

Novel directional patterns and a Generalized Supervised Dimension Reduction System (GSDRS) for facial emotion recognition

A. Sherly Alphonse¹  · Dejey Dharma¹

Received: 19 January 2017 / Revised: 10 July 2017 / Accepted: 22 August 2017 /
Published online: 19 September 2017
© Springer Science+Business Media, LLC 2017

Abstract This paper presents two novel directional patterns, a Maximum Response-based Directional Texture Pattern (MRDTP) and a Maximum Response-based Directional Number Pattern (MRDNP), for recognizing the facial emotions in constrained as well as unconstrained situations. The intensity information obtained from the maximum of the edge responses, after applying eight Kirsch masks, is used for the calculation of facial features in MRDTP. In MRDNP, instead of intensity information, the direction number of the maximum response is used. After dividing MRDNP and MRDTP code images into grids, feature vectors are created from the concatenated histograms obtained from the grids. This paper also proposes an effective Generalized Supervised Dimension Reduction System (GSDRS) and uses Extreme Learning Machine with Radial Basis Function (ELM-RBF) classifier for rapid and efficient classification of emotions. Both the proposed patterns are more effective than the existing ones in removing random noise and providing good structural information using prominent edges which help to achieve high classification accuracy when tested with seven datasets.

Keywords Emotion · dimension reduction · feature extraction · ELM · classification

1 Introduction

Facial expression recognition is inevitable nowadays in social networking, fraud detection by Police department and psychological studies. Emotions are better conveyed by human face expressions, and about seven emotions are vital while dealing with human faces [57, 61].

✉ A. Sherly Alphonse
sherls82@gmail.com

Dejey Dharma
dejeytilak@gmail.com

¹ CSE Department, Regional Campus, Anna University, Tirunelveli 627007, India

Ekman and Friesen [22, 23] have classified human emotions as Happiness, Sadness, Anger, Disgust, Fear and Surprise after experimenting with adults and children. A good feature extraction technique should be capable of extracting the exact features for facial expression recognition. It should also be robust to noise, illumination, pose and several transformations on the face. The best feature descriptor should have a simple way of extraction. It should be compact and also be very low in dimension to reduce the classification time. The feature descriptors should also produce excellent results under constrained as well as unconstrained environments.

Considering all these factors, the contributions of the proposed work in facial expression recognition are summarized as follows:

- Two very compact and robust feature extraction techniques are proposed. The first feature extraction technique Maximum Response-based Directional Texture Pattern (MRDTP) is based on the intensity information of the maximum response of each pixel. The second feature extraction technique Maximum Response-based Directional Number Pattern (MRDNP) is based on the direction number of the maximum response of each pixel. The proposed feature descriptors are compared with existing directional patterns using compass masks, so as to prove their suitability in classifying facial expression under varying illumination, noise and poses.
- Random noise restraining step is included in MRDTP, which is not present in the other existing directional patterns.
- Performance of the proposed techniques is tested under constrained and unconstrained environments. Their superiority over other existing techniques under unconstrained conditions where there are high variations in scaling, rotation and illumination is also highlighted.
- A Generalized Supervised Dimension Reduction System (GSDRS) based on Pearson General Kernel (PGK) is introduced for reducing the time involved in the optimization procedure, with the selection of kernel. Also, the use of Extreme Learning Machine with Radial Basis Function (ELM-RBF) classifier for a fast and accurate classification is proved with experimental results.

The paper is organized as follows. The related works and motivational factors are explained in Section 2. A detailed description of the proposed method is presented in Section 3. Experimental results obtained are discussed in Section 4. Conclusion and considerations for future works are given in Section 5.

2 Related works and motivation

This work has three stages while classifying emotions from the detected face. The three stages are (i) Feature extraction (ii) Dimension reduction and (iii) Classification. A survey on the related works is presented here.

2.1 Feature extraction techniques

Feature extraction techniques are grouped as geometric information-based features and appearance-based features.

2.1.1 Geometric information- based features

These types of features are based on the shape information of face. Kotsia and Pitas [39] have used shape information of face to place landmarks. The grids on the landmarks are used to find the displacement between two frames and thus for emotion recognition. Bourbakis et al. [12] have extracted the meta-features from face. Berretti et al. [10] have utilized SIFT features on landmarks to extract the information from face. These types of landmarks based feature extractions are heavy in computation. Anisetti et al. [6] have used Facial Action Coding System (FACS) and Russel's circumplex model, which is unable to accurately track the shape information. The changes in the shapes of the face can be easily captured by Histogram of Oriented Gradients (HOG) that creates high dimensional feature vectors and also takes more time for extraction [16]. The proposed patterns exhibit very low time complexity and feature vector dimension compared to SIFT and HOG.

2.1.2 Appearance- based features

Appearance-based features are either extracted from the whole face or from the individual components of face which are later combined to form a single feature vector. The features which are extracted from the face as a whole are called holistic, and the features extracted from the components of face are called component-based.

The holistic methods for feature extraction are mainly PCA-based like, Kernel Principal Component Analysis + Linear Discriminant Analysis (KPCA + LDA) [70], Two Dimensional Principal Component Analysis (2DPCA) [69] and Eigen faces [58]. The local descriptors are more robust to illumination and pose variations when compared to these methods. Among the component-based methods, an Emotion Avatar Image is created by using LBP and Local Phase Quantization (LPQ) features by Yang et al. [68]. Gabor [1, 8, 41, 66] has achieved good recognition rate in facial emotion recognition applications. But the high dimension of feature vector restricts its usage. Gabor features also capture the edges from all the orientations as well as in the noisy regions. Local Binary Pattern (LBP) [54] is commonly used in literature, but is very sensitive to noise and non-monotonic illumination variations. Local Phase Quantization (LPQ) [46], Pyramid Local Phase Quantization (PLPQ) [63], Local Ternary Pattern (LTP) [56], Local Principal Texture Pattern (LPTP) [50], Gradient Local Ternary Patterns (GLTP) [3], Elastic Bunch Graph [11, 25] and Dual Tree-Complex Wavelet Transforms (DT-CWT) [55] are some other component-based feature extraction techniques. The drawback of these methods is that they are sensitive to the grayscale transformations of pixels.

