The Visual Computer
https://doi.org/10.1007/s00371-020-01988-1

**ORIGINAL ARTICLE**

# Facial emotion recognition using subband selective multilevel stationary wavelet gradient transform and fuzzy support vector machine

**R. Jeen Retna Kumar[1] · M. Sundaram[2] · N. Arumugam[3]**

Accepted: 2 October 2020



## Abstract

Facial emotion recognition finds a major role in affective computing. Recognizing emotion by facial expression is an extremely important activity to design control oriented and human computer interactive applications especially in cognitive science and neuroscience. For a precise and robust recognition, feature extraction is one of the major challenges in facial expression recognition system. Wavelet transform is one of the major key methods utilized for feature extraction in facial emotion recognition. In this paper, the statistical parameters from the proposed subband selective multilevel stationary wavelet gradient transform are calculated and are utilized as features for efficacious recognition of emotion. The features of the wavelet transform contain both spatial and spectral domain information which is best suited for identifying human emotions through facial expression. The introduction of gradient transform to find the gradient of subband avails to estimate the edges in images for the quality amelioration of subbands. The dimension reduction in the extracted features is done by using Pearson–kernel–principal component analysis method. The classification of emotion using the selected features is done by the proposed Gaussian membership function fuzzy SVM classifier. Experiments were performed on the well-known database for facial expression such as JAFEE database, CK + database and FG Net database and obtained promising emotion classification results.


**Keywords** Facial emotion recognition · Fuzzy support vector machine · Kernel principal component analysis · Stationary biorthogonal wavelet transform · Wavelet gradient transform

## 1 Introduction

Human emotions can be recognized by facial expression, body posture, gesture [1] and speech [2]. Emotion recognition is subsidiary in the evaluation of stress, intelligent learning [3], medicine, monitoring, marketing, etc. One of the most natural ways of humans to convey their emotional state to observers is through facial expressions. Human face conveys their messages or intentions by the movements and position of the muscles in the skin of face. For people to communicate with each other, facial expressions are the most subsidiary nonverbal way and it was first suggested by Darwin [4]. Six basic expressions such as happy, sad, anger, surprise, fear and disgust were suggested by Ekman and Friesen [5]. The recognition of emotions of a human through expressions in face has gained more attention in various areas of research including pattern recognition, affect computing, computer vision and human computer interaction. It performs an imperative role in plentiful applications such as mood prediction, telecommerce, patient monitoring system, humanoid robots, clinical and crime psychology and sentiment analysis. Face has the inherent characteristics with different facial expressions which are signalized by distinct appearance of facial features due to the facial muscle motions. Hence, an explicit feature extraction is the prime influential and challenging key component for efficacious recognition of emotion. The disparity in the face attributes due to pose vari-

✉ R. Jeen Retna Kumar
jejinrsrch@gmail.com

M. Sundaram
cm.sundaram2011@gmail.com

N. Arumugam
arms.ece@gmail.com

1 Department of ECE, Bethlahem Institute of Engineering, Karungal, Tamil Nadu 629157, India

2 Department of ECE, VSB Engineering College, Karur, Tamil Nadu 639111, India

3 Department of ECE, National Engineering College, Kovilpatti, Tamil Nadu 628503, India

Published online: 21 December 2020

ation, illumination variation, background noise and uneven lighting makes the feature extraction task more challenging. An emotion recognition system consists of a preprocessing stage, a feature extraction stage, a feature dimension reduction stage and a classification stage. An emotion recognition system to identify one of the seven emotions from a facial image utilizing the proposed method is expounded in this paper.

The overall content of this paper is abridged as follows

1. The feature extraction of the face expression image is done by using the proposed subband selective multilevel stationary wavelet gradient transform (SM-SWGT) method.
2. The edges of the images are estimated by the gradient transform which makes the quality of face image better, to get more accuracy and sensitivity for classification.
3. The redundancy in the subband features is minimized using the subband selective feature selection, based on the energy values and the combination of subband gradients that make the features more prominent.
4. The dimension reduction in extracted features is made by using a novel Pearson–kernel–principal component analysis method, and it gives the framework of this work an added quality.
5. The classification of the emotions is done by the proposed Gaussian membership function fuzzy support vector machine classifier (GMF-FSVM). The Gaussian membership function yields a promising result when added to fuzzy SVM owing to its smoothness and computational efficiency.

The flow for rest of this paper is organized as follows. Section 2 outlines the various related existing works. Section 3 provides the methodology for feature extraction using the statistical parameters obtained from the proposed subband selective multilevel stationary wavelet gradient transform (SM-SWGT) and describes the classification method using the proposed Gaussian membership function fuzzy SVM classifier. Section 4 illustrates the experimental setups and the results with discussions. Finally, Sect. 5 concludes the paper.

## 2 Related works

In contemporary years, recognition of facial expressions realized a challenging field in the area of research. Several works on emotion recognition using facial expression have been proposed by researchers in the literature. In facial expression recognition system, the face feature extraction is carried by geometric-based approach and appearance-based approach. The geometric-based approach exploits the relationships among human face components such as nose, eyebrows, lips, eyes in respect of location, distance and shape. Contrary, the appearance of a face is described in appearance-based approach. Recognition of emotion from face images using discrete wavelet transform (DWT) and convolution neural networks is presented by Bendjillali et al. [6]. The face images are enhanced using contrast enhanced adaptive histogram equalization, and the features are extracted using DWT. A 2 level DWT is performed and the decimation operation in DWT makes the coefficients inconsistent due to the shift invariance. The classification was done by using deep convolution neural network which makes the system more complex. Zhang et al. [7] computed the biorthogonal wavelet entropy from the face image to extract the multiscale features. The fuzzy SVM is used as classifier for classifying the emotion. The 2 level biorthogonal wavelet transform (BWT) used to impose the shift invariance problem leads to inconsistent coefficients. Also entropy is the only statistical parameter estimated which confines the generalization principle. Feature extraction for facial expression recognition utilizing the horizontal and vertical subband of stationary wavelet transform is made by Qayyum et al. [8]. Wang et al. [9] proposed a facial emotion recognition system by estimating the stationary wavelet entropy from face images. Moreover, Jaya algorithm is used for training and the classification is done by feed forward neural network which makes the system more complex. An optimum feature set for facial emotion recognition is obtained by combining the magnitude and phase features of Gabor wavelet transform and CNN [10]. The use of CNN necessitates many convolutional layers which makes the network structure complex.

A new approach in facial expression recognition based on the graph signal processing is presented by Meena et al. [11]. In this approach, the high dimensional histogram of gradient features of facial expression images is reduced into a relatively low dimension data using graph signal processing and the classification was done by kNN classifier. The extraction of local features from facial images plays a vital role in appearance-based approach. Facial expression recognition based on the Pyramid of Local Binary pattern is proposed by Khan et al. [12]. Here, reduced features are extracted from the salient face regions and are subjected to expression recognition by various classifiers. The perceptual salient region of expression calculation may lead to transpire wrong stimuli resulting in performance degradation of classifier. Another facial expression recognition system based on the local feature extraction using Local Prominent directional pattern is proposed by Makhmudkhujaev et al. [13]. A new feature descriptor formed by calculating the gradients in four direction of the reference pixel is given by Uma Maheswari et al. [14]. Here, expression-related micro-level features are extracted from the salient face regions and are subjected to expression recognition by various classifiers. A new approach using HOS features and SVM is presented by Ali et al. [15]. The HOS features are extracted from 1D facial signal which is obtained from successive projections of 2D spatial domain facial images by means of Radon transform. Principal component analysis (PCA) is applied on HOS-based features to reduce high dimensional features into a lower dimensional features space. The reduced features are then fed as input to SVM classifier to classify the facial emotions based on significant HOS features.

Highly discriminative binary sparse feature vectors are extracted using gentle boost decision trees algorithm by Gogić et al. [16]. The expression specific features extracted are classified using shallow neural networks which engender better performance and lesser execution time. The training process in the classification is improved by using a genetic algorithm called Bee Royalty Offspring Algorithm (BROA) proposed by Jamshidnezhad e al. [17]. In this method, the fuzzy knowledge base is improved by estimating and tuning the fuzzy membership function. A multitask global network which performs a multitask learning is proposed by Yu et al. [18]. By this approach, the local–global and the spatio-temporal features are extracted for recognition. The global and the local features are fused together to form a weakly supervised local global attention network proposed by Zhang et al. [19]. The attention map mechanism is used for part separation and local-global feature fusion. A multiple attention network to extract the discriminative features from the face region is proposed by Gan et al. [20]. The expression-related critical regions are extracted using region aware subnet masks, and subsequently expression recognition subnet with multiple attention blocks is used to learn discriminative features. Facial expression recognition model constructed by joint partial image and deep metric learning is proposed by Yu and Bai [21]. The action unit which is closely related to expression is cropped to generate partial image, and the intraclass similarities and interclass variations are enhanced using expression metric loss function. The expression metric loss and classification loss are jointly optimized to attain better performance.

