CrossMark

Face expression recognition system based on ripplet transform type II and least square SVM

Nikunja Bihari Kar[1] · Korra Sathya Babu[1] ·
Arun Kumar Sangaiah[2] · Sambit Bakshi[1]

Abstract This paper discusses the development of an efficient and automated system for the recognition of facial expressions, which is essentially an application augmented with many multimedia computing systems. The proposed scheme works in three stages. In the first stage, ripplet transform type II (ripplet-II) is employed to extract the features from facial images because of its efficiency in representing edges and textures. In the next stage, a principal component analysis (PCA)+linear discriminant analysis (LDA) approach is utilized to obtain a more compact and discriminative feature set. In the final stage, classification is performed using a least squares variant of support vector machine (LS-SVM) with radial basis function (RBF) kernel. The proposed system is validated on two benchmark datasets namely the Extended Cohn-Kanade (CK+) and Japanese female facial expression (JAFFE). The experimental results demonstrate that the propose system yields superior performance as compared to other state-of-the-art schemes.

Keywords Face expression recognition · Linear discriminant analysis · Principal component analysis · Ripplet transform

✉ Nikunja Bihari Kar
   nikunjakar@gmail.com

   Korra Sathya Babu
   ksathyababu@nitrkl.ac.in

   Arun Kumar Sangaiah
   sarunkumar@vit.ac.in; arunkumarsangaiah@gmail.com

   Sambit Bakshi
   sambitbaksi@gmail.com; bakshisambit@nitrkl.ac.in

[1]   Department of Computer Science and Engineering, National Institute of Technology, Rourkela, Odisha 769 008, India

[3]   School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu 632 014, India

⚛ Springer

1 Introduction

Face expression recognition (FER) has long been a research area of great interest. The work of analysis of face expressions was started in the early nineties which indicates the significance of the research area. At first, the psychologists were investigated the different facial emotions. The six basic emotions are first formalized in the work of [11] and these emotions are: anger, disgust, fear, happiness, sadness, and surprise. A set of facial muscle movement that represents the facial expressions termed as facial action units (FAUs), this helps the recognition process to be more standardized. This FAUs forms facial action coding system (FACS). FAUs are combined to formulate rules for different facial expressions. The analysis of facial expressions are the primary research area of psychology, and much of the work and literature is published in this area. However, the interest broadened with the publication of the work [35], which presents an automatic face expression recognition system from a sequence of face expressive images.

Facial expression recognition has been a potential research area in recent years due to the advancement of its related research area especially machine learning, computer vision, image processing, and cognitive science. FER with high accuracy remains to be troublesome and still addresses a dynamic investigation area. These wide area applications such as next generation human computer interface (HCI), behavioral science, video games, robotics, psychology, touchy music juke boxes and TVs, telecommunications, animations, automobile safety, and educational software [3] are effectively used the face expression investigation for better utilization of resources.

In emotion recognition system, the facial features are categorized into two types, such as appearance based feature extraction and geometric based feature extraction. In appearance based feature extraction, the Gabor wavelets [1] image filters are used on total or a part of the face for feature extraction. Generally, the existing methods such as local ternary pattern (LTP), local binary pattern (LBP) [14], histograms of oriented gradients (HOG) [19], discrete wavelet transform (DWT) [10], and stationary wavelet transform (SWT) [45] are also used for appearance based feature extraction technique. The geometrically based feature extraction has used the location and shape of facial components such as eyes, eyebrows, nose, and mouth etc. for feature representation.

In this work, ripplet transform type II (ripplet-II) algorithm [49] is harnessed for feature extraction. Hybrid approaches (combination of geometry and appearance features) that promise better results are deployed for feature extraction. Finally, a classifier is utilized to detect emotion from facial features. Wavelet transform (WT) has been observed as the most popular feature extraction method in earlier systems [10, 29, 32, 45, 52]. However, it has a deficiency of limited directionality, non-supportiveness to anisotropy, etc. Therefore, WT cannot capture the subtle and intrinsic details of the face expression images which are required for segregating emotions.

The primary motivation of this proposed work is to develop an automatic and robust face expression recognition system which minimizes the computational complexity and maximizes the recognition accuracy. The other motivation of this work is to make the algorithm more general so that it can be used with other face expression datasets by varying emotion classes.

In summary, the proposed system has the following characteristics,

– An automatic and robust face expression recognition system is proposed.
– The proposed system uses ripplet-II transform for feature extraction as it has better localization capability in spatial as well as in frequency domain. In addition, ripplet-II efficiently captures 2D singularities along different curves in any direction.

– Principal component analysis (PCA)+linear discriminant analysis (LDA) method is applied to the higher dimensional ripplet-II coefficients resulting in a reduced feature vector with 6 features which ensures low computational overhead.
– The LS-SVM is used as the classifier which is computationally faster than conventional SVM.
– The proposed system is compared against other state-of-the-art methods over two standard datasets and the results confirm the superiority of the proposed scheme.

The remainder of this paper is structured as follows: Section 2 provides a brief overview of the literature survey. Section 3 contains the working procedure of each step of the proposed system. Section 4 describes the experimental results and discussions. Finally, Section 5 concludes the paper.

2 Literature survey

In recent years, there has been a significant progress in FER system development. The DWT, SWT, LBP, HOG, LTP, and Gabor wavelets are some feature extraction techniques based on appearance method, which extracts features from the whole face or some part of the face without any previous knowledge. So in this appearance based technique, the size of the feature vector is the whole face image or the component of the face image. In case of geometric based feature extraction technique, the features are dependent upon some fiducial points. As the fiducial points are set by manual method, so it increases the complexity. In this approach, accuracy is completely dependent upon the face feature points.

Uddin et al. [40–42] have proposed several new FER systems, which recognize emotions from facial expression depth videos. Deshmukh et al. [9] reviewed current FER systems and have evaluated them concerning their application to real-time. Bartlett et al. [1] have utilized a set of Gabor filters to extract features from a face image. Adaboost is utilized to select 200 features among them for each action unit. A linear SVM is used to identify facial action unit. A novel face emotion recognition approach is proposed by Tsai et al. [38] based on shape and texture features. In this method, the face area of the image is detected by self-quotient image (SQI) and Haar-like features (HFs) filters. Subsequently, the features are extracted by using discrete cosine transform (DCT), angular radial transformation (ART), and Gabor filters(GF) techniques. Finally, SVM is harnessed as classifier to recognize emotions. Michel and Kaliouby [24] employed a face tracker to locate the position of 22 facial features. A displacement vector is calculated by taking the Euclidean distance of neutral and peak facial expressions of these 22 facial features. The displacement vector is fed to the SVM classifier with RBF kernel for emotion recognition.

A hybrid feature extraction technique is proposed by Chen et al. [5], which includes local texture displacement and facial feature point displacement between neutral to peak face expression images. In this method, seven facial expressions are recognized by multi-class SVM approach. A face detector based on Gabor-feature-based boosted classifier is proposed by Valster and Pantic [43]. In this method, 20 facial fiducial points are located through a series of images using particle filtering with factorized likelihoods and SVM, GentleBoost, and hidden Markov models are used for action unit (AU) recognition. The six semantic feature extraction technique is proposed by Hsieh et al. [17] using Laplacian of Gaussian (LoG) and directional gradient operators like GFs. In this approach, the human face and calibrate facial components are detected by active shape model (ASM). Later, Gabor and LoG extract semantic features by edge detection technique.

Dahmane and Meunier [8] used a dynamic grid and HOG to extract facial features. A face image is divided into 48 cells with eight rows and six columns. A local histogram is processed for each cell. A block containing four cells, and local histograms of each cell was concatenated to get a single histogram of each block with L2-norm block normalization. Combining all the 12 block histograms resulted in a 432-dimensional feature vector. A SVM with RBF kernel was used to detect the emotion class. Chen et al. [6] employed HOG to various components of face like eyes, eyebrows, and mouth to extract the features. At the beginning, the features of HOG are extracted from all the face components, and then integrated later to produce the resultant feature vector. Gritti et al. [14] used HOG, LTP, and LBP descriptors for facial representations. They were using a linear SVM scheme with 10-fold cross validation testing in their recognition experiments. Wang et al. [46] introduced a novel hybrid feature extraction technique that combine weber local descriptor (WLD) with HOG features. These features contain both texture and shape information. A k-nearest neighbors (kNN) classifier is suggested for face emotion recognition. Mlakar and potocinik [25] proposed a FER system based on HOG descriptor. Genetic algorithm is used to determine the optimal parameters for HOG. A difference vector was obtained by subtraction of HOG features from neutral and peak expressive subjects image. The resultant difference vector is used by a SVM classifier to recognize emotion.

