Soft Computing (2021) 25:5483–5501
https://doi.org/10.1007/s00500-020-05550-y

METHODOLOGIES AND APPLICATION

Face feature extraction for emotion recognition using statistical parameters from subband selective multilevel stationary biorthogonal wavelet transform

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

Published online: 9 January 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

Facial expression recognition is an extensive aspect in the field of pattern recognition and affective computing. Recognizing emotions by facial expression is an imperative action to design control-oriented and human computer interactive applications. Facial expression recognition is probable by the motion of facial muscles resulting in the appearance variation of face features. Accurate feature extraction is one of the extreme challenges that should be scrutinized for an admirable facial expression recognition system. One of the extensive key techniques used for feature extraction mechanism in facial expression recognition is wavelet transform. The features extracted from the wavelet transform incorporate both spatial and spectral domain information which is best adequate for identifying human emotions through facial expressions. In this paper, the statistical parameters from the proposed subband selective multilevel stationary biorthogonal wavelet transform are estimated and are used as features for effective recognition of emotion. The potency of the feature extraction algorithm is boosted by calculating the mean and maximum local energy wavelet subband of stationary biorthogonal wavelet transform. SVM classifier is used for classification of emotion using the preferred chosen features. Protracted experiments with well-known database for facial expression such as JAFEE database, CK + database, FEED database, SFEW database and RAF database demonstrates a better promising results in emotion classification.

Keywords Facial emotion recognition · Wavelet transform · Stationary biorthogonal wavelet transform · Local energy wavelet · Support vector machine

1 Introduction

Emotions of a human can be inferred by facial expression, attitude of body, body language, gesture (Gavrilescu 2015) and speech (Li et al. 2007). Recognition of human emotion is more convenient in evaluation of stress, smart learning (Cambria 2016), mood prediction, patient monitoring system, sentiment analysis, telemedicine, marketing, etc. The facial expression is one among the most natural ways by which humans express their emotional state to observers. The messages or intentions of a human face are revealed by the motion or positions of the muscles in the skin of face. Among all the ways exercised in routine, facial expressions may be the most suited nonverbal way for humans to communicate with each other, and it was first suggested by Darwin (1872). Happy, sad, anger, surprise, fear and disgust were the six basic expression suggested by Ekman and Friesen (1971). The emotion recognition of a human by the consideration of expression in face has achieved more intentness in various areas of research including pattern recognition, affect computing, computer vision and human computer interaction. Human face is liable with different facial expressions which are well distinguished by the variant face features characterized by the motion of facial muscles. For an efficacious recognition of these facial expressions, effectual feature extraction is a more challenging task that has to be considered in an expression recognition system. This paper explicates an emotion

✉ R. Jeen Retna Kumar
jejinrsrch@gmail.com

[1] Bethlahem Institute of Engineering, Karungal 629157, India

[2] VSB Engineering College, Karur 639111, India

[3] National Engineering College, Kovilpatti 628 503, India

recognition system to determine one of the seven emotions from a facial image.

The overall aspect of this paper is compiled as follows

1. Preprocessing of input image is done using contrast limited adaptive histogram equalization method, and the face is detected using viola jones algorithm.
2. The feature extraction of the face expression image is done by using the proposed subband selective multilevel stationary biorthogonal wavelet transform (SM-SBWT) method. The mean and maximum local energy wavelet subband is estimated to increase the effectiveness of the proposed feature extraction method.
3. The dimension reduction of extracted features is made by using block DCT and principal component analysis method.
4. The classification of the emotions is done by the support vector machine classifier.

Due to the enormous importance of facial expression recognition, numerous methods have been developed. The face feature extraction method is broadly categorized into geometric based approach and appearance-based approach. The geometric-based approach is executed by utilizing the relationship in the face components in concurrence with shape, distance and location. Meanwhile the appearance of face features is taken into account in the appearance-based approach. An automatic recognition of facial expressions from face images utilizing the features acquired by discrete wavelet transform (DWT) is presented by S B Kazmi et al. (2012). A bank of seven support vector machines are used for classification in which each support vector is trained to identify a particular facial expression, such that the identified expression is eminent sensitive to that expression. A new approach in facial expression recognition is presented by Meena et al. (2019) based on the graph signal processing and kNN classifier. In this approach, the high dimensional data of HOG features are downsized into a relatively low dimension data using graph signal processing, and the classification was done by nearest neighbor algorithm. Facial expression recognition by extracting the Pyramid of Local Binary pattern features from the salient face region alone is proposed by (Khan et al. 2013). Here reduced features extracted are subjected to expression recognition by five different classifiers. Amani Alfakih et al. (2020) propose a FER system by a multi-view deep convolution network and learned using cooperative learning. A hybrid statistical feature extractor which computes local mean binary pattern from 1-level Haar approximation coefficients is proposed by Goyani and Patel (2017). By this approach, the approximation subband is divided into small region, and from each region LHMBP histogram is computed to derive texture and shape of face. Facial expression recognition using local directional ternary pattern is proposed by Byungyong Ryu et al. (2017). The emotion-related features are encoded efficiently with the directional information and ternary patterns.

Geometric features extracted from facial image are also used for facial expression recognition. Ghimire and Lee (2013) present a recognition method using geometric features extracted from facial image based on facial landmarks. A multiclass adaboost classifier with dynamic time warping and SVM on the boosted features is proposed. A new region attention network which extracts the face region for pose and occlusion variant is proposed by Kai Wang et al. (2020). The face regions extracted are aggregated by region attention network which in turn runs the convolution neural network. Ye Tian et al. (2019) propose a FER system by designing the secondary Information-aware facial expression network which analyzes the inherent components. Facial expression recognition system with SQI filter for face detection and a simultaneous implementation of angular radial transform, discrete cosine transform and Gabor filter for feature extraction is described by Tsai and Chang (2017). A metric learning method for facial expression recognition is proposed by Sadeghi and Raie (2019). For the histogram-based features extracted, Chi-square distance is estimated for classification using kNN classifier. Sub-regions are derived from face image and accurate histogram-based features are extracted. An optimum feature set for facial emotion recognition is procured by using multilevel Haar wavelet-based approach proposed by (Goyani and Patel 2017). The most geometric components are segmented, divided into regions prior to the extraction of local Haar features.

A new method based on local binary patterns (LBP) and local Fisher discriminant analysis (LFDA) for facial expression recognition is presented by Zhang et al. (2012). The high-dimensional LBP features extracted from the original facial expression images are converted into a low-dimensional discriminative embedded data using LFDA. The features extracted using higher-order spectra is presented by (Ali et al. 2015). 1D facial signal obtained from Radon transformation is utilized to extract the HOS features. The high-dimensional features are reduced into lower-dimensional features using principal component analysis (PCA). The reduced features then are fed to SVM classifier for further classification. The facial feature extraction based on integrating radial basis function kernel and multidimensional scaling analysis is given by Shaowei Wang et al. (2013). A genetic algorithm called bee royalty offspring algorithm (BROA) proposed by Jamshidnezhad and Nordin (2013) improves the training process in classification. By estimating and tuning the fuzzy membership function, the fuzzy knowledge base is improved in this method.

The presence of weak and distorted edges due to noise is overcome by neighborhood-aware edge directional pattern descriptor proposed by Iqbal et al. (2018). A wider neighborhood is explored by inspecting the center pixel as well as its neighboring pixel. The consistency in local region is attained by generating pattern code by introducing a template orientation of neighboring pixel. A facial expression recognition using an edge-based descriptor local prominent directional pattern was proposed by F Makhmudkhujaev et al. (2019). A pattern code is encoded using the statistical information of neighborhood pixel by this method. A new feature descriptor namely local directional maximum edge pattern descriptor is proposed by Uma Maheswari et al. (2020). More features for better recognition are evoked by calculating the gradient in four directions. Facial expression recognition using fusion features of multi-scale block local binary pattern uniform histogram and HOG is proposed by Yan et al. (2019). The holistic structure features are obtained by filtering the facial image using MB-LBPUH operator and normalizing the filtered image. The fusion features obtained resemble local appearance and shape–texture along with the global structure patterns. Fusion of feature set using edge-enhanced bidimensional empirical mode decomposition is proposed by Arghya Bhattacharya et al. (2018). The appropriate fusion feature subset for classification is selected by a recursive feature elimination based algorithm. The dimensionality reduction is performed by principal component analysis and linear discriminant analysis. The features are trained for classification by multi-class SVM, ELM with RBF kernel and kNN classifier independently.

Xu et al. (2020) proposed weakly supervised facial expression recognition. A double active layer-based CNN and a two-stage transfer learning method is adopted in this method to transfer information to the recognition system in order to exploit the inadequate training data in deep convolutional neural network. Xiao Sun et al. (2020) proposed a robust convolutional neural network that generates the features in the region of interests by introducing an attention mechanism. The ROIs-related convolution calculation is performed by adopting the attention mechanism in the first layer of neural network. A multi-task global–local network for facial expression recognition is designed by Yu et al. (2020). The local regions are constructed using part-based module, and the global appearance features are extracted using global face module. Yanling Gan et al. (2020) propose a multiple attention network to extract the discriminative features from the face region for facial expression recognition. The region-aware subnet masks are used to extract the expression-related critical regions, and the discriminative features are learned using expression recognition subnet mask with multiple attention blocks. The fusion of HOG and CNN features for facial expression recognition is presented by Pan (2020). The useful temporal and spatial representations of the facial image are used to obtain the CNN and HOG features. The discriminative features and handcraft features of shape and appearance are combined by feature fusion for facial expression recognition which was given by Fan and Tjahjadi (2019). Kuan Li et al. (2020) presented an approach which makes CNN simple by increasing the size of the database with random rotation and horizontal flipping of face image. The extracted features by this method are best suited for classification. The accuracy is improved by the creation of simple structure with the absence of hidden fully connected layer in CNN.