An attention mechanism adopted in the convolution neural network in order to extract the region of interest in face image is presented by Sun et al. [22]. The ROI in face image is marked in the first layer of the convolution network, and more robust features are extracted. The facial feature fusion made by combining the discriminative features and handcraft features of shape and appearance is given by Fan et al. [23]. The facial expression recognition based on the fusion of HOG and CNN is presented by Pan [24]. The CNN features and HOG features are obtained for the useful temporal and spatial representation of the facial image. Hence, the facial appearance and corresponding facial motions are detected which results in increased recognition rate. Li et al. [25] presented an approach in which the size of the database is increased by random rotation and horizontal flipping of face image. This makes the CNN simple to extract the useful features which are best suited for classification. The absence of hidden fully connected layer in CNN creates a simple structure with improved accuracy. Discovering region of interest from face image and utilizing the relationship among ROI to train a robust deep CNN model is presented by Sun et al. [26]. The reliability of the predicted target was very much intensified resulting in performance proliferation. Reddy et al. [27] developed a recognition system with amalgamation of facial landmark points and XceptionNet features. The fusion of domain specific handcrafted features and efficient gradient informative deep learnt features improves the performance of recognition process. A novel local feature extraction method named neighborhood aware edge directional pattern was proposed by Iqbal et al. [28] which prevails the unstable feature descriptions due to weak and distorted edges. This method inspects the gradients at the center pixel and neighborhood pixels to generate pattern codes attaining consistency in local regions. Facial emotion detection by predicting the changes in face geometry is proposed by Joseph et al. [29]. A combination of DWT and fuzzy is adapted to enhance the face image, and the facial geometry is estimated with the help of landmarks using the modified eyemap and mouthmap algorithm. The classification is done by neural network availing the use of tensor flow.

Though various terminologies are conceded on facial expression recognition, still it is a challenging task due to the fact of getting stable features for classification. Copious deep learning methods are proposed in this field recently with some of them is quoted above, and getting vast dataset for training seems major problem with most complexity. Determination of accurate and reliable position of facial component is essential in geometric method, and it is difficult to resolve in case of disparity in face traits. This leads to the occurrence of poor results under distinct situations. The local feature extraction methods depicted suffers in the effect of noise and delicate local distortions. The aforementioned problems are tackled by proposing the novel feature extraction method subband selective multilevel stationary gradient transform. The wavelet transform is one among the best feature extraction method in emotion recognition approach. The discrete wavelet transform (DWT), curvelet transform (CT), shearlet transform (ST) and contourlet transform (CNT) use up samplers and down samplers resulting unwanted artifacts. The nonsubsampled contourlet transform (NSCT), nonsubsampled shearlet transform (NSST) and stationary wavelet transform (SWT) are suitable for feature extraction due to the shift-variant property. The implementation of NSCT and NSST is more computationally complex compared to stationary wavelet transform (SWT). The good localization characteristics of SWT in both spectral and spatial domains make the utilization of this transformational technique wider.

Various researchers developed facial expression recognition using SWT and exploited a significant result. In most of the works [7–9], the statistical values are extracted directly from the subbands which is inconsistent, incompatible and less discriminating. In this work, we devised a technique of segregating the subbands from different levels of stationary wavelet transform retaining only the expression-related significant features. The subband procured from the multilevel SWT is selected predicating the energy values of subband and a sequence of subband is evolved by subband combination. The statistical parameters calculated from this subband are more compatible and consistent which yields a significant accuracy in recognition.

# 3 Proposed facial emotion recognition method

The proposed method encompasses preprocessing, feature extraction, dimension reduction and classification. The preprocessing is entailed using the CLAHE algorithm. The feature extraction is contrived by the proposed SM-SWGT method, and the dimensionality reduction is devised using Kernel PCA method. Ultimately, the classification is performed using the proposed Gaussian membership function FSVM. The complete flow of the proposed work is illustrated in Fig. 1.

## 3.1 Preprocessing and face detection

Every face database requires preprocessing. In order to constitute the extraction process more fast, accurate and efficient preprocessing is performed on images before feature extraction. Conventionally, the image preprocessing step comprises of operations like image scaling, image brightness, image contrast adjustment and other image enhancement operations. Images with unequal value of brightness can be enhanced by using histogram equalization process. Contrast limited adaptive histogram equalization (CLAHE) which is a modified version of adaptive histogram equalization is done in this work. Rather computing global equalization by histogram equalization, an adaptive method redistributes the low intensity values of the image by dividing into tiles, in consequence computing several histograms. Each computed histogram corresponds to a distinct section of the image [30]. The contrast enhancement to each distinct section is done in such a manner that the histogram procured is used to generate intensity remapping function which is accomplished by the bilinear interpolation to smooth inter-distinct section boundaries. CLAHE confines over amplifying noise by clipping the histogram at predefined values in relatively homogeneous regions of an image. In CLAHE method, the histogram equalization is applied to each distinct region and a redistributed histogram is generated with each pixel intensity has been confined to a selected maximum [31].

Face detection is a technique used for extracting the face region alone from unwanted backgrounds for further processing. The Viola–Jones algorithm is a widely used and most accurate mechanism for the detection of face region [32] which is intensely used in the proposed work. The advantage of this algorithm is fast detection though the training process is slow. Haar basis feature filters are used in this algorithm; hence, it does not use multiplications. The first step in this algorithm is the Haar feature selection. Next is the generation of the integral image. The Haar extractors are calculated by adding only four numbers of the integral image by allowing integrals. Face detection is initiated using the detection window created by the integral image. While moving the detection window across the image, a set of face recognition filter is applied. The cascade connected classifiers of face recognition filters looks at the rectangular subset of the detection window. If it determines a face, the next classifier is applied. Ultimately face is detected only if the entire classifier filter yields a positive answer.

## 3.2 Proposed feature extraction method: subband selective multilevel stationary wavelet gradient transform (SM-SWGT)

Feature extraction is the most sensitive and critical stage in facial emotion recognition. The technique utilized in this stage is critical as it is much influenced with the efficiency of facial expression recognition. The common feature extraction techniques are principal component analysis (PCA), Gabor filters, local binary patterns (LBP), independent component analysis (ICA), linear discriminant analysis (LDA), etc. In this work, a modified form of stationary wavelet transforms, a subband selective multilevel stationary wavelet gradient transform (SM-SWGT) is tailored and is revealed in Fig. 2. The statistical parameters determined from subband selected multilevel stationary wavelet gradient transform of the input facial image constitute the face features extracted under feature extraction in this paper.

### 3.2.1 Subband selective multilevel stationary wavelet gradient transform

Wavelet transform gives the frequency domain and time domain information of a signal. The wavelet transform decompose the images into different frequency ranges by permitting the isolation of the frequency components into different subbands. In stationary wavelet transform, the input signal is convolved with low and high pass filter and the number of coefficients obtained after SWT is same as that of the samples in the input signal since there is no decimation involved. In order to make the transformation of SWT sym-

