal amount of discriminative information to predict the expression class. The detection and localization of facial regions that have significant contribution to expression recognition and extraction of highly discriminative feature distribution from those regions are not fully explored. The key contributions of the proposed work are developing novel feature distribution upon combining the discriminative power of shape and texture feature; determining the contribution of facial regions and identifying the prominent facial regions that hold abstract and highly discriminative information for expression recognition. The shape and texture features taken into consideration are Local Phase Quantization (LPQ), Local Binary Pattern (LBP), and Histogram of Oriented Gradients (HOG). Multiclass Support Vector Machine (MSVM) is used while one versus one classification. The proposed work is implemented on CK+, KDEF, and JAFFE benchmark facial expression datasets. The recognition rate of the proposed work is 94.2% on CK+ and 93.7% on KDEF, which is significantly more than the existing handcrafted feature-based methods.

1 | INTRODUCTION

Communication in multitude of forms, verbal, non-verbal, or any other plays a vital role in successful completion of one's various daily routine tasks. Facial expression is one such form of communication which being non-verbal in nature precedes verbal communication in both origin and conception. Facial expressions hold the rich information on individuals' behaviour and play a predominant role in expressing emotions. Facial Expression Recognition as a process involves interpreting exhibited expressions on the face in identifying emotional state of a subject by pinpointing and measuring highly discernible features from the subject's facial image. When this happens automatically and in a naturally induced way, it can be termed Automatic Facial Expression Recognition (AFER). Since primary mode of interaction for human beings happens via face, there has been a steady increase in demand for an effective AFER system capable of recognizing subjects' emotional state relying solely on face images, irrespective of age, race, or gender.

Herein, facial expressions taken into consideration in designing, testing, and validating a novel approach to AFER are Anger, Disgust, Fear, Happiness, Sadness, and Surprise, which apart from being basic are also universal in nature.

In recent years, AFER is beginning to gain more and more importance in the emerging field of human-computer interaction (HCI) by virtue of being an overwhelming presence in related areas such as assistive health care technologies [1, 2], interactive agent [3], fatigue measurement [4], lie detection [5], education [6], security [7], and so on. As the name itself suggests, interaction plays an important role in effective realization of HCI systems. But herein the challenge lies in facilitating interaction happen as though it is happening naturally, and not feel like something which is artificially manufactured. Even though a vast amount of progress has been made towards righteous recognition of expressions, it still is an arduous task due to the sheer amount of complexity and variability associated with expressions on the face. Therefore, an efficient facial representation technique pushes itself as a vital requirement in

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.



achieving improvement over the existing recognition rate of an AFER system. In addition to this, while addressing AFER-related problems there arises a need for expression-specific information to be emphasized, to address the issue of variability associated with facial expressions. In AFER, visual information is vital, and more so assists us in analysing and identifying changes in facial features as correlated to various expressions.

The proposed work aims at exploring prominent facial regions along with associated feature descriptors that provide highly discriminative information in relation to facial expression classification problem as belonging to static face images. Herein, highly discriminative information refers to those which by virtue of their nature cause reduction in variance within expression class and accretion in variance across expression class. Apart from being highly discriminative, there is an inherent need for the information to be robust against variations in subject's age, race, image resolution, illumination, and partial occlusion, which makes AFER implementation on static face images highly challenging as information available herein is very limited in nature. In addressing the same, the key contributions of the proposed work are:

- Developing novel feature distribution upon combining the discriminative power of shape and texture feature to improve the recognition rate of the AFER system.
- Determining the contribution/s of facial regions towards recognition of each basic expression.
- Identifying prominent facial regions which are having significant contributions in developing a highly discernible and abstract feature distribution that offers robustness to changes in subjects age, culture, illumination levels, image resolution, and partial occlusion.
- The proposed work is implemented on three standard benchmark datasets (CK+ [8], KDEF [9], and JAFFE [10]) to test its generalization capability.

The rest of the sections are organized as follows: Section 2 presents the related work in the field of AFER. Section 3 describes the proposed system. Experimental analysis of the proposed system is described in Section 4 and Section 5 concludes the paper and provides insight into further enhancement.

2 | RELATED WORK

In a broad sense, AFER can be viewed as extraction of measurable features from the face and training of the same with a classifier of specific interest in obtaining a model which can perform emotion classification upon samples from a test dataset. The survey of the existing methods for AFER can be broadly classified into still image-based and image sequence-based methods. In case of still image (spatial representation) based AFER, the image frame exhibiting hitherto expression is processed [11], whereas in image sequence based (spatio-temporal representations) methods the sequence of frames

exhibiting different layers of an expression (from neutral to a specific expression) are processed [12]. Each of aforesaid methods can be further classified into geometric feature, appearance feature, and deep feature methods based on feature extraction scheme taken into consideration [13].

In case of geometrical feature extraction, the feature set includes information on location and shape of facial components and is also used in training and testing phase of the classifier. The features in relation to facial regions such as distance between facial components, orientation of facial components, are used in facial feature representation. Geometrical feature extraction demands accurate detection and localization, along with tracking of prominent landmarks on the face [14]. This is particularly challenging to achieve in real-life scenarios. In Action Unit (AU)-based representation the relation between facial feature and AUs; relation between AUs; relation between AUs and expressions need to be determined; hence, it is more complex in nature and difficult to implement by considering spatial information only. AU-based expression recognition methods demand action units whose data related to training has been previously labelled by experts [15]. This is a process which is quite time-consuming and labour-intensive.

Deep neural networks are being widely used for a lot of pattern-recognition tasks, namely, image recognition, face recognition, gender recognition, human pose estimation, just to name a few. Liu et al. [16] proposed research findings on action unit localisation and encoding for facial expression analysis, wherein discriminative part-based representation is obtained by applying learning from deformable parts model onto 3D CNN, to detect particular action units. A. Mollahosseini et al. [17] proposed deep neural networks for AFER which comprise two convolution layers and four inception layers. In [18] the robustness to pose and identity is achieved by the dynamic multi-channel metric learning network. The end-to-end Co-attentive Multi-task CNN (CMCNN) comprises the Channel Co-Attention Module (CCAM) and the Spatial Co-Attention Module (SCAM) is proposed for AFER [19]. The min–min game with joint and marginal distribution sampling is used to maximize the difference between expression and identity information [20]. The approach is independent of adversarial game to bypass identity information. Face frontalization is used to remove the back projection noise and shortcut CNN is proposed for AFER [21]. The selection of optimal weights for various models and its effective combination is achieved by the adaptive exponential ensemble rule. In [22], local binary convolution layer is combined with standard convolution layer to address AFER. Islets loss function is used to enhance the discriminative power of deep representation. Sparse deep feature-based regularization with good generalization capability is proposed in [23]. Deep networks demand humungous volume of data for learning and computation complexity is comparatively more in case of most deep convolutional neural network-based approaches [24]. Benitez-Garcia et al. [25] presented a handcrafted local feature descriptor which provides better recognition rates compared to existing deep neural network approaches.

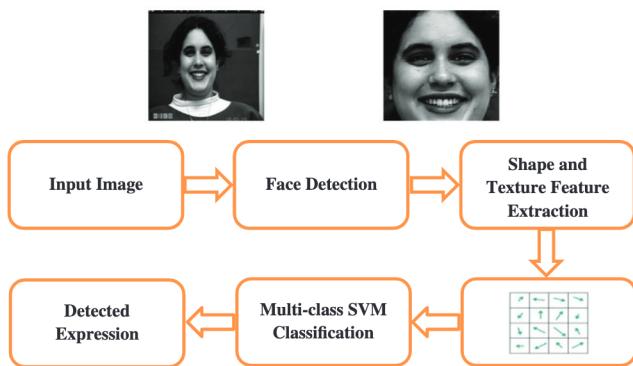


FIGURE 1 Holistic approach for AFER. AFER, automatic facial expression recognition.

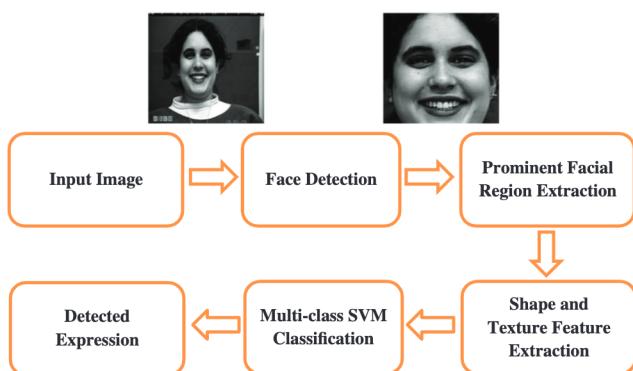


FIGURE 2 Component-based approach for AFER

3 | PROPOSED METHODOLOGY

AFER aims at recognizing the emotional state of a subject as exhibited through facial expressions from the subject's facial image. Depending on the application of feature descriptor, AFER approaches can be broadly categorized as either belonging to holistic approach or component-based approach. In the holistic approach, the whole face image is used to predict the expression class of the subject, whereas in component-based approach facial regions only those which are deemed as prominent are used to predict the expression class of the subject. The block diagrams showcasing holistic approach and component-based approach in relation to expression recognition are depicted in Figures 1 and 2, respectively. The important stages involved while expression recognition such as face detection, feature extraction, emotion classification, remains same across both approaches; afore-mentioned three stages are elaborated in the following section.