Various direction-based feature extraction techniques exist in literature, to overcome the drawbacks of LBP like, Local Directional Pattern (LD_iP) [33], Local Directional Number Pattern (LDN) [49], Local Directional Texture Pattern (LDTP) [52], Directional Ternary Pattern (DTP) [4], Local Sign Directional Pattern (LSDP) [14], Local Gaussian Directional Pattern (LGDP) [51] and Directional Binary Codes (DBC) [71] which create robust codes compared to LBP because they use the information from eight directions around a pixel. The information used by the directional patterns is more stable compared to the pixel intensities used by LBP. Among the various direction-based feature extraction techniques, LDTP is a more effective feature extraction technique that provides excellent results for facial expression recognition than LD_iP and LDN because of its ability to code both the prominent direction information as well as the intensity information. In LBP method of feature extraction, only sparse points are used. Even though MRDTP and MRDNP utilize all the eight directions of

neighborhood, they are more compact than LBP. The proposed patterns are also more compact than the existing directional patterns like LDTP and LDN by encoding any one among the two types of information (i.e., either pixel intensity or direction information). Also, the existing directional patterns still suffer from some of the random noise within the edges. Both the proposed patterns are robust to noise as they use only the maximum response information and eliminate the noisy and redundant information obtained from other responses. This makes them superior to other existing directional patterns as they use only the information needed for emotion recognition. In MRDTP, the noise restraining process further limits the edges from random noise and retains only the prominent edges of the face, thus improving accuracy. Compared to Gabor, the proposed patterns create a very low dimension feature vector with good structural information.

Most of the existing component-based feature descriptor techniques provide good accuracy when used under constrained environment. But their recognizing capability decreases considerably under scaling, rotation and illumination variations. However, while capturing face images with camera under unconstrained environments, the feature descriptor representing the image should be robust to scaling, rotation and lighting variations. Both MRDTP and MRDNP are scaling-invariant because of the histogram-based feature vector creation technique used. They both use the maximum of the responses obtained using eight directional masks which make them rotation-invariant. They remove the illumination artifacts by using grids in feature vector formation. These make MRDTP and MRDNP perform well under constrained and unconstrained situations.

2.2 Dimension reduction

Reducing the dimension of the feature vector obtained using MRDTP and MRDNP will minimize the time and memory requirements. It also improves the efficiency of the machine learning algorithm used for the classification purpose. Discriminant Laplacian Embedding (DLE) [64], Principal Component Analysis (PCA) [19], Linear Discriminant Analysis (LDA) [13] and Locality Preserving Projection (LPP) [53] are the various dimension reduction techniques existing in literature. DLE involves large number of computations, and PCA is commonly used for directional pattern-based feature extraction techniques, but it incorporates the variations due to the lighting conditions, while reducing dimensions. In LDA, the projection matrix depends on S_W^{-1} (S_W is the scatter within classes) which is not present in small number of training samples and LPP is sensitive to noise. GDA generalizes the dimension reduction as a non-linear mapping technique by selecting only the discriminative features, and the projection matrix is not dependent on S_W^{-1} . Generalized Discriminant Analysis (GDA) has been already experimented with Gabor features in some of the existing works. It has also not been applied for directional patterns in literature. GDA [9] considerably reduces the dimension, but the drawback is that it spends much time in the selection of the kernel. This demands the need for a dimension reduction system that suggests the suitable kernel for achieving the best dimension reduction. The Pearson VII function has been used as a universal kernel for SVM to achieve good classification rate in a work proposed by Üstün et al. [59]. But that kernel has not been used in any other dimension reduction systems. This paper proposes a Pearson VII function-based Generalized Supervised Dimension Reduction System (GSDRS) which is an inspirational work from GDA that completely eliminates the need to experiment with the other kernels for each dataset. The

Pearson VII function works as a single general kernel PGK that performs good dimension reduction for MRDTP and MRDNP irrespective of the nature of the datasets used.

2.3 Classifiers used

Support Vector Machine (SVM) [17, 18], Convolutional Neural Network (CNN) [28], K-Nearest Neighbor (KNN) [24], and Deep learning [36] are the classification techniques that have been used for emotion recognition from face in literature. SVM, CNN and Deep learning algorithms produce good classification accuracy compared to KNN. The disadvantage is that they consume considerable training time. But the classification algorithm used for emotion recognition should be fast having better generalization performance so that it could be used in real-time environments. Iosifidis et al. [32] have applied Extreme Learning Machine (ELM) on JAFFE and CK datasets. Because of the milder constraints in optimization as well as the rapidness, ELM is chosen as the base classifier in the experiments carried out here.

3 The proposed system

The complete map of the proposed method is illustrated in Fig. 1. Facial expression images from datasets are given as input to the face detector. From it, the cropped images are given as input for MRDTP or MRDNP to form a feature vector. This feature vector is then given to the proposed GSDRS which is based on PGK. This reduces the feature vector to size $N-1$, where N is the number of emotion categories. If seven emotion categories are considered for classification, the dimension is reduced to six. GSDRS is a motivation from Generalized Discriminant Analysis (GDA) proposed by Baudat and Anouar [9]. PGK acts as a substitute to all other kernels stated in literature and used with GDA to form the proposed GSDRS. PGK saves the time for selecting the best kernel among the existing kernels while reducing dimensions. This produces good results because the testing samples are reduced in dimension using the discrimination analysis on training samples and is explained in detail in Section 3.3. Then, the reduced feature vectors are classified using ELM-RBF which is faster than RBF kernel-based SVM. It classifies emotions into anyone of the categories, namely Anger, Fear, Disgust, Happiness,

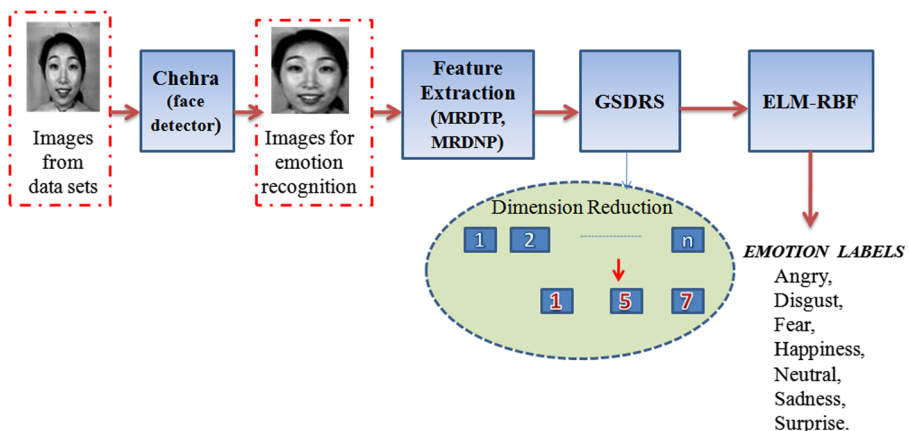


Fig. 1 Complete map of the proposed approach

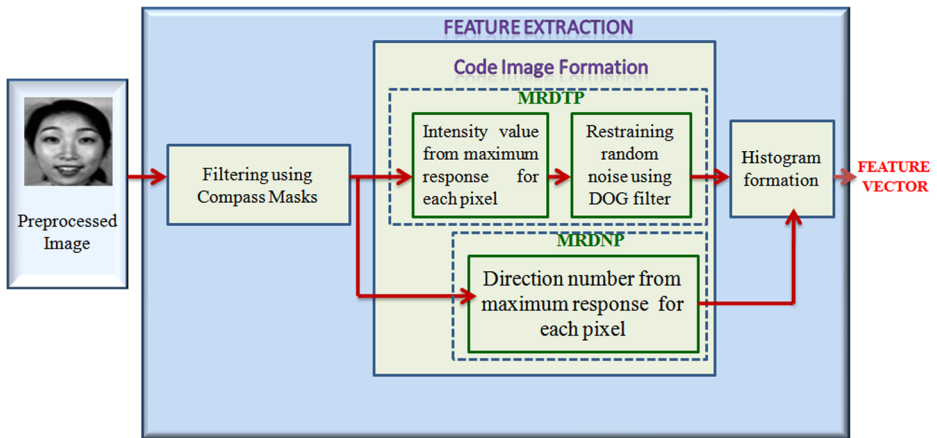


Fig. 2 Steps in the feature extraction of MRDTP and MRDNP

Neutral, Sadness, Surprise. Fig. 2 explains the procedural steps within the proposed feature extraction techniques.