Guo et al. [15] harnessed extended local binary pattern (ELBP) to extract face features. The covariance matrix transform in K-L transform (KLT) is applied to reduce the feature dimension while preserving only main features. SVM classifier is used for best recognition performance. Uccar et al. [39] applied curvelet transform to each local face regions to avoid huge curvelet coefficients. Feature matrix is then generated by calculating entropy, standard deviation, mean for curvelet coefficient of each region. Finally, online sequential extreme learning machine-spherical clustering (OSELM-SC) is employed for classification of face expression images. Siddiqi et al. [33] employed stepwise linear discriminant analysis (SWLDA) for choosing the localized features from expression frames using partial F-test values, that reduce within class variance and increase the between class variance. Then the hidden conditional random fields (HCRIs) model is utilized for recognition. In the work of Zhang and Tjondronegoro [51] facial features are obtained by extracting patch-based 3D Gabor features, by selecting 'salient' patches with patch matching operations. Facial movement and muscle movements are considered for better recognition results. This method shows promising results for face registration errors with fast processing time. Elaiwat et al. [12] introduced a novel spatio-temporal model capable of capturing different transformations. The proposed model employed two hidden sets, to encode the relationship between image pairs. The first hidden layer performs facial expression transformations, while the second layer performs the task of other transformations namely pose and occlusion. Happy and routray [16] investigated the relevance of different facial patches for FER. Appearance features are extracted from the selected facial patches. These selected patches are further processed to get the salient patches, which contain the discriminant information for classification of each pair of expressions. SVM with RBF kernel is adopted for classification. Jung et al. [18] reported a deep learning based feature extraction technique to reduce the effort of manual detection of face features. The proposed deep network is based on two models, first one extracts geometry features from facial landmark points, while other model extracts appearance features from image sequences. Kar et al. [19] proposed a novel hybrid FER system, which uses HOG as feature extraction tool and PCA+LDA for feature reduction. Back-propagation neural network (BPNN) classifier is harnessed to detect facial expressions.

Dongcheng and Jieqing [10] used a combination of DWT and principal component analysis (PCA)/Fisher linear discriminant analysis (FLD) for feature extraction. Initially, the original image is decomposed using DWT. Then, FLD is then applied to extract relevant features from low-frequency and high-frequency components. At last, the nearest neighbor classifier (NNC) is employed to detect the facial expressions. Siddiqi and Lee [32] proposed a new FER system using symlet wavelet for feature extraction. In their work, an input image is partitioned into many regions, and the distance between the two pixels are calculated for each region. The feature dimension is reduced by using LDA. Finally HMM is employed for labeling expressions. Zhang et al. [52] presented biorthogonal wavelet entropy (BWE) to extract multiscale features and then used fuzzy multiclass SVM as the classifier. Kazmi et al. [20] used the 3-level 2D-DWT decomposition of an image for feature representation. The PCA is used for feature reduction. Finally, a bank of SVMs is employed for classification. Wang et al. [45] explore an FER algorithm by combining stationary wavelet entropy (SWE) and single hidden layer feed-forward neural network (SHLFNN) with Jaya algorithm. A summary of state-of-the-art face expression recognition methods explained briefly in Table 1.

From the above mentioned literature, it has been observed that most of the systems use appearance based feature extraction schemes which leads to high dimensional feature space. Further, some methods used DWT as a feature extraction tool, which is translation-variant. Moreover, standard DWT has disadvantages of limited directional selectivity and shift variance [27]. SWT can handle the shift invariance problem [28], but it offers more redundancy and does not perform well in representing higher dimensional singularities. To address the above mentioned problems, we deployed ripplet-II transform for feature extraction which represents the image at different scales and different directions and in addition to this, it efficiently describes the images with edges. Further, SVM is used in most cases for recognition which needs more computationally overhead. Hence, a computational efficient variant of SVM called LS-SVM is used for classification in this study.

3 Proposed work

The proposed method includes four major steps: preprocessing of face images, feature extraction, feature length reduction, and classification. The block diagram of the proposed system is shown in Fig. 1. Detail description of each block of the proposed system is described below.

3.1 Preprocessing

At first, the face images are converted to a gray scale image. Then the contrast of the image was adjusted so that 1% of the information is immersed at low and high intensities. Thereafter, the face is detected using popular Viola and Jones face detection algorithm [44]. The detected face region is cropped from the original face image and the cropped face is reshaped to size of 128×128.

3.2 Feature extraction using ripplet-II

Current image representation schemes have limited capability of representing 2D singularities (e.g. edges in an image). Wavelet transform has better performance in representing 1D singularities than Fourier transform. Wavelet transform overcomes the problem of Fourier

Table 1 Different approaches for FER system

| Approach	Authors	Feature extraction	Classifier	Perfomace results	Data with size
Geometric based	Michel & Kaliouby [24] (2003)	Facial points	RBF SVM	87.9%	CK & Live-video
	Chen et al. [5] (2012)	Facial points & local textures	RBF SVM	95%	CK (327 images)
	Valstar & Pantic [43] (2012)	Facial points	RBF SVM & HMM	91.7% 95.3%	CK (153 videos) MMI (244 videos)
	Hsieh et al. [17] (2015)	Semantic features ASM+GF+LoG	RBF SVM	94.7%	CK (341 images)
	Saeed et al. [30] (2014)	Geometrical features	RBF SVM	83.01%	CK (327 images)
Appreance based	Bartlett et al. [1] (2006)	Gabor wavelets	Linear SVM & AdaBoost	90.9%	Ekman-Hagar & CK (2568 images)
	Gritti et al. [14] (2008)	HOG	Linear SVM	92.7%	CK (310 images)
	Tsai et al. [38] (2010)	SQI+Sobel+ DCT +ART+GF	RBF SVM	98.59%	CK, JAFFE, & FG-NET (3624 images)
	Dongcheng & Jieqing [10] (2010)	DWT+PCA/FLD	KNN	88.89%	JAFFE (213 images)
	Dahmane & Meunier [8] (2011)	Dynamic grid & HOG	RBF SVM	70%	GEMEP-FERA
	Chen et al. [5] (2012)	Facial points & local textures	RBF SVM	95%	CK (327 images)
	Wang et al. [46] (2013)	WLD & HOG	kNN	95.86% 93.97%	CK (1344 images) JAFFE (213 images)
	Siddiqi & Lee [32] (2013)	Symlet wavelet+LDA	HMM	94%	CK (2880 images)
	Chen et al. [6] (2014)	Facial components & HOG	Linear SVM	88.7% 94.3%	CK (327 images) JAFFE (213 images)
	Mlakar & Potocinik [25] (2015)	HOG feature vector differences	SVM	95.64% 87.82%	CK (1232 images) JAFFE (192 images)
	Happy & Routray [16] (2015)	Salient facial patches	RBF SVM	94.09% 91.79%	CK (329 images) JAFFE (213 images)
	Siddiqi et al. [33] (2015)	Stepwise LDA	HRCF	96.83% 96.33%	CK (309 images) JAFFE (195 images)
	Uccar et al. [39] (2016)	Local Curvlet transform	OSELM-SC	95.17% 94.65%	CK (1184 images) JAFFE (213 images)
	Zhang et al. [52] (2016)	BWE	FSVM	96.77%	Own (700 images)
	Kar et al. [19] (2017)	HOG+PCA+LDA	BPNN	99.51%	CK (327 images)
	Wang et al. [45] (2017)	SWE	SHLFNN +Jaya	96.80%	Own (700 images)
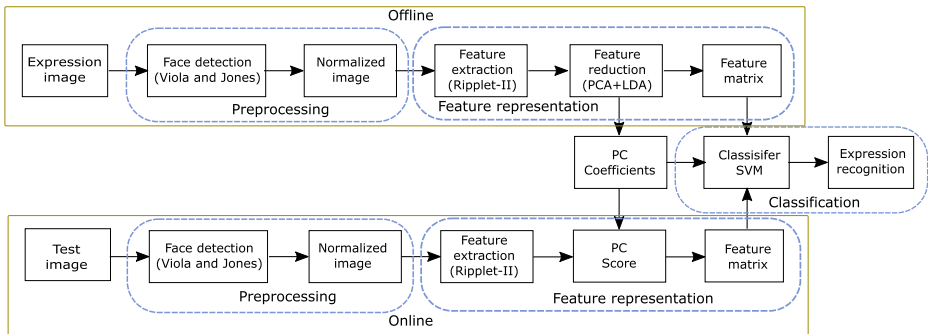
Fig. 1 Block diagram of proposed face expression recognition system