Several methods are developed on facial expression recognition, but getting stable features for classification is still a challenging task. The need for vast dataset for training in deep learning methods looks more complex and makes the system less efficient. Due to the disparity in face traits, it is arduous to determine accurate and reliable position of facial component. The feebleness in the geometric method of feature extraction under distinct situations leads to poor results. The effect of noise and elegant local distortions makes the local feature extraction method less adept in recognition. The problems mentioned above are tackled by proposing a novel feature extraction method subband selective multilevel stationary wavelet transform. Due to good localization in both spectral and spatial domains, SWT is widely used in facial expression recognition and attained promising results. In most of the works of Yu-Dong Zhang et al. (2016), Huma Qayyum et al. (2017), Shui-Hua Wang et al. (2017), the statistical values are extracted directly from the subbands which are inconsistent, incompatible and less discriminating. In this proposed work, we demonstrate a method which separates the subbands obtained from different levels of stationary biorthogonal wavelet transform in consideration of expression-related significant features. The subband retrieved from the multilevel SBWT is selected in significance with the entropy values of subband, and a sequence of subband is formed by subband combination. A significant accuracy in recognition is attained by calculating a more consistent and compatible statistical parameters from this subband.

The flow for rest of this paper is organized as follows. Section 2 presents the preprocessing and face detection method of the facial images. Section 3 provides the methodology for feature extraction using the statistical parameters obtained from the proposed subband selective multilevel stationary biorthogonal wavelet transform (SM-SBWT). Section 4 describes the classification method using SVM classifier. Section 5 illustrates the experimental setup and the results with discussions. Finally, Section 6 concludes the paper.

2 Preprocessing and face detection

2.1 Preprocessing

Preprocessing is commonly a necessitous process which should be ensued to every face database. The feature extraction process eventually is more fast, accurate and efficient due to preprocessing performed on images. Principally, the image preprocessing step encompasses the various operations such as brightness correction, scaling, contrast adjustment, gray level transformation and other enhancement operations. Image scaling resizes the image to be scaled to the same size as in the face gallery so that the correlation between the images is maximized. To enrich the interpretability or perception for human viewers, the image enhancement operations are performed. The contrast adjustment in images with unequal intensity values can be improved by using histogram equalization process. Preferably an adaptive method of histogram equalization, which redistributes the lower intensity values of the image by computing several histograms, is incorporated in this work. In this contrast-limited adaptive histogram equalization (CLAHE), each computed histogram denotes a distinct section of the image (Reza 2004). The contrast enhancement to each distinct section is done in such a manner that the histogram obtained for the output region approximately matches the histogram specified by the distribution parameter. A contrast-limited adaptive histogram equalization limits over-amplifying noise in relatively analogous regions of an image. Consequently, the histogram equalization is applied to each distinct section and a redistributed histogram is developed with each pixel intensity limited to a selected maximum (Jinxiang et al. 2018).

2.2 Face detection

Face detection is a requisite approach in facial expression recognition used to segregate the required face region from unwanted backgrounds. In the proposed work, the most widely used Viola–Jones algorithm is used for face detection (Viola and Jones 2004). Haar basis feature filters are used in this algorithm which comprises the Haar feature selection and generation of the integral image. The detection window created by integral image initiates the face detection. The detection window is drifted across the image with implementation to a set of face recognition filter in it. Each face recognition filter looks at the rectangular subset of the detection window. A face is detected only if all the classifiers yield a positive answer unless next classifier is applied.

3 Proposed feature extraction method-SM-SBWT

Feature extraction is the process of extracting the face features from a face image. This is the most challenging and imperative stage in facial emotion recognition. The efficiency of FER is very much emulated by performance of feature extraction technique. The feature extraction is extremely important to the whole classification process. For facial expression recognition, the feature extraction step extracts remarkably a large number of features. In correspondence with certain optimal criteria, a smaller subset of eminent features needs to be selected. The prevalent feature extraction approaches in facial expression recognition are Wavelet transform, local descriptors, geometric descriptors, component analysis, etc. Subband selective multilevel stationary biorthogonal wavelet transform (SM-SBWT) is the proposed feature extraction method adopted in this work and is shown in Fig. 1. The local energy of the coefficient in wavelet subband is estimated to enhance the edge details and local texture information. Subsequently limited number of subbands from multilevel stationary biorthogonal wavelet transform is selected based on the entropy values of subbands. Finally, the statistical parameters are measured from the selected subbands of the input facial image which is considered as facial features.

In recent times, for disparate applications in image processing such as image denoising, image compression, face recognition, image super-resolution, etc., the wavelets have been used in feature extraction. Wavelet transform delivers the frequency domain and time domain information of a signal. The wavelet transform decomposes the images into different frequency components with different frequency ranges known as subbands. The approximations of original image are obtained using low pass filtering and the detailed features such as edges are obtained using high pass filtering (Tim Edwards 1992). The row-wise decomposition and column-wise decomposition using 1-D DWT is performed to obtain 2-D wavelet decomposition. The row-wise decomposition is done by applying 1-D DWT along the rows of the input image, and column-wise decomposition is achieved by applying 1-D DWT along the columns. As a result, the input image is decomposed into four subbands.

Images are referred to as low–low (LL1), low–high (LH1), high–low (HL1), and high–high (HH1). The LL band of previous level is further decomposed by DWT in case of multi-resolution analysis.

DWT can be mathematically expressed in Eq. (1) as

$$\mathrm{DWT}_{x(n)} = \begin{cases} D_{j,k} = \sum x(n) h_j^*(m - 2^j k) \\ A_{j,k} = \sum x(n) l_j^*(m - 2^j k) \end{cases} \quad (1)$$

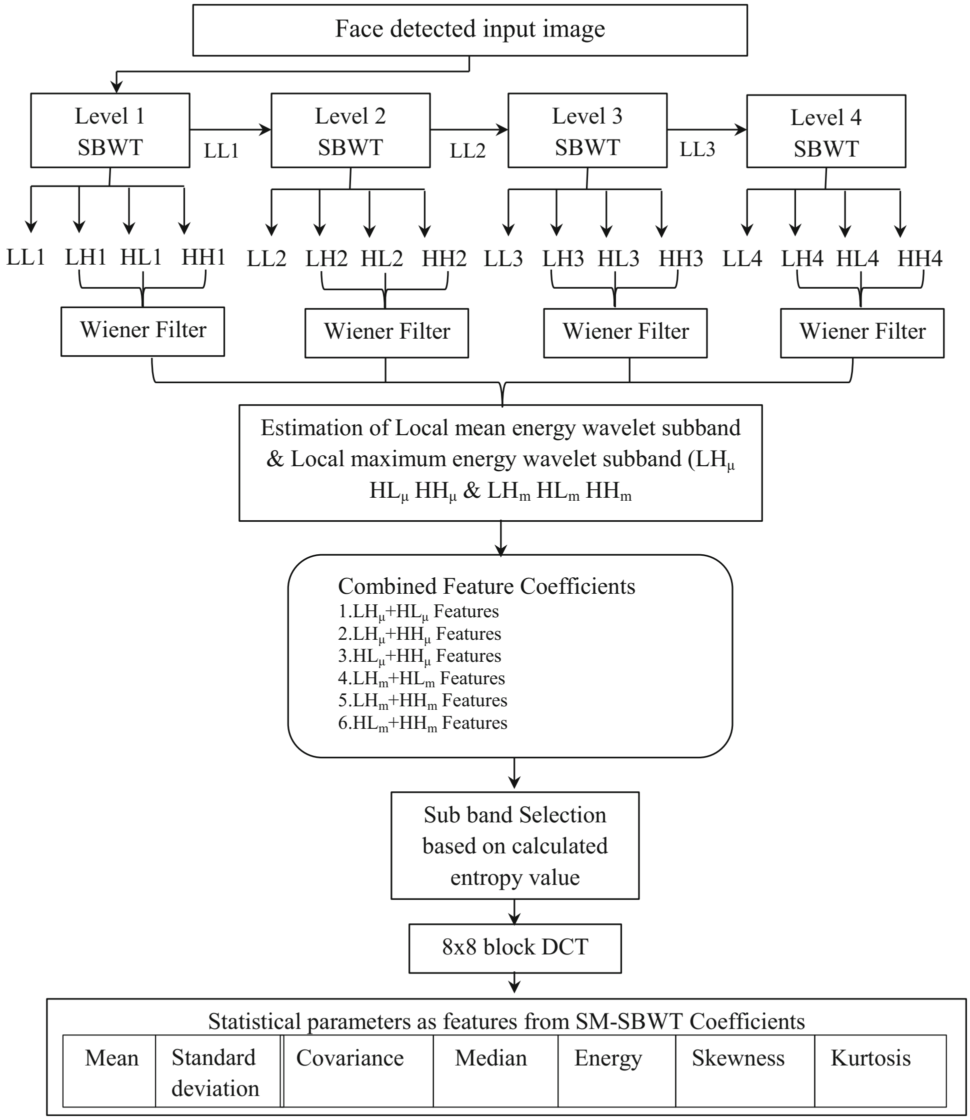

Face detected input image
Level 1 SBWT
Level 2 SBWT
Level 3 SBWT
Level 4 SBWT
LL1
LL2
LL3
LL1 LH1 HL1 HH1
LL2 LH2 HL2 HH2
LL3 LH3 HL3 HH3
LL4 LH4 HL4 HH4
Wiener Filter
Wiener Filter
Wiener Filter
Wiener Filter
Estimation of Local mean energy wavelet subband & Local maximum energy wavelet subband (LH_μ HL_μ HH_μ & LH_m HL_m HH_m
Combined Feature Coefficients
1.LH_μ+HL_μ Features
2.LH_μ+HH_μ Features
3.HL_μ+HH_μ Features
4.LH_m+HL_m Features
5.LH_m+HH_m Features
6.HL_m+HH_m Features
Sub band Selection based on calculated entropy value
8x8 block DCT
Statistical parameters as features from SM-SBWT Coefficients
Mean
Standard deviation
Covariance
Median
Energy
Skewness
Kurtosis

Fig. 1 Proposed feature extraction method—SM-SBWT

where $D_{j,k}$ is the detail components in the signal x(n) and $A_{j,k}$ is the approximation components in the signal x(n). h(m) and l(m) are the high-pass and low-pass filter coefficients.

3.1 Stationary wavelet transform

Because of the decimation operation deployed in DWT, the shifted version of a signal and the shift in DWT of a signal is not equal. Hence, DWT is a spatial variant transform. This problem of shift invariance is solved by using stationary wavelet transform (Nason and Silverman 1995). The decimation operation performed in DWT is ignored in SWT. The input signal is convolved with low l[m] and high h[m] pass filter in SWT which is more identical to DWT. The coefficients' size obtained after SWT is same as that of the input signal size due to the lack of decimation in SWT. In SWT, the input image is decomposed into four subbands as shown in Fig. 2. The four subbands obtained are LL1, LH1, HL1 and HH1 which represents approximation, horizontal, vertical and diagonal information of the input image. The LL1 coefficient contains VLF and LF component. The LH1 coefficient and the HL1 coefficient contain HF components and some amount of VHF components. The HH1 coefficient contains some HF components and mostly VHF components.