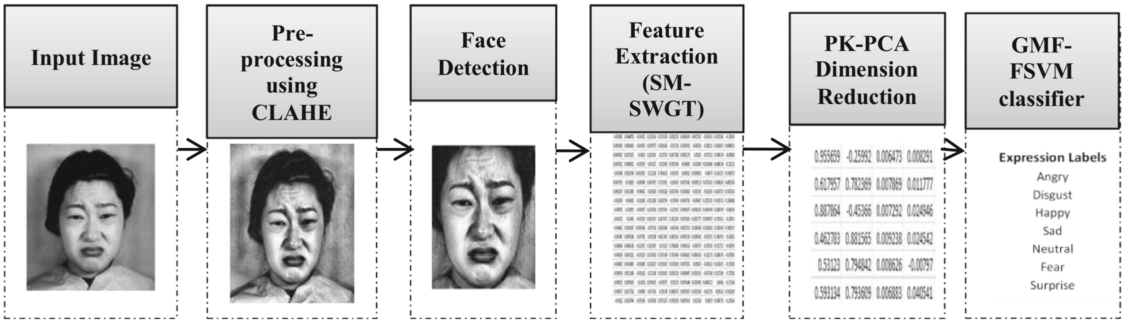
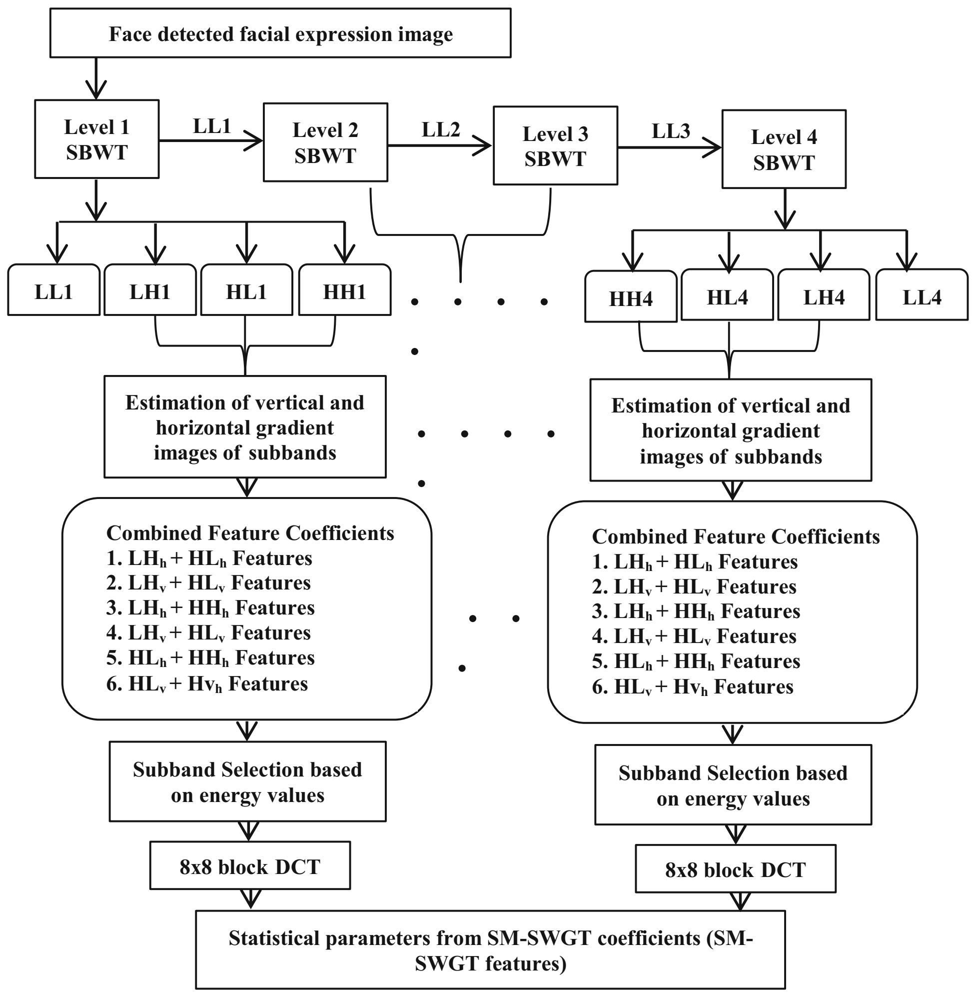



**Fig. 1** Illustration of proposed work

**Fig. 2** Proposed feature extraction method—SM-SWGT




metry, and to preserve more energies in the low frequency subbands, a non-orthogonal which means a biorthogonal wavelet transform is used in this paper. In facial expression identification, the detailed coefficients play a vital role rather than approximation coefficients [8]. Hence, the detailed coefficients LH, HL and HH coefficients are considered here.

The LH, HL and HH components are subjected to image gradient transform to obtain the horizontal and vertical gradient vectors. The edges of the images are obtained by estimating the horizontal and vertical gradient vector. The gradient operators are applied to sharpen or smoothen images. Gradient images are enumerated by convolving the original image with a gradient filter. The gradient filter chosen in the proposed work is the Sobel filter. The Sobel filter is selected because the high intensity point variation in the image does not distress its performance since the local averaging reduces this. Also in addition to the estimation of edge direction, the more informative edge magnitude is also esti-

| | | |
|---|---|---|
| -1 | 0 | +1 |
| -2 | 0 | +2 |
| -1 | 0 | +1 |

| | | |
|---|---|---|
| +1 | +2 | +1 |
| 0 | 0 | 0 |
| -1 | -2 | -1 |

**Fig. 3** Sobel convolutional kernel $3 \times 3$ (vertical and horizontal)

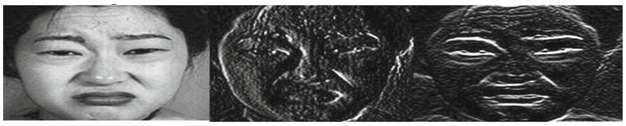


**Fig. 4** Subband face image and its horizontal and vertical gradient image

mated using Sobel operator. The changes in the images which are best suited for expression classification are transferred to the gradient domain [33]. The Sobel convolution kernel is shown in Fig. 3.

In the gradient image, each pixel appraises the transform in the intensity of that same pixel in the original image in a particular direction. The gradient images in the horizontal and vertical directions are computed in order to acquire the full range of direction. The strength of the edge corresponds to the rate of change of intensity of pixel. The pixels with higher gradient values constitute the possible edges, and the edges are discovered in a direction which is perpendicular to the gradient direction.

The horizontal and vertical gradient images are obtained using the formula

$$G_{hj}(a,b) = \sum\sum \Delta H_j \cdot X_j^k(a,b) \tag{1}$$

$$G_{vj}(a,b) = \sum\sum \Delta V_j \cdot X_j^k(a,b) \tag{2}$$

where $\Delta H_j$ and $\Delta V_j$ are the horizontal and vertical gradient filters at scale j and $X_j^k$ is the subband at scale j with k direction. The horizontal and vertical gradient image of a face image from jaffe dataset are shown in Fig. 4. The gradient images of the subbands obtained from the four levels of SWGT are filtered using an adaptive Wiener filter. The Wiener filter is used to reduce the high frequency component effects in the image.

Here, a subband combination method is proposed based on the weight value calculated from each subband. Hence, six subbands from each level are obtained after combination of coefficients as given in Eq. (3) and are subjected for feature extraction.

$$S_{kj}(a,b) = \begin{cases} p_j LH_{hj}(a,b) + q_j HL_{hj}(a,b), & \text{for} k = 1 \\ p_j LH_{vj}(a,b) + q_j HL_{vj}(a,b), & \text{for} k = 2 \\ p_j LH_{hj}(a,b) + r_j HH_{hj}(a,b), & \text{for} k = 3 \\ p_j LH_{vj}(a,b) + r_j HH_{vj}(a,b), & \text{for} k = 4 \\ q_j HL_{hj}(a,b) + r_j HH_{hj}(a,b), & \text{for} k = 5 \\ q_j HL_{vj}(a,b) + r_j HH_{vj}(a,b), & \text{for} k = 6 \end{cases} \tag{3}$$

where j = 1,2,3,4 gives the number of levels of SWGT, k gives the combination formed in each level using the detailed coefficient. $p_j$, $q_j$, and $r_j$ are the jth level weight value calculated by the Pearson correlation coefficient between approximation and individual detailed coefficients.

$$p_j = \frac{\operatorname{cov}\left(LL_j, LH_j\right)}{\sigma_{LL_j} \cdot \sigma_{LH_j}}, \quad q_j = \frac{\operatorname{cov}\left(LL_j, HL_j\right)}{\sigma_{LL_j} \cdot \sigma_{HL_j}}, \quad \&r_j = \frac{\operatorname{cov}\left(LL_j, HH_j\right)}{\sigma_{LL_j} \cdot \sigma_{HH_j}} \tag{4}$$

where cov is the covariance and $\sigma$ is the standard deviation which is calculated from the four levels of decomposition. The weight value calculated by the Pearson correlation coefficient gives more stability to the combinational measures.

### 3.2.2 Subband selection in SM-SWGT

From each level of SWGT, six subbands are obtained and a total of twenty-four subbands are determined from four levels of SWGT. The subbands requisite for the feature extraction are selected based on the calculated weighted energy value of each subband. Out of the 24 subbands determined from each image, 15 subbands with maximum weighted energy values are selected for feature extraction. The weighted energy values are calculated from norm 1, norm 2 energy and entropy values which is given in Eqs. (5), (6), (7) and (8).

$$\text{norm 1 energy is NE1} = \frac{1}{N}\sum_{n=1}^{N}|X| \tag{5}$$

$$\text{norm 2 energy is NE2} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}|X|^2} \tag{6}$$

$$\text{entropy is EN} = -\sum_k P_k log_2 P_k \tag{7}$$

$$\text{Weighted energy value is EV} = \sqrt{NE1^2 + NE2^2 + EN^2} \tag{8}$$

Discrete cosine transform is then applied to the above selected subband coefficients to transform the spatial domain into frequency domain. An $8 \times 8$ block DCT is applied, and the majority of the energy of that subband lie in the first coefficient (i.e., the DC component) of each block.

Hence, it is selected as feature from each subband. Ultimately, this combination of SM-SWGT and DCT results in improved classification. From the above SM-SWGT subband coefficients, the statistical parameters like mean, standard deviation, covariance, median, energy, skewness and kurtosis are calculated and is used for training the database.