transform and has gained popularity among researchers because of its localization and multiscale property. Wavelet transform performs better in 2D i.e., it can deal with point singularities; however, it fails to represent higher dimensional singularities like line, curves, etc., effectively due to the poor directional selectivity. Therefore ripplet-II transform can be utilized as a feature extraction tool because of its efficiency in representing edges and textures. The ripplet transform provide efficient representation of images that contain edges [48]. Ripplet-II transform has the capability to record structure information along with arbitrary curves by adjusting parameters like scale, orientation, location, and degree [49].

In this work, the input for the ripplet-II transform is an image. Therefore there is a need to apply a discrete version of ripplet-II transform instead of continuous version. Ripplet-II transform is based on generalized Radon transform (GRT). The GRT convert curves into points, which creates peaks located at the corresponding curve. Ripplet-II transforms consists of two phases: (1) It uses GRT to turn curve singularities into point singularities in a generalized random domain, and (2) uses wavelet transform (WT) to settle point singularities into a generalized random domain. A digital image processing need to use discrete ripplet-II transform instead of continuous ripplet-II transform. We need to compute the ripplet-II function first, to represent ripplet-II transform. Given a smooth univariate function $\omega : \mathbb{R} \to \mathbb{R}$ with $\int \omega(t)dt = 0$, a bivariate function $\varphi_{s,t,d,\phi} : \mathbb{R}^2 \to \mathbb{R}^2$ in a polar coordinate system is obtained as follows,

$$\varphi_{s,t,d,\phi}(\varrho, \theta) = s^{-1/2}\omega((\varrho \cos^d((\phi - \theta)/d) - t)/s) \tag{1}$$

where $s > 0$ denotes scale, $t \in \mathbb{R}$ denotes translation, $d \in \mathbb{N}$ denotes degree and $\phi \in [0, 2\pi]$ denotes orientation. The function $\varphi_{s,t,d,\phi}$ is called ripplet-II function.

Ripplet-II transform of a real-valued 2D function is defined by the inner product of the function f and ripplet-II function,

$$\mathcal{R}_f(s, t, d, \phi) = \int \int \bar{\varphi}_{s,t,d,\phi}(\varrho, \theta)f(\varrho, \theta)\varrho d\varrho d\theta \tag{2}$$

Where, $\bar{\varphi}$ denotes the complex conjugate of φ and $f(\varrho, \theta)$ is the polar coordinate system.

Ripplet-II transform has the capability of capturing structure information along arbitrary curves by tuning the scale, location, orientation, and degree parameters. It can be obtained by the inner product between GRT and 1D wavelet. The ripplet-II transform of a function f can be computed by first calculating GRT of f, and then computing 1D wavelet of the GRT of f as below,

$$f(\varrho, \theta) \stackrel{GRT}{\Rightarrow} GR_d[f](r, \phi) \stackrel{1d-WT}{\Rightarrow} \mathcal{R}_f(s, t, d, \phi) \tag{3}$$

where 1D-WT is with respect to r.

3.2.1 Forward ripplet-II transform with $d = 2$

1. Convert Cartesian coordinate to polar coordinate i.e. to convert $f(x, y)$ to $f(\varrho, \theta)$.

   For $(\varrho'^2, 2 \times \theta')$ substitute (ϱ, θ) in $f(\varrho, \theta)$.

   Convert the polar coordinate (ϱ', θ') to Cartesian coordinates (x, y) and obtain a new image $f_1(x, y)$ by interpolation process. Where x and y are integers.

2. Classical Radon transform is applied to $f_1(x, y)$ resulting $R(r', \phi')$.

   Substitute (r', ϕ') with $(\sqrt{r}, \phi/2)$ in function $R(r', \phi')$.

   The generalized Radon coefficients can be obtained as follows,

$$
\begin{aligned}
GR_2(r, \phi) &= 2\sqrt{r} \int \int \varrho' f(\varrho'^2, 2\theta') \times \delta(\varrho' \cos(\theta' - \tfrac{1}{2}\phi) - \sqrt{r}) d\theta' d\varrho' \\
&= 2\sqrt{(r)} R[f(\varrho'^2)](\sqrt{r}, \phi/2)
\end{aligned}
\tag{4}
$$

Where $R[f(x, y)](r, \phi)$ denotes the classical Radon transform that maps $f(x, y)$ to $R(r, \phi)$. We can convert $f(x, y)$ to $f(\varrho, \theta)$ in polar coordinate system. Then the classical Radon transform of $f(\varrho, \theta)$ can be obtained by,

$$
R(r, \phi) = \int \int f(\varrho, \theta) \delta(\varrho \cos(\theta - \phi) - r) \varrho \, d\varrho \, d\theta
\tag{5}
$$

In general the GRT of a function f can be calculated via Fourier series and defined by,

$$
\begin{aligned}
GR_d(r, \phi) = 2\sum_{n=-\infty}^{+\infty} [\int_r^\infty \int f(\varrho, \theta) e^{-in\theta} d\theta \\
\times (1 - (r/\varrho)^{2/d})^{-1/2} \times T_{nd}((r/\varrho)^{1/d}) d\varrho] e^{in\theta}
\end{aligned}
\tag{6}
$$

3. 1D-WT is applied to $GR_2(r, \phi)$ with respect to r to get ripplet-II coefficients.

3.3 Feature reduction using PCA+LDA

LDA has been effectively associated with different classification problems like face recognition, speech recognition, cancer detection, multimedia information retrieval, etc [50]. LDA searches for vectors in underlying space that best discriminate among classes rather than data. More precisely, data is described by many independent features; LDA finds a linear combination of these that leads to a largest mean difference between classes [23]. The fundamental target of LDA is to discover projection F that boosts the proportion of between class scatter S_b to within class scatter S_W.

PCA usually deals with orthogonal projection of data which are in higher dimension to lower dimensional linear space [4]. This mapping maximizes the variance of the projected data. The problem arises if the number of samples are much lesser than the original feature space. This issue is called small sample size (SSS) problem. To resolve this problem, PCA method is utilized to the feature vector before LDA [37].

In our work, the original N-dimensional space is projected onto an intermediate $X - 1$ dimensional space using PCA, where X is the total number of samples. Then, LDA is connected to project $X - 1$ dimensions onto M-dimensional space. The feature matrix along with a vector holding the class labels of all samples is fed as an input to the classifier.

$$
RFV = (F_{PCA})_{LDA}
\tag{7}
$$

Where, F is the input feature vector, and RFV is the resultant reduced feature vector after applying the PCA+LDA technique.

The process of finding the suitable number of features is described as follows. At first, we arrange the eigenvalues of the features in decreasing order. Thereafter, a measure called

the normalized cumulative sum of variances (NCSV) corresponding to each feature is calculated. The NCSV value for j^{th} feature is defined as

$$NCSV(j) = \frac{\sum_{u=1}^{j} \alpha(u)}{\sum_{u=1}^{N} \alpha(u)} \qquad ; \ 1 \leq j \leq N \tag{8}$$

where, $\alpha(u)$ represents the eigenvalue of the uth feature and N denotes the dimensionality of the feature vector. Finally, the NCSV value that surpasses the threshold is considered.