3.1 | Face detection

Irrespective of the approach, the very first step in AFER involves detecting faces from an input image. The face detec-

tion algorithms employed in performing the same must be capable of classifying faces as against non-faces from an input image. The Viola-Jones algorithm is employed herein to achieve face detection [26]. It is preferred mostly because of its detection accuracy, speed, robustness, along with ease of real-time implementation. The detected face images are resized to 200×200 pixel resolution to achieve spatial normalization. The Viola-Jones algorithm is applied on the detected face image to extract the facial regions (eyes and eyebrows, nose, mouth, and forehead).

3.2 | Feature extraction

Feature extraction is a process which involves transforming raw image data in making it suitable for processing in computers. Feature description stage involves assigning quantitative attributes for the features. With respect to the AFER problem, feature extraction stage plays a vital and important role in overall efficacy of the solution. Studies as of late opine that the approaches relying on appearance features can provide similar or sometimes even better performance when compared with action units-based expression recognition methods [27]. Herein, texture information related to facial expressions is obtained through appearance-based features. The descriptors used in extracting texture information can be applied either to entire face or in part to particular facial regions. In the appearance-based approach, 'emotion labels' by themselves are sufficient for training process and as such do not need accurate facial landmarks detection and localization, thus making it more appropriate for AFER as against geometric features. Appearance-based AFER system comprises different stages, as depicted in the flowchart shown in Figure 3. Each stage in Figure 3 has its own indelible impact on the recognition rate of AFER system. The feature distribution of training samples is used to train Multiclass Support Vector Machine (MSVM) classifier. Feature distribution of test sample is applied on the model built upon training to predict the expression class of the test sample. Abstract and discriminative feature extraction serves as one of the crucial stages in AFER.

The characteristics of feature descriptor about the AFER problem are

- Independent of subjects age, gender, and culture
- High discriminative power (reduces variation within the class and increases variation across the class)
- Represents unique characteristics of each expression class
- Amplify expression-specific information and suppress expression irrelevant information (identity)
- Insensitive or less sensitive to illumination changes, small pose variations, partial occlusion, and head movements

The features taken into consideration herein in the proposed research work are Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), and Local Phase Quantization (LPQ).

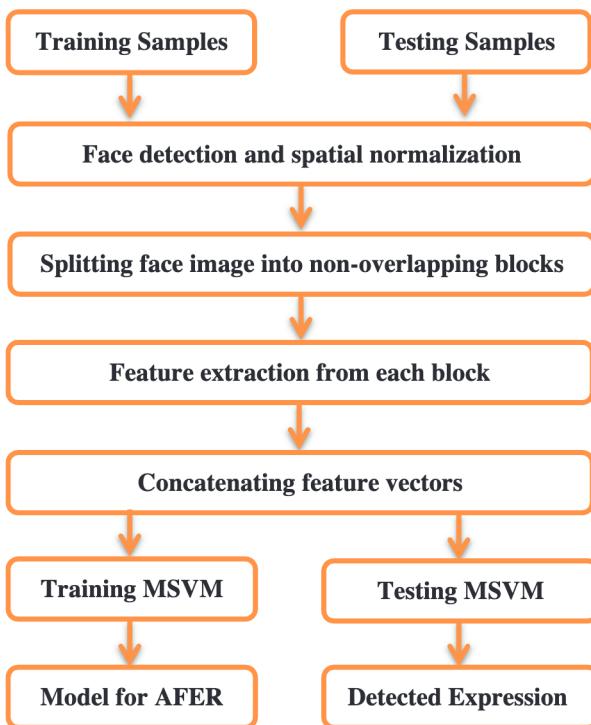


FIGURE 3 Flow chart depicting appearance feature-based AFER system

3.2.1 | Histogram of oriented gradients (HOG) feature descriptor

Gradient represents directional changes in intensity. HOG counts occurrence of such oriented gradients from subset of an image known as cell. Before applying HOG, the image is divided into cells; HOG feature is computed for each cell and all such features comprising of whole image are concatenated to form a feature vector for the image; so extracted HOG features are used as shape and appearance features in defining an expression while training and testing a classifier [28].

Let $f(x,y)$ represent an expressive face image being divided into nine cells, before computing gradients. Equation (1) represents the image being split into nine cells.

$$f(x,y) = \begin{bmatrix} f_{c1}(x,y) & f_{c2}(x,y) & f_{c3}(x,y) \\ f_{c4}(x,y) & f_{c5}(x,y) & f_{c6}(x,y) \\ f_{c7}(x,y) & f_{c8}(x,y) & f_{c9}(x,y) \end{bmatrix} \quad (1)$$

Shape and appearance feature is extracted from each cell as follows:

Let us consider $f_{CN}(x,y)$ as the N^{th} cell of an image $f(x,y)$. Image gradient is computed by convolving the image with the gradient operators.

Gradient operator in the horizontal and vertical directions is represented by Equations (2) and (3), respectively.

$$b_g = [-1 \ 0 \ 1] \quad (2)$$

$$v_g = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

Image gradient in the horizontal and vertical directions is computed by Equations (4) and (5), respectively.

$$f_b(x,y) = f_{CN}(x,y) * b_g \quad (4)$$

$$f_v(x,y) = f_{CN}(x,y) * v_g \quad (5)$$

where '*' represents the 2-D convolution operator

Orientation of gradients is computed by Equation (6).

$$f_{OG}(x,y) = \tan^{-1} \left(\frac{f_v(x,y)}{f_b(x,y)} \right) \quad (6)$$

The size of the oriented gradient pattern is the same as cell size of the image taken into consideration. The HOG is obtained by counting the occurrence of oriented gradients. The normalized histogram depicts the feature vector of a cell. The concatenation of feature vector comprising of all cells defines the shape feature vector for an expressive face. Let H_{CN} represent the normalized feature vector of the N^{th} cell of an image $f(x,y)$. If there are M cells in an image, the normalized feature vector for the whole face image is defined by Equation (7). The HOG feature visualization for images representing various emotional expressions is shown in Figure 4.

$$\text{feature vector } r_{HOG} = [H_{C1}, H_{C2}, \dots, \dots, H_{CM}] \quad (7)$$

3.2.2 | Local binary pattern (LBP) feature descriptor

Gabor filters and LBP variants are among the most widely used texture feature descriptors for expression recognition [29]. LBP is computed by thresholding the N neighbour pixels upon a circle having radius R with respect to centre pixel. Binary weights are assigned to each bit position; sum of all these weights ranges from 0 to 255 (decimal). The image is split into non-overlapping blocks prior to computation of LBP. Let f represent one such block in an expressive face image. The process of LBP computation with respect to a pixel in a subset of an image is defined by Equation (8).

$$LBP_{q,R} = \sum_{q=0}^{N-1} 2^q W \left(f_{(q,R)} - f_{(\text{centre})} \right) \quad (8)$$

The thresholding function is represented by Equation (9)