For an input image I of size MxN, the SWT at jth level is given in Eqs. 2–5.

$$\mathrm{LL}_{j+1}(a,b) = \sum\sum l_x^j l_y^j \mathrm{LL}_j(a+x, b+y) \qquad (2)$$

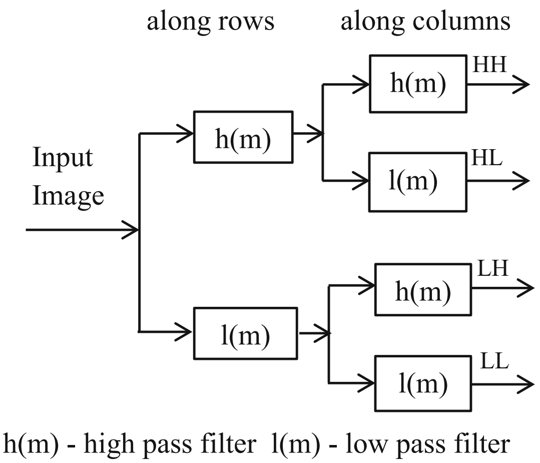


Fig. 2 Single-level SWT decomposition of an image into four subbands

$$\mathrm{LH}_{j+1}(a,b) = \sum\sum h_x^j l_y^j \mathrm{LL}_j(a+x, b+y) \tag{3}$$

$$\mathrm{HL}_{j+1}(a,b) = \sum\sum l_x^j h_y^j \mathrm{LL}_j(a+x, b+y) \tag{4}$$

$$\mathrm{HH}_{j+1}(a,b) = \sum\sum h_x^j h_y^j \mathrm{LL}_j(a+x, b+y) \tag{5}$$

Here a = 1,2,….*M*, *b* = 1,2,…..*N*, *h* & *l* are the low-pass and high-pass filters. LL, LH, HL and HH are SWT coefficients.

So as to maintain symmetry in SWT, and for preserving more energy in subbands, a biorthogonal wavelet transform is used in SWT (Zhang et al. 2016). In biorthogonal wavelet transform, the functions used for calculations are very simple and are easier to build. Based on the merits of the above method, the stationary biorthogonal wavelet transform (SBWT) is used in this paper. At first, a one level decomposition for the input image is made by using SBWT to get LL_1, LH_1, HL_1 and HH_1 subbands. The approximation component LL_1 obtained in the first-level SBWT is again decomposed using second-level SBWT to obtain the components LL_2, LH_2, HL_2 and HH_2. Here, LL_2 is the approximation coefficient and LH_2, HL_2, HH_2 represents the horizontal, vertical and detailed coefficients. Similarly, the components from four levels of SBWT are calculated and are subjected to further processing. Figure 3 shows the schematic representation of four levels of SWT decomposition, and Fig. 4 shows the output subbands obtained from four levels of SWBT.

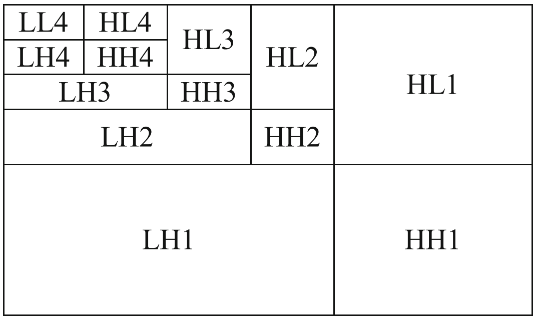


Fig. 3 SWT 4th level of decomposition

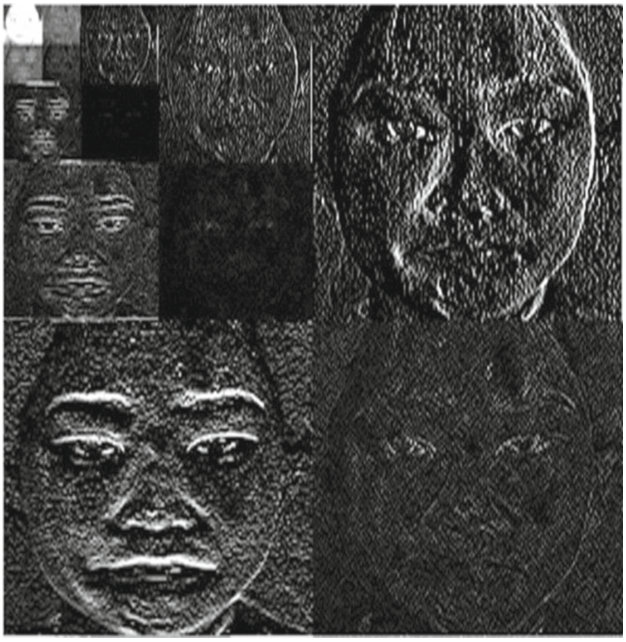


Fig. 4 Level 4 SWBT of JAFEE sample face

3.2 Subband selective multilevel stationary biorthogonal wavelet transform

In the output subbands from four levels of SBWT, the LL subband gives the overall significant information of the image which is the approximation coefficient. The changes that occur in the image orientation are reflected in LH, HL and HH which are the detailed coefficients. In facial expression identification, the detailed coefficients plays vital role rather than approximation coefficients (Wang et al. 2017). Hence, the detailed coefficients LH, HL and HH coefficients are considered here.

The LH, HL and HH components obtained from four levels of SBWT are filtered using an adaptive wiener filter. The wiener filter is used to reduce the high-frequency component effects in the image. The VHF components spread over a wider area with the use of wiener filter. After wiener filtering operation, the VHF components tend to become small since it has very low amplitudes (Sanket et al. 2014).

The output of the wiener filter is denoted as $x[n]$ which is given by the expression

$$x(n) = \sum_{i=0}^{N} a_i w(n-i) \qquad (6)$$

where w(n) is the signal and N is the order of filter. a_0,a_1,a_2,….. are the coefficients.

The edge and the texture information elicited in the wavelet subband are enhanced by calculating the local energy in the wavelet subband coefficients. A linear transform operation, as given in Eq. 7, is to convert negative coefficients to corresponding positive coefficients.

$$S_i = \left|\min\left\{S_i,\ i \int \{1, 2, \ldots, a\,x\,b\}\right\}\right| + S_i + 1 \qquad (7)$$

where S_i denotes the subband wavelet coefficient, a and b give the number of rows and columns in the wavelet subband. In order to avoid zero values, one is added in equation. The energy of a pixel "i" in a subband is given by

$$E_i(a,b) = S_i^2(a,b) \qquad (8)$$

Subsequently from each wavelet subband, the mean local energy wavelet subband and maximum local energy wavelet subband are estimated. The mean local energy wavelet subband is estimated by calculating the mean energy value of a center pixel in a *w* x *w* window and is given in Eq. 9.

$$\mathrm{LE}_{\mu}(a,b) = \frac{1}{n}\sum_{a=1}^{w}\sum_{b=1}^{w} E(a,b) \qquad (9)$$

where E(a,b) is the energy coefficient of subband. n denotes the total number of energy coefficient in the wxw window (n = wxw). In the proposed method, we have considered the window size be 3x3. Equivalently the maximum local energy wavelet subband is estimated by calculating the maximum energy value in a w x w window and is given in Eq. 10

$$LE_m(a,b) = \max\left\{E(a,b)\,|\,a, b \int \{1, 2, \ldots w\}\right\} \qquad (10)$$

Accordingly for every detailed subband (LH, HL & HH), mean local energy wavelet subband (μ) and maximum local energy wavelet (m) subband is derived. (i.e., LH_{μ} and LH_{m} for LH subband, HL_{μ} and HL_{m} for HL subband and HH_{μ} and HH_{m} for HH subband). Consequently, more descriptive features pertaining to horizontal, vertical and diagonal directions are obtained by the local energy wavelet subband combination. Here a subband combination of local energy wavelet subband is proposed with consideration to the weight value calculated for each detailed subband. Therefore, for each level of wavelet transform six subbands obtained after subband combination.

The subband combination is made using the following equation

$$S_{kj}(a,b) = \begin{cases} u_j LH\mu_j(a,b) + v_j HL\mu_j(a,b), for\ k = 1 \\ u_j LH\mu_j(a,b) + w_j HH\mu_j(a,b), for\ k = 2 \\ v_j HL\mu_j(a,b) + w_j HH\mu_j(a,b), for\ k = 3 \\ u_j LHm_j(a,b) + v_j HLm_j(a,b), for\ k = 4 \\ u_j LHm_j(a,b) + w_j HHm_j(a,b), for\ k = 5 \\ v_j HLm_j(a,b) + w_j HHm_j(a,b), for\ k = 6 \end{cases} \qquad (11)$$

where *j* = 1,2,3,4 results the number of levels of SBWT and k denotes the combination formed using the detailed coefficient in each level. u_j, v_j, and w_j are the *j*th level weight value calculated based on the association measure of detailed subbands (LH, HL and HH) subbands with the approximation subband LL.

$$u_j = \sum\left(\frac{1 + \mathrm{cor}\left(\mathrm{LL}_j, \mathrm{LH}_j\right)}{2}\right)^{\beta} \qquad (12)$$

$$v_j = \sum\left(\frac{1 + \mathrm{cor}\left(\mathrm{LL}_j, \mathrm{HL}_j\right)}{2}\right)^{\beta} \qquad (13)$$

$$w_j = \sum\left(\frac{1 + \mathrm{cor}\left(\mathrm{LL}_j, \mathrm{HH}_j\right)}{2}\right)^{\beta} \qquad (14)$$

where *β* is a soft threshold constant. In this proposed work, β is considered as one. The weight value calculated based on the association measure of subbands endures robust, stable and more reliable support for the combinational measures.