**Algorithm**: Proposed method for feature extraction

1) Let $LH_j(a,b)$, $HL_j(a,b)$, $HH_j(a,b)$ (j=1,2,3,4) be the subbands obtained by applying four level of stationary biorthogonal wavelet transform from the preprocessed input image.
2) For each subband obtained in step 1, the horizontal and vertical gradient images are calculated using equation 1 & 2.
3) Apply wiener filter to gradient images of all subbands to reduce the high frequency components effect.
4) The pixel level fusion is made to the gradient images subbands to obtain a combination set of subbands $SS_{kj}(a,b)$ (k=1,2…..,6 & j=1,2,3,4) by equations 3 & 4.
5) The weighted energy value of all subbands is calculated by using equation 5, 6, 7 & 8.
6) The subbands with maximal energy values $SS_e(a,b)$ (e=1,2….,15) are selected for further processing.
7) An 8x8 block DCT is applied to all selected subbands and the dc coefficient obtained in each block is considered and the size of subband is changed to $SS_r(m,n)$. Here mxn = d << axb, d is the number of dc coefficients obtained in each subband.
8) Compute the statistical parameters such as mean, standard deviation, covariance, median, energy, skewness and kurtosis from each subband and convert the parameters from all subbands into an one dimensional vector with size of 1x105 (seven parameters from 15 subbands 7x15=105) which form feature vector for the input image.

For each input image, the output of the SM-SWGT consists of 15 subbands from the four levels of decomposition. From each subband, seven parameters are selected and hence 105 (15 × 7) feature values are selected from each image.

### 3.3 Dimension reduction using PK-PCA

To effectively reduce the dimension of feature values, principal component analysis (PCA) is employed. A low dimensional representation of the data is constructed using principal component analysis which describes as much variance in the data as possible. A linear basis of dimensionality reduction for the data is found by choosing the maximal amount of variance in the data. Kernel principal component methods are adopted in order to transform the input space nonlinearly in a high dimensional feature space [34]. The KPCA overcomes the difficulty of computing PCA in the high dimensional feature space by computing dot products in the low dimensional input space by the use of kernel functions.

For an input sample x, g(x) is the mapping of test point in feature space S. The projection in feature vectors $P^k$ is given by

$$\left(\mathrm{P}^{\mathrm{k}} \cdot \mathrm{g}(\mathrm{x})\right) = \sum_{i=1}^{m} \alpha_i^k (\mathrm{g}(\mathrm{x_i}) \cdot \mathrm{g}(\mathrm{x})) = \sum_{i=1}^{m} \alpha_i^k \mathrm{K}(\mathrm{x_i}, \mathrm{x}) \tag{9}$$

where α is the eigen vector coefficient and K is the kernel which is denoted as $K(x_i,x_j) = g(x_i)g(x_j)$. Thus, the kernel function avoids the problem of calculation in the high dimensional feature space. The kernel function used here is the Pearson kernel. The Pearson kernel is given by the equation

$$\mathrm{K}\left(\mathrm{x_i}, \mathrm{x_j}\right) = \frac{1}{\left[1 + \left(\frac{\sqrt[2]{\|x_i - x_j\|^2 \sqrt{2^{\left(\frac{1}{\omega}-1\right)}}}}{\sigma}\right)^2\right]^{\omega}} \tag{10}$$

In the above equation, tuning the parameters ω and $\sigma$ makes the kernel suitable for the dimension reduction system.

### 3.4 Classification using proposed GMF-FSVM

Support vector machine is a machine learning algorithm based on statistical learning theory that is used to analyze the data. SVM is a supervised learning algorithm and is widely used in both classification and regression tasks. If we have two classes of +1 and − 1 and the data set is

$$D = \{(u_i, v_i) | u_i R^P\}, \quad i = 1, \ldots n \tag{11}$$

where n is the number of samples, p gives the length of input feature coefficients, $u_i$ is the feature coefficients of ith sample and $v_i$ is its corresponding target label.

The SVM algorithm finds a hyperplane with maximum margin which classifies one class from other class. For each feature vector $u_i$ either

$$w^T \cdot u_i + b \geq 1 \text{ for } u_i \text{ having the class } 1 \tag{12}$$

or

$$w^T \cdot u_i + b \leq -1 \text{ for } u_i \text{ having the class } -1 \tag{13}$$

The unique constraint equation is given by

$$v_i\left(w^T \cdot u_i + b\right) \geq 1 \text{ for all } 1 \leq i \geq n \tag{14}$$

The cost function to find the optimal hyperplane is the following optimization function

$$\min_w \frac{1}{2} w^T w + C \sum_{i=1}^{P} \xi(w, x_i, y_i) \tag{15}$$

The fuzzy support vector machines (FSVM) apply a membership function known as fuzzy membership function to every training data and the fuzzy training data consists of input, target and the fuzzy altitude [35]. The objective function to find the optimal hyperplane using fuzzy membership function is

$$\min_w \frac{1}{2} w^T w + C \sum_{i=1}^{P} \pi_i \, \xi(w, x_i, y_i) \tag{16}$$

subject to $v_i((w.u_i) + b) \geq 1 - \xi_i \;\&\; \xi_i \geq 0, \;\; 1 \geq \pi_i \geq 0,$

The membership function used in this proposed GMF-FSVM is the Gaussian membership function and is specified by the membership function mentioned below with center and width denoted as c and $\sigma$.

$$\mathrm{GMF}(u, c, \sigma) = e^{-1/2\left(\frac{u-c}{\sigma}\right)^2} \tag{17}$$

Within each sample point, the +1 and $-1$ class category to the hyperplane distances are

$$\omega_{i+} = \mathrm{e}^{-1/2\left(\frac{(u-u_+)-c}{\sigma}\right)^2} \tag{18}$$

$$\omega_{i-} = \mathrm{e}^{-1/2\left(\frac{(u-u_-)-c}{\sigma}\right)^2} \tag{19}$$

where $u_+$ and $u_-$ are the mean sample points of class +1 and $-$ 1. The membership function for the fuzzy support vector machine is formulated as

$$\pi_\mathrm{i} = 1 - \frac{\omega_{i+}}{A_+ + \delta}, \quad \mathrm{v_i} = +1 \text{ class} \tag{20}$$

$$\pi_\mathrm{i} = 1 - \frac{\omega_{i-}}{A_- + \delta}, \quad \mathrm{v_i} = -1 \text{ class} \tag{21}$$

In the above equation, $A_+ = \max\{\omega_{i+}\}$, $A_- = \max\{\omega_{i-}\}$ and $\delta$ is a small value to make $\pi_i$ nonzero. In this work, we have a multiclass problem with seven expressions to be recognized. The detailed comparison made by Yu Dong Zhang et al. [7] explains that the one-against-one-based method is a competitive approach for multiclass SVM. In one-against-one-based multiclass SVM method, for each class one SVM classifier is constructed and hence, 7 individual SVM's need to be created. The classification can be done by testing a new test data to determine whether the test data belong to a class or not. A score is obtained as the output of each classifier and the classifier that obtains the maximum score is identified as the test class.

The not separable problem is converted to separable problem in SVM by transforming a low dimensional input space to a higher dimensional space with a technique called the kernel trick. The different kernels functions such as the linear, polynomial and radial basis function kernels make the hyperplane decision boundary different between the classes. The performance of the classifier is studied based on the consideration of different kernel functions and parameters. The mathematical formula for the kernel functions is given in the below equations

$$\text{Linear kernel function, } K\left(u_i, u_j\right) = u_i^T \cdot u_j \tag{22}$$

$$\text{Polynomial function with degree } p,\; K\left(u_i, u_j\right) = \left(1 + u_i^T \cdot u_j\right)^p \tag{23}$$

$$\text{Radial basis function using Gaussian, } \mathrm{K(u_i, u_j)} = \exp\left(\frac{\left\|u_i - u_j\right\|^2}{2\sigma^2}\right) \tag{24}$$

The Pearson kernel is given in Eq. (**10**).

## 4 Results and discussion

### 4.1 Datasets

The proposed work has been implemented using MATLAB software for JAFEE database [36], CK + database [37] and FG net database [38]. In the JAFFE dataset used, there are 213 gray scale images with seven different expressions of 10 different females. Each image has a spatial resolution of $256 \times 256$. The FG-NET dataset with facial expressions is an image database which contains face images of a number of persons delivering different emotions. The database contains images of 18 different individuals. Each image has a spatial resolution $320 \times 240$. The CK + dataset includes facial expression images of 210 adults with both posed and non-posed position. The posed images are well suited for this work, and hence, posed images from CK + dataset are considered for this work. The CK + dataset has face images with seven different emotions. The spatial resolution of each image used is $640 \times 480$.

## 4.2 Performance analysis experiments and results

The preprocessing of facial image is done using CLAHE. Here, the size of distinct section is 8x8 and the clip limit is 0.01. The face detection is made by Viola–Jones algorithm, and face detected image is normalized to 128x128 resolution. A tenfold stratified cross-validation is used in this work. The stratification splits the total number of images into 10 different folds in such a way that each fold contains the same number of class labels. For training eight folds are used, for validation onefold is used and for testing onefold is used in each trail. The sensitivity results after tenfold cross-validation is given in Table 1. The sensitivity results show that the images with happy expression are the one that can be identified easily. The images with anger expression are the second easiest to identify and then the neutral, fear, sad, disgust and surprise.