3.1 Preprocessing

In preprocessing, the face is detected from the background, cropped to a predetermined size so as to be suitable for applying MRDTP and MRDNP. In most of the emotion recognition applications, Viola Jones [62] face detector is used. It has a series of classifiers arranged as cascade, but when new samples arrive, each classifier depends on the previous one. But here Chehra face detector is used in such a way that it can be extended to unconstrained situations too. This is because in Chehra [7] when new training samples arrive, incremental training is performed on the generic model using regression functions arranged in cascade. It performs better in unconstrained situations as each regression function does not depend on the previous function. ‘Viola Jones’ uses Haar features, while ‘Chehra’ uses SIFT (Scale Invariant Feature Transform) features for face detection.

3.2 Feature extraction

The detected faces from the preprocessing stage are given as input to either MRDTP or MRDNP to form the feature vector for the face. The feature extraction process is composed of three stages, i.e., (i) Filtering face images using compass masks (ii) Code image formation based on maximum response and (iii) Histogram formation and construction of feature vector. The MRDTP and MRDNP differ in the second step, i.e., the code image formation step as in Fig. 2. The feature vectors obtained as output are given as input to the ELM-RBF- based classification.

3.2.1 Filtering face images using compass masks

The magnitudes from the edges are very invariant to illumination changes, and so in this method the edge information calculated from the compass masks is used for the formation of code image and feature vector. Here, eight directional masks $\{M_{\theta_0}, M_{\theta_1}, \dots, M_{\theta_7}\}$, i.e., the masks for North, South, North East, South East, South West, North West, East and West

directions are used. The response obtained from each mask is considered as $\{R_{\theta_0}, R_{\theta_1}, \dots, R_{\theta_7}\}$ respectively for eight directions. In this paper, asymmetric Kirsch mask [38] is considered. An angle of 45° is used for the rotation of the Kirsch mask and to obtain eight directional masks as in Fig. 3.

The eight directional Kirsch masks are then used for the filtering of edges from face by convolving the 3×3 neighborhoods of the image with the Kirsch masks. These eight directional masks result in eight responses for each pixel. If all the eight responses are used for the feature vector formation, then the length of the feature vector becomes large. So in the proposed MRDTP and MRDNP, the feature vector formation is based only on the maximum response.

3.2.2 Code image formation

Let the eight responses obtained be denoted by $\{R_{\theta_0}, R_{\theta_1}, \dots, R_{\theta_7}\}$. All the positive and negative responses obtained for a pixel are taken altogether, and the maximum response value for each pixel among the eight responses is chosen to form the code image. This significantly reduces the complexity within the code when compared to other existing approaches. MRDTP and MRDNP differ in the code image formation step. MRDTP uses the pixel intensity information of the maximum response which is a decimal value, and MRDNP uses the direction information which creates a 3-bit code. The difference of the proposed patterns from the previous works is that the LDN uses the sign information and assigns the direction number of the top positive response as the three most significant bits and the direction number of the top negative response as the three least significant bits thus forming a 6-bit code. In LDTP, the code is formed as a single number using the most prominent direction and the difference in intensity from the opposite pixels of the two prominent directions.

1) Code Image formation for MRDTP

The maximum response image $C(x, y)$ is shown in (1).

$$C(x, y) = \max(R_{\theta_i}(x, y) | 0 \leq i \leq 7) \tag{1}$$

where $R_{\theta_i}(x, y)$ denotes the response obtained at a particular pixel position (x, y) for a directional mask M_{θ_i} and $\theta_i, 0 \leq i \leq 7$ correspond to the eight directions of the masks equally spaced at an interval of 45° such that $0 \leq \theta \leq 360^\circ$. The code $C(x, y)$ from (1) is using the intensity information of the maximum response of each pixel, among the eight responses as in Fig. 4(b). Then the DOG filter is calculated using

$$\begin{matrix}
 \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} & \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \\
 M_{\theta_0} & M_{\theta_1} & M_{\theta_2} & M_{\theta_3} & M_{\theta_4} & M_{\theta_5} \\
 & & \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix} & & \\
 & & M_{\theta_6} & M_{\theta_7} & &
 \end{matrix}$$

Fig. 3 Eight directional Kirsch mask



(a) *JAFFE image* (b) *Maximum response Image C(x, y)* (c) *Code Image D(x, y)*

Fig. 4 Code image obtained using MRDTP (a) JAFFE image, (b) Maximum response Image C(x, y), (c) Code Image D(x, y)

$$X = DOG((x, y); \sigma_1, \sigma_2) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \tag{2}$$

where σ_1 is the standard deviation that should be higher than σ_2 . DOG filter is calculated using (2) and represented as X . Then X is convoluted with $C(x, y)$ to get the code image $D(x, y)$, which has only the strong edges that are robust against illumination and random noise as in Fig. 4(c).

$$D(x, y) = C(x, y) * X \tag{3}$$

Although the MRDTP uses high response information, it is still having random noise. The convolution with DOG filter removes the random noise and also sharpens the edges so that better structural information is represented in the final code image $D(x, y)$. It also removes the illumination artifacts and enhances the features which increase the classification accuracy.