3.4 Classification using LS-SVM

Support vector machine (SVM) has been generally utilized as a part of different pattern recognition tasks for classifications. Hence, the SVM is trained to perform facial expression classification using ripplet-II features. In general, SVM constructs a hyperplane to isolate the high-dimensional space. SVM is found to be the most suitable classifier for orientation classification by many researchers [13, 31] and variations to SVM like cost sensitive SVM (CS-SVM) [34], one class SVM (OC-SVM) [2], asymmetric SVM (ASVM) [47], risk area SVM (RA-SVM) [26], least square SVM (LS-SVM) [36] have also been derived which bear special ability.

In general, SVM is used for classification of two class problem [7]. SVM normally works on the following idea: the feature vectors are non-linearly mapped to a higher dimensional space, called a feature space. Then, a linear hyperplane is constructed in this feature space for two-class classification problem. Originally, SVM is a binary classifier. To solve the multiclass problem we need deploy a multiple numbers of support vector machines. However, we can take the one-against-all (OAA) strategy to perform the multi-class classification. OAA method constructs k binary classifiers, where each classifier is trained using the data from two classes. Given training data of N samples $\{x_i, y_i\}_{i=1}^{N}$. where, $x_i \in R^n$ is the ith input sample and $y_{i \in R}$ is the ith output sample.

Classical SVM results to high computational complexity when it deals with large dimension samples. To reduce the computational overhead, least squares support vector machine (LS-SVM) is harnessed as a classifier in this work. In LS-SVM the solutions came out by solving a set of linear equations instead of quadratic equations as in classical SVM's. It is mainly due to the formulation of LS-SVM includes equality type constraints.

$$\min_{\omega, b, \ell} J(\omega, b, \ell) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^{n} \ell_i^2 \tag{9}$$

Subject to equality constraint

$$y_i [\omega^T \varphi(x_i) + b] - 1 + \ell_i, \quad i = 1, 2, ..., N \tag{10}$$

Where $\varphi(.)$ is the mapping function which maps input space into higher dimensional space, ω is the weight vector, ℓ is the error, and γ is the regularization factor.

The Lagrangian is defined as

$$L(\omega, b, \ell; \alpha) = J(\omega, b, \ell) - \sum_{i=1}^{N} \alpha_i y_i [\omega^T \varphi(x_i) + b] - 1 + \ell_i \tag{11}$$

The conditions for optimality are as follows,

$$\frac{\partial L}{\partial \omega} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i \varphi x_i]$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \ell_i} = 0 \rightarrow \alpha_i = \gamma \ell_i \quad i = 1, 2, ..., N$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i [\omega^T \varphi(x_i) + b] - 1 + \ell_i, \quad i = 1, 2, ..., N \tag{12}$$

The conditions for optimality are used to solve the following set of linear equations,

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ A & Y & I & 0 \end{bmatrix} \begin{bmatrix} \omega \\ b \\ \ell \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{1} \end{bmatrix} \tag{13}$$

where, $A = [\varphi(x_1)^T y_1, \varphi(x_2)^T y_2, ..., \varphi(x_N)^T y_N]$, $Y = [y_1, y_2, ..., y_N]$, $\mathbf{1} = [1, 1, ..., 1]$, $\ell = [\ell_1, \ell_2, ..., \ell_N]$, $\alpha = [\alpha_1, \alpha_2, ..., \alpha_N]$.

The solution can also be given by,

$$\begin{bmatrix} 0 & -Y^T \\ Y & AA^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \tag{14}$$

where, $\Omega = AA^T$, The meher's condition states that,

$$\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j) \qquad = y_i y_j \psi(x_i, x_j) \tag{15}$$

Where, x_i is the training data mapped to a higher dimensional space by a function φ. Several experiments were conducted with different classifiers and with different kernels of LS-SVM namely linear, polynomial, and RBF. However, we choose the SVM with RBF kernel for its superior classification performance.

The LS-SVM classifier function can be obtained as,

$$f(x) = sign \left[\sum_{i=1}^{N} \alpha_i y_i \kappa(x, x_i) + b \right] \tag{16}$$

where, $\kappa(., .)$ is the kernel function.

The different kernel functions used for training the LS-SVM classifier is demonstrated in Table 2. The parameters θ indicates the degree of polynomial and σ used to control the shape of the kernel.

Table 2 Kernel functions used by LS-SVM

Kernel function	Definiton
Radial Basis Function (RBF)	$\kappa(x, x_i) = exp(-\gamma \|x - x_i\|^2)$
Linear	$\kappa(x, x_i) = x_i^T x$
Polynomial	$\kappa(x, x_i) = (x_i^T x + 1)^\theta$

Table 3 Samples from each emotion type for CK+ and JAFFE dataset

	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral	Contempt	All
CK+	45	59	25	69	28	83	123	18	450
JAFFE	30	29	32	31	31	30	30	–	213

4 Experimental results and discussion

4.1 Dataset

Two benchmark datasets, CK+ [21] and JAFFE [22] were used to evaluate the proposed method. The number of image sequences for each emotions of two datasets are listed in Table 3. The Cohn-Kanade database contains image sequences of both male and female for eight emotions. The CK+ dataset contains the facial expressions of 210 adults. The individuals were 18 to 50 years of age, among them 81% Euro-American, 69% female, 13% Afro-American, and 6% diverse groups. We have 593 posed facial expressions from 123 subjects. Among 593 posed facial expressions 327 were labeled with seven essential emotion categories. In our experiment, we choose the last image from each image sequence in which the expressions are at its the peak intensity. The number of face images for each expression varies according to its availability. In our experiment on CK+ dataset we have used 450 images from eight emotions in total: anger (45), disgust (59), fear (25), happy (69), sad (28), surprise (83), neutral (123), and contempt (18).

JAFFE dataset consists of 213 images from 10 subjects. All subjects has seven emotion expressions. The seven emotions are anger, disgust, contempt, fear, happy, neutral, sad, and surprise. In this database, each subject has at least 3–4 expression images from each emotion label. In our experiment on JAFFE dataset we have used 213 images from seven emotions in total: anger (30), disgust (29), fear (32), happy (31), sad (31), surprise (30), and neutral (30). The preprocessed sample images from CK+ and JAFFE dataset of the experiment are shown in Figs. 2 and 3, respectively. A detailed description of the datasets is presented briefly in Table 4.

4.2 Feature extraction and reduction

Initially, the face region is detected by Viola and Jones algorithm. Then, the face region is cropped from the original face image and normalized. The ripplet-II transform is utilized to extract features from face expression images. The total number of features extracted by

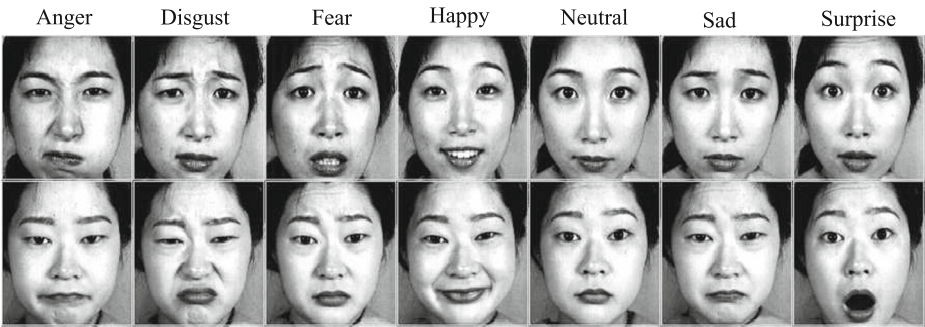**Fig. 2** Preprocessed sample images from CK+ dataset of the experiment

Anger   Disgust   Fear   Happy   Neutral   Sad   Surprise

Fig. 3 Preprocessed sample images from JAFFE dataset of the experiment

ripplet-II coefficients is $128 \times 128 = 16384$, which is quite large for any existing classifier to learn. The PCA+LDA is utilized to reduce the features to merely 6 for CK+ and JAFFE dataset. These are the initial principal components (PCs) with maximum variance.