The happy expression is easy to identify because of the four distinguished features of the happy image which includes the eye corners, lip corners, forehead muscles and the eyebrows. The distinguished features of the angry image are the eyebrows, wrinkled nose, narrowed eyes and jaws. The neutral expression relaxes the facial muscles, while the other expression uses the facial muscles around the ear, muscles around eyelid, muscles around nose and muscles around the mouth.

The CRR with different levels from $n = 1$ to $n = 5$ are shown in Table 2. It clearly shows that the best performance is attained at level $n = 4$. The horizontal and the vertical gradient image obtained by the gradient transform shows the edges of the image. The combination of subband gradients with the weight value calculated by Pearson correlation coefficient makes the subbands more effective in terms of sensitivity. The weighted energy value for the 24 subbands obtained from the four levels of SWGT for a face image is plotted in Fig. 5. Out of the 24 subbands, a maximum of 15

**Table 1** Sensitivity results based on statistical analysis

| | Anger | Disgust | Fear | Happy | Neutral | Sad | Surprise |
|---|---|---|---|---|---|---|---|
| Run1 | 100 | 98 | 100 | 100 | 100 | 99 | 97 |
| Run2 | 100 | 97 | 98 | 100 | 99 | 97 | 99 |
| Run3 | 100 | 99 | 100 | 99 | 99 | 100 | 97 |
| Run4 | 99 | 100 | 100 | 100 | 98 | 100 | 99 |
| Run5 | 100 | 98 | 98 | 100 | 99 | 99 | 99 |
| Run6 | 100 | 97 | 98 | 100 | 100 | 99 | 99 |
| Run7 | 99 | 98 | 100 | 100 | 99 | 97 | 96 |
| Run8 | 100 | 97 | 97 | 99 | 100 | 99 | 98 |
| Run9 | 100 | 99 | 98 | 100 | 98 | 98 | 96 |
| Run10 | 99 | 98 | 100 | 100 | 99 | 97 | 99 |
| Average | 99.7 | 98.1 | 98.9 | 99.8 | 99.1 | 98.5 | 97.9 |

**Table 2** Classification performance with different levels of SM-SWGT

| Run | $N = 1$ | | | $N = 2$ | | | $N = 3$ | | | $N = 4$ | | | $N = 5$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jafee dataset | CK + dataset | FG net dataset | Jafee dataset | CK + dataset | FG net dataset | Jafee dataset | CK + dataset | FG net dataset | Jafee dataset | CK + dataset | FG net dataset | Jafee dataset | CK + dataset | FG net dataset |
| 1 | 89.1 | 90.1 | 88.2 | 92.8 | 92.8 | 89.1 | 95.7 | 96.1 | 94.9 | 97.3 | 98.4 | 97.1 | 94.3 | 94.8 | 94.1 |
| 2 | 91.6 | 91.9 | 89.2 | 91.2 | 91.5 | 90.6 | 96.2 | 96.8 | 95.2 | 96.7 | 98.9 | 95.5 | 95.5 | 95.7 | 94.9 |
| 3 | 90.2 | 90.6 | 89.6 | 93.5 | 93.8 | 92.8 | 94.9 | 94.5 | 93.8 | 97.6 | 98.7 | 95.8 | 94.7 | 95.3 | 94.6 |
| 4 | 87.3 | 89.1 | 87.1 | 94.1 | 94.4 | 93.8 | 95.3 | 95.1 | 94.9 | 97.3 | 100 | 97.1 | 96 | 95.9 | 95.6 |
| 5 | 89.5 | 89.4 | 88.5 | 92.3 | 92.1 | 91.9 | 96.4 | 96.5 | 95.3 | 97.4 | 98.6 | 95.6 | 95.4 | 95.8 | 95.1 |
| 6 | 88.4 | 88.9 | 88.6 | 93.2 | 93.4 | 92.8 | 96.5 | 96.7 | 96.1 | 96.6 | 97.8 | 94.6 | 93.6 | 94.1 | 93.5 |
| 7 | 88.2 | 88.1 | 87.9 | 91.4 | 91.6 | 91.2 | 95.6 | 96.1 | 95.4 | 96.4 | 99.1 | 94.1 | 93.2 | 93.5 | 93.5 |
| 8 | 90.5 | 92.4 | 90.2 | 92.9 | 93.1 | 92.8 | 95.8 | 96.4 | 94.2 | 96.9 | 100 | 95.3 | 94.1 | 93.9 | 93.9 |
| 9 | 89.7 | 90.6 | 89.7 | 93.9 | 94.2 | 93.8 | 96.1 | 96.5 | 94.8 | 97.9 | 98.9 | 97.1 | 94.7 | 94.9 | 94.1 |
| 10 | 90.1 | 91.1 | 88.6 | 94.4 | 94.4 | 94.1 | 94.9 | 95.5 | 95.1 | 97.1 | 98.5 | 96.5 | 95.1 | 95.5 | 94.9 |
| Average | 89.5 | 90.2 | 88.7 | 92.9 | 93.1 | 92.3 | 95.7 | 96.5 | 94.9 | 97.1 | 98.9 | 95.8 | 94.6 | 94.9 | 94.4 |

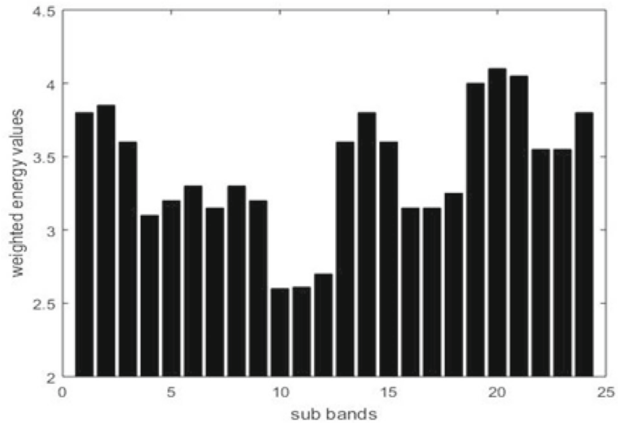



**Fig. 5** Weighted energy values obtained for a face image

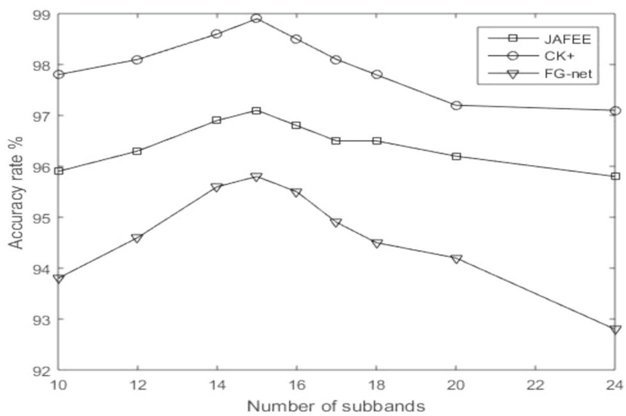



**Fig. 6** Average accuracy rate obtained at different number of subbands

subbands with greater energy values are selected. The best accuracy rate is obtained when 15 numbers of subband are selected which is clearly indicated by Fig. 6.

The CRR improvement attained by using PK-PCA dimension reduction technique is given in Fig. 7. It is evident that the PK-PCA yields greater recognition rate in all the three datasets. The Pearson kernel produces good result when compared with the existing kernel such as linear, RBF and polynomial kernel. The smoothness of the kernel mostly depends on the value of $\sigma$ and ω. The value of $\sigma$ is chosen from the set $\{ 2^{10}, 2^{9}, \ldots, 2^{-9}, 2^{-10}\}$, and the value of ω is chosen from the set $\{2^{1}, 2^{2}, \ldots, 2^{9}, 2^{10}\}$. The values are chosen using grid search method along with cross-validation. The classification accuracy attained for certain values of $\sigma$ and ω is given in Table 3. For this work, the value of $\sigma$ and ω chosen experimentally is $2^{10}$ and $2^{6}$. The classification performance considering different kernel functions by the proposed fuzzy support vector machine for different database is shown in Tables 4, 5 and 6. The number of subbands selected after stationary wavelet gradient transform and their performance obtained is shown in table. It is observed that the best result is obtained when 15 number of subband is selected. It is observed that the Pearson Kernel functions produce high accuracy for most of the cases (Table 6).