2) *Code image formation for MRDNP*

The direction information from the maximum response of each pixel is used to form a code image which is actually a direction map, from which also a feature vector can be formed using the histogram formation step. It is indicated as follows:

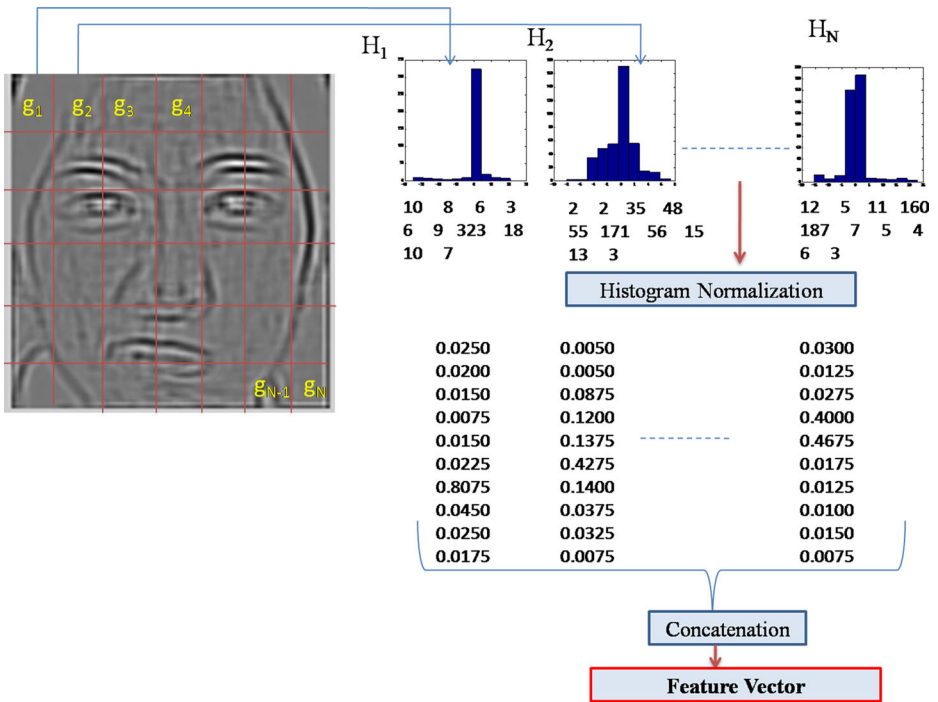
$$THETA(x, y) = \arg \max_i (R_{\theta_i}(x, y) \ 0 \leq i \leq 7) \tag{4}$$

where $R_{\theta_i}(x, y)$ indicates the response for a directional mask M_{θ_i} at particular pixel position (x, y) , and $\theta_i, 0 \leq i \leq 7$ represent the eight directions of the Kirsch masks respectively. Here, i is the direction number of the particular response. Thus, $THETA(x, y)$ of MRDNP is formed using the direction numbers of the maximum response of each pixel which excludes all the noisy edge information and is very robust.

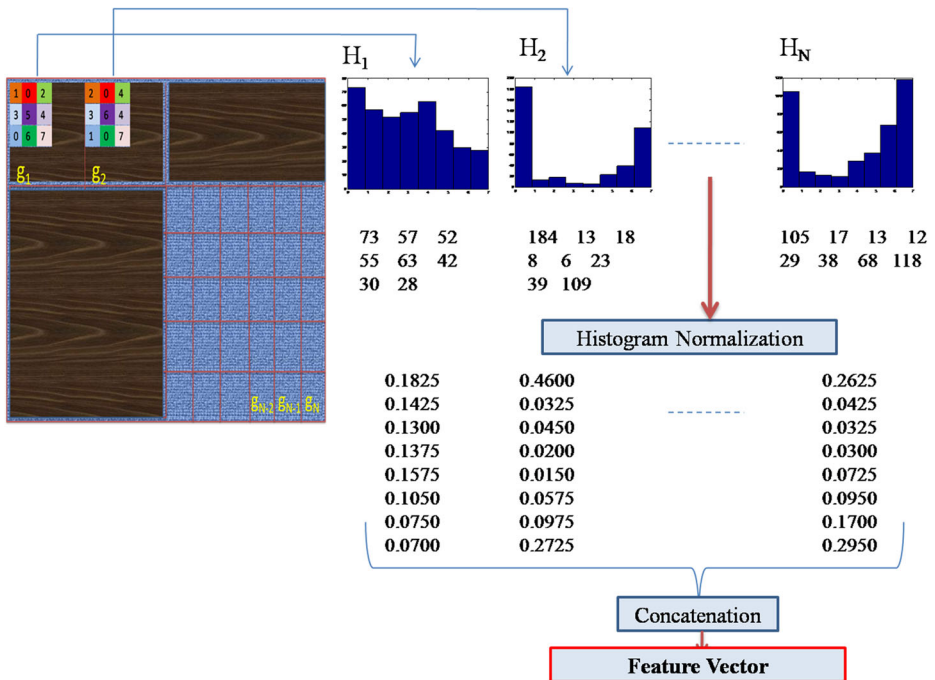
3.2.3 *Histogram Formation and construction of feature vector*

The histogram formation and construction of feature vector for both patterns are illustrated in Fig. 5(a) and (b). Here, the code image is divided into N equally sized grids $g_i, 1 \leq i \leq N$ such that the normalized histogram H_i is computed for each grid g_i and is concatenated to form the final feature vector. The final feature vectors in both MRDTP and MRDNP are the concatenated histograms of each sub region as in (5).

$$feature\ vector = \langle H_1, H_2, \dots, H_N \rangle \tag{5}$$



(a)



(b)

Fig. 5 Histogram formation (a) for MRDTP (b) for MRDNP

where N is the total number of smaller grids formed in the code image. This method of feature vector formation helps to extract the information of smaller to larger edges and corners of face. The dimension of the feature vector can be reduced using the proposed GSDRS system, which is explained in the next section.

3.3 Generalized supervised dimension reduction system technique

This technique aims to use PGK for all datasets, in the proposed dimension reduction system using MRDTP and MRDNP for facial expression recognition. This function is used by Gupta in curve fitting the scans [26].

3.3.1 Pearson VII function in the formation of PGK

The general form of the Pearson VII function is given as

$$f(x) = \frac{P}{\left[1 + \left(\frac{2(x-x_0)\sqrt{2^{(1/\omega)}-1}}{\sigma} \right)^2 \right]^\omega} \quad (6)$$

To satisfy Mercer conditions, (6) is rewritten as

$$k(x_i, x_j) = \frac{1}{\left[1 + \left(\frac{2\sqrt{\|x_i-x_j\|^2}\sqrt{2^{(1/\omega)}-1}}{\sigma} \right)^2 \right]^\omega} \quad (7)$$

In (6), P is the peak height at x_0 the center, and x is a variable which is self regulating. Here, σ and ω are the width of the peak and the tailing factor. By tuning σ and ω parameters, various shapes from Gaussian to Lorentzian can be formed. (7) is formed from (6) to satisfy the Mercer's conditions, where x is replaced by two vectors x_i, x_j and their formula to calculate distance. Then, x_0 is deleted and P is replaced by 1. Tuning the parameters of PGK makes it suitable to replace any other kernel. Thus, PGK can be used in the place of any other kernel in a kernel-based dimension reduction system.