3.3 Subband selection in SM-SBWT

The subband combination in each level of SBWT estimates six subbands, and a total of twenty-four subbands are retrieved from four levels of SBWT. The subbands needed for the feature extraction are selected based on the calculated entropies of each subband. The subbands with maximum entropies are selected for feature extraction. The entropy is calculated using the formula

$$H = -\sum_{k} P_k \log_2 P_k \tag{15}$$

The increase in dimension of feature vector leads to degradation in the performance of the classifier. In order to adhere this, the dimension of feature vector is reduced with the help of a feature selection scheme in which the subbands which are appropriate alone are selected, removing the remaining subbands. Thereby, the subbands are selected based on entropy values. Entropy is a statistical measure which precisely describes the randomness of image, analyzes the local characteristics of image and symbolizes the texture of image. By this reason, entropy is chosen here for subband selection. Out of the 24 subbands estimated for each image, only 15 subbands stationed with maximum entropies are chosen for the estimation of statistical parameters. The selected subbands obtained by specific combination with greater entropy values are the subband selective multilevel SBWT subbands. An 8x8 block DCT is then applied to the above selected subband coefficients to convert the spatial domain into frequency domain. The DC component of the subband lies in the first coefficient of each 8x8 block. Henceforth, the DC component is retained and a reduced feature vector length is utilized by considering the DC coefficients alone. Ultimately, this combination of SM-SBWT and DCT results in improved classification.

Finally, from the above SM-SBWT subband coefficients, the statistical parameters like mean, standard deviation, covariance, median, energy, skewness and kurtosis are estimated. The statistical measures are widely used in various social and scientific researches (Crouse et al. 1998), and hence, the statistical parameters of the subbands are considered as features for this proposed work.

Algorithm of the proposed feature extraction method

Input : Training Image, I_{train}
Output : Statistical Features F_{img}

1) Preprocess the input image using CLAHE algorithm.
2) Detect the face region of input image by Viola-Jones algorithm.
3) For each face detected image, calculate four levels of SWBT approximation and detailed coefficients or subbands. (LL_{x4}, LH_{x4}, HL_{x4} & HH_{x4})
4) Let $LH_j(a,b)$, $HL_j(a,b)$, $HH_j(a,b)$ (j=1,2,3,4) be the detailed subbands obtained by applying four level of SWBT. a, b is the row and column size of the subband coefficients respectively.
5) Apply wiener filter to all subbands by equation 6 to reduce the high frequency components effect.
6) Estimate the local energy in each coefficient of detailed subbands by equation 8.
7) For each subband, calculate the mean local energy wavelet subband and maximum local energy wavelet subband using equation 9 & 10.
8) Evaluate pixel level fusion to the local energy wavelet subbands to obtain a combination set of subbands $S_{kj}(a,b)$ (k=1,2…..,6 & j=1,2,3,4) by equations 11-14.
9) Calculate the entropy value of all subbands using equation 15. The selections of subbands are made by the consideration of maximal entropy values. $S_e(a,b)$ (e=1,2….,15)
10) Apply 8x8 block dct to all selected subbands and the dc coefficient obtained in each block alone is retained so that the size of each subband is changed to $S_r(m,n)$. Here mxn = d<<axb, d is the number of dc coefficients obtained in each subband.
11) The statistical parameters such as mean, standard deviation, covariance, median, energy, skewness and kurtosis are estimated from each subband and is switched to one dimensional vector with size of 1x105 (seven parameters from 15 subbands 7x15=105) which forms feature vector F_{img}.

As a result of feature extraction by the proposed method, 15 subbands are selected from four levels of SWBT decomposition. Seven statistical parameters are selected from each selected subbands; hence, 105 (15 × 7) feature coefficients are selected from each image. The dimensions of retrieved features are adequately reduced by manipulating principal component analysis (PCA). PCA constructs a low-dimensional representation of the data at which as much as variance in the data is exemplified. The utmost amount of variance in the data is chosen for a linear basis of dimensionality reduction.

4 Classification

4.1 SVM classifier

Support vector machine is a maximal bound hyperplane machine learning algorithm for classification purpose. The data are analyzed using statistical learning theory which assures high generalization performance. SVM is a supervised learning algorithm and exhibits good classification accuracy even with fewer amounts of training data. It is widely used as dynamic and associative methods for recognition related tasks. If we have two classes of +1 and -1, and the data set is

$$U = \left\{(x_i, y_i)^P\right\},\ i = 1, \ldots n \tag{16}$$

where n is the number of samples, p gives the length of input feature coefficients, x_{i} is the feature coefficients of ith sample, and y_{i} is its corresponding target label.

The SVM algorithm quest an optimum hyperplane with maximum margin. The hyperplane classifies one class from other class. For each feature vector x_{i}, either

$$w^T \cdot x_i + b \geq 1 \text{ for } x_i \text{ having the class } 1 \tag{17}$$

$$w^T \cdot x_i + b \leq -1 \text{ for } x_i \text{ having the class } -1 \tag{18}$$

The unique constraint equation is given by

$$y_i\left(w^T \cdot x_i + b\right) \geq 1 \quad \text{for all } 1 \leq i \geq n \tag{19}$$

The following optimization function explicates the cost function to determine the hyperplane

$$\min_w \frac{1}{2} w^T w + C \sum\nolimits_{(i=1)}^{P} \xi(w, xi, yi) \tag{20}$$

A multiclass problem with seven expressions to be recognized are analyzed here. The explanation given by Yu Dong Zhang et al. (2016) denotes that the one-against-one-based method is a competitive and best-suited approach for multiclass SVM. One SVM classifier is constructed for each class in one-against-one-based multiclass SVM method, and hence, seven individual SVMs need to be constructed. The test data are tested at each classifier and determined whether it belongs to a class or not. A calculated score is obtained at each classifier, and the classifier that gives the maximum score is identified as the test class.

The input data which may not be linearly separable is effectively mapped to a higher dimensional space with a technique called the kernel trick. The hyperplane decision boundary between the classes is made different by the different kernels functions such as the linear, polynomial and radial basis function. By considering the different kernel functions and parameters, the performance of the classifier is analyzed. The mathematical formula for the kernel functions is given as follows.

Linear kernel function

$$K\left(z_i, z_j\right) = z_i^T \cdot z_j \tag{21}$$

Polynomial function with degree p

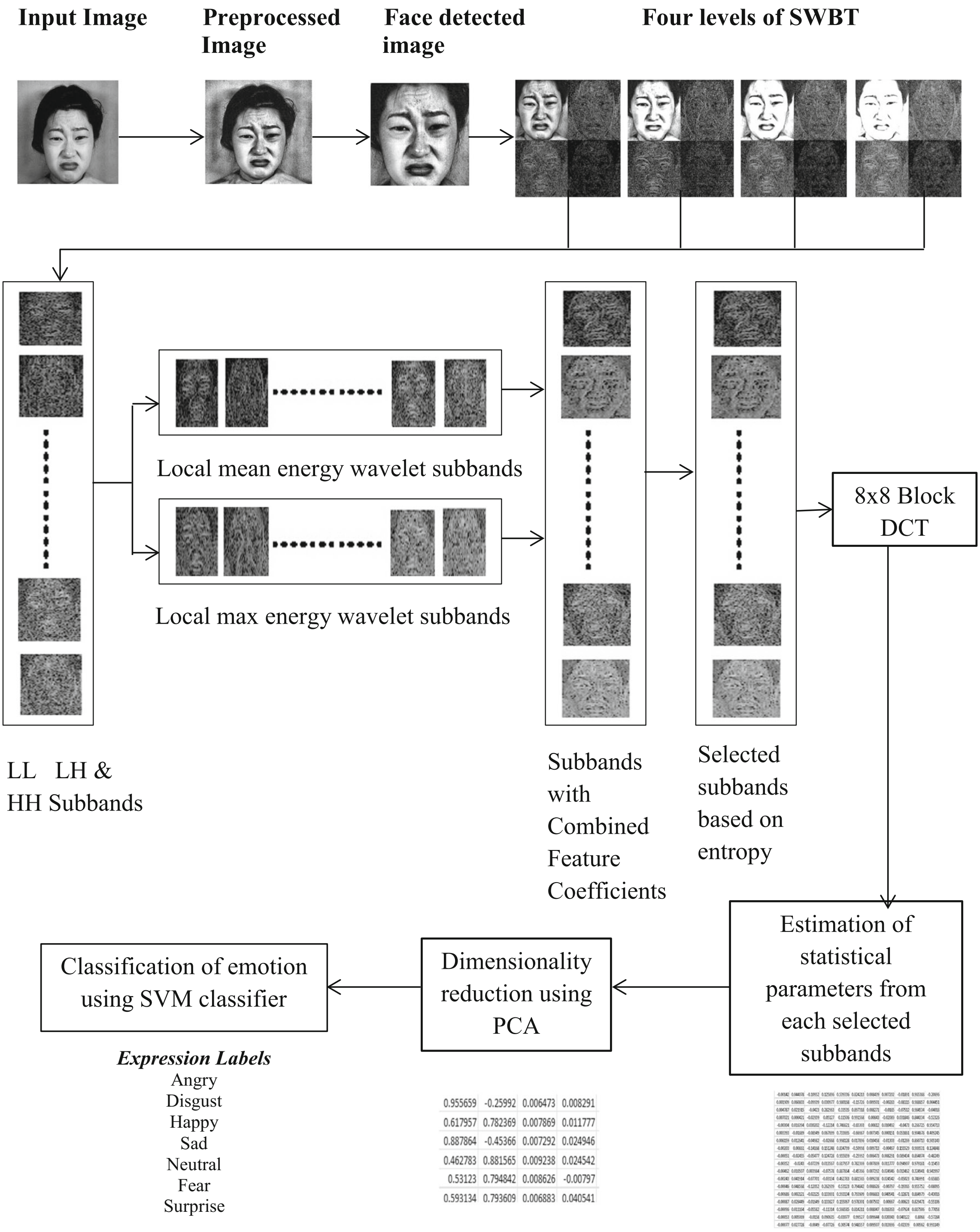
$$K\left(z_i, uz_j\right) = \left(1 + z_i^T \cdot z_j\right)^p \tag{22}$$

Radial basis function using Gaussian

$$K(zi, zj) = \exp\left(-\frac{\left\|z_i - z_j\right\|^2}{2\sigma^2}\right) \tag{23}$$