4.3 Classification comparison

The reduced feature set not only makes the the calculation complexity simple but also makes the classifier more feasible. Five-fold stratified cross-validation (CV) strategy is utilized for the experiments to make the classifier more generalize and stable to independent datasets. The accuracy of the classifier is then evaluated only with the testing image set. Classification accuracy is denoted as the ratio of correctly classified samples to a total number of samples. The performance of the proposed system was measured using CK+ 8 expressions, and JAFFE 7 expressions datasets. The optimal principal components have been determined based on NCSV values. The threshold value is set to 0.9. Table 5 reported different measures that are used to evaluate the performance of the proposed system.

Where *TP* represents true positive, *TN* represents true negative, *FP* represents false positive, and *FN* represents false negative. These parameters for a specific emotion category, say anger are as follows: *TP* (images of 'anger' type classified correctly to the same type), *TN* (images of 'non-anger' type classified correctly as 'non-anger' type), *FP* (images of

Table 4 Description of the datasets: CK+ and JAFFE dataset

Name of the database	# Subjects	# Images	# Expressions	Color model	Size of image	Specialties
CK+	123	10,708	8	RGB	640×480 pixels	●Illumination variation: Yes
				Gray	720×480 pixels	●Distance variation: No
					640×490 pixels	●Pose & tilt variation: Yes
						●Demographic variation: Yes
JAFFE	10	213	7	Gray	256×256 pixels	●Illumination variation: No
						●Distance variation: No
						●Pose & tilt variation: No
						●Demographic variation: No

Table 5 Different classification performance measures

Measures	Definition
Sensitivity	$TP/(TP + FN)$
Specificity	$TP/(TP + FN)$
Precision	$TP/(TP + FP)$
F-Measure	$2TP/(2TP + FP + FN)$
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$

'non-anger' type classified wrongly as 'anger' type), *FN* (images of 'anger' type classified wrongly as 'non-anger' type). Therefore, *TP* and *TN* represented as correctly classified samples and *FP* and *FN* represented as misclassified samples.

Performance measure

– **Class Specific Accuracy**: Class specific accuracy is an important performance metric in multi-class classification. It is defined as proportion of correctly classified instances out of all the instances for classification in a specific class. The formula for class specific accuracy is given by

$$Accuracy = \frac{TP}{TP + FP} \tag{17}$$

where,

TP: True Positive
FP: False Positive

Other than *TP* and *FP*, subsequent parameters uses the notation *TN* and *FN* where,

TN: True Negative
FN: False Negative

– **Balanced Accuracy**: A class specific accuracy may largely vary from class to class and to evaluate a classifier for overall classification over a dataset, balanced accuracy is used and is given by the arithmetic mean of class specific accuracies.

$$Balanced\ accuracy = \frac{\sum_{all\ classes} Class\ specific\ accuracy}{Number\ of\ classes} \tag{18}$$

– **Sensitivity**: Sensitivity is the proportion of correctly classified positive instances out of all instances in a class. Sensitivity is an important measure in binary classification, in multi-class classification they are equivalent to class specific accuracy. It is also known as recall or True Positive Rate (TP Rate) and is given by

$$Sensitivity = \frac{TP}{TP + FN} \tag{19}$$

In multi-class classification, as in the proposed case the corresponding information is given by average sensitivity, which is arithmetic mean of all positive cases that were correctly identified and equivalent to balanced accuracy.

$$Average\ Sensitivity = \frac{\sum_{all\ classes} Sensitivity}{Number\ of\ classes} \tag{20}$$

- **Precision**: Precision is the proportion of correctly classified positive instances out of all instances classified to that particular class and is given by

$$Precision = \frac{TP}{TP + FP} \tag{21}$$

A more suitable parameter in case of multi-class classification is the average precision which is defined as the arithmetic mean of precisions for each of the classes. It is given by

$$Average\ Precision = \frac{\sum\limits_{all\ classes} Precision}{Number\ of\ classes} \tag{22}$$

- **F Measure**: F Measure is the harmonic mean of precision and sensitivity. This is a measure to judge the accuracy of a classifier. With equal weight to both precision and sensitivity, it is more specifically called as F_1 Measure.

$$F_1 Measure = \frac{2.Precision.Sensitivity}{Precision + Sensitivity} \tag{23}$$

In multi-class classification, F Measure of classification can be defined with a more suitable parameter called average F Measure which is given by:

$$Average\ F_1\ Measure = \frac{\sum\limits_{all\ classes} F\ Measure}{Number\ of\ classes} \tag{24}$$

- **False Positive Rate**: FP Rate is the proportion of positive cases that were incorrectly identified and is given by

$$FP\ Rate = \frac{FP}{FP + TN} \tag{25}$$

To find False Positive Rate in multi-class classification, it is split into multiple *1:rest* binary classes and average False Positive Rate is given by:

$$Average\ FP\ Rate = \frac{\sum\limits_{all\ classes} FP\ Rate}{Number\ of\ classes} \tag{26}$$

- **Error Rate**: Error rate is given by $(1 - Class\ specific\ accuracy)$. In the proposed multi-class classification, error rate is presented as percentage of incorrectly classified instances and is given by the average of misclassification for each individual classes.

$$Average\ Error\ Rate = \frac{\sum\limits_{all\ classes} Error\ Rate}{Number\ of\ classes} \tag{27}$$

The performance of PCs with the accuracy on CK+ and JAFFE dataset is presented in Fig. 4. It is observed that the proposed system used only 6 feature to achieved highest classification accuracy with LS-SVM and RBF kernel for both CK+ and JAFFE datasets. The resultant 6 features along with all the class labels are combined to form a resultant dataset. The performance of the LS-SVM with kernels like linear, polynomial, and RBF are calculated along with KNN, RF, and BPNN classifiers. Several experiments are carried out to find the optimal parameter settings of these classifiers. The parameters with lowest error rate are chosen to train these classifiers. Table 6 shows the parameter settings for LS-SVM RBF, LS-SVM Poly, KNN, RF, and BPNN classifiers.

There are two tunable parameters C and γ, while training of LS-SVM with RBF kernel. The kernel parameter γ controls the shape of the kernel and the regularization parameters C controls the margin maximization and error minimization. Initially, the values of C and
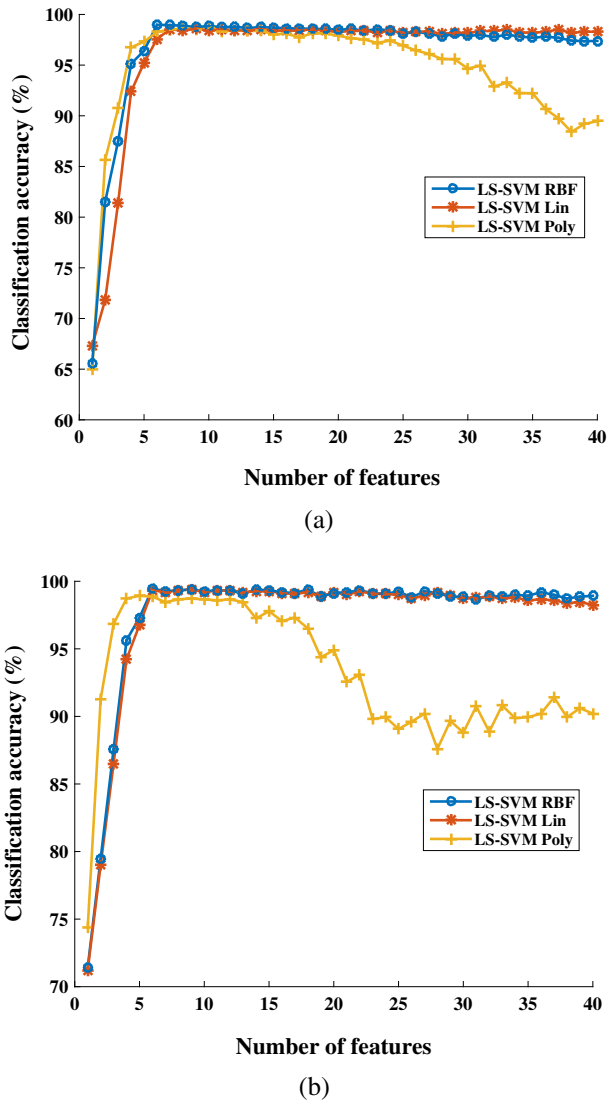
Fig. 4 Classification performance with respect to number of features over two datasets: **a** CK+. **b** JAFFE

γ are unknown. We need to find the best combination of C and γ values which give the classification results for a given problem. Therefore, various combination of C and γ are tired over the CV procedure. The values of C and γ with least CV error rate is picked for training the classifier, where C and $\gamma \in [1, 10]$.