$$W(s) = \begin{cases} 1, & s \geq 0 \\ 0, & s < 0 \end{cases} \quad (9)$$

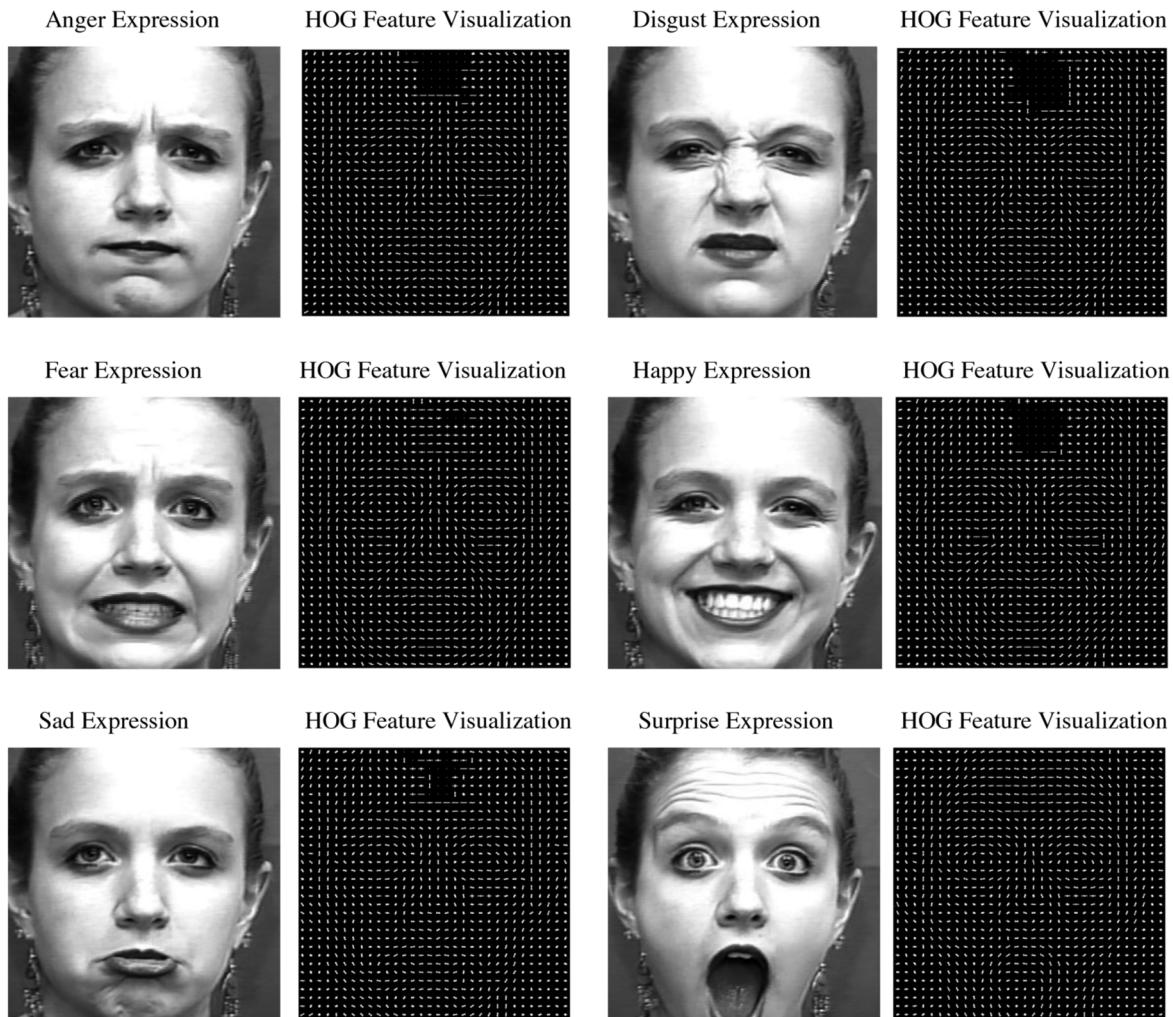


FIGURE 4 HOG feature visualization for images representing various emotional expressions. HOG, histogram of oriented gradients.

$f_{(q,R)}$: Neighbouring pixel with in radius R

$f_{(\text{centre})}$: Centre pixel with in radius R

Thresholding:

$$\begin{bmatrix} 7 & 10 & 16 \\ 2 & 7 & 5 \\ 3 & 4 & 5 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow [11100000]$$

Assigning binary weights example:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 4 \\ 128 & 8 & 0 \\ 64 & 32 & 16 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 7$$

(*) represents element-wise matrix multiplication.

Uniform LBP counts the occurrence of uniform patterns. In uniform patterns, the number of bitwise transitions ($0 \rightarrow 1$ or

$1 \rightarrow 0$) must not exceed 2. Most of the binary patterns extracted from local sections of the face images are uniform in nature. Hence, uniform LBP is a powerful texture descriptor when applied over local region of a face image. Examples representing uniform and non-uniform pattern are shown below.

Uniform pattern:

$$\begin{bmatrix} 7 & 10 & 16 \\ 2 & 7 & 5 \\ 3 & 4 & 5 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow [11100000]$$

Non-uniform pattern:

$$\begin{bmatrix} 7 & 10 & 16 \\ 2 & 7 & 10 \\ 3 & 8 & 5 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow [11110100]$$

The number of uniform patterns that exist from 0 (00000000) to 255 (11111111) is 58; all non-uniform patterns

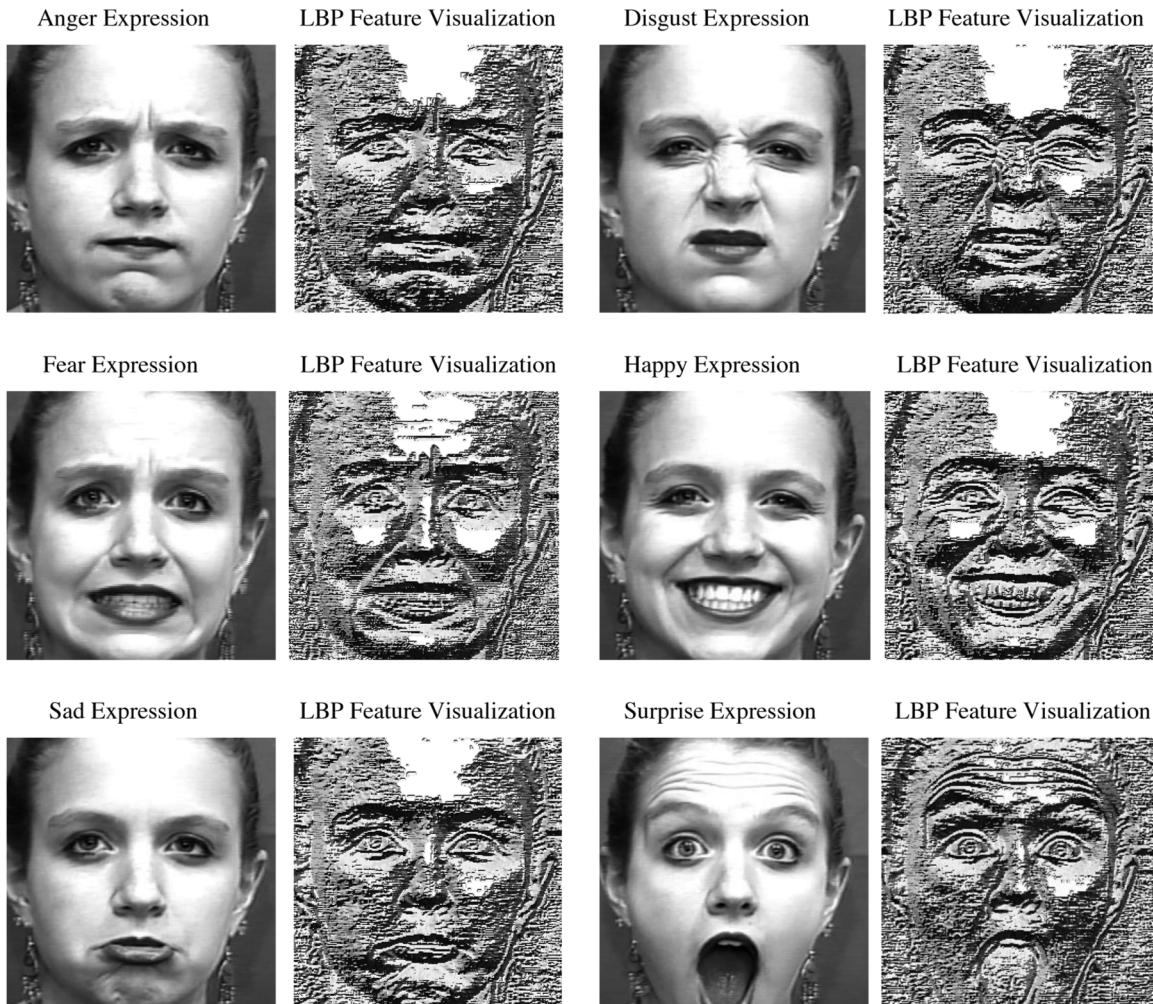


FIGURE 5 LBP feature visualization for images representing various emotional expressions

are represented as one pattern. Uniform LBP counts the occurrence of all these 59 patterns in an image block and represents them in the form of normalized histogram. Let G_{BN} represent normalized feature vector of the N^{th} block of an image $f(x, y)$. If there are M blocks in an image, the normalized feature vector for a face image is defined by Equation (10). The LBP feature visualization for images representing various emotional expressions is shown in Figure 5.

$$\text{feature vector}_{LBP} = [G_{B1}, G_{B2}, \dots, \dots, G_{BM}] \quad (10)$$

3.2.3 | Local phase quantization (LPQ) feature descriptor

Spatial texture information can be extracted by quantizing the local phase information. LPQ feature descriptor is applied upon Discrete Fourier Transform (DFT) of local sections of the facial image to extract texture features [30]. Quantization of phase information in the local sections of the image makes LPQ feature robust against blur and low resolution. In the proposed work, two different approaches are used in extracting frequency

information from the local section of an image, namely, Short-time Fourier Transform (STFT) with Uniform Window, and STFT with Gaussian Window [31]. Let $f(x, y)$ represent a local section of a face image. The LPQ feature computation on $f(x, y)$ is as follows:

$$\text{Window size: } W_s = N$$

$$\text{Radius: } R = \frac{N-1}{2}$$

$$\text{Spatial coordinates in window: } x = -R : R$$

1-D filters for STFT with uniform window are represented in Equation (11).

$$w_0 = 1, \quad w_1 = e^{-j2\pi x/N} \quad \& \quad w_2 = e^{j2\pi x/N} \quad (11)$$

The Gaussian function and 1-D filters for STFT with Gaussian window are represented in Equations (12) and (13) respectively.

$$G_w = \frac{e^{-\left(\frac{1}{2}\right)\left(\frac{x}{0.25(N-1)}\right)^2}}{\left(\sqrt{2\pi}\right)0.25(N-1)} \quad (12)$$

$$\mathbf{w}_0 = G_w \mathbf{w}_0, \quad \mathbf{w}_1 = G_w \mathbf{w}_1 \quad \& \quad \mathbf{w}_2 = G_w \mathbf{w}_2 \quad (13)$$

Normalizing filter coefficients to zero mean, we get

$$\mathbf{w}_1 = \mathbf{w}_1 - \left(\frac{1}{N} \right) \sum_{i=1}^N \mathbf{w}_1^i \quad \& \quad \mathbf{w}_2 = \mathbf{w}_2 - \left(\frac{1}{N} \right) \sum_{i=1}^N \mathbf{w}_2^i \quad (14)$$

The new set of filters to compute the frequency response is represented in Equations (15) and (16), respectively.

$$f_1 = (f(x, y) * \mathbf{w}_0^T) * \mathbf{w}_1, \quad f_2 = (f(x, y) * \mathbf{w}_2^T) * \mathbf{w}_0 \quad (15)$$

$$f_3 = (f(x, y) * \mathbf{w}_1^T) * \mathbf{w}_1, \quad f_4 = (f(x, y) * \mathbf{w}_1^T) * \mathbf{w}_2 \quad (16)$$

where '*' represents convolution.

$$\text{Frequency response : } f_{resp} = \left\{ R(f_1), I(f_1), R(f_2), I(f_2), R(f_3), I(f_3), R(f_4), I(f_4) \right\} \quad (17)$$

In Equation (17), R and I represent real and imaginary components of the frequency response. Furthermore, discriminative information can be extracted by suppressing the redundant information from the frequency response. Whitening transformation is used to obtain uncorrelated vectors whose co-variance is identity matrix.