3.3.2 GSDRS

The inter-class scattering is maximized, and the intra-class scattering is minimized by GDA [27]. In the proposed approach, the usage of PGK provides good classification after dimension reduction for different datasets. A denotes the total number of categories among the samples, and N_a represents the number of data samples within each class a . $\{x_{ab}, a = 1, 2 \dots A; b = 1, 2 \dots N_a\}$ denote the set of data considered for training. The training set after application of GDA process is denoted by $\{\phi(x_{ab}), a = 1, 2 \dots A; b = 1, 2 \dots N_a\}$, ϕ denotes the non-linear function for mapping the features from high dimension space G to low dimension space H . Then, $\phi: G \rightarrow H, x \rightarrow \phi(x)$.

Then, S_W which is the scattering within the same category, and S_B which is the scattering between different categories for the training set are calculated as in (8) and (9).

$$S_W = \frac{1}{A} \sum_{a=1}^A \frac{1}{N_a} \sum_{b=1}^{N_a} \phi(x_{ab}) \phi(x_{ab})^T \tag{8}$$

$$S_B = \frac{1}{A} \sum_{a=1}^A (\mu_a - \mu)(\mu_a - \mu)^T \tag{9}$$

where μ_a , is the mean of the samples belonging to class a .

λ is the Eigen value, and V is the Eigen vector estimated in GDA process respectively so that it satisfies (10).

$$\lambda S_W V = S_B V \tag{10}$$

$$\lambda = \frac{V^T S_B V}{V^T S_W V} \tag{11}$$

The Eigen vector solution is denoted by (12).

$$V = \sum_{a=1}^A \sum_{b=1}^{N_a} \alpha_{ab} \phi(x_{ab}) \tag{12}$$

where $\phi(x_{11}) \dots \phi(x_{ab})$ is the span, and α_{ab} is the Eigen vector coefficient. The kernel function can be used to represent the dot product calculated between sample data i and j from two different classes p and q in the feature space H as in (13) so that the discriminant analysis is generalized to a non-linear case.

$$(k_{ij})_{pq} = \phi(x_{pi}) \cdot \phi(x_{qj}) = k(x_{pi}, x_{qj}) \tag{13}$$

Pearson VII function given in (14) is the generalized kernel used as PGK in the proposed work for representing the dot product. Since the performance of the other existing kernels varies for different datasets, PGK works as a standard replacement for all other kernels that can cope with GDA. This leads to direct use of PGK, instead of experimenting with all other kernels for GDA. PGK with GDA forms the GSDRS, which saves a lot of experimenting time as it is directly applicable to different datasets.

$$k(x_{pi}, x_{qj}) = \frac{1}{\left[1 + \left(\frac{2\sqrt{\|x_{pi} - x_{qj}\|^2} \sqrt{2^{(1/\omega)} - 1}}{\sigma} \right)^2 \right]^\omega} \tag{14}$$

where K is a $C \times C$ matrix that is defined on the members of the class by $((K_{pq})_p = 1 \dots A, q = 1 \dots A)$. K_{pq} is the matrix composed of dot products between the samples belonging to class p and q . Then,

$$K_{pq} = (k_{ij})_{\substack{i = 1 \dots N_p \\ j = 1 \dots N_q}} \tag{15}$$

Assume, D is a $C \times C$ block diagonal matrix as in (16).

$$D = (D_a)_{a=1 \dots A} \tag{16}$$

where D_a is an $N_a \times N_a$ matrix and all the elements = $\frac{1}{N_a}$. After substituting (8), (9), and (12)

into (10), an inner product of (10) with $\phi(x_{ab})$ is computed. In the solution obtained after doing inner product, two terms D, K are substituted to obtain (17).

$$\lambda KK e = KDK.e \quad (17)$$

Here, e represents a column vector with elements α_{ab} , $a = 1, 2, \dots, A$; $b = 1, 2, \dots, N_a$. From (17), the matrix $(KK)^{-1}KDK$ is formed. Then, the Eigen vector of $(KK)^{-1}KDK$ is found which is the solution of e . If matrix K is not reversible, then K has to be diagonalized first before finding solution of e [9]. Using M Eigen vectors, the projection matrix L is created as in (18).

$$L = [e_1 e_2 \dots e_M] \quad (18)$$

where M is the total number of Eigen vectors. Thus, x which is a test sample is mapped on to the M dimensional space H using L as in (19).

$$y = K_x L, K_x = [k(x, x_{11}) \dots k(x, x_{ab}) \dots k(x, x_{AN_a})] \quad (19)$$

Thus, the length of the feature vectors becomes $A - 1$, where A is the number of unique labels that denote the number of categories to which the training samples belong. GSDRS is an LDA-based method where the maximum number of the reduced dimensions is $A - 1$.

The overall steps in GSDRS are as follows:

- (i) Compute K and D using (15) and (16).
- (ii) Compute Eigen vector from $(KK)^{-1}KDK$.
- (iii) Compute the projection matrix from the most significant Eigen vectors using (18) which is used to project a test sample to a low dimension space H .

3.4 ELM-RBF for classification

Huang et al. [30, 31] have reported that Extreme Learning Machine (ELM) with Single hidden Layer Feed forward Neural networks (SLFNs) performs classification faster than SVM. ELM is capable of doing both binary classification as well as multi-classification. The kernel-based ELM with RBF kernel provides good results when used for 6-class as well as 7-class emotion recognition which is evident from our experimental results discussed in the next section.

4 Experiments and performance evaluation

The experiments with the proposed approach are conducted using MATLAB R2014a and Intel® core(TM) i5-4210 U CPU @1.70GHz with 4GB RAM.

4.1 Datasets used

4.1.1 JAFFE

There are totally 213 images taken from 10 subjects. All the images are of size 256×256 . Seven classes of emotions are Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise [41]. All 213 images are considered in the experiments conducted.

4.1.2 CK+

CK+ version has better expressions than CK. The sizes of the images are 640×490 . 123 subjects are used to capture 593 sequences out of which only 327 sequences are annotated. Each sequence has about 10 to 60 static images. From each of the 327 sequences, 3 to 4 images showing the peak of the expressions are considered for the experiments to reduce the computation time. The first frame from each sequence is considered for the neutral category. Totally 1281 images are considered in the experiments [35, 40].

4.1.3 MUG

Multimedia Understanding Group (MUG) dataset has 50 to 60 images per sequence of expression. 1462 sequences captured from 86 subjects are in the dataset. The images are of 896×896 pixel resolution. 52 subjects are chosen for the experiments. 1 or 2 peak expression images are chosen from each sequence of those subjects. Images at the beginning of each sequence are chosen for neutral category [5]. 81 images are considered under each category for the experiments. So, 567 images from MUG are used in the experiments.

4.1.4 SFEW

SFEW is a dataset gathered under unconstrained settings with the aim to extend the facial expression recognition to real-time environments. The images are having very high illumination variations, pose variations and noise. The images are of size 720×576 . From the available 1394 images in the dataset, 881 images are taken in the training set and 474 images in the testing set. All the experiments conducted here are strictly person- independent. The classification process is repeated ten times and the average of the performance measures is considered [20, 21].