5 Results and Discussion

The implementation of the proposed work was done using MATLAB software for JAFEE database (Lyons et al. 1998), CK + database (Lucey et al. 2010), FEED database (Frank Wallhoff et al. 2006), SFEW database (Dhall et al. 2011) and RAF database (Li et al. 2017). The recognition rate retrieved using the proposed method is correlated with state-of-the-art facial expression recognition methods. The schematic flow diagram of the proposed method is depicted in Fig. 5. The JAFFE dataset contains 213 gray scale images of 10 different females for seven different expressions. The seven different emotions incorporated are anger, disgust, fear, happy, neutral, sad and surprise. Each JAFEE image has a spatial resolution of 256 × 256. The FEED dataset with facial expressions contains face images of 18 different individuals delivering different emotions. The different emotions include happy, disgust, anger, fear, sad, surprise and neutral. The spatial resolution of FEED dataset image is 320 × 240. The CK + dataset constitute both posed and non-posed facial expression images of 210 adults. Since the posed images are well adequate for this work, the posed images from CK + dataset are considered for this work. The CK + dataset has face images with seven different emotions and each image has a spatial resolution of 640 × 480. The static facial expression in the wild (SFEW) dataset consists of 700 images labeled with


Input Image
Preprocessed Image
Face detected image
Four levels of SWBT
Local mean energy wavelet subbands
Local max energy wavelet subbands
LL LH &
HH Subbands
Subbands with Combined Feature Coefficients
Selected subbands based on entropy
8x8 Block DCT
Estimation of statistical parameters from each selected subbands
Dimensionality reduction using PCA
Classification of emotion using SVM classifier
Expression Labels
Angry
Disgust
Happy
Sad
Neutral
Fear
Surprise

0.955659	-0.25992	0.006473	0.008291
0.617957	0.782369	0.007869	0.011777
0.887864	-0.45366	0.007292	0.024946
0.462783	0.881565	0.009238	0.024542
0.53123	0.794842	0.008626	-0.00797
0.593134	0.793609	0.006883	0.040541

Fig. 5 Schematic flow diagram of the proposed method

six basic expressions and neutral expression. The images in the dataset constitute varied head poses, different focus and distinct resolution of face. The Radboud Faces Database (RaFD) contains facial expression images obtained from 67 persons with eight different emotions. For this work seven emotions are considered from RAF database. The sample images representing the seven expressions from Jaffe, CK + , FEED, SFEW and RAF database are shown in Fig. 6. The face detected images using viola jones algorithm is shown in Fig. 7.

In this work, a 10-fold stratified cross-validation is exercised for performance analysis. The applied stratification divides the total number of images into 10 different folds such that each fold encompasses the same number of

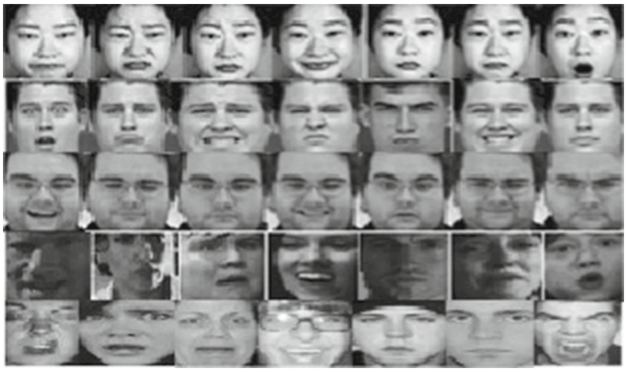


Fig. 6 Sample images representing seven expressions taken from JAFEE, CK +, FEED, SFEW and RAF-DB databases

(a) FEED **(b)** CK+ **(c)** JAFEE **(d)**SFEW**(e)**RAF-DB

Fig. 7 Face-detected images

class labels. Out of 10 folds, eight folds are used for training, one fold is used for validation, and one fold is used for testing in each trial. The training fold is used to fit the model, and the validation fold is used to select the parameter set by evaluating the model. The test fold is used to analyze the performance of the model with the selected parameter set. Figure 8 shows the validation result for an image in a trial. The searching for an optimal parameters on a discrete set of σ^2 = [0.05, 0.1, 0.5, 1, 5] and C = [1, 10, 100, 10^3, 10^4] is done by validation for RBF kernel in SVM. From Fig. 8, the optimal parameters selected are σ^2 = 1 and C = 100 for the particular trial in JAFEE dataset. The same procedure is applied to each validation session of each fold to find out the optimal parameter set.

The sensitivity and overall accuracy for a cost matrix C_{ij} (i = 1,2….7 & j = 1,2…7) is obtained using Eqs. 24 and 25.

Sensitivity of class S is

$$\mathrm{SEN}(S) \; = \frac{C_{SS}}{\sum_j C_{Sj}} \tag{24}$$

Overall accuracy is

$$\mathrm{OA} \; = \frac{\sum_i C_{ii}}{\sum_j \sum_i C_{ij}} \tag{25}$$

The sensitivity results for each expression after the implementation of 10 fold cross-validation to JAFEE dataset is given in Table 1. As per the sensitivity results, it is inferred that the expression that has been identified easily is happy. The anger expression is the second easiest expression to identify followed by neutral, sad, fear, disgust and surprise, respectively. The distinguished features of the happy image implicated by eye corners, lip corners, forehead muscles and the eyebrows made it an easiest expression to identify. Equivalently the distinguished features of anger image involved by the eyebrows, wrinkled nose, narrowed eyes and face jaws make it the second easiest expression to identify. The facial muscle relaxes while enacting the neutral expression. The facial muscles around the ear, muscles around eye lid, muscles around nose and muscles around the mouth are utilized for other expressions. The overall accuracies manifesting the correct recognition rate of all runs in JAFEE dataset are listed in Table 2. An overall accuracy of 97.3% is attained. The selected values of SVM kernel parameters, those that yield maximum accuracy from the validation results, are summarized in Table 3.

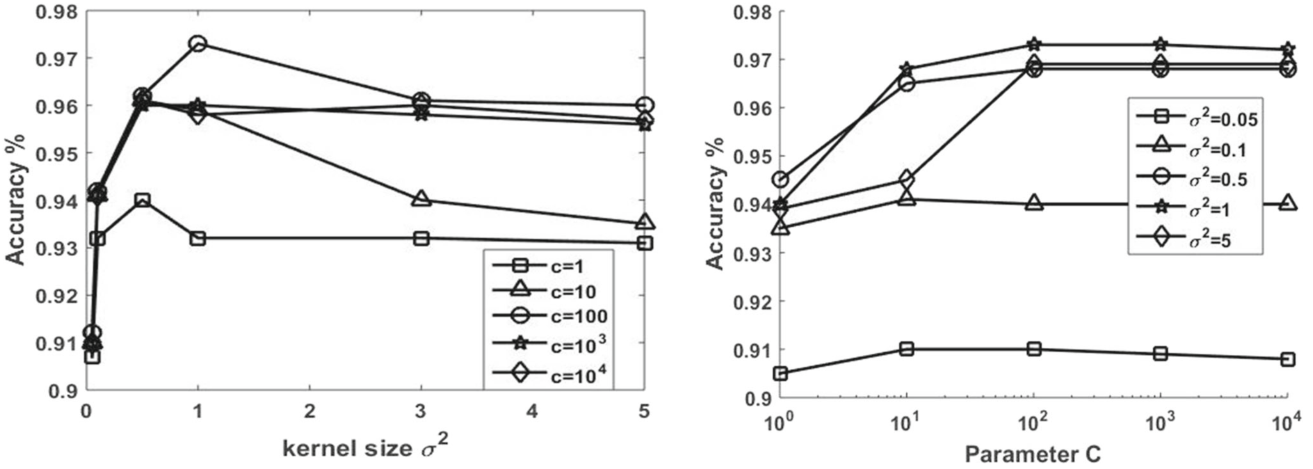


Fig. 8 Tenfold cross-validation results for SVM parameter selection

Table 1 Sensitivity results based on statistical analysis in JAFEE dataset

	Anger	Disgust	Fear	Happy	Neutral	Sadness	Surprise
Run1	99	96	98	99	99	97	94
Run2	98	95	96	100	97	95	97
Run3	99	98	98	100	98	98	95
Run4	97	98	99	99	97	99	98
Run5	98	97	95	100	98	96	97
Run6	98	96	96	98	99	97	96
Run7	97	95	98	100	97	96	94
Run8	100	94	95	98	99	97	96
Run9	100	98	96	100	96	96	95
Run10	98	96	98	99	98	95	97
Average	98.4	96.3	96.9	99.3	97.8	96.6	95.9

Table 2 Overall accuracy analysis in JAFEE dataset

Run	Overall Accuracy in %
1	97.5
2	97.2
3	97.6
4	97.6
5	97.4
6	96.7
7	96.7
8	96.8
9	98.1
10	97.4
Average	97.3

Table 3 Selection of SVM Kernel parameters

Dataset	Linear	RBF	Polynomial
JAFEE	$C = 150$	$C = 100$ $\sigma = 1$	$C = 180$ p = 2
CK+	$C = 100$	$C = 100$ $\sigma = 1$	$C = 150$ p = 2
FEED	$C = 150$	$C = 100$ $\sigma = 0.5$	$C = 200$ p = 2
RAF	$C = 180$	$C = 100$ $\sigma = 0.5$	$C = 150$ p = 2
SFEW	$C = 180$	$C = 100$ $\sigma = 1$	$C = 150$ p = 2

The correct recognition rate (CRR) achieved with different levels from $n = 1$ to $n = 5$ is outlined in Table 4. At level $n = 1$, the subbands used are selected from single level SBWT. Consequently at level $n = 2$, the subbands used are selected from both level 1 and level 2. Likewise the subbands from higher levels are selected in similar manner. From Table 4, it is evident that the maximum performance is procured at level $n = 4$. This is because at lower levels, the subbands are insufficient to generate compatible and consistent statistical parameters. Meanwhile at higher levels, the subbands become redundant which makes decease in performance. Hence, the value of n selected is 4.