2-D filters for co-variance matrix computation can be computed by the following equations:

$$k_1 = \mathbf{w}_0^T \mathbf{w}_1 \quad \& \quad k_2 = \mathbf{w}_1^T \mathbf{w}_0 \quad (18)$$

$$k_3 = \mathbf{w}_1^T \mathbf{w}_1 \quad \& \quad k_4 = \mathbf{w}_1^T \mathbf{w}_2 \quad (19)$$

$$c_1 = CV\{R(k_1)\} \quad c_2 = CV\{I(k_1)\} \quad (20)$$

$$c_3 = CV\{R(k_2)\} \quad c_4 = CV\{I(k_2)\} \quad (21)$$

$$c_5 = CV\{R(k_3)\} \quad c_6 = CV\{I(k_3)\} \quad (22)$$

$$c_7 = CV\{R(k_4)\} \quad c_8 = CV\{I(k_4)\} \quad (23)$$

where ' CV ' represents column vector

$$V = \{c_1^T; c_2^T; c_3^T; c_4^T; c_5^T; c_6^T; c_7^T; c_8^T\} \quad (24)$$

$$\begin{array}{ccccccccc} & 1 & 2 & \dots & N & 1 & 1 & \dots & 1 \\ & 1 & 2 & \dots & N & 2 & 2 & \dots & 2 \\ x_c = & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ & \cdot & \cdot & \dots & \cdot & N & N & \dots & N \end{array}$$

Co-variance between pixel positions is computed by Equation (25), where x_c and y_c are the grids used to compute co-variance between pixel positions.

$$[x_v y_v] = [CV(x_c) \quad CV(y_c)] \quad (25)$$

The distance between the pixel positions is obtained by using Equation (26).

$$D_e(i, j) = \sqrt{\left(\sum_{i,j=1}^{N^2} (x_v(i) - y_v(j))^2 \right)} \quad (26)$$

$$U = (C_c)^{D_e} \quad (27)$$

where ' C_c ' represents correlation coefficient.

$$W = VUV^T \quad (28)$$

$$[P, Q, R] = SVD(AW\mathcal{A}) \quad (29)$$

where ' \mathcal{A} ' represents identity matrix, ' T ' represents matrix transform, and ' SVD ' represents singular value decomposition. Equation (30) represents the result of whitening transform.

$$f_{res} = \left(R^T f_{res}^T \right)^T \quad (30)$$

Equation (31) represents the LPQ feature of an image segment. Normalized LPQ feature is obtained for each sub-region in a face exhibiting expression and finally the concatenation of normalized features from all sub-regions represents the LPQ feature vector of the expressive face image [32]. The LPQ feature extracted from the face image is used in training and testing the classifier in relation to the AFER problem. LPQ feature visualization using uniform window and Gaussian window are shown in Figures 6 and 7, respectively (window size used is five).

$$LPQ = \sum_{k=1}^8 (f_{res}(k) > 0) \cdot 2^{k-1} \quad (31)$$

3.3 | Classification

SVM is one among the most popular supervised learning methods [33]. In addressing the multi-class classification problem,

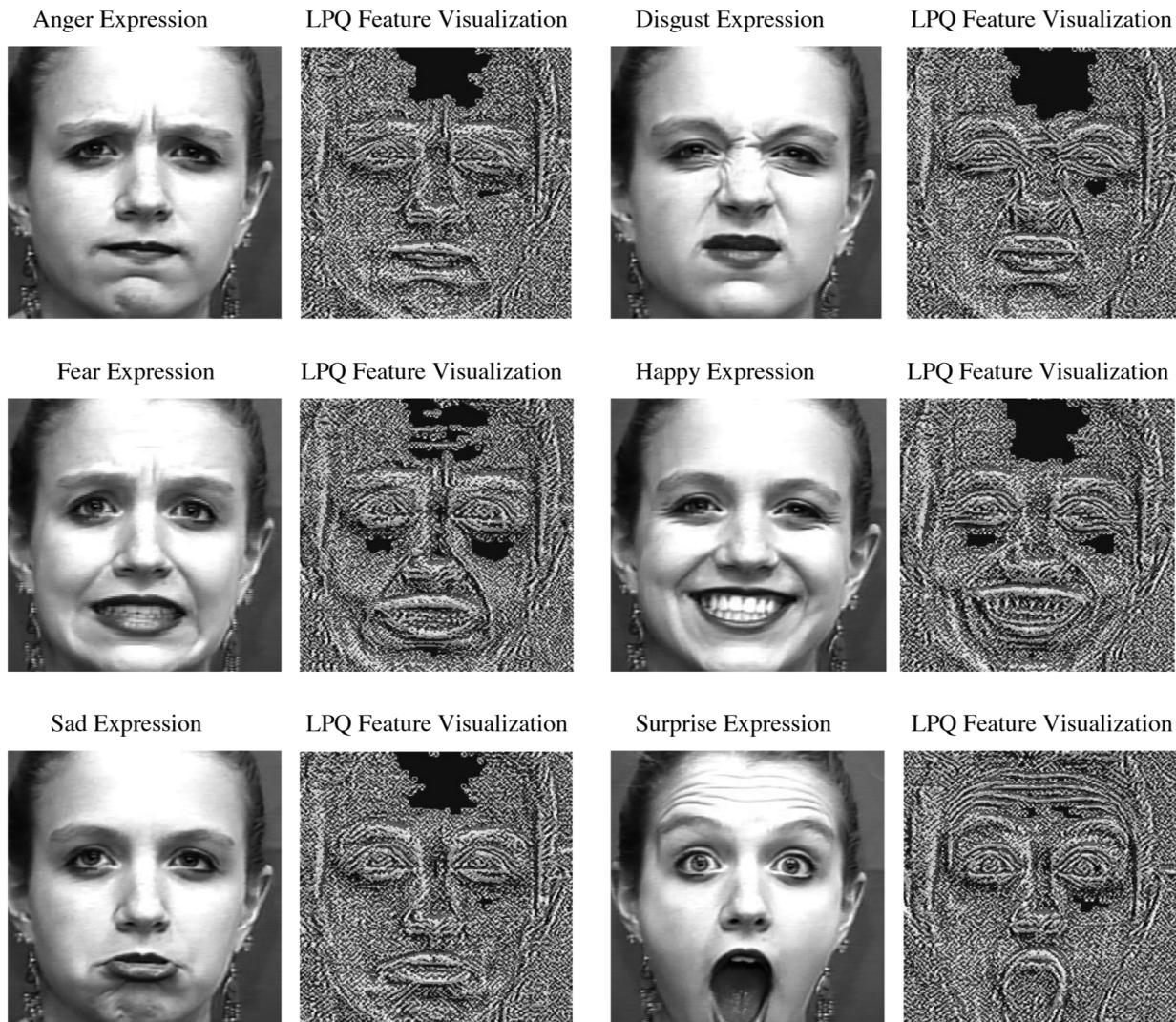


FIGURE 6 LPQ feature visualization using uniform window in STFT for images representing various emotional expressions. LPQ, local phase quantization; STFT, short-time Fourier transform.

SVM can be used as one versus one classification or one versus all classification. The one versus one classification approach uses $15(6c_2)$ classifiers for a six-class AFER problem, whereas one versus all approach uses a mere six classifiers for the same [34]. SVM is used as one versus one classifier and 5-fold cross validation scheme is employed in all the experiments. In 5-fold cross validation approach each dataset is divided into five subsets based on subjects [35]. Subject independency is achieved by applying a method wherein expressive face images pertaining to the same subject are being used either in training or in testing phase. The abstract and highly discriminative features extracted from the face image samples belong to training set and are used in training the classifier. Upon training, the classifier builds a model for AFER. The features extracted from face image samples belong to test set and are used in testing the classifier; recognition rate is measured for each basic expression from confusion matrix obtained for the test data.

4 | RESULTS AND DISCUSSION

4.1 | Experimental setup

The datasets used in the experiments include Extended Cohn-Kanade Dataset (CK+) [8], Japanese Female Facial Expressions (JAFFE) [10], and Karolinska Directed Emotional Faces (KDEF) [9]. The characteristics of the datasets are tabulated in Table 1.