4.1.5 MMI

This is a Man Machine Interaction (MMI) dataset having 312 sequences of images and 30 subjects. A total of 11,500 images are available in the dataset. From each sequence, 3 to 4 of the peak expression images are taken. The end frame is considered for neutral emotion category. About 1050 images are used in the experiments conducted [47, 60].

4.1.6 DISFA, DISFA+

Denver Intensity of Spontaneous Facial Actions (DISFA) database has 27 subjects and is FACS coded. It has 12 Action Units with intensity levels coded between 0 and 5. It does not have emotion labels. So, emotion FACS (EMFACS) is used to obtain the emotion labels. It has about 89,000 images after converting the video sequences to image frames, where most of them are neutral images, and the distribution of images between the classes is not uniform. Happiness and neutral expressions are having more images than the other classes. So in the experiments, out of the 28,404 images available for the happiness emotion only 5000 images are selected. Also from the 48,582 available images of neutral emotion, 5000 images are selected. For the remaining expressions all the images are selected for our experiments. DISFA+ is an extended dataset of DISFA. It has posed

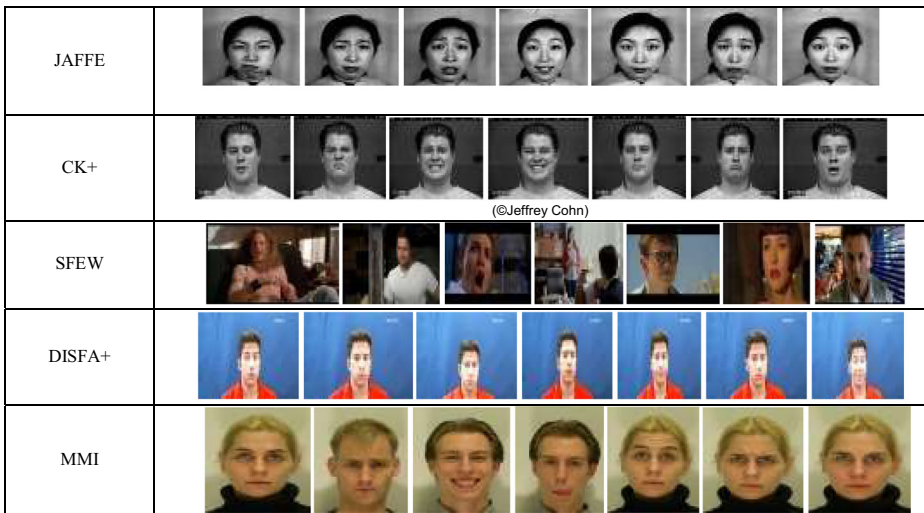


Fig. 6 Sample images from some of the datasets used

expressions of nine subjects present in the DISFA dataset. There are over 57,000 images that are annotated. Each sequence starts with a neutral expression. So for each subject about 100 images are selected from peak expressions for each emotion category. Altogether, 6300 images are chosen for DISFA+ and added to the DISFA dataset and used for further experiments as combined DISFA and DISFA+ [42–44] (Fig. 6).

4.2 Experiments and results

Table 1 indicates the number of samples from each dataset selected under each emotion category for our experiments. All the experiments are conducted using 10-fold cross validation on the available images. Images are taken in the size of 162×122 for all the experiments. The 10-fold cross validation is done by choosing each fold as a testing set while all the remaining nine folds form the training set. The overall classification accuracy is the average of the performance measures obtained on the ten folds. The complete process is repeated 10 times and the average of the accuracies obtained is displayed in the results. For SFEW dataset already the dataset is divided into two independent folds.

From Table 2 it can be seen that both MRDTP and MRDNP are performing better than the existing feature extraction techniques like LBP [54], LD_iP [33], LTP [56], GABOR [1], LPQ

Table 1 Number of samples in each emotion category for the datasets

Category	Angry (AN)	Disgust (DI)	Fear (FE)	Happy (HA)	Neutral (NE)	Sad (SA)	Surprise (SU)
JAFFE	30	29	33	31	30	31	29
CK+	84	114	153	261	291	153	225
MUG	81	81	81	81	81	81	81
SFEW	229	99	153	228	242	249	155
MMI	150	150	150	150	150	150	150
DISFA	436	5326	4073	5000	5000	1024	1365
DISFA & DISFA+	1336	6226	4973	5900	5900	1924	2265

Table 2 Comparison of classification accuracy for seven-class emotion recognition without dimension reduction

Feature Extraction	JAFFE (%)	MUG (%)	CK+ (%)	SFEW (%)	MMI (%)	DISFA (%)	DISFA & DISFA+ (%)
LBP	89.9	93.4	93.1	25.1	78.9	62.4	65.3
LD ₁ P	90.1	92.1	94.8	32.1	79.1	68.4	68.2
LTP	88.2	93.1	94.6	25.3	76.3	66.3	66.8
GABOR	92.0	93.3	96.6	30.0	79.3	67.9	68.6
LPQ	88.9	92.1	94.6	24.7	76.1	65.1	67.3
LDTP	91.1	94.1	96.0	33.5	79.3	68.4	68.4
LDN	90.1	93.7	94.2	33.5	80.5	68.1	68.2
SIFT	87.6	92.6	95.7	32.0	78.1	68.6	68.5
HOG	91.1	92.5	96.1	31.9	78.0	68.5	68.2
MRDTP without noise restraining	92.1	94.2	97.0	33.8	81.6	70.0	69.9
MRDTP with noise restraining	94.1	95.3	98.2	34.0	82.0	70.3	70.4
MRDNP	92.2	94.3	97.1	33.8	81.8	70.1	70.1

[46], LDTP [52], LDN [49], SIFT [10] and HOG [16]. This is because LBP is more sensitive to illumination variations, while LPQ is somewhat robust to illumination, but in highly varying lightings, LPQ's performance is also low, when compared to MRDTP and MRDNP. Conventional Gabor takes noisy edges also into account. HOG is sensitive to scaling and rotations of image and SIFT is very sensitive to luminance variations. MRDTP performs well than the other directional patterns, because of its ability to extract the robust structural information from the images due to Kirsch mask and noise restraining step. MRDNP takes only the most prominent directional numbers into account, and therefore highly robust to noise. In order to show the impact of noise restraining step in MRDTP, the experiments are also conducted without noise restraining step and the results are included in Table 2. It can be seen that the inclusion of the noise restraining step in MRDTP has helped significantly to improve the classification accuracy.

The performances of the proposed patterns are analyzed at different resolutions for seven-class emotion recognition using ELM-RBF without dimension reduction in Table 3. The JAFFE images were resized to different resolutions, and the MRDTP and MRDNP code image based on Maximum response is found. For each input image of size 162×122 , each code image is divided into block size of 20. Then, each sub block is used to form the histogram of bin size 10 in MRDTP and 8 in MRDNP. Thus for an image of size 162×122 , the proposed method results in a feature vector of size 480 in MRDTP and 384 in MRDNP, which is very low when compared to the dimension of the Gabor feature vector which is 19,764. The above results indicate that for high resolution images, the dimension size increases, with accuracy. But for low resolution images, the dimension decreases with the accuracy as in Table 3.