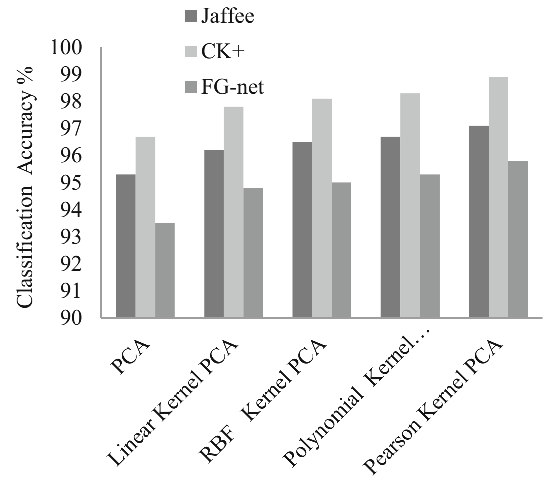



**Fig. 7** Classification accuracy improvement using dimension reduction PK-PCA

**Table 3** Parameter selection for Pearson Kernel

| PGK parameters ($\sigma$,ω) | Jaffe dataset | CK + Dataset | FG-net Dataset |
|---|---|---|---|
| $2^{2}, 2^{4}$ | 96.6 | 98.1 | 95.4 |
| $2^{-3}, 2^{4}$ | 96.5 | 98.5 | 95.6 |
| $2^{-4}, 2^{5}$ | 96.8 | 98.6 | 95.6 |
| $2^{-6}, 2^{6}$ | 96.9 | 98.8 | 95.6 |
| $\mathbf{2^{10}, 2^{6}}$ | **97.1** | **98.9** | **95.8** |

The learning parameters of a classifier have to be chosen prudent to achieve good classification accuracy. The penalty parameter C has a robust effect in the performance of SVM. Larger value of C leads to less training error but consumes more training time. Moreover, smaller value of C generates larger margin but generate more errors. Also, optimal kernel parameters and fuzzy values need to be selected for better performance accuracy. In this procedure, the parameter of the fuzzy membership function is set first (i.e., center $c = 5$ and width $\sigma = 2$). Then, the penalty parameter is searched with the desired kernel parameters as described in Table 7. For each kernel, the optimal penalty parameter C varies and the best performance is achieved with $C = 200$ for Pearson kernel. The bold font in Table 3, 4, 5, 6 and 7 indicates best performance results. The confusion matrix obtained by the proposed fuzzy support vector machine classifier for the seven emotions of JAFFE dataset is given in Table 8. Fear and sad emotion having CRR of 96.1 and 95.4 percent, respectively, and are difficult to recognize compared to other expressions. Also, it is observed that fear and anger can be confused easily. This is because fear and anger expression evokes same muscular activities. Again, it is observed

**Table 4** Correct recognition rate for JAFEE database with different kernel function

| Number of subband selected | Feature vector dimension (subband selected × 7) | Linear | RBF | Polynomial | Pearson |
|---|---|---|---|---|---|
| 24 | 168 | 93.1 | 93.3 | 94.6 | 95.8 |
| 20 | 140 | 95.1 | 95.4 | 95.9 | 96.2 |
| **15** | 105 | 95.5 | 96.2 | 96.8 | **97.1** |
| 10 | 70 | 93.9 | 94.9 | 95.3 | 95.9 |

**Table 5** Correct recognition rate for CK + database with different kernel function

| Number of subband selected | Feature vector dimension (subband selected × 7) | Linear | RBF | Polynomial | Pearson |
|---|---|---|---|---|---|
| 24 | 168 | 96.1 | 96.5 | 97.8 | 98.1 |
| 20 | 140 | 97.5 | 97.9 | 97.1 | 97.2 |
| **15** | 105 | 98.0 | 98.1 | 98.3 | **98.9** |
| 10 | 70 | 97.3 | 97.1 | 97.3 | 97.8 |

**Table 6** Correct recognition rate for FG-net database with different kernel function

| Number of subband selected | Feature vector dimension (subband selected × 7) | Linear | RBF | Polynomial | Pearson |
|---|---|---|---|---|---|
| 24 | 168 | 93.3 | 93.2 | 93.6 | 92.8 |
| 20 | 140 | 93.5 | 93.8 | 93.9 | 94.2 |
| **15** | 105 | 94.9 | 95.1 | 95.6 | **95.8** |
| 10 | 70 | 94.4 | 94.2 | 94.3 | 93.8 |

**Table 7** Hyperparameter selection of FSVM on datasets with center (c = 5) and width ($\sigma = 2$)

| Kernel | C | Jaffe dataset | CK + dataset | FG-net dataset |
|---|---|---|---|---|
| Linear | 10 | 95.1 | 97.6 | 94.7 |
| | 50 | 95.1 | 97.7 | 94.7 |
| | 150 | 95.5 | 98.0 | 94.9 |
| | 500 | 95.3 | 97.9 | 94.9 |
| Gaussian RBF ($\sigma = 2$) | 10 | 95.8 | 97.9 | 94.8 |
| | 100 | 95.9 | 98.1 | 94.8 |
| | 250 | 96.2 | 98.1 | 95.1 |
| | 500 | 95.9 | 97.8 | 95.0 |
| Polynomial (p = 2) | 10 | 96.1 | 97.9 | 95.1 |
| | 100 | 96.2 | 98.1 | 95.3 |
| | 180 | 96.8 | 98.3 | 95.6 |
| | 500 | 96.6 | 98.2 | 95.5 |
| Pearson kernel ($\sigma = 2^{10}$), ($\omega = 2^{6}$) | 10 | 96.5 | 98.5 | 95.3 |
| | 50 | 96.7 | 98.6 | 95.3 |
| | 100 | 96.7 | 98.6 | 95.5 |
| | **200** | **97.1** | **98.9** | **95.8** |
| | 600 | 96.9 | 98.7 | 95.7 |
| | 1000 | 96.9 | 98.7 | 95.7 |

that sad and neutral are also confused easily. This is because the sad expression has little different muscle activities than neutral expression. The confusion matrix for CK + database is presented in Table 9. The happy expression is recognized easily because the forehead muscles relax and the eyebrows slightly get pulled up in this expression. The disgust is difficult to recognize. This is because for disgust expression, the eyebrows and the upper lips are pulled up which is just similar to expression of surprise. It is observed that disgust and surprise are confused easily because of the above said reason. The confusion matrix of FG-net database is shown in Table 10. It shows that happy expression is recognized well,

**Table 8** Confusion matrix of the proposed method on jaffe database

| | Anger | Disgust | Fear | Happy | Neutral | Sad | Surprise |
|---|---|---|---|---|---|---|---|
| Anger | 98.4 | 0 | 0.9 | 0 | 0 | 0.7 | 0 |
| Disgust | 0 | 97.3 | 0.7 | 0.9 | 0 | 1.1 | 0 |
| Fear | 1.8 | 0 | 96.1 | 1.2 | 0.9 | 0 | 0 |
| Happy | 0.5 | 0 | 0.2 | 99.3 | 0 | 0 | 0 |
| Neutral | 0 | 0 | 1.5 | 0 | 96.7 | 1.8 | 0 |
| Sad | 0 | 1.5 | 2.4 | 0 | 0.7 | 95.4 | 0 |
| Surprise | 1.3 | 0 | 1.8 | 0 | 0 | 0.7 | 96.2 |

**Table 10** Confusion matrix of the proposed method on FG-net database

| | Anger | Disgust | Fear | Happy | Neutral | Sad | Surprise |
|---|---|---|---|---|---|---|---|
| Anger | 96.7 | 0 | 1.1 | 0 | 0 | 1.2 | 0 |
| Disgust | 0 | 94.8 | 1.2 | 0 | 2.4 | 1.6 | 0 |
| Fear | 0.8 | 0 | 95.8 | 1.2 | 1.6 | 0 | 0.6 |
| Happy | 1.2 | 0 | 0.5 | 98.3 | 0 | 0 | 0 |
| Neutral | 0 | 0.4 | 1.5 | 0 | 96.3 | 1.8 | 0 |
| Sad | 0 | 1.5 | 1.2 | 0 | 1.7 | 95.6 | 0 |
| Surprise | 3.2 | 0 | 0 | 0 | 2.2 | 0.7 | 93.9 |

while surprise is less recognized. Also, it is observed that sad and neutral get confused easily. Overall, when we compare the accuracy among three datasets, the CK + yields a high accuracy than jaffe and FG-net.

### 4.3 Comparative analysis

In this section, a comparison is made among various wavelet transform approaches related to the facial emotion recognition strategy in CK + dataset. The various methods compared are BWE + FSVM [7], SWT + DCT + NN [8], SWE + SHLFNN [9], DWT + CNN [6], GWT + CNN [10] and the proposed method. In [7–9], the stationary wavelet transform is used for feature extraction. Table 11 shows the comparison of the above methods in terms of correct recognition rate for various expressions. We can observe that in the overall recognition accuracy our proposed method yield high CRR and this proves the effectiveness of this method. The proposed method outperforms Zhang et al. [7] and Wang [9] methods in which the biorthogonal wavelet entropy calculated from two levels is used as features in the former case and stationary wavelet entropy from four levels are used as features in the latter case. The statistical parameters calculated and the selections of subbands make our method better compared to the above methods. Combining the features of LH and HL, subband of SWT to attain better accuracy is shown in [8] by Qayyum et al. which utilize DCT for feature reduction. The face traits in face image leads to misrecognition of expression, but the image gradient used in the proposed method signifies the edge in face image which gives better performance in terms of accuracy. The proposed method also outperforms [6, 10] which uses DWT and GWT for feature extraction which is clearly evident in Table 11.