The performance of LS-SVM classifier with RBF, linear and polynomial kernels on CK+ dataset are presented in Tables 7, 8, and 9. The classification performances with three different kernels such as RBF, linear, and polynomial over JAFFE dataset are reported in Tables 10, 11, and 12 respectively. Better classification accuracy is achieved with LS-SVM and RBF kernel as compared to linear and polynomial, which is evident from Tables 7 and 10.

Table 6 Parameter settings for different classifiers

Classifier	Parameter	Value
KNN	Nearest neighbors (K)	1
	Nearest neighbor search algorithm	Euclidean distance
	Number of folds	5
RF	Number of trees	100
	Seed	1
	Number of folds	5
BPNN	Learning rate, momentum	0.05, 0.4
	Hidden neurons	6 (JAFFE), 7 (CK+)
	Number of folds	5
LS-SVM Poly	Order	2
	Kernel	Poly
	Number of folds	5
LS-SVM RBF	C, γ	$[1-10]$
	Kernel	RBF
	Number of folds	5

Table 7 Performance of LS-SVM with RBF kernel on CK+ 8 expressions

Average performances: Sensitivity = 97.00%, Specificity = 99.00%, Accuracy = 98.97%

Emotions	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	Accuracy(%)
Anger	45	399	6	0	100	99.00	98.67
Contempt	16	425	7	2	89.00	98.00	98.00
Disgust	58	391	0	1	98.00	100	99.78
Fear	25	425	0	0	100	100	100
Happy	69	381	0	0	100	100	100
Sad	27	419	3	1	96.00	99.00	99.11
Surprise	82	367	0	1	99.00	100	99.78
Neutral	121	313	14	2	98.00	96.00	96.44

Table 8 Performance of LS-SVM with Linear kernel on CK+ 8 expressions

Average performances: Sensitivity = 96.00%, Specificity = 98.00%, Accuracy = 97.58%

Emotions	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	Accuracy(%)
Anger	43	399	6	2	96.00	99.00	98.22
Contempt	16	422	10	2	89.00	98.00	97.33
Disgust	57	391	0	2	97.00	100	99.56
Fear	25	425	0	0	100	100	100
Happy	69	381	0	0	100	100	100
Sad	27	418	4	1	96.00	99.00	98.89
Surprise	82	367	0	1	99.00	100	99.78
Neutral	113	278	49	10	92.00	85.00	86.89

Table 9 Performance of LS-SVM with Polynomial kernel on CK+ 8 expressions

Emotions	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	Accuracy(%)
Anger	45	393	12	0	100	97.00	97.33
Contempt	16	422	10	2	89.00	98.00	97.33
Disgust	58	387	4	1	98.00	99.00	98.89
Fear	25	425	0	0	100	100	100
Happy	69	381	0	0	100	100	100
Sad	25	414	8	3	89.00	98.00	97.56
Surprise	82	367	0	1	99.00	100	99.78
Neutral	120	309	18	3	98.00	94.00	95.33

Average performances: Sensitivity = 97.00%, Specificity = 98%, Accuracy = 98.28%

Table 10 Performance of LS-SVM with RBF kernel on JAFFE 7 expressions

Emotions	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	Accuracy(%)
Anger	30	183	0	0	100	100	100
Disgust	29	183	1	0	100	99.00	99.53
Fear	32	178	3	0	100	98.00	98.59
Happy	30	181	1	1	97.00	99.00	99.06
Neutral	30	181	2	0	100	99.00	99.06
Sad	31	182	0	0	100	100	100
Surprise	30	183	0	0	100	100	100

Average performances: Sensitivity = 100%, Specificity = 99%, Accuracy = 99.46%

Table 11 Performance of LS-SVM with Linear kernel on JAFFE 7 expressions

Emotions	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	Accuracy(%)
Anger	30	183	0	0	100	100	100
Disgust	29	183	1	0	100	99.00	99.53
Fear	32	178	3	0	100	98.00	98.59
Happy	30	181	1	1	97.00	99.00	99.06
Neutral	30	181	2	0	100	99.00	99.06
Sad	31	180	2	0	100	99.00	99.06
Surprise	30	183	0	0	100	100	100

Average performances: Sensitivity = 100%, Specificity = 99.00%, Accuracy = 99.33%

Table 12 Performance of LS-SVM with Polynomial kernel on JAFFE 7 expressions

Emotions	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	Accuracy(%)
Anger	29	183	0	1	97	100	99.53
Disgust	29	183	1	0	100	99.00	99.53
Fear	32	175	6	0	100	97.00	97.18
Happy	31	181	1	0	100	99.00	99.53
Neutral	29	180	3	1	97.00	98.00	98.12
Sad	30	179	3	1	97.00	98.00	98.12
Surprise	30	183	0	0	100	100	100

Average performances: Sensitivity = 99.00%, Specificity = 99.00%, Accuracy = 98.86%

Table 13 Performance matrices of different classifiers (CK+ 8 expressions)

Classifier	Sensitivity(%)	Specificity(%)	Accuracy(%)
LS-SVM RBF	98.00	99.00	98.97
LS-SVM Linear	96.00	98.00	97.58
LS-SVM Poly	97.00	98.00	98.28
KNN	95.00	98.00	95.77
RF	94.00	98.00	94.66
BPNN	96.00	99.00	96.00

Table 14 Performance matrices of different classifiers (JAFFE 7 expressions)

Classifier	Sensitivity(%)	Specificity(%)	Accuracy(%)
LS-SVM RBF	100	99.00	99.46
LS-SVM Linear	100	99.00	99.33
LS-SVM Poly	99.00	99.00	98.86
KNN	96.00	99.00	96.24
RF	96.00	99.00	96.71
BPNN	95.00	99.00	95.77

Table 15 Comparison of proposed system with existing schemes on CK+ database

Approaches	Facial feature	Classifier	Accuracy(%)
Zhang et al. [51] (2011)	Patch based Gabor	SVM	94.48
Wang et al. [46] (2013)	HOG + WLD	KNN	93.97
Siddiqi et al. [32] (2013)	Wavelet + LDA	HMM	94.00
Hsieh et al. [17] (2015)	Semantic features	ASM + RBF SVM	94.70
Mlakar et al. [25] (2015)	HOG difference vector	SVM	95.64
Happy et al. [16] (2015)	Salient facial patches	RBF SVM	94.09
Siddiqi et al. [33] (2015)	SWLDA	HCRF	96.83
Uccar et al. [39] (2016)	Local Curvlet transform	OSELM-SC	95.17
Proposed method	Ripplet-II + PCA + LDA	LS-SVM RBF	98.97