Table 3 Accuracy and dimension variations for different sizes of JAFFE images

Resolution	Dimension of MRDTP	Accuracy of MRDTP (%)	Dimension of MRDNP	Accuracy of MRDNP (%)
242×242	1440	94.3	1152	92.9
162×122	480	94.1	384	92.2
102×102	250	90.2	200	90.1
42×62	60	75.9	48	75.2

The experiments are conducted with several facial expression datasets like JAFFE, MUG, CK+, SFEW, MMI, DISFA and DISFA+ initially with the existing Generalized Discriminant Analysis. MRDTP is used for extracting features from images and ELM-RBF is used for classification. Experiments confirm that different kernels are finalized as the optimum kernels when using existing GDA, after seeing the good classification results in different datasets. But more time is required in carrying out the experiments and finalizing the optimum kernel. In the proposed approach GSDRS, a common PGK produces good results when used for dimension reduction for all datasets as in Table 4. Already existing kernels like Linear, Poly and RBF are considered for the existing GDA calculations in dimension reduction. The linear kernel is considered as $k(x, y) = (x, y)$, Polynomial kernel as $k(x, y) = (x, y)^d$ where d denotes the degree and the Gaussian RBF kernel as $k(x, y) = \exp\left(-\frac{\|x-y\|^2}{\sigma}\right)$, where σ is the width of the Gaussian peak respectively. The results obtained for both seven-class and six-class emotion recognition (without neutral expression) is represented. In Table 4, best results are only depicted after experimenting with $d = 2, 3, \dots, 8$ and selecting the best results. For RBF kernel, σ is selected from the set $\{2^{10}, 2^9, \dots, 2^{-9}, 2^{-10}\}$ by using a linear search. For GSDRS, Pearson kernel is used as in (14). σ is selected from the set $\{2^{10}, 2^9, \dots, 2^{-9}, 2^{-10}\}$ and ω is selected from the set $\{2^0, 2^1, \dots, 2^{10}\}$ by using grid search method along with cross validation on the samples taken as training set. The set of parameters that maximizes the classification accuracy is chosen as the best set of parameters for test set. In classification purposes using ELM, RBF kernel is used. The parameter σ is selected from the set $\{2^{10}, 2^9, \dots, 2^{-9}, 2^{-10}\}$. The parameter R which is the regularization parameter in ELM is selected from the range $R = 10^l, l = -3, \dots, 3$ using grid search along with cross validation.

It can be seen from the columns 3,4 and 5 results of Table 4, that for six-class emotion recognition Linear kernel performs better with JAFFE and DISFA datasets, while Polynomial kernel performs better with CK+,SFEW and MMI datasets. RBF kernel is better with MUG, DISFA, combined DISFA and DISFA+ datasets. The performance of each existing kernel differs in each dataset. Similar problem arises in seven-class emotion recognition too as in Table 4. So to select the best kernel for dimension reduction using GDA, it requires various experimentations with different kernels. To avoid this, a PGK is used with GSDRS here. It is

Table 4 Comparison of classification accuracy obtained after reducing dimensions of MRDTP feature vector using GSDRS and GDA

Datasets	No. of Classes	Dimension Reduction			
		Linear (GDA)	Poly (GDA)	RBF (GDA)	PGK (GSDRS)
JAFFE	6	95.9	91.9	93.5	96.7
	7	92.9	92.9	94.3	94.3
CK+	6	97.2	98.9	98.1	98.9
	7	96.0	98.4	97.6	98.4
MUG	6	93.4	92.2	95.3	95.9
	7	95.7	93.4	94.9	95.7
SFEW	6	35.0	35.2	34.7	35.9
	7	33.1	33.8	34.5	34.8
MMI	6	80.0	82.4	82.3	82.5
	7	78.0	82.0	79.5	82.1
DISFA	6	71.2	69.7	68.5	71.2
	7	70.5	68.1	68.3	70.7
DISFA & DISFA+	6	67.4	68.1	71.4	71.5
	7	67.3	67.8	70.2	70.4

*Best Classification results are indicated using bold font

Table 5 The similarity measure (E_K/PGK_K) for JAFFE images

Common kernels for E_K	PGK Parameters (σ, ω)	(E_K/PGK_K)
Linear	$2^2, 2^4$	0.99–1.00
Poly(d = 2)	$2^{-3}, 2^4$	0.98–0.99
Poly2(d = 3)	$2^{-4}, 2^5$	0.98–0.99
Poly3(d = 4)	$2^{-6}, 2^6$	0.98–0.99
RBF1($\sigma = .5$)	$2^{10}, 2^6$	0.99–1.00

used as the generalized kernel in GSDRS to produce small dimensional feature vectors with no reduction in accuracy. It can also be seen in the last column results of Table 4 that in most of the datasets PGK produces better results than other existing kernels. The low classification rate of SFEW with MRDTP when compared to other datasets is because of the highly wild conditions of the image in the database. The GSDRS has produced better classification results compared to the results produced by MRDTP without dimension reduction in Table 2 which proves the efficiency of GSDRS in emotion recognition. In Table 4, it can also be seen that the six-class emotion category has higher accuracy than the seven-class emotions because of the absence of the neutral category.

Results of Table 4 indicate that the mapping of the feature space created by the PGK kernel is more or less similar to the mapping created by three existing kernels (Linear, Poly and RBF) through the classification accuracy results produced by the PGK. This is further proved by (E_K/PGK_K) which is the similarity measure, used to calculate the similarity between the kernel matrix created by the existing kernels and PGK. E_K denotes the kernel matrix of any one existing kernel, and PGK_K denotes the kernel matrix created by PGK. Table 5 displays the results obtained. For this purpose, the leave-one-out technique is used to divide the JAFFE