The entropy value for the 24 subbands obtained from a facial image is plotted in Fig. 9. Out of the 24 subbands, a maximum of 15 subbands with greater entropy values are selected. The combination of subbands made after calculating mean and maximum energy wavelet subbands and the selection of subbands using entropy values make the performance of the system more effective in terms of efficiency. Figure 10 gives the graph drawn between the number of selected subbands and accuracy and between feature dimensions and accuracy. It is apparently evident from the figure that maximum accuracy is attained when 15 subbands are selected for feature extraction. Obviously, it is precisely noticeable in the graph drawn with dimensions versus accuracy. The classification performance achieved by considering different SVM kernel functions for different database are shown in Table 5. The bold font in Table 5 indicates best performance results. The number of subbands selected after stationary biorthogonal wavelet transform, the feature dimension obtained by the selected subbands and their performance are shown in the table. It is observed that the accuracy value is maximum when 15 subbands are selected. It is also noticed that the polynomial functions produce high accuracy for most of the cases.

Table 6 summarizes the confusion matrix acquired using tenfold stratified cross-validation for the seven emotions on JAFFE dataset with the proposed algorithm. Happy and anger emotions are easily recognized with an accuracy of 99.3% and 98.3%. Sad and fear have the least CRR of 95.8% and 96.8%, respectively. This is because the

Table 4 Classification performance with different levels of SM-SBWT

Run	$N = 1$			$N = 2$			$N = 3$			$N = 4$			$N = 5$		
	JAFEE dataset	CK + dataset	FEED dataset	JAFEE Dataset	CK + dataset	FEED dataset	JAFEE dataset	CK + dataset	FEED dataset	JAFEE dataset	CK + dataset	FEED dataset	JAFEE dataset	CK + dataset	FEED dataset
1	89.1	90.1	88.2	92.8	92.8	89.1	95.7	96.1	94.9	97.5	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	97.2	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	99.7	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	97.6	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	99.6	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	96.7	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	96.7	99.5	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	96.8	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	98.1	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	97.4	99.5	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	97.3	99.2	95.9	94.6	94.9	94.4

fear is often confused and misclassified as anger and sad often misclassified as disgust. This is due to the fact that fear and anger have the same face muscle activities. Meanwhile, sad has only a little different muscle activities than disgust emotions. The confusion matrix for CK + database is presented in Table 7. The happy expression is recognized easily with an accuracy of 100%, and disgust is difficult to recognize with an accuracy of 98.6%. It is observed that surprise and fear is confused easily. This may be expected because of the similar appearance features that can be shared by both expressions. The confusion matrix of FEED database is shown in Table 8. It shows that happy expression is recognized well while surprise is recognized less. Also it is observed that the anger and surprise has been confused easily. This is because the upper and lower eyelids are open for both anger and surprise emotion.

The confusion matrix of RAF database is shown in Table 9. It shows that happy expression is recognized well with an accuracy of 92.6% while disgust is recognized less with an accuracy of 82.7%. Also it is noticed that the disgust and fear have been confused easily. The reason is evident from the aspect that the facial crucial areas for disgust and fear are little, and hence, it creates more confusion during recognition. The confusion matrix of SFEW database is shown in Table 10. It shows that happy and sad expressions are recognized well with an accuracy of 64.6% and 61.8%, respectively. Anger and surprise are recognized less with an accuracy of 51.5% and 52.9%, respectively. Also it is revealed that the anger is confused with disgust and fear, while surprise is confused with fear. The wrinkles of the forehead in anger expression reveal same for disgust and fear expression. The entire eyebrows of face pulled up for both surprise and fear emotion which eventually makes those expressions get confused more.

The classification accuracy achieved with different resolution of face images from datasets is shown in Fig. 11. The classification accuracy reduces with decrease in the resolution of images. Table 11 shows the comparison of SWT + SVM and SM-SWBT + SVM in terms of sensitivity of each expression in JAFEE database. The change in facial muscles plays an essential role in facial expression recognition for different expressions. Motion of facial muscles delivers more visual malformation in expression such as happy and anger, whereas it creates less visual effect on expressions like sad and fear and is evidently depicted in the accuracy reported in Table 11. Thus, happy and anger register high recognition rate while slightly low recognition is obtained for other expressions. It is also observed that in all facial emotion classes our proposed method yield high CRR, and this proves the effectiveness of our proposed method.

When compared with the state-of-the-art methods, our proposed method gives better results. For all the five

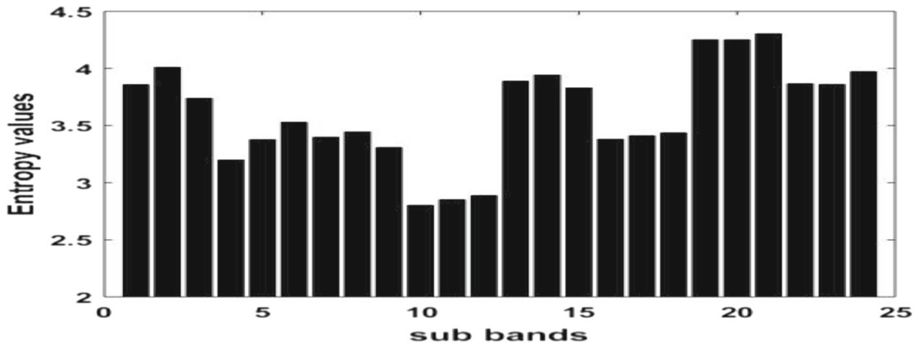


Fig. 9 Entropy values obtained from 4 level of SM-SBWT subbands for a face image

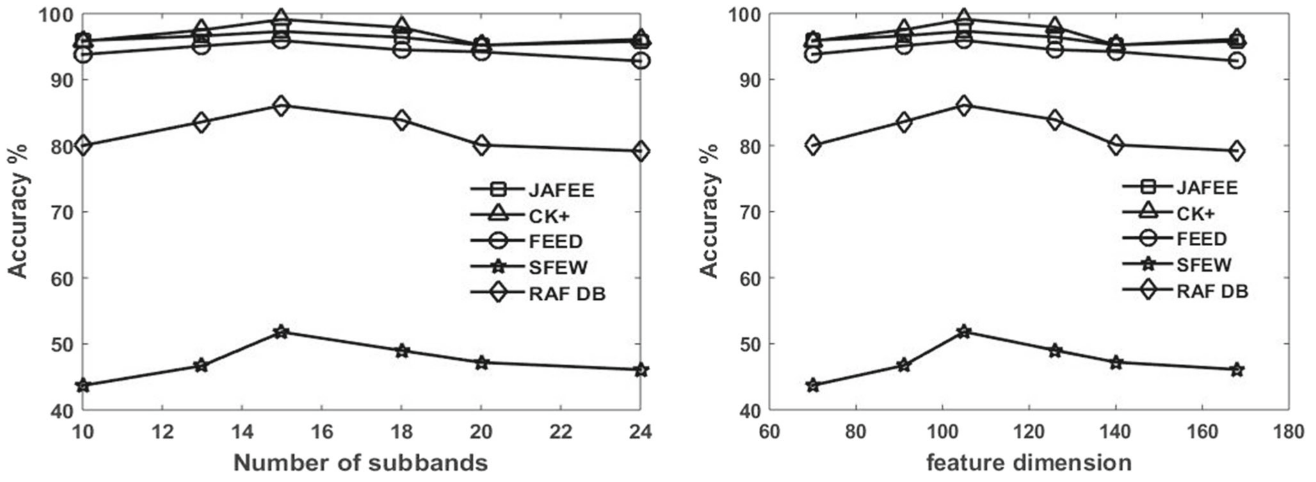


Fig. 10 Average accuracy rate obtained at different number of subbands and different feature dimension

Table 5 Correct recognition rate with different kernel functions

Number of subband selected	Feature vector dimension (subband selected x7)	JAFEE			CK+			FEED			SFEW			RAF-DB		
		Linear	RBF	Poly	Linear	RBF	Poly	Linear	RBF	Poly	Linear	RBF	Poly	Linear	RBF	Poly
24	168	93.3	94.6	95.8	94.5	95.8	96.1	93.2	93.6	92.8	44.1	45.3	46.1	78.5	78.3	79.2
20	140	94.4	94.9	95.2	95.9	95.1	95.2	93.8	93.9	94.2	46.2	46.4	47.2	79.4	79.6	80.1
18	126	95.9	96.2	96.4	96.4	97.2	98.1	94.2	94.5	94.5	48.1	48.6	49.0	82.3	82.5	83.9
15	**105**	96.8	97.1	**97.3**	98.3	98.7	**99.2**	95.1	95.8	**95.9**	49.2	49.5	**56.5**	84.4	84.8	**86.1**
13	91	96.1	96.4	96.6	96.7	96.9	97.5	94.8	94.8	95.1	48.3	48.3	46.7	83.5	83.8	83.6
10	70	94.6	95.1	95.9	95.1	95.3	95.8	94.2	94.3	93.8	42.0	43.7	43.7	79.4	79.8	80.0

Table 6 Confusion matrix of the proposed method for JAFEE database

	Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
Anger	98.3	0	0.9	0	0	0.6	0.2
Disgust	0.3	97.8	0.4	0.6	0	0.9	0
Fear	1.8	0	96.5	1.0	0.7	0	0
Happy	0.4	0	0.3	99.3	0	0	0
Neutral	0	0.3	1.1	0	97.1	1.5	0
Sad	0	2.0	1.5	0	0.7	95.8	0
Surprise	1.2	0	1.2	0	0	0.7	96.9

Table 7 Confusion matrix of the proposed method for CK + database

	Anger	Contempt	Disgust	Fear	Happy	Sad	Surprise
Anger	99.5	0	0.5	0	0	0	0
Contempt	0.2	99.1	0.7	0	0	0	0
Disgust	0	0	98.6	0.9	0.2	0.3	0
Fear	0	0.3	0	98.9	0	0	0.8
Happy	0	0	0	0	100	0	0
Sad	0	0.6	0.6	0	0	98.8	0
Surprise	0.1	0	0	0.9	0	0	99.0

Table 8 Confusion matrix of the proposed method for FEED 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	2.4	0	1.6	0
Fear	0.8	0	95.8	1.8	1.6	0	0
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	0	95.6	1.7
Surprise	3.2	0	0	0	2.2	0.7	93.9