Table 16 Comparison of proposed system with existing schemes on JAFFE database

| Approaches	Facial feature	Classifier	Accuracy(%)
Zhang et al. [51] (2011)	Patch based Gabor	SVM	92.93
Kazmi et al. [20] (2012)	2D-DWT	SVM	96.00
Wang et al. [46] (2013)	HOG+WLD	KNN	95.86
Mlakar et al. [25] (2015)	HOG difference vector	SVM	87.82
Happy et al. [16] (2015)	Salient facial patches	RBF SVM	91.79
Siddiqi et al. [33] (2015)	SWLDA	HCRF	96.33
Guo et al. [15] (2016)	K-ELBP	SVM	93.3
Uccar et al. [39] (2016)	Local Curvlet transform	OSELM-SC	94.65
Proposed method	Ripplet-II + PCA + LDA	LS-SVM RBF	99.46
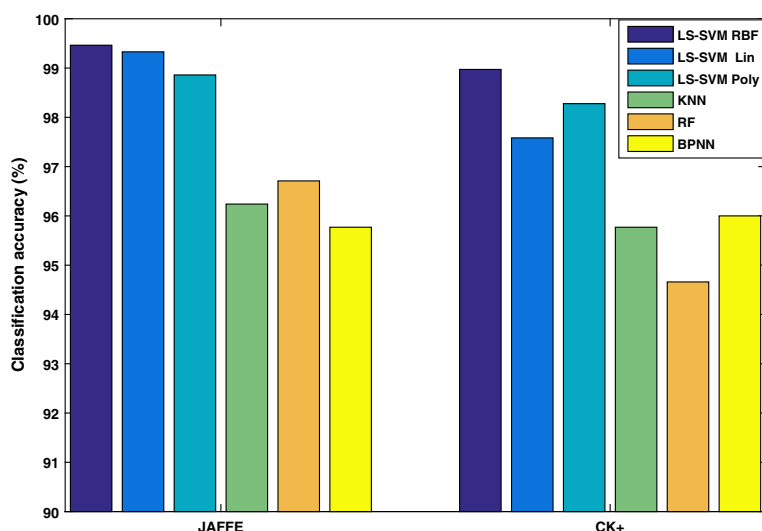
Fig. 5 Classification accuracy with JAFFE and CK+ dataset

The average sensitivity, precision, F-measure, and accuracy of the proposed method with five fold stratified CV is reported in Tables 13 and 14 on CK+ and JAFFE dataset respectively. The result shows that the proposed method perfectly worked on JAFFE dataset. For CK+ dataset the proposed method achieves average sensitivity of 0.97, specificity of 0.99, and accuracy of 98.97%. However, the proposed system yields average sensitivity of 1, specificity of 0.99, and accuracy of 99.46% on JAFFE dataset.

4.4 Comparision with state-of-the-art methods

It may be observed that the classification performance achieved by LS-SVM with RBF kernel is better as compared to other classifiers with same number of features. The proposed scheme has been compared with other existing competent schemes in terms of classification accuracy and the results for both the dataset are listed in Tables 15 and 16. From the tables, it is seen that our scheme yields better accuracy as compared to other schemes on both the datasets.

Figure 5 depicts the recognition accuracy of the proposed method with RBF, linear, and polynomial kernels of LS-SVM along with KNN, RF, and BPNN classifier. It is evident that the RBF kernel leads to better classification accuracy as compared to the other classifiers with same feature set.

5 Conclusion

In this work, we proposed an efficient system based on ripplet-II transform for face expression recognition. Initially, Viola and Jones face detection algorithm is harnessed to detect the face region from the input image. The detected face region is then cropped and resized to maintain a uniform dimension of all the input image samples. Thereafter, ripplet-II transform to extract features from the preprocessed face image followed by a PCA+LDA

approach to find the discriminative features from the high dimensional ripplet-II features. Finally, LS-SVM is applied to perform classification. Simulation results on two benchmark datasets namely JAFFE and CK+ to demonstrate the superiority of the proposed approach over other state-of-the-art methods on datasets, respectively. The proposed scheme achieves recognition accuracy of 98.97% and 99.46% on CK+ and JAFFE datasets respectively. This surely advocates the strength of the proposed methodology. However, the method is tested on images captured in constrained environment with controlled lighting conditions, thus not hindering to the performance of the proposed system. Application on real world images may partially affect the performance of the system. In future, other powerful feature extraction and representation tools, including deep learning, can also be utilized to improve the overall performance of the FER system, or toward making the FER system work in real-time.

Abbreviations FER, Face expression recognition Ripplet-II, Ripplet transform type II CK+, Extended Cohn-Kanade JAFFE, Japanese female facial expression PCA, Principal component analysis LDA, Linear discriminant analysis HOG, Histograms of Oriented Gradient DWT, Discrete wavelet transform SWT, Statyionary wavelet transform RBF, Radial basis function SVM, Support vector machine LS-SVM, Least square support vector machine

References

1. Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J (2006) Fully automatic facial action recognition in spontaneous behavior. In: 7th international conference on automatic face and gesture recognition. IEEE, pp 223–230. https://doi.org/10.1109/FGR.2006.55
2. Ben-Hur A, Horn D, Siegelmann HT, Vapnik V (2002) Support vector clustering. J Mach Learn Res 2:125–137
3. Bettadapura V (2012) Face expression recognition and analysis: the state of the art. arXiv:1203.6722
4. Bishop CM (2006) Pattern recognition and machine learning. Springer
5. Chen J, Chen D, Gong Y, Yu M, Zhang K, Wang L (2012) Facial expression recognition using geometric and appearance features. In: Proceedings of the 4th international conference on internet multimedia computing and service. ACM, pp 29–33. https://doi.org/10.1145/2382336.2382345
6. Chen J, Chen Z, Chi Z, Fu H (2014) Facial expression recognition based on facial components detection and HOG features. In: International workshops on electrical and computer engineering subfields
7. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/BF00994018
8. Dahmane M, Meunier J (2011) Emotion recognition using dynamic grid-based HoG features. In: International conference on automatic face & gesture recognition and workshops (FG 2011). IEEE, pp 884–888. https://doi.org/10.1109/FG.2011.5771368
9. Deshmukh S, Patwardhan M, Mahajan A (2016) Survey on real-time facial expression recognition techniques. IET Biom 5(3):155–163. https://doi.org/10.1049/iet-bmt.2014.0104
10. Dongcheng S, Jieqing J (2010) The method of facial expression recognition based on DWT-PCA/LDA. In: International congress on image and signal processing (CISP), vol 4. IEEE, pp 1970–1974. https://doi.org/10.1109/CISP.2010.5648166
11. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Pers Soc Psychol 17(2):124. https://doi.org/10.1037/h0030377
12. Elaiwat S, Bennamoun M, Boussaid F (2016) A spatio-temporal RBM-based model for facial expression recognition. Pattern Recogn 49:152–161. https://doi.org/10.1016/j.patcog.2015.07.006
13. Gandhi T, Trivedi MM (2008) Image based estimation of pedestrian orientation for improving path prediction. In: The proceedings of the IEEE intelligent vehicles symposium, pp 506–511. https://doi.org/10.1049/10.1109/IVS.2008.4621257

14. Gritti T, Shan C, Jeanne V, Braspenning R (2008) Local features based facial expression recognition with face registration errors. In: 8th IEEE international conference on automatic face & gesture recognition, FG'08. IEEE, pp 1–8. https://doi.org/10.1109/AFGR.2008.4813379

15. Guo M, Hou X, Ma Y, Wu X (2016) Facial expression recognition using ELBP based on covariance matrix transform in KLT. Multimedia Tools and Applications 1–16. https://doi.org/10.1007/s11042-016-3282-9

16. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. IEEE Trans Affect Comput 6(1):1–12. https://doi.org/10.1109/TAFFC.2014.2386334

17. Hsieh C-C, Hsih M-H, Jiang M-K, Cheng Y-M, Liang E-H (2015) Effective semantic features for facial expressions recognition using SVM. Multimedia Tools and Applications 1–20. https://doi.org/10.1007/s11042-015-2598-1

18. Jung H, Lee S, Park S, Lee I, Ahn C, Kim J (2015) Deep temporal appearance-geometry network for facial expression recognition. arXiv:1503.01532

19. Kar NB, Babu KS, Jena SK (2017) Face expression recognition using histograms of oriented gradients with reduced features. In: Proceedings of international conference on computer vision and image processing. Springer, pp 209–219. https://doi.org/10.1007/978-981-10-2107-7_19

20. Kazmi SB, Jaffar MA et al. (2012) Wavelets-based facial expression recognition using a bank of support vector machines. Soft Comput 16(3):369–379. https://doi.org/10.1007/s00500-011-0721-4

21. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The Extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: IEEE Computer society conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 94–101. https://doi.org/10.1109/CVPRW.2010.5543262

22. Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with Gabor wavelets. In: Third IEEE international conference on automatic face and gesture recognition. IEEE, pp 200–205. https://doi.org/10.1109/AFGR.1998.670949

23. Martínez AM, Kak AC (2001) PCA Versus LDA. IEEE Trans Pattern Anal Mach Intell 23(2):228–233. https://doi.org/10.1109/34.908974

24. Michel P, El Kaliouby R (2003) Real time facial expression recognition in video using support vector machines. In: Proceedings of the 5th international conference on multimodal interfaces. ACM, pp 258–264. https://doi.org/10.1145/958432.958479

25. Mlakar U, Potočnik B (2015) Automated facial expression recognition based on histograms of oriented gradient feature vector differences. SIViP 1–9. https://doi.org/10.1007/s11760-015-0810-4

26. Moraes D, Wainer J, Rocha A (2016) Low false positive learning with support vector machines. J Vis Commun Image Represent 38:340–350. https://doi.org/10.1016/j.jvcir.2016.03.007

27. Nayak DR, Dash R, Majhi B (2016) Brain mr image classification using two-dimensional discrete wavelet transform and adaboost with random forests. Neurocomputing 177:188–197. https://doi.org/10.1016/j.neucom.2015.11.034

28. Nayak DR, Dash R, Majhi B (2017) Stationary wavelet transform and adaboost with svm based pathological brain detection in mri scanning. CNS Neurol Disord Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders) 16(2):137–149. https://doi.org/10.2174/1871527315666161024142036

29. Qayyum H, Majid M, Anwar SM, Khan B (2017) Facial expression recognition using stationary wavelet transform features. Math Probl Eng 2017. https://doi.org/10.1155/2017/9854050

30. Saeed A, Al-Hamadi A, Niese R, Elzobi M (2014) Frame-based facial expression recognition using geometrical features. Advances in Human-Computer Interaction 2014:4

31. Shimizu H, Poggio T (2003) Direction estimation of pedestrian from images. In: AI Memo 2003-020, Massachusetts Institute of Technology, pp 1–11

32. Siddiqi MH, Lee S (2013) Human facial expression recognition using wavelet transform and hidden markov model. In: International workshop on ambient assisted living. Springer, pp 112–119. https://doi.org/10.1007/978-3-319-03092-0

33. Siddiqi MH, Ali R, Khan AM, Park Y-T, Lee S (2015) Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans Image Process 24(4):1386–1398. https://doi.org/10.1109/TIP.2015.2405346

34. Sun Y, Kamel MS, Wong AKC, Wang Y (2007) Cost-sensitive boosting for classification of imbalanced data. Pattern Recogn 40(12):3358–3378. https://doi.org/10.1016/j.patcog.2007.04.009

35. Suwa M, Sugie N, Fujimora K (1978) A preliminary note on pattern recognition of human emotional expression. In: International joint conference on pattern recognition, pp 408–410

36. Suykens JAK, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300. https://doi.org/10.1023/A:1018628609742

37. Theodoridis S, Koutroumbas K (1999) Pattern recognition. Academic Press, New York

38. Tsai H-H, Lai Y-S, Zhang Y-C (2010) Using SVM to design facial expression recognition for shape and texture features. In: International conference on machine learning and cybernetics, vol 5. IEEE, pp 2697–2704. https://doi.org/10.1109/ICMLC.2010.5580938

39. Uçar A, Demir Y, Güzeliş C (2016) A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. Neural Comput & Applic 27(1):131–142. https://doi.org/10.1007/s00521-014-1569-1

40. Uddin MZ (2016) A depth video-based facial expression recognition system utilizing generalized local directional deviation-based binary pattern feature discriminant analysis. Multimedia Tools and Applications 75(12):6871–6886. https://doi.org/10.1007/s11042-015-2614-5

41. Uddin MZ (2016) Facial expression recognition using depth information and spatiotemporal features. In: 18th international conference on advanced communication technology (ICACT), 2016. IEEE, pp 726–731. https://doi.org/10.1109/ICACT.2016.7423536

42. Uddin MZ, Hassan MM, Almogren A, Alamri A, Alrubaian M, Fortino G (2017) Facial expression recognition utilizing local direction-based robust features and deep belief network. IEEE Access 5:4525–4536. https://doi.org/10.1109/ACCESS.2017.2676238

43. Valstar MF, Pantic M (2012) Fully automatic recognition of the temporal phases of facial actions. IEEE Trans Syst Man Cybern B Cybern 42(1):28–43. https://doi.org/10.1109/TSMCB.2011.2163710

44. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, vol 1, pp I–511. https://doi.org/10.1109/CVPR.2001.990517

45. 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. https://doi.org/10.1016/j.neucom.2017.08.015

46. Wang X, Jin C, Liu W, Hu M, Xu L, Ren F (2013) Feature fusion of HOG and WLD for facial expression recognition. In: International symposium on system integration (SII). IEEE, pp 227–232. https://doi.org/10.1109/SII.2013.6776664

47. Wu SH, Lin KP, Chien HH, Chen CM, Chen MS (2013) On generalizable low false-positive learning using asymmetric support vector machines. IEEE Trans Knowl Data Eng 25(5):1083–1096. https://doi.org/10.1109/TKDE.2012.46

48. Xu J, Wu D (2008) Ripplet transform for feature extraction. In: SPIE defense and security symposium, pages 69700x–69700x. International society for optics and photonics. https://doi.org/10.1117/12.777302

49. Xu J, Wu D (2012) Ripplet transform type II transform for feature extraction. IET Image Process 6(4):374–385. https://doi.org/10.1049/iet-ipr.2010.0225

50. Yu H, Yang J (2001) A direct LDA algorithm for high-dimensional data—with application to face recognition. Pattern Recogn 34(10):2067–2070

51. Zhang L, Tjondronegoro D (2011) Facial expression recognition using facial movement features. IEEE Trans Affect Comput 2(4):219–229. https://doi.org/10.1109/T-AFFC.2011.13

52. Zhang Y-D, Yang Z-J, Lu H-M, 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 4:8375–8385. https://doi.org/10.1109/ACCESS.2016.2628407

Nikunja Bihari Kar is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India. His current research interests include Computer Vision, Machine Learning, and Deep Learning.

Korra Sathya Babu received his PhD degree from National Institute of Technology, Rourkela, India, in 2013. He is currently working as an assistant professor in the Department of Computer Science and Engineering at National Institute of Technology, Rourkela, India. His current research interests include social computing, natural language processing, and data engineering.

Arun Kumar Sangaiah has received his Master of Engineering (ME) degree in Computer Science and Engineering from the Government College of Engineering, Tirunelveli, Anna University, India. He had received his Doctor of Philosophy (PhD) degree in Computer Science and Engineering from the VIT University, Vellore, India. He is presently working as an Associate Professor in School of Computer Science and Engineering, VIT University, India. His area of interest includes software engineering, computational intelligence, wireless networks, bio-informatics, and embedded systems. He has authored more than 100 publications in different journals and conference of national and international repute. His current research work includes global software development, wireless ad hoc and sensor networks, machine learning, cognitive networks and advances in mobile computing and communications. Also, he was registered a one Indian patent in the area of Computational Intelligence. Besides, Prof. Sangaiah is responsible for Editorial Board member / Associate Editor of various international journals.

Sambit Bakshi is currently with Centre for Computer Vision and Pattern Recognition of National Institute of Technology Rourkela, India. He also serves as Assistant Professor in Department of Computer Science & Engineering of the institute. He earned his PhD degree in Computer Science & Engineering in 2015. He serves as associate editor of International Journal of Biometrics (2013 -), Innovations in Systems and Software Engineering (2016 -), IEEE Access (2016 -), and Plos One (2016 -). He is technical committee member of IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence. He received the prestigious Innovative Student Projects Award - 2011 from Indian National Academy of Engineering (INAE) for his master's thesis. He has more than 30 publications in journals, reports, conferences.