The proposed method is also compared with some of state-of-the-art approaches in facial emotion recognition which uses technique rather than wavelet transform in feature extraction. Table 12 describes the comparison made with the methods in terms of accuracy rate. For all the three datasets used in this work, the proposed method yields better results and it is clearly shown in Table 12 as bold font. It is a fact that the proposed method cannot be directly compared to the other methods in Table 12 as the experimental setup differs. Hence, the accuracy rate of the methods is just enumerated in Table 12. The proposed method surpasses the appearance-based approaches such as HOG + GSP [11] and HOS [15] since statistical analysis and feature selection are not adopted in those methods. The comparison result shows that the proposed method hugely outruns with a high margin in the accuracy rate. The comparison result with some of the local feature extraction method such as NEDP [28], PLBP [12], LPDP [13] and LDMEP [14] is also displayed in Table 12. The proposed method achieves better performance than the listed local descriptors. The statistics of the weighted distribution of directional information accumulated in the local shape is represented as local features in the above methods. Due to the ability of the proposed method to enhance edges by gradient calculation and the suppression of redundancy by the selective subband structure, better accuracy is attained.

**Table 9** Confusion matrix of the proposed method on CK + Database

| | Anger | Contempt | Disgust | Fear | Happy | Sad | Surprise |
|---|---|---|---|---|---|---|---|
| Anger | 99.7 | 0 | 0.3 | 0 | 0 | 0 | 0 |
| Contempt | 0.3 | 99.1 | 0.6 | 0 | 0 | 0 | 0 |
| Disgust | 0 | 0 | 98.1 | 0 | 0.2 | 0.5 | 1.2 |
| Fear | 0 | 0.5 | 0 | 98.9 | 0 | 0 | 0.6 |
| Happy | 0 | 0 | 0 | 0 | 100 | 0 | 0 |
| Sad | 0 | 0.8 | 0.7 | 0 | 0 | 98.5 | 0 |
| Surprise | 0.3 | 0 | 0 | 0.5 | 0.9 | 0 | 98.3 |

**Table 11** Analyzing the effect of proposed method by comparison

| | BWE + FSVM (Zhang et al. [7]) | SWT + DCT + NN (Qayyum et al. [8]) | SWE + SHLFNN (Wang et al. [9]) | DWT + CNN (Bendjillali et al. [6]) | GWT + CNN (Qin et al. [10]) | SM-SWGT + GMF-FSVM (proposed) (%) |
|---|---|---|---|---|---|---|
| Anger | 98.0 | 98.65 | 98.1 | 95.0 | 90.38 | 99.7 |
| Disgust | 95.6 | 93.7 | 95.5 | 98.5 | 96.58 | 98.1 |
| Fear | 96.1 | 95.4 | 96.3 | 93.7 | 96.15 | 98.9 |
| Happy | 98.5 | 96.85 | 98.3 | 93.1 | 97.71 | 100 |
| Neutral | 97.3 | 94.62 | 97.1 | 100 | 100 | 99.1 |
| Sad | 96.6 | 97 | 96.8 | 99.4 | 93.22 | 98.5 |
| Surprise | 95.3 | 100 | 95.5 | 99.7 | 100 | 97.9 |
| Overall | 96.77% | 96.61% | 96.8% | 97.05% | 96.81% | 98.9 |

Supplemental to this, the inclusion of dimension reduction technique and fuzzy classification furtherance the proposed method good in classification compared to other state-of-the-art methods.

The accuracy of some of the recent deep learning feature extraction methods proposed by Yu et al. [18], Zhang et al. [19], Gan et al. [20], Sun et al. [22], Fan et al. [23], Pan et al. [24], Li et al. [25] and Xiao et al. [26] is also shown in Table 12. In the recent years, facial emotion recognition using deep learning strategies expresses progressed result. Nonetheless, the inadequacy of sufficient data for training causes degradation of performance in progress. To discord this issues, the number of training samples is artificially increased by using data augmentation or by using the samples from other domain to pre train the network. This makes the deep learning approaches more complex compared to other methods. It is clearly evident in Table 12 that the performance of the proposed method is above those deep learning methods in terms of classification accuracy.

Although our facial expression recognition system achieved better results, there exists some problem. It is a bit sensitive to illumination changes and the geometric transformation of faces is not considered in this work. The database used for this analysis contains images only, and it is not analyzed with video databases. Our future research work aims at solving this problem. In future, this method can be applied to the micro-expression recognition system [39] which is similar to our facial expression system.

**Table 12** Performance comparison to the state-of-the-art approaches

| References | Dataset | Techniques used | Accuracy rate (%) |
|---|---|---|---|
| Meena et al. [11] | Jaffe | HOG + GSP + kNN | 88.57 |
| Ali et al. [15] | Jaffe | RT + HOS + SVM | 86.09 |
| Sun et al. [22] | Jaffe | AM/CNN | 92 |
| Iqbal et al. [28] | Jaffe | NEDP + SVM | 67.97 |
| Khan et al. [12] | Fg-Net | PLBP +SVM | 92.3 |
| Jamshidnezhad et al. [17] | Fg-Net | BROA | 89.8 |
| Meena et al. [11] | Ck+ | HOG + GSP + kNN | 97.61 |
| Khan et al. [12] | Ck+ | PLBP + SVM | 96.7 |
| Makhmudkhujaev et al. [13] | Ck+ | LPDP + SVM | 94.5 |
| Uma Maheswari et al. [14] | Ck+ | LDMEP | 94.85 |
| Yu et al. [18] | Ck+ | MGLN + LSTM | 98.7 |
| Zhang et al. [19] | Ck+ | WS-LGAN | 98.06 |
| Gan et al. [20] | Ck+ | MA/CNN | 96.28 |
| Yu et al. [21] | Ck+ | PSE | 98.83 |
| Sun et al. [22] | Ck+ | AM/CNN | 87.2 |
| Fan et al. [23] | Ck+ | Fused features | 92.5 |
| Pan et al. [24] | Ck+ | HOG + CNN | 97.01 |
| Li et al. [25] | Ck+ | CNN | 97.38 |
| Xiao et al. [26] | Ck+ | CNN | 94.67 |
| Iqbal et al. [28] | Ck+ | NEDP + SVM | 92.97 |
| Proposed method | Jaffe | **SM-SWGT + GMF-FSVM** | **97.1** |
| | Fg-Net | **SM-SWGT + GMF-FSVM** | **95.8** |
| | Ck+ | **SM-SWGT + GMF-FSVM** | **98.9** |

## 5 Conclusion

In this paper, a new face feature extraction method for a facial emotion recognition system is proposed. Different emotions generate varying muscular movements on human face. It is observed that the majority of the emotions are expressed with the horizontal and vertical muscle movements on the face. Therefore, the LH, HL and HH subbands of stationary wavelet transform which represents the horizontal, vertical and diagonal directional information alone is considered, rejecting the approximation subband. The gradient image of the subbands is estimated to generate more effective features that influences more accurate recognition. In order to avoid redundancy, the subbands are selected based on the weighted energy values of subbands and then DCT is performed utilizing the more significant information in the selected subbands. The statistical parameters are calculated for the selected subbands, and they are subjected to Pearson Kernel PCA for feature dimensionality reduction. A GMF-fuzzy SVM is used as the classifier to classify the different facial emotions. Experimental results show that the method proposed in this paper gives a promising result in facial emotion recognition compared to existing methods in the literature, and this proposed method is robust for different facial emotional expressions. About 1056 images showing different facial emotions from various datasets like JAFEE database, CK + database and FG Net database have been used for the classification. The future research will be on analyzing images with different postures, images with non-uniform illumination and images with occlusion and real-time video stream that will lead to meeting the necessity of many engineering applications.

**Author contributions** All authors have made significant contributions to the research reported and have read and approved the submitted manuscript.