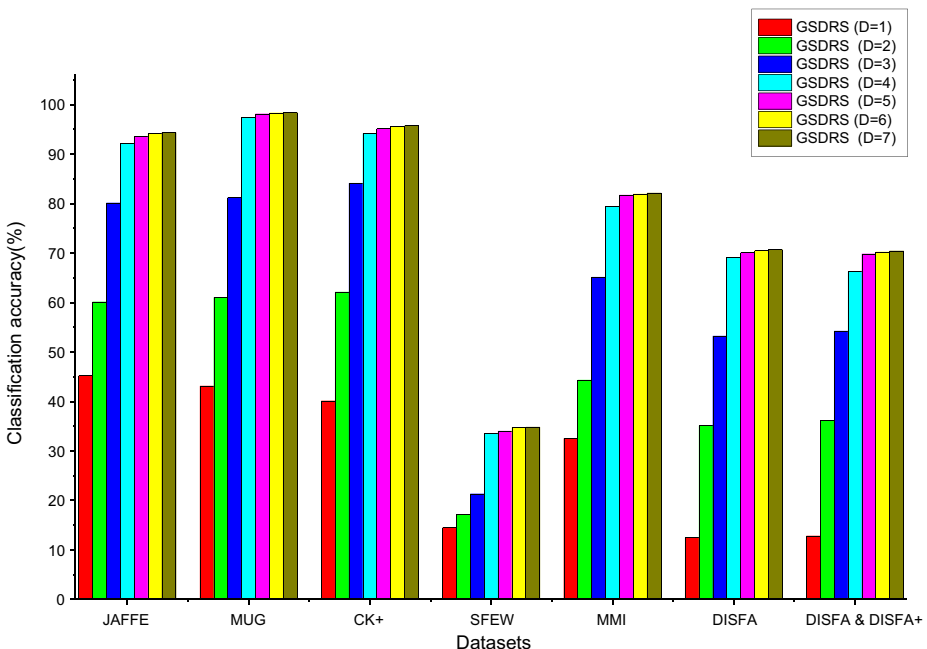


Fig. 7 Classification accuracy using GSDRS with different values for dimension number (D)

Table 6 Confusion Matrix for JAFFE using MRDTP + GSDRS + ELM-RBF

	An	Di	Fe	Ha	Ne	Sa	Su
An	27	1	1				1
Di	1	28					
Fe		1	29		1	1	1
Ha				31			
Ne					30		
Sa			2	2		27	
Su							29

dataset into training and testing samples. A sample image from each subject-expression combination is used to create the test set, while the remaining samples construct the training set. Totally 143 samples are used for training and 70 for testing. The similarity measures are calculated as in Table 5. The grid search algorithm is used to decide the hyper parameter values.

In Table 5, E_K/PGK_K denotes that the ratios are distributed between 0.98 and 1.00 which shows the high similarity between the kernel matrices E_K and PGK_K . The calculated similarity measures indicate that for particular values of σ , ω the PGK kernel is very much similar to other existing kernels. As the values of σ , ω vary, the PGK kernel evolves from linear to RBF. Fig. 7 indicates classification accuracy results obtained using MRDTP with different dimension numbers that are set for GSDRS. The maximum number of reduced dimensions is six for seven emotion categories. When six is taken as the reduced dimension number the classification accuracy is good. But as the dimension reduces further the accuracy is affected as seen from the results obtained.

The confusion matrix in Table 6 displays the number of instances predicted under each emotion category of JAFFE. From this matrix the number of instances that are accurately and inaccurately predicted is easily known. Fear, Anger and Sadness are the expressions that mainly cause the misclassifications.

The confusion matrix in Table 7 depicts the number of instances predicted under each emotion category of CK+. Anger, Disgust and Neutral expressions are affecting the classification rate in CK+.

The confusion matrix in Table 8 depicts the number of instances predicted under each emotion category of MUG. Here, Happiness expression is heavily confused among the other expressions.

The confusion matrix in Table 9 depicts the number of instances predicted under each emotion category of SFEW. The low classification accuracy is due to the imbalance of samples in different classes of the dataset. It also requires more training samples because of high variance in illumination, noise, pose and transformations in images. Both MRDTP and MRDNP perform very well while classifying SFEW when compared to the other existing

Table 7 Confusion Matrix for CK+ using MRDTP + GSDRS + ELM-RBF

	An	Di	Ha	Fe	Ne	Sa	Su
An	77	1	1		1	1	3
Di	2	111				1	
Ha			151			1	1
Fe				259	2		
Ne			2	1	287	1	
Sa	1					151	1
Su			1				224

Table 8 Confusion Matrix for MUG using MRDTP + GSDRS + ELM-RBF

	An	Di	Ha	Fe	Ne	Sa	Su
An	78		1		2		
Di		77	1	1		1	1
Ha	2	1	74	1	2		1
Fe	1	1	2	77			
Ne	1	1			79		
Sa			1			80	
Su	1		1			1	78

Table 9 Confusion Matrix for SFEW for test set using MRDTP + GSDRS + ELM-RBF

	An	Di	Fe	Ha	Ne	Sa	Su
An	32		1	31	8	14	
Di	11	7	1	4	6	4	
Fe	4		9	21	8	11	2
Ha	14		1	43	2	11	1
Ne	18		1	23	30	17	3
Sa	12	1		24	5	34	1
Su	11		2	19	4	13	10

feature extraction techniques, thus proving their efficiency in unconstrained situations. Disgust, Fear and Surprise expressions are difficult to be recognized under unconstrained situations. The confusion matrices obtained in Tables 10, 11 and 12 for MMI and DISFA datasets indicate that surprise emotion is poorly recognized in these datasets.

The dimension reduction techniques are substituted in the proposed approach by PCA, LDA and LPP and classified using ELM-RBF that are given in Table 13. From Table 13, it can be seen that GSDRS achieves better results than PCA [34] at a very low dimension for all the datasets. PCA has been applied to retain 95% of the variance and the reduced dimensions are as in Fig. 8. LDA [13], LPP [53] and GDA [9] reduce dimension of feature vector to 6. MRDTP + GSDRS achieves good results for all the datasets at dimension 6 which proves the efficiency of GSDRS compared to other existing dimension reduction techniques.

For images of size 162×122 , the features are obtained using proposed patterns, and the dimension number results obtained after reducing dimensions using PCA of variance 95% are displayed in Fig. 8. The reduced dimensions obtained using PCA are very much greater compared to the proposed GSDRS.

The proposed method is run for different types of dimension reduction algorithms and also without any dimension reduction algorithms and the running times are recorded. Various dimension reduction algorithms like PCA, LDA, LPP and GDA are compared with the

Table 10 Confusion Matrix for MMI using MRDTP + GSDRS + ELM-RBF

	An	Di	Fe	Ha	Ne	Sa	Su
An	123	7	5	2	7	4	2
Di	5	122	7	7	4	5	
Fe	3	6	122	3	9	7	
Ha	1	3	9	125	7	4	1
Ne	2	2	5	10	120	10	1
Sa	1	3	11	13	7	115	
Su	2		7	6			135

Table 11 Confusion Matrix for DISFA using MRDTP + GSDRS + ELM-RBF

	An	Di	Fe	Ha	Ne	Sa	Su
An	324		56		23		33
Di	354	3200	708	474	236	354	
Fe		153	3800	120			
Ha		370		4026	274	330	
Ne		636			3426	574	364
Sa	343		258			423	
Su						840	525

Table 12 Confusion Matrix for combined DISFA and DISFA+ using MRDTP + GSDRS + ELM-RBF

	An	Di	Fe	Ha	Ne	Sa	Su
An	1245			46	22	23	
Di		5692