Experiments are carried out on MATLAB 2017 platform (Intel i3 processor with 8 GB RAM). After face detection, all resulting face images are resized to arrive at a uniform resolution of 200×200 pixels [36]. Furthermore, various experiments that have been conducted on the aforesaid three datasets are classified as belonging into either holistic approach or component-based approach; the same is summarized in the following section. Recognition rate herein refers to the average

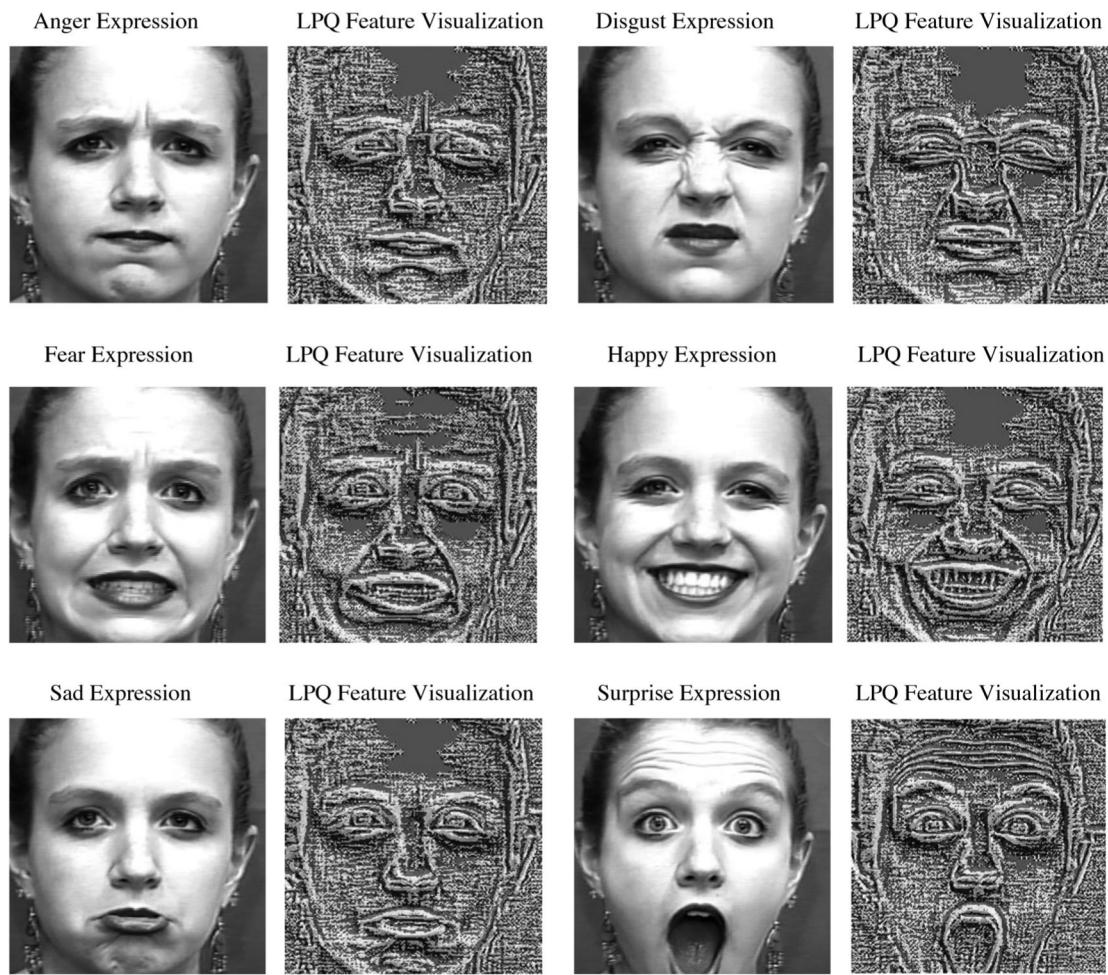


FIGURE 7 LPQ feature visualization using Gaussian window in STFT for images representing various emotional expressions

TABLE 1 Characteristics of the datasets

| Dataset characteristics | CK+ | JAFFE | KDEF |
|-----------------------------------|---|------------------|-----------------------------|
| Number of subjects | 123(69% female and 31% male) | 10 (all females) | 70(35 females and 35 males) |
| Age group | 18–50 years | 18–50 years | 20–30 years |
| Selection of subjects | Euro-American:81% Afro-American:13% Others:6% | Japanese: 100% | Differing region |
| Image resolution | 640 × 490 pixels | 256 × 256 pixels | 562 × 762 pixels |
| Number of basic expressions | 7 | 7 | 7 |
| Number of image Samples/sequences | 593 | 213 | 490 |
| Nature of expressions | Posed | Posed | Posed |

recognition rate, that is, recognition rate arrived at when all six expressions are taken into account.

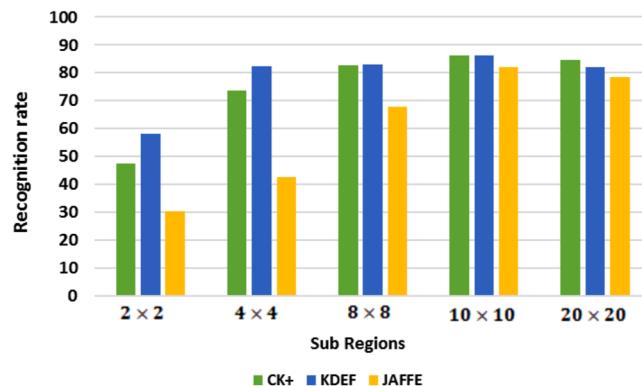
4.2 | Holistic approach

In this approach, LBP feature descriptor is applied on whole face region and features obtained thus are used to train and test

the classifier. Once face detection is done, block processing is employed wherein the facial image under consideration is split into a number of sub-regions determined prior through a set of experiments. This is all done prior to extraction of texture features. Recognition rate of AFER system obtained via LBP feature extraction scheme (training and testing of SVM classifier using LBP feature only) for varying size of sub-regions is tabulated in Table 2. Figure 8 depicts the recognition rate of

TABLE 2 Recognition rate when LBP feature descriptor is employed

| No. of sub-regions | 2×2 | 4×4 | 8×8 | 10×10 | 20×20 |
|--------------------|--------------|--------------|--------------|----------------|----------------|
| CK+ | 47.6 | 73.6 | 82.7 | 86.3 | 84.5 |
| KDEF | 58.2 | 82.3 | 82.9 | 86.2 | 82 |
| JAFFE | 30.4 | 42.6 | 67.7 | 82.1 | 78.6 |

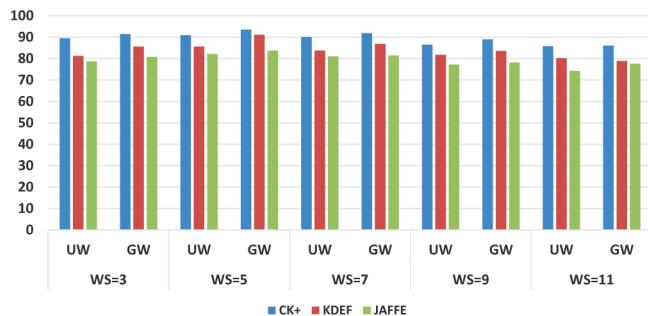
**FIGURE 8** Recognition rate when LBP feature descriptor is employed**TABLE 3** Recognition rate when LPQ feature descriptor is employed

| | WS = 3 | | WS = 5 | | WS = 7 | | WS = 9 | | WS = 11 | |
|-------|--------|------|-------------|-------------|--------|------|--------|------|---------|------|
| | UW | GW | UW | GW | UW | GW | UW | GW | UW | GW |
| CK+ | 89.5 | 91.4 | 90.9 | 93.5 | 90.1 | 91.9 | 86.5 | 89 | 85.8 | 86.1 |
| KDEF | 81.3 | 85.6 | 85.6 | 91.2 | 83.7 | 86.9 | 81.8 | 83.6 | 80.3 | 78.9 |
| JAFFE | 78.8 | 80.9 | 82.1 | 83.7 | 81 | 81.5 | 77.2 | 78.2 | 74.3 | 77.7 |

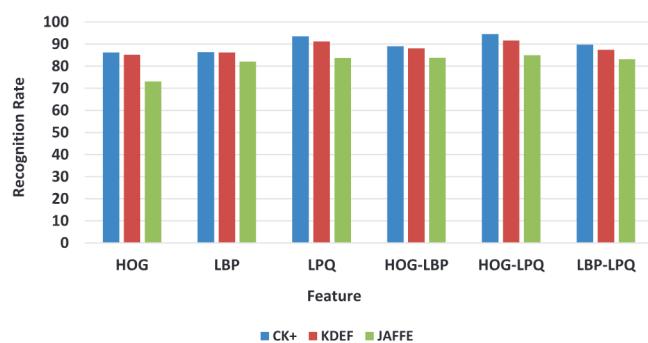
GW, Gaussian window; STFT, short-time Fourier transform; UW, uniform window ; WS, window size.

AFER system obtained via the LBP feature extraction scheme for varying size of sub-regions. The highest recognition rate is obtained for those facial images split into 10×10 sub-regions using LBP texture feature descriptor.