**Funding** Not applicable.

## Compliance with ethical standards

**Conflict of interest** No potential conflict of interest was reported by the authors.

## References


1. Gavrilescu, M.: Recognizing emotions from videos by studying facial expressions, body postures and hand gestures. In: 23rd Telecommunications ForumTelfor, Belgrade, SERBIA (2015). https://doi.org/10.1109/telfor.2015.7377568
2. Li, W., Zhang, Y., Fu, Y.: Speech Emotion recognition in elearning system based on affective computing. In: Proceedings of Natural Computation (ICNC 2007), Aug 2007 (2007). https://doi.org/10.1109/icnc.2007.677
3. Cambria, E.: Affective computing and sentiment analysis. IEEE Intell. Syst. **31**(2), 102–107 (2016). https://doi.org/10.1109/MIS.2016.31
4. Darwin, C.: The expression of the emotions in man and animals. J. Murray, London (1872)
5. Ekman, P., Friesen, W.V.: Constant across cultures in face and emotions. J. Personal. Soc. Psychol. **17**(2), 124–129 (1971)
6. Bendjillali, R.I., Beladgham, M., Merit, K., Taleb-Ahmed, A.: Improved facial expression recognition based on DWT feature for deep CNN. Electronics **8**(3), 324 (2019). https://doi.org/10.3390/electronics8030324
7. Zhang, Yu.-Dong., Yang, Zhang.-Jing., Hui-Min, Lu., Zhou, Xing.-Xing., Phillips, Preetha., Liu, Qing.-Ming., Wang, Shui.-Hua.: Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified crossvalidation. IEEE Access (2016). https://doi.org/10.1109/ACCESS.2016.2628407.4
8. Qayyum, H., Majid, M., Anwar, S.M., Khan, B.: Facial expression recognition using stationary wavelet transform features. Hindawi Math. Probl. Eng. (2017). https://doi.org/10.1155/2017/9854050
9. Wang, S.H., Phillips, P., Dong, Z.C., Zhang, Y.D.: Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. Neurocomputing **10**, 15–20 (2017). https://doi.org/10.1016/j.neucom.2017.08.015
10. Qin, Shu., Zhu, Zhengzhou., Zou, Yuhang., Wang, Xiaowei.: Facial expression recognition based on Gabor wavelet transform and 2-channel CNN, Int. J. Wavelets Multiresol. Inf. Process. (2020). https://doi.org/10.1142/S0219691320500034
11. Meena, H.K., Joshi, S.D., Sharma, K.K.: Facial expression recognition using graph signal processing on HOG. IETE J. Res. (2019). https://doi.org/10.1080/03772063.2019.1565952
12. Khan, R.A., Meyer, A., Konik, H., Bouakaz, S.: Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit. Lett. **34**, 1159–1168 (2013). https://doi.org/10.1016/j.patrec.2013.03.022
13. Makhmudkhujaev, F., Abdullah-Al-Wadud, M., Iqbal, M.T.B., Ryu, B., Chae, O.: Facial expression recognition with local prominent directional pattern. Signal Process. Image Commun. (2019). https://doi.org/10.1016/j.image.2019.01.002
14. Uma Maheswari, V., Varaprasad, G., Viswanadha Raju, S.: Local directional maximum edge patterns for facial expression recognition. J. Ambient Intell. Hum. Comput. (2020). https://doi.org/10.1007/s12652-020-018863
15. Ali, H., Hariharan, M., Yaacob, S., Adom, A.H.: Facial emotion recognition based on higher-order spectra using support vector machines. J. Med. Imaging Health Inform. **5**(6), 1272–1277 (2015). https://doi.org/10.1166/jmihi.2015.1527
16. Gogić, I., Manhart, M., Pandžić, I.S., et al.: Fast facial expression recognition using local binary features and shallow neural networks. Vis. Comput. **36**, 97–112 (2020). https://doi.org/10.1007/s00371-018-1585-8
17. Jamshidnezhad, A., Nordin, M.J.: Bee royalty offspring algorithm for improvement of facial expressions classification model. Int. J. Bio-Inspired Comput. (2013). https://doi.org/10.1504/IJBIC.2013.055092
18. Yu, M., Zheng, H., Peng, Z., Dong, J., Du, H.: Facial expression recognition based on a multi-task global-local network. Pattern Recognit. Lett. (2020). https://doi.org/10.1016/j.patrec.2020.01.016
19. Zhang, H., Su, W., Wang, Z.: Weakly supervised local-global attention network for facial expression recognition. IEEE Access (2020). https://doi.org/10.1109/ACCESS.2020.2975913
20. Gan, Yanling., Chen, Jingying., Yang, Zongkai., Luhui, Xu.: Multiple attention network for facial expression recognition. IEEE Access (2019). https://doi.org/10.1109/ACCESS.2020.2963913
21. Yu, N., Bai, D.: Facial expression recognition by jointly partial image and deep metric learning. IEEE Access **8**, 4700–4707 (2019). https://doi.org/10.1109/ACCESS.2019.2963201
22. Sun, X., Zheng, S., Fu, H.: ROI-attention vectorized CNN model for static facial expression recognition. IEEE Access (2020). https://doi.org/10.1109/ACCESS.2020.2964298
23. Fan, X., Tjahjadi, T.: Fusing dynamic deep learned features and handcrafted features for facial expression recognition. J. Vis. Commun. Image Represent. (2019). https://doi.org/10.1016/j.jvcir.2019.102659

24. Pan, X.: Fusing HOG and convolutional neural network spatial–temporal features for video-based facial expression recognition. IET Image Process **14**(1), 176–182 (2020). https://doi.org/10.1049/iet-ipr.2019.0293
25. Li, K., Jin, Y., Akram, M.W., et al.: Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. Vis. Comput. **36**, 391–404 (2020). https://doi.org/10.1007/s00371-019-01627-4
26. Sun, X., Xia, P., Zhang, L., Shao, L.: A ROI-guided deep architecture for robust facial expressions recognition. Inf. Sci. **522**, 35–48 (2020). https://doi.org/10.1016/j.ins.2020.02.047
27. Reddy, G.V., Savarni, C.D., Mukherjee, S.: Facial expression recognition in the wild, by fusion of deep learnt and hand-crafted features. Cognit. Syst. Res. **62**, 23–34 (2020). https://doi.org/10.1016/j.cogsys.2020.03.002
28. Iqbal, M.T.B., Abdullah-Al-Wadud, M., Ryu, B., Makhmudkhujaev, F., Chae, O.: Facial expression recognition with neighborhood-aware edge directional pattern (NEDP). IEEE Trans. Affect. Comput **11**(1), 125–137 (2020). https://doi.org/10.1109/taffc.2018.2829707
29. Joseph, A., Geetha, P.: Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. Vis. Comput. **36**, 529–539 (2020). https://doi.org/10.1007/s00371-019-01628-3
30. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (CLAHE) for real time image enhancement. J. VLSI Signal Process. Syst. Signal Image Video Technol. **38**(1), 35–44 (2004). https://doi.org/10.1023/B:VLSI.0000028532.53893.82
31. Ma, J., Fan, X., Yang, S.X., Zhang, X., Zhu, X.: Contrast limited adaptive histogram equalization-based fusion in YIQ and HSI color spaces for underwater image enhancement. Int. J. Pattern Recognit Artif Intell. **32**(07), 1854018 (2018). https://doi.org/10.1142/S0218001418540186
32. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004). https://doi.org/10.1023/B:VISI.0000013087.49260
33. Nezhadarya, E., Ward, R.K., & Wang, Z.J.: Wavelet-based gradient transform and its applications. In: IEEE 14th International Workshop on Multimedia Signal Processing (MMSP) (2012). https://doi.org/10.1109/mmsp.2012.6343424
34. Datta, A., Ghosh, S., Ghosh, A.: PCA, Kernel PCA and dimensionality reduction in hyperspectral images. In: Advances in Principal Component Analysis, pp. 19–46 (2017) https://doi.org/10.1007/978-981-10-6704-4_2
35. Sevakula, R.K., Verma, N.K.: Compounding general purpose membership functions for fuzzy support vector machine under noisy environment. IEEE Trans. Fuzzy Syst. (2017). https://doi.org/10.1109/tfuzz.2017.2722421
36. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp 200–205 (1998). https://doi.org/10.1109/afgr.1998.670949
37. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z.: The extended cohn–kanadedataset (CK +): a complete dataset for action unit and emotion-specified expression. In: Proceedings of the Third International Workshop on CVPR for Human Communicative Behaviour Analysis (CVPR4HB 2010) (2010). https://doi.org/10.1109/cvprw.2010.5543262
38. Wallhoff, F., Schuller, B., Hawellek, M., Rigoll, G.: Efficient Recognition of authentic dynamic facial expressions on the feedtum database. In: IEEE ICME, IEEE Computer Society, pp. 493–496 (2006). https://doi.org/10.1109/icme.2006.262433
39. Goh, K.M., Ng, C.H., Lim, L.L., Sheikh, U.U.: Micro-expression recognition: an updated review of current trends, challenges and solutions. Vis. Comput. **36**(3), 445–468 (2020). https://doi.org/10.1007/s00371-018-1607-6




**R Jeen Retna Kumar** is working as an Assistant Professor in Bethlahem Institute of Engineering, Karungal, Tamil Nadu. He obtained his M.E. degree from MIT Anna University, Chennai. His area of interest includes image processing, pattern recognition and machine learning.

**M. Sundaram** is currently working as a Professor in VSB College of Engineering, Karur, TamilNadu. He completed his Ph.d. degree from Anna University, Chennai, and he is a recognized supervisor for research under Anna University. He has published his research works in many national and international conferences and journals. His research areas includes thermal image processing for detection of breast abnormalities, quality analysis of edible items using Microwave energy , Nail image and Tongue image analysis for medical diagnosis, Pattern recognition and Machine Learning.

**N. Arumugam** is working as an Associate Professor in National Engineering College, Kovilpatti, Tamil Nadu. He has 32 years of teaching Experience. His area of interest includes image processing, VLSI and Sensor Networks.