Table 9 Confusion matrix of the proposed method for RAF-DB database

	Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
Anger	83.2	4.8	3.9	2.0	2.9	1.4	1.8
Disgust	3.2	82.7	5.4	1.9	3.8	1.5	1.5
Fear	3.8	2.7	85.1	1.6	3.6	1.5	1.7
Happy	2.1	1.1	2.3	92.6	0.6	0.8	0.5
Neutral	4.5	2.9	2.8	1.4	84.3	0.6	3.5
Sad	2.6	2.4	3.1	2.1	1.7	86.3	1.8
Surprise	2.0	2.2	3.1	1.2	3.5	0.8	87.2

Table 10 Confusion matrix of the proposed method for SFEW database

	Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
Anger	51.5	9.2	9.4	7.1	7.9	7.5	7.4
Disgust	10.3	54.7	8.1	5.6	6.3	6.9	8.1
Fear	8.0	7.4	55.9	6.6	7.3	8.0	6.8
Happy	5.7	5.8	6.3	64.6	5.1	7.2	5.3
Neutral	9.1	7.3	9.3	5.8	54.3	6.0	8.2
Sad	7.1	6.7	8.5	5.1	6.0	61.8	4.8
Surprise	10.5	6.3	11.2	6.2	6.3	6.6	52.9

datasets used in this work, our proposed methods produce improved results, and it is undoubtedly exposed in Table 12. Since the experimental setup varies for different methods, it is onerous to directly compare them with our proposed method. Our proposed method outperforms some of the methods mentioned in table which use wavelet transform approaches. Yu-Dong Zhang et al. (2016) and Huma Qayyum et al. (2017) adapt stationary wavelet transform for feature extraction. The biorthogonal wavelet entropy from two levels are calculated as features for the former case, and the combinations of LH and HL subbands obtained from stationary wavelet transform are used as features in the latter case. The formation of local mean and maximum energy wavelet subbands and combination and selection of subbands makes the proposed method superior in terms of accuracy compared with the above methods. The disparity in face attributes of facial image experience misrecognition of expression, but the advancement made

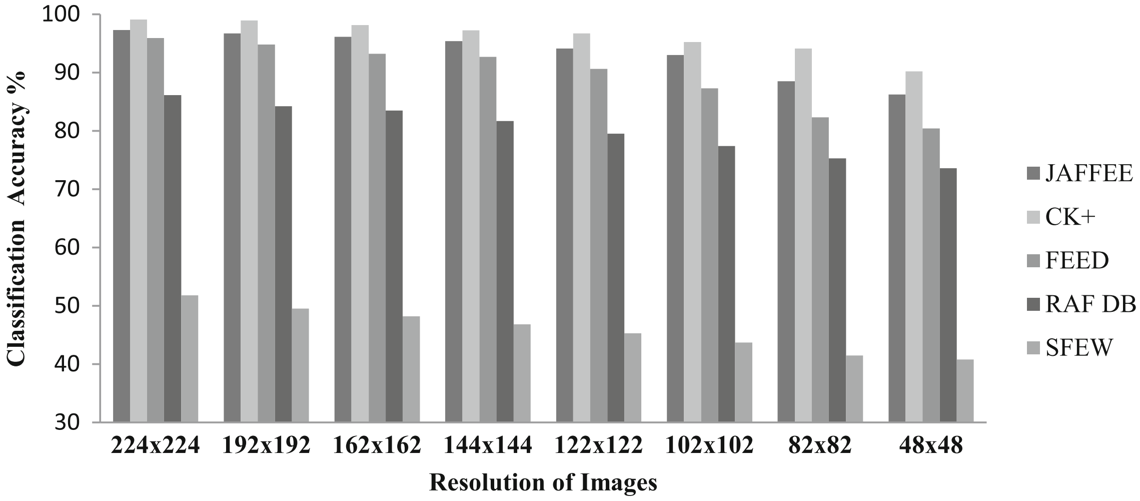


Fig. 11 Classification accuracy with different resolution of image

Table 11 CRR of SWT + SVM vs SM-SBWT + SVM

	SWT + SVM (%)	SM-SBWT + SVM (proposed) (%)
Anger	96.7	98.4
Disgust	93.5	96.3
Fear	94.2	96.9
Happy	97.5	99.3
Neutral	96.1	97.8
Sad	96.4	96.6
Surprise	97.2	95.9
Overall	95.9	97.3

by calculating the local mean and maximum energy calculation proclaim the edge and texture information in facial image which in turn gives better accuracy.

Our proposed method is also compared with some of the methods uses appearance-based approaches and obtained better comparative performance as described in Table 12. For all the datasets used in our work, the proposed method attains high accuracy comparatively with other methods in table. Our method achieves better performance when compared with appearance based approaches such as HOG + GSP by Meena et al. (2019), DWT by Kazmi and Arfan (2012) and LHMBP by Goyani and Patel (2018). The statistical analysis and the feature selection adapted in our method make our method surpass the state-of-the-art methods. Comparison is also made with some of the method which use feature descriptors such as NEDP (Iqbal et al. 2018), PLBP (Khan et al. 2013), LDTP (Ryu et al. 2017), LPDP (Makhmudkhujaev et al. 2019) and LDMEP (Uma Maheswari et al. 2020). Our proposed method enacts better performance compared with the listed local descriptors.

In the recent years, deep learning approach explicit tolerant results in the field of facial expression recognition. The performances of some of the deep learning methods are listed in Table 12. Our proposed method shows superior in accuracy compared with deep learning methods proposed by Xu et al. (2020), Gan et al. (2019), Alfakih et al.(2020), Wang et al. (2020), Xu et al. (2020), etc. In deep learning methods, the inadequate training data leads to data augmentation for artificial generation of training samples. By this considerable effect, the system gets more complex, and this will leads to the degradation of performance. It is certainly obvious in Table 12 that our proposed method surpasses the above deep learning methods in terms of classification accuracy.

Even if our recognition system accomplish enhanced results, there endure few issues. The geometric features are not considered here, and also our proposed system is bit sensitive to illumination changes. Again video database is not analyzed in our work. The aforesaid issues will be expounded in our future works. In future, we shall apply our method to the micro-expression recognition system (Kam Meng Goh et al. 2018) which resembles similar to facial expression recognition system. Moreover, we shall analyze our system with 3D dataset models (Yuping Ye et al. 2020). Further, we shall impose our method to other fields such as big data affective analysis (Shoumya et al. 2019), intelligent tutoring system (Ramon Zatarain Cabada et al. 2019) and emotion-aware robots (Jun Yanga et al. 2019).

Table 12 Comparison with state-of-the-art methods

References	Database	Technique used	Class	Accuracy rate (%)
Sidra BatoolKazmi and Arfan (2012)	JAFEE	DWT + SVM	7	96.4
H.K Meena et al. (2019)	JAFEE	HOG + GSP	7	88.57
Yu-Dong Zhang et al. (2016)	JAFEE	BWE +FSVM	7	96.77
Mahesh Goyani and Patel (2018)	JAFEE	LHMBP + HNAD	7	88.0
B. Ryu et al. (2017)	JAFEE	LDTP + SVM	7	93.2
Iqbal et al. (2018)	JAFEE	NEDP	7	67.97
Uma Maheswari et al. (2020)	JAFEE	LDMEP	7	68.95
Sun, Zheng and Fu (2020)	JAFEE	AM/CNN	7	92
The proposed Method	JAFEE	SM-SBWT + SVM	7	97.3
H.K Meena et al. (2019)	CK+	HOG + GSP	6	97.61
HumaQayyum et al. (2017)	CK+	SWT + DCT + NN	7	96.61
Rizwan Ahmed Khan et al. (2013)	CK+	PLBP + SVM	6	96.7
Mahesh Goyani and Patel (2018)	CK+	LHMBP + HNAD	7	98.2
B. Ryu et al. (2017)	CK+	LDTP + SVM	7	94.2
Sadeghi and Raie (2019)	CK+	LCML	7	98.17
Amir Jamshidnezhad and Nordin (2013)	CK+	BROA	4	93.5
Iqbal et al. (2018)	CK+	NEDP	7	92.97
Makhmudkhujaev et al. (2019)	CK+	LPDP	7	94.5
Uma Maheswari et al. (2020)	CK+	LDMEP	7	93.89
Sun et al. (2020)	CK+	AM/CNN	7	87.2
Yanling Gan et al. (2019)	CK+	MA/CNN	7	96.28
Pan (2020)	CK+	HOG + CNN	6	97.01
The proposed method	CK+	SM-SBWT + SVM	7	99.2
Rizwan Ahmed Khan et al. (2013)	FEED	PLBP + SVM	6	92.3
Amir Jamshidnezhad and Nordin (2013)	FEED	BROA	4	89.8
The proposed Method	FEED	SM-SBWT + SVM	7	95.9
Amani Alfakih et al.(2020)	RAF	DCNN	7	83.08
Kai Wang et al. (2020)	RAF	RAN + CNN	7	59.5
Sadeghi and Raie (2019)	RAF	LCML	7	80.74
Yanling Gan et al. (2019)	RAF	MA/CNN	7	85.69
The proposed method	RAF	SM-SBWT + SVM	7	86.1
Mahesh Goyani and Patel (2018)	SFEW	LHMBP + HNAD	7	45.0
Kai Wang et al. (2020)	SFEW	RAN + CNN	7	54.1
Sadeghi and Raie (2019)	SFEW	LCML	7	55.5
Ying Xu et al. (2020)	SFEW	CNN	7	51.9
The proposed Method	SFEW	SM-SBWT + SVM	7	56.5

6 Conclusion

In this paper, a novel facial emotion recognition system is proposed. In human face, distinct muscle movements generate disparate emotions. It is noticed such that the horizontal and vertical muscle movement on the face precise majority of the emotions. Accordingly, the expression-specific LH, HL and HH subband features of SBWT are combined to generate more effective features which influence more accurate recognition. The edge and texture information in facial image is embellished by calculating the local energy in wavelet subbands. Since the decimation operation is not involved in SWT, it results in a large number of coefficients in the subband. In order to achieve compatible and consistent statistical features, different levels of SWT are estimated. Since entropy is an efficacious statistical parameter, subbands are selected based on the entropy values calculated for each subband. A combination of subband is enforced continual to the estimation of local mean and maximal energy of wavelet subband. Then DCT is performed to gain the most energy effective information in the selected subbands. Subsequently, the statistical parameters calculated from the selected subbands are the extracted features and are imposed to PCA for dimension reduction. The classification mechanisms for categorizing different expressions are performed using multiclass SVM classifier. Our proposed method gives an aspirant result in facial emotion recognition when compared with the existing methods in literature. Also, our proposed method is more effective and is robust for different facial expressions. Face images with facial emotions from various database like JAFEE database, CK + database, FEED database, RAF database and SFEW database have been used to evaluate the qualitative performance of classification by our proposed method. It is inspected that capturing the spatial–temporal information is difficult due to the movement of facial critical areas. Hence, the movement of facial critical areas should be modeled using more powerful models by utilizing specific methods such as metric learning in the future. Moreover, our future research will be on analyzing images with pose variation, images with occlusion images with non-uniform illumination as well as real-time video stream that will lead to meet the necessity of many engineering applications.