In the following approach, the LPQ texture feature extracted from the whole face region is used to train and test the classifier. In the case of LPQ feature descriptor, the parameters taken into consideration are window used to extract frequency information, number of sub-regions, and window size [37]. STFT using uniform window and Gaussian window is used to extract frequency information. Like LBP descriptor, the LPQ also provides good recognition rates for facial images split into 10×10 sub-regions. The recognition rate of the AFER system obtained via the LPQ feature extraction scheme for varying window size is tabulated in Table 3 (number of sub-regions = 10×10). Figure 9 depicts the recognition rate of the AFER system obtained via the LPQ feature extraction scheme for varying window sizes. Experiments carried out on LPQ texture feature descriptor point towards best results being obtained when STFT uses Gaussian window to extract frequency and phase

**FIGURE 9** Recognition rate when LPQ feature descriptor is employed**TABLE 4** Recognition rate in holistic approach for different feature extraction schemes

| | HOG | LBP | LPQ | HOG-LBP | HOG-LPQ | LBP-LPQ |
|-------|------|------|-------------|---------|-------------|---------|
| CK+ | 86.2 | 86.3 | 93.5 | 89 | 94.5 | 89.7 |
| KDEF | 85.2 | 86.2 | 91.2 | 88.1 | 91.6 | 87.4 |
| JAFFE | 73.1 | 82.1 | 83.7 | 83.8 | 84.9 | 83.2 |

**FIGURE 10** Recognition rate in holistic approach for different feature extraction schemes

information. The computation complexity and response time increase with increase in window size. It is found that the optimum window size to extract LPQ feature is 5 with consideration being made towards recognition rate, computation complexity, and response time.

The error analysis while implementing the aforesaid approaches motivated us to consider feature combinations for the AFER problem. Error analysis points towards the fact that images being misclassified in one feature extraction scheme may not necessarily bear same representation in another feature extraction scheme, which is the case here, that is, images classified as misclassifications in one approach being rightly classified in another approach. In line with the same, several experiments have been conducted wherein for the first few sets of experiments, the classifier is trained and tested upon a single feature and then upon a combination of features. Results obtained while HOG, LBP, LPQ are individually employed, along with combinations of aforesaid features, are tabulated in Table 4.

Figure 10 depicts the recognition rate in holistic approach for different feature extraction schemes. Results obtained from

TABLE 5 Performance analysis of different feature extraction schemes in holistic approach

| | Recognition rate | | | Feature length | Training time (s) | | | |
|---------|------------------|------|-------|----------------|-------------------|-------|-------|--|
| | | | | | CK+ | KDEF | JAFFE | |
| | CK+ | KDEF | JAFFE | | | | | |
| HOG | 86.2 | 85.2 | 73.1 | 20,736 | 46.7 | 16.5 | 10.3 | |
| LBP | 86.3 | 86.2 | 82.1 | 5900 | 222.8 | 63.3 | 38.5 | |
| LPQ | 93.5 | 91.2 | 83.7 | 25,600 | 287 | 92.3 | 45.2 | |
| HOG-LBP | 89 | 88.1 | 83.8 | 26,636 | 227.3 | 79.9 | 34.9 | |
| HOG-LPQ | 94.5 | 91.6 | 84.9 | 46,336 | 300.9 | 99.9 | 47.4 | |
| LBP-LPQ | 89.7 | 87.4 | 83.2 | 31,500 | 467.4 | 155.3 | 65.3 | |

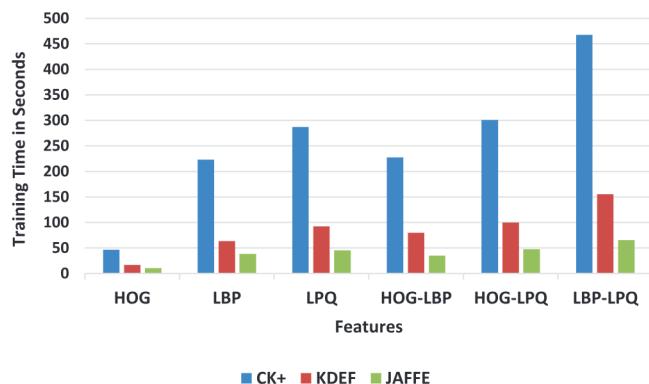


FIGURE 11 Training time analysis of different feature extraction schemes in holistic approach

experiments thus carried out reveal some prominent observations: when classifier is trained and tested upon a single feature (HOG/LBP/LPQ), recognition rate for LPQ is found to be superior compared to HOG and LBP as can be seen from Table 4; when the classifier is trained and tested upon a combination of features, significant improvement in recognition rate is observed for the combination of HOG and LPQ feature compared to other feature combinations. Results obtained so far show that HOG-LPQ combination outperforms other feature combinations for all the datasets taken into consideration.

The feature extraction schemes taken into consideration are further analyzed with respect to recognition rate, feature length, and training time (computation complexity). Performance analysis of aforesaid feature descriptors in holistic approach is tabulated in Table 5. Figure 11 depicts training time analysis of different feature extraction schemes in the holistic approach

Block processing is employed to extract texture feature (LBP and LPQ), wherein the face image is split into sub-regions. The number of sub-regions used to extract texture feature is 10×10 . Furthermore, the training time required to train SVM classifier (one versus one classification) using 5-fold cross validation approach is measured for each dataset. Results show that time taken to train in CK+ dataset is much higher compared to JAFFE and KDEF datasets, as CK+ consists of a greater number of samples compared to JAFFE and KDEF. To apply

texture feature descriptor (LBP & LPQ), block processing is used, due to which number of computations tends to increase which in turn leads to increase in training time. The best result in terms of recognition rate is obtained for the HOG-LPQ feature combination at the expense of higher feature dimension and processing time. Hence, achieving higher recognition rate along with reduction in both feature dimension and processing time becomes an arduous task in holistic approach.

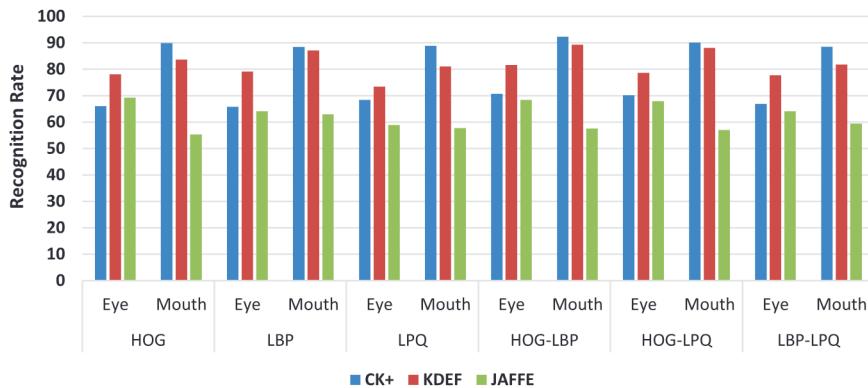
4.3 | Component-based approach

In this approach, an attempt has been made to explore facial regions and associated feature descriptors in providing discriminative information for specific expression recognition. The set of experiments conducted on various facial regions (eye and eyebrows, mouth and chin, nose, forehead, and cheeks) reveals that the eye and mouth regions contribute abstract, and highly discriminative information for expression recognition. Herein, eye region refers to the facial region covering eye and eyebrows, and mouth region refers to the facial region covering mouth and chin, respectively. In the component-based approach, features extracted from prominent facial regions are used in training and testing of the classifier instead of features from the whole face region. Facial region segmentation is carried out on the detected face to extract prominent facial regions. The contribution of prominent facial regions towards expression recognition has been analyzed through a series of experiments wherein expression recognition is achieved via employing feature extraction; herein, the exception lies in the fact that instead of whole face, only a single prominent facial region is selected. The result of experiments thus carried out is tabulated in Table 6. Figure 12 depicts recognition rate of different feature extraction schemes in the component-based approach. Block processing has been employed to extract texture features (LPQ and LBP). Various block sizes are tested, and results obtained thus point towards LBP feature operating at best with block size of 20×20 , and LPQ feature for block size of 10×10 (eye image size is 50×150 pixels and mouth image size is 60×120 pixels). Recognition rates are at their best for combination of HOG-LBP feature in case of expression recognition from single prominent facial region (eye/mouth). Recognition rate obtained using features extracted from mouth region is close to recognition rate obtained from the holistic approach.

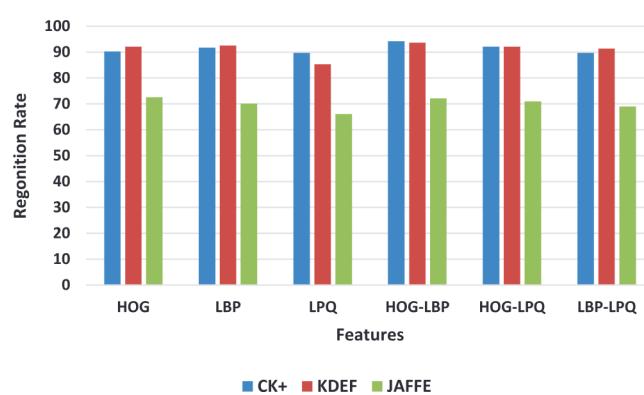
In the next phase of experimentation, features extracted from both eye and mouth region are combined to train and test the classifier. Several experiments are conducted, wherein for the first set of experiments the classifier is trained and tested on a single feature and for the remaining experiments the classifier is trained and tested on a combination of features. The results of the experiments conducted in relation to all three datasets are tabulated in Table 7. Figure 13 depicts recognition rate of different feature extraction schemes upon deriving feature distribution from both eye and mouth region (components-based approach). Results obtained thus point towards combination of HOG-LBP feature being best across different combinations in relation to all datasets taken into consideration. Results of

TABLE 6 Recognition rate of different feature extraction schemes in component-based approach