Compliance with ethical standards

Conflict of interest All authors of the paper declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was not required as no humans or animals were involved.

References

Alfakih A, Yang S, Hu T (2020) Multi-view cooperative deep convolutional network for facial recognition with small samples learning. Advances in intelligent systems and computing, vol 1003. Springer, Cham. https://doi.org/10.1007/978-3-030-23887-2_24

Ali H, Hariharan M, Yaacob S, Adom AH (2015) Facial emotion recognition based on higher-order spectra using support vector based on higher-order spectra using support vector machines. J Med Imag Health Inf 5:1272–1277

Bhattacharya A, Choudhury D, Dey D (2018) Edge-enhanced bi-dimensional empirical mode decomposition-based emotion recognition using fusion of feature set. Soft Comput 22:889–903. https://doi.org/10.1007/s00500-016-2395-4

Cabada RZ, Rangel HR, Estrada MLB, Lopez HMC (2019) Hyper-parameter optimization in CNN for learning centered emotion recognition for intelligent tutoring systems. Soft Comput. https://doi.org/10.1007/s00500-019-04387-4

Cambria E (2016) Affective computing and sentiment analysis. IEEE Intell Syst 31(2):102–107

Crouse MS, Nowak RD, Baraniuk RG (1998) Wavelet-based statistical signal processing using hidden Markov models. IEEE Trans Signal Proc 46(4):886–902

Darwin C (1872) The expression of the emotions in man and animals. J. Murray, London

Dhall A, Goecke R, Lucey S, Gedeon T (2011) Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark. In: Proceedings of IEEE international conference on computer vision workshops, pp 2106–2112

Edwards T (1992) Discrete wavelet transforms: theory and implementation. Technical report, Stanford University, 1991

Ekman P, Friesen WV (1971) Constant across cultures in face and emotions. J Pers Soc Psychol 17(2):124–129

Fan X, Tjahjadi T (2019) Fusing dynamic deep learned features and handcrafted features for facial expression recognition. J Vis Commun Image Represent. https://doi.org/10.1016/j.jvcir.2019.102659

Gan Y, Chen J, Yang Z, Luhui X (2019) Multiple attention network for facial expression recognition. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2963913

Gavrilescu M (2015) Recognizing emotions from videos by studying facial expressions, body postures and hand gestures. In: 23rd Telecommunications Forum Telfor, Belgrade, SERBIA, pp 720–723

Ghimire G, Lee J (2013) Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. J Sens 13:7714–7734

Goh KM, Ng CH, Lim LL, Sheikh UU (2018) Micro-expression recognition: an updated review of current trends, Challenges and Solutions. Vis Comput Springer 2018:1–24

Goyani M, Patel N (2017) Template matching and machine learning-based robust facial expression recognition system using multi-level Haar wavelet. Int J Comput Appl. https://doi.org/10.1080/1206212X.2017.1395134

Goyani M, Patel N (2018) Robust facial expression recognition using local haar mean binary pattern. J Inf Sci Eng 34:1237–1249. https://doi.org/10.6688/JISE.201809_34(5)0008

Iqbal MTB, Abdullah-Al-Wadud M, Ryu B, Makhmudkhujaev F, Chae O (2018) Facial expression recognition with neighborhood-aware edge directional pattern (NEDP). IEEE Trans Affect Comput 11(1):125–137. https://doi.org/10.1109/taffc.2018.2829707

Jamshidnezhad A, Nordin MJ (2013) Bee royalty offspring algorithm for improvement of facial expressions classification model. Int J Bio-Inspired Comput 5(3):175–191

Kazmi SB, Arfan QJ (2012) Wavelets-based facial expression recognition using a bank of support vector machines. Soft Comput 16:369–379. https://doi.org/10.1007/s00500-011-0721-4

Khan RA, Meyer A, Konik H, Bouakaz S (2013) Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recogn Lett 34:1159–1168

Li W, Zhang Y, Fu Y (2007) Speech emotion recognition in Elearning system based on affective computing. In: Proceedings of natural computation, 2007, (ICNC 2007), pp 809–813

Li S, Weihong D, JunPing D (2017) Reliable crowdsourcing and deep locality preserving learning for expression recognition in the wild. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2584–2593

Li K, Jin Y, Akram MW (2020) Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. Vis Comput 36:391–404. https://doi.org/10.1007/s00371-019-01627-4

Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z (2010) The extended Cohn-Kanade Dataset (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), pp 94–101

Lyons MJ, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with Gabor wavelets. In: 3rd IEEE international conference on automatic face and gesture recognition, pp 200–205

Ma J, Fan X, Yang SX, Zhang X, Zhu X (2018) Contrast limited adaptive histogram equalization based fusion in YIQ and HIS color spaces for underwater image enhancement. Int J Pattern Recognit Artif Intell 32(07):1854018

Makhmudkhujaev F, Abdullah-Al-Wadud M, Iqbal MTB, Ryu B, Chae O (2019) Facial expression recognition with local prominent directional pattern. Signal Process Image Commun. https://doi.org/10.1016/j.image.2019.01.002

Meena HK, Joshi SD, Sharma KK (2019) Facial expression recognition using graph signal processing on HOG. IETE J Res. https://doi.org/10.1080/03772063.2019.1565952

Nason GP, Silverman BW (1995) The stationary wavelet transform and some statistical applications. Wavelet at Statistics, Lecture Notes in statistics, Vol. 103, Springer, New York. pp 281–299

Pan X (2020) Fusing HOG and convolutional neural network spatial–temporal features for video-based facial expression recognition. IET Image Process 14(1):176–182. https://doi.org/10.1049/iet-ipr.2019.0293

Qayyum H, Majid M, Anwar SM, Khan B (2017) Facial expression recognition using stationary wavelet transform features. Hindawi Math Probl Eng, Vol 2017

Reza AM (2004) 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

Ryu B, Rivera AR, Kim J, Chae O (2017) Local directional ternary pattern for facial expression recognition. IEEE Trans Image Process 26(12):6006–6018. https://doi.org/10.1109/TIP.2017.2726010

Sadeghi H, Raie AA (2019) Histogram distance metric learning for facial expression recognition. J Vis Commun Image Represent 62:152–165. https://doi.org/10.1016/j.jvcir.2019.05.004

Sanket NJ, Vyshak AV, Manikantan K, Ramachandran S (2014) Face recognition using adaptive filter wavelet transform based feature extraction. In: International conference on Science Engineering and Management Research, (ICSEMR) 2014 IEEE Stanford University

Shoumya NJ, Angb L-M, Sengc KP, Motiur Rahamana DM, Ziaa T (2019) Multimodal big data affective analytics: a comprehensive survey using text, audio, visual and physiological signals. J Netw Comput Appl. https://doi.org/10.1016/j.jnca.2019.102447

Sun X, Zheng S, Fu H (2020) ROI-attention vectorized CNN model for static facial expression recognition. IEEE Access 8:7183–7194. https://doi.org/10.1109/ACCESS.2020.2964298

Tian Y, Cheng J, Li Y, Wang S (2019) Secondary information aware facial expression recognition. IEEE Signal Process Lett 26(12):1753–1757. https://doi.org/10.1109/LSP.2019.2942138

Tsai HH, Chang YC (2017) Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Comput 22(13):4389–4405. https://doi.org/10.1007/s00500-017-2634-3

Uma Maheswari V, Varaprasad G, Viswanadha Raju S (2020) Local directional maximum edge patterns for facial expression recognition. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-018863

Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154

Wallhoff F, Schuller B, Hawellek M, Rigoll G (2006) Efficient recognition of authentic dynamic facial expressions on the feedtum database. In: IEEE international conference on multimedia and expo, IEEE Computer Society, pp 493–496

Wang S, Zhuo Z, Yang H, Li H (2013) An approach to facial expression recognition integrating radial basis function kernel and multidimensional scaling analysis. Soft Comput 18(7):1363–1371. https://doi.org/10.1007/s00500-013-1149-9

Wang S-H, Phillips P, Dong Z-C, Zhang Y-D (2017) Intelligent facial emotion recognition based on stationary wavelet entropy and jaya algorithm. Neurocomputing 272:668–676

Wang K, Peng X, Yang J, Meng D, Qiao Yu (2020) Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans Image Process 29:4057–4069

Xu Y, Liu J, Zhai Y et al (2020) Weakly supervised facial expression recognition via transferred DAL-CNN and active incremental learning. Soft Comput 24:5971–5985. https://doi.org/10.1007/s00500-019-04530-1

Yan W, Ming L, Congxuan Z, Hao C, Yuming L (2019) Weighted-fusion feature of MB-LBPUH and HOG for facial expression recognition. Soft Comput. https://doi.org/10.1007/s00500-019-04380-x

Yanga J, Wanga R, Guanb X, Hassanc MM, Almogrenc A, Alsanadc A (2019) AI-enabled emotion-aware robot: the fusion of smart clothing, edge clouds and robotics. Future Gener Comput Syst 102:701–709. https://doi.org/10.1016/j.future.2019.09.029

Ye Y, Song Z, Guo J, Qiao Y (2020) SIAT-3DFE: a high-resolution 3D facial expression dataset. IEEE Access 8:48205–48211. https://doi.org/10.1109/ACCESS.2020.2979518

Yu M, Zheng H, Peng Z, Dong J, Du H (2020) Facial expression recognition based on a multi-task global-local network. Pattern Recognition Lett. https://doi.org/10.1016/j.patrec.2020.01.016

Zhang S, Zhao X, Lei B (2012) Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans Signal Process 8:21–31

Zhang Y-D, Yang Z-J, Hui-Min L, Zhou X-X, Phillips P, Liu Q-M, Wang S-H (2016) Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access 4(2016):8375–8385

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.