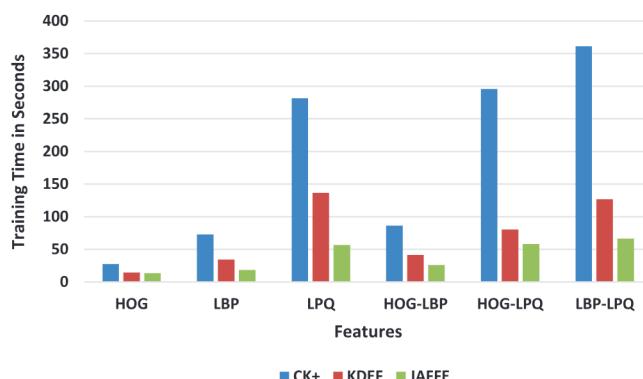
| HOG | | LBP | | LPQ | | HOG-LBP | | HOG-LPQ | | LBP-LPQ | |
|-------|-------|------|-------|------|-------|---------|-------|---------|-------|---------|-------|
| Eye | Mouth | Eye | Mouth | Eye | Mouth | Eye | Mouth | Eye | Mouth | Eye | Mouth |
| CK+ | 66.1 | 89.9 | 65.8 | 88.4 | 68.4 | 88.9 | 70.7 | 92.3 | 70.2 | 90.1 | 66.9 |
| KDEF | 78.1 | 83.7 | 79.1 | 87.1 | 73.4 | 81.1 | 81.6 | 89.3 | 78.6 | 88.1 | 77.7 |
| JAFFE | 69.2 | 55.3 | 64.1 | 63 | 58.9 | 57.8 | 68.4 | 57.6 | 67.9 | 57 | 64.1 |
| | | | | | | | | | | | 59.4 |

**FIGURE 12** Recognition rate of different feature extraction schemes in component-based approach**TABLE 7** Recognition rate of different feature extraction schemes in components-based approach (features are derived from both eye and mouth region)

| | HOG | LBP | LPQ | HOG-LBP | HOG-LPQ | LBP-LPQ |
|-------|------|------|------|---------|---------|---------|
| CK+ | 90.2 | 91.7 | 89.7 | 94.2 | 92.1 | 89.7 |
| KDEF | 92.1 | 92.5 | 85.3 | 93.7 | 92.1 | 91.4 |
| JAFFE | 72.6 | 70.1 | 66.1 | 72.1 | 71 | 69 |

**FIGURE 13** Recognition rate of different feature extraction schemes in components-based approach (Features are derived from both eye and mouth region)

experiments in which the classifier is trained and tested on a single feature point towards LBP feature descriptor's performance being superior compared to HOG and LPQ (except JAFFE). As can be seen from Table 7, significant improvement over recognition rate is observed when a classifier is trained

**FIGURE 14** Training time analysis of different feature extraction schemes in components-based approach

and tested employing a combination of shape and texture feature (HOG-LBP and HOG-LPQ), compared to FER systems where the classifier is trained and tested on a single feature (shape/texture).

The feature extraction schemes taken into consideration are further analyzed with respect to recognition rate, feature length, and training time (computation complexity). Performance analysis of components-based approach with respect to afore-mentioned parameters is tabulated in Table 8. Figure 14 depicts training time analysis of different feature extraction schemes in a components-based approach. Significant reduction in feature dimension and training time is achieved for most of the feature extraction schemes employed via components-based approach. Feature dimension in relation to LPQ feature is comparatively high in components-based approach when compared with holistic approach; this is partly because components-based approach works best for block size 10×10 , whereas

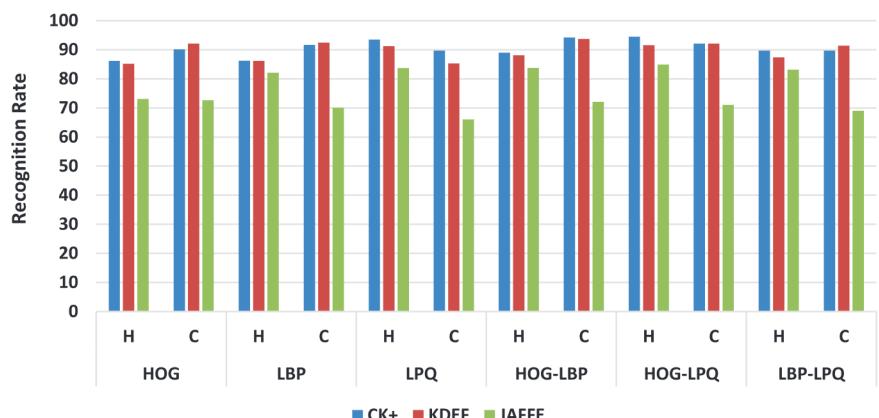
TABLE 8 Performance analysis of different feature extraction schemes in components-based approach

| | Recognition rate | | | Feature length | Training time (s) | | |
|----------------|------------------|------|-------|----------------|-------------------|-------|-------|
| | CK+ | KDEF | JAFFE | | CK+ | KDEF | JAFFE |
| HOG | 90.2 | 92.1 | 72.6 | 6084 | 27.3 | 14.4 | 13.3 |
| LBP | 91.7 | 92.5 | 70.1 | 2478 | 72.7 | 34.5 | 18.4 |
| LPQ | 89.7 | 85.3 | 66.1 | 37,632 | 281.3 | 136.5 | 56.6 |
| HOG-LBP | 94.2 | 93.7 | 72.1 | 8562 | 86.2 | 41.4 | 25.9 |
| HOG-LPQ | 92.1 | 92.1 | 71 | 43,716 | 295.6 | 80.4 | 57.9 |
| LBP-LPQ | 89.7 | 91.4 | 69 | 40,110 | 361.3 | 126.7 | 66.5 |

TABLE 9 Comparative analysis of holistic and components-based approach

| | HOG | | LBP | | LPQ | | HOG-LBP | | HOG-LPQ | | LBP-LPQ | |
|--------------|------|------|------|------|------|------|---------|-------------|-------------|------|---------|------|
| | H | C | H | C | H | C | H | C | H | C | H | C |
| CK+ | 86.2 | 90.2 | 86.3 | 91.7 | 93.5 | 89.7 | 89 | 94.2 | 94.5 | 92.1 | 89.7 | 89.7 |
| KDEF | 85.2 | 92.1 | 86.2 | 92.5 | 91.2 | 85.3 | 88.1 | 93.7 | 91.6 | 92.1 | 87.4 | 91.4 |
| JAFFE | 73.1 | 72.6 | 82.1 | 70.1 | 83.7 | 66.1 | 83.8 | 72.1 | 84.9 | 71 | 83.2 | 69 |

C, components-based approach; H, holistic approach.

FIGURE 15 Comparative analysis of holistic and components-based approach

holistic approach works best for block size 20×20 . Hence, achieving higher recognition rate along with reduction in both feature dimension and processing time is possible in case of components-based approach.

Comparative analysis of the results obtained via holistic and components-based approach is tabulated in Table 9. Figure 15 depicts comparative analysis of holistic and components-based approach. The holistic approach works best for a combination of HOG and LPQ features, whereas components-based approach works best for combination of HOG and LBP features. In most of the experiments that have been conducted, slight improvement in recognition rate is observed when the classifier is trained and tested on a combination of features rather than being trained with a single feature. In general, the combination of shape and texture features has caused significant improvement in the performance of FER system with respect to recognition rate. Further analysis shows us that LPQ feature

descriptor exhibiting better performance in holistic approach fails to provide same results in components-based approach.

5 | CONCLUSION

The proposed work aims at exploring prominent facial regions to extract the abstract and highly discriminative features for AFER problem. HOG feature distribution is used for shape feature representation and LBP and LPQ feature distributions are used for texture feature representation. The new set of convolution filters is used in LPQ to derive the texture information. The 5-fold cross validation is employed using MSVM in all the experiments. The feature distribution from all the facial regions is taken into consideration in holistic approach irrespective of its contribution. Based on the comprehensive experiments conducted on the three benchmark datasets

(CK+, JAFFE, and KDEF), eye and mouth regions are identified as prominent facial regions that hold abstract and highly discriminative information for AFER. The abstract and highly discriminative feature distribution is derived from the prominent facial regions in components-based approach. The error analysis motivated us to perform feature fusion to improve the recognition accuracy. The recognition rate 94.2% and 93.7% is obtained for HOG-LBP feature fusion in components-based approach for CK+ and KDEF datasets, respectively. The recognition rate is improved by combining the discriminative power of shape and texture features derived from the prominent facial regions. Reduction in both feature dimension and computation complexity along with robustness to occlusion (except eye and mouth region) is achieved in the proposed work by retaining the higher recognition rate. The work is implemented on three benchmark datasets which showcase the generalization capability of the feature distribution. In the future, the temporal information can be considered to further improve the recognition rate. The work can be implemented on more number of datasets (both constrained and unconstrained) to test its generalization capability. AFER in uncontrolled environments has many unresolved issues, which need to be addressed.

AUTHOR CONTRIBUTIONS

Naveen H N Kumar: Conceptualization; Formal analysis; Investigation; Methodology; Validation; Writing-original draft. Suresh A Kumar: Supervision; Writing-review & editing. Guru M S Prasad: Data curation; Investigation; Validation; Writing-review & editing. Mohd Asif Shah: Investigation; Visualization; Writing-review & editing.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING INFORMATION

The authors declare that no funds, grants, or other support were received for this manuscript.

DATA AVAILABILITY STATEMENT

Data derived from public domain resources The data that support the findings of this study are available in reference number [8, 9, and 10]. These data were derived from the following resources available in the public domain: <https://www.kaggle.com/datasets> <https://www.kdef.se/download-2/register.html>

ORCID

Naveen Kumar HN  <https://orcid.org/0000-0001-6922-8095>
Suresh Kumar  <https://orcid.org/0000-0001-7145-6337>
Guru Prasad M S  <https://orcid.org/0000-0002-1811-9507>
Mohd Asif Shah  <https://orcid.org/0000-0002-0351-9559>

REFERENCES

1. Leo, M., Carcagni, P., Mazzeo, P.L., Spagnolo, P., Cazzato, D., Distante, C.: Analysis of facial information for healthcare applications: A survey on computer vision-based approaches. *MDPI J. Inf.* 11(3), 128 (2020). <https://doi.org/10.3390/info11030128>
2. Guha, T., Yang, Z., Grossman, R.B., Narayanan, S.S.: A computational study of expressive facial dynamics in children with autism. *IEEE Trans. Affective Comput.* 9(1), 14–20 (2018). <https://doi.org/10.1109/TAFFC.2016.2578316>
3. Spezialetti, M., Placidi, G., Rossi, S.: Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Front. Rob. AI* 7, 532279 (2020). <https://doi.org/10.3389/frobt.2020.532279>
4. Hooda, R., Joshi, V., Shah, M.: A comprehensive review of approaches to detect fatigue using machine learning techniques. *Chronic Dis. Transl. Med.* 13, 415–425 (2021). <https://doi.org/10.1016/j.cdtm.2021.07.002>
5. Pan, H., Xie, L., Wang, Z., Liu, B., Yang, M., Tao, J.: Review of micro-expression spotting and recognition in video sequences. *Virtual Reality Intell. Hardware* 3(1), 1–17 (2021) <https://doi.org/10.1016/j.vrih.2020.10.003>
6. Utami, P., Hartanto, R., Soesanti, I.: A study on facial expression recognition in assessing teaching skills: Datasets and methods. *Procedia Comput. Sci.* 2011, 1–6 (2019). ISSN 1877-0509 <https://doi.org/10.1016/j.procs.2019.11.154>
7. Butalia, A., Ingle, M., Kulkarni, P.: Facial expression recognition for security. *Int. J. Mod. Eng. Res.* 2(4), 1449–1453 (2012)
8. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (Cat. No. PR00580), pp. 2037–2041 (2000). <https://doi.org/10.1109/AFGR.2000.840611>
9. Lundqvist, D., Litton, J.E.: The averaged karolinska directed emotional faces – AKDEF. CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, ISBN 91-630-7164-9 (1998)
10. Lyons, M., Kamachi, M., Gyoba, J.: The Japanese female facial expression (JAFFE) database (1998). <https://doi.org/10.5281/zenodo.3451524>
11. Naveen Kumar H N, Jagadeesha, S., Jain, A.K.: Human facial expression recognition from static images using shape and appearance feature. In: *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 598–603 (2016). <https://doi.org/10.1109/ICATCCCT.2016.7912070>
12. Golzadeh, H., Faria, D.R., Manso, L.J., Elkárt, A., Buckingham, C.D.: Emotion recognition using spatiotemporal features from facial expression landmarks. In: *2018 International Conference on Intelligent Systems (IS)*, pp. 14–20 (2018). <https://doi.org/10.1109/IS.2018.8710573>
13. Michael Revina, I., Sam Emmanuel, W.R.: A survey on human face expression recognition techniques. *J. King Saud Univ. Comp. Inf. Sci.* 8(6), 26–35 (2021) ISSN 1319—1578. <https://doi.org/10.1016/j.jksuci.2018.09.002>
14. Durmuşoğlu, A., Kahraman, Y.: Facial expression recognition using geometric features. In: *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 1–5 (2016). <https://doi.org/10.1109/IWSSIP.2016.7502700>
15. Yao, L., Wan, Y., Ni, H., et al.: Action unit classification for facial expression recognition using active learning and SVM. *Multimed. Tools Appl.* 161, 544–552 (2021). <https://doi.org/10.1007/s11042-021-10836-w>
16. Liu, M., Li, S., Shan, S., Wang, R., Chen, X.: Deeply learning deformable facial action parts model for dynamic expression analysis. In: *Asian Conference on Computer Vision*, Springer, pp. 143–157 (2014)
17. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, pp. 1–10 (2016). <https://doi.org/10.1109/WACV.2016.7477450>
18. Liu, Y., Dai, W., Fang, F., Chen, Y., Huang, R., Wang, R., Wan, B.: Dynamic multi-channel metric network for joint pose-aware and identity-invariant facial expression recognition. *Inf. Sci.* 578, 195–213 (2021) ISSN 0020–0255 <https://doi.org/10.1016/j.ins.2021.07.034>
19. Yu, W., Xu, H.: Co-attentive multi-task convolutional neural network for facial expression recognition. *Pattern Recognit.* 33, 619 (2022) ISSN 0031—3203. <https://doi.org/10.1016/j.patcog.2021.108401>
20. Liu, X., Jin, L., Han, X., You, J.: Mutual information regularized identity-aware facial expression recognition in compressed video. *Pattern Recognit.* 119, 108105 (2021) ISSN 0031—3203. <https://doi.org/10.1016/j.patcog.2021.108105>
21. Tsai, K.-Y., Tsai, Y.-W., Lee, Y.-C., Ding, J.-J., Chang, R.Y.: Frontalization and adaptive exponential ensemble rule for deep-learning-based facial

- expression recognition system. *Signal Process. Image Commun.* 80, 24287 (2021) ISSN 0923—5965. <https://doi.org/10.1016/j.image.2021.116321>
- 22. Li, H., Wang, N., Yu, Y., Yang, X., Gao, X.: LBAN-IL: A novel method of high discriminative representation for facial expression recognition. *Neurocomputing* 432, 159–169 (2021) ISSN 0925—2312. <https://doi.org/10.1016/j.neucom.2020.12.076>
 - 23. Xie, W., Jia, X., Shen, L., Yang, M.: Sparse deep feature learning for facial expression recognition. *Pattern Recognit.* 96, 106966 (2019) ISSN 0031—3203. <https://doi.org/10.1016/j.patcog.2019.106966>
 - 24. Li, S., Deng, W.: Deep facial expression recognition: A survey. *IEEE Trans. Affective Comput.* <https://doi.org/10.1109/TAFFC.2020.2981446>
 - 25. Benitez-Garcia, G., Nakamura, T., Kaneko, M.: Facial expression recognition based on local Fourier coefficients and facial Fourier descriptors. *J. Signal Inf. Process.* 08, 132–151 (2017)
 - 26. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vision* 12, 311–322 (2004)
 - 27. Liu, Y., Zhang, X., Lin, Y., Wang, H.: Facial expression recognition via deep action units graph network based on psychological mechanism. *IEEE Trans. Cogn. Dev. Syst.* 57(2), 137–154 (2020) <https://doi.org/10.1109/TCDS.2019.2917711>
 - 28. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, vol. 60, pp. 4065–4080 (2005). <https://doi.org/10.1109/CVPR.2005.177>
 - 29. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 26–35 (2006)
 - 30. Ojansivu, V., Heikkilä, J.: Blur insensitive texture classification using local phase quantization. In: International Conference on Image and Signal Processing, Springer, pp. 236–243 (2008)
 - 31. Pei, S., Huang, S.: STFT with adaptive window width based on the chirp rate. *IEEE Trans. Signal Process.* 578(8), 195–213 (2012). <https://doi.org/10.1109/TSP.2012.2197204>
 - 32. Turan, C., Lam, K.-M.: Histogram-based local descriptors for facial expression recognition (FER): A comprehensive study. *J. Visual Commun. Image Represent.* 123, 108401 (2018) <https://doi.org/10.1016/j.jvcir.2018.05.024>
 - 33. Hsu, C.-W., Lin, C.-J.: A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* 119(2), 415–425 (2002) <https://doi.org/10.1109/72.991427>
 - 34. Liu, Y., Wang, R., Zeng, Y.-S.: An improvement of one-against-one method for multi-class support vector machine. In: 2007 International Conference on Machine Learning and Cybernetics, pp. 2915–2920 (2007). <https://doi.org/10.1109/ICMLC.2007.4370646>
 - 35. Tsujitani, M., Tanaka, Y.: Cross-validation, bootstrap, and support vector machines. *Adv. Artif. Neural Syst.* 432, Article ID 302572 159 (2011). <https://doi.org/10.1155/2011/302572>
 - 36. Kumar, G., Bhatia, P.K.: A detailed review of feature extraction in image processing systems. In: 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 5–12 (2014). <https://doi.org/10.1109/ACCT.2014.74>
 - 37. Yin, Q., Shen, L., Lu, M., Wang, X., Liu, Z.: Selection of optimal window length using STFT for quantitative SNR analysis of LFM signal. *J. Syst. Eng. Electron.* 55(1), 331–341 (2013). <https://doi.org/10.1109/JSEE.2013.00004>

How to cite this article: Kumar H N, N., Kumar, S.A., Prasad, G.M.S., Shah, M.A.: Automatic facial expression recognition combining texture and shape features from prominent facial regions. *IET Image Process.* 17, 1111–1125 (2023). <https://doi.org/10.1049/ipr2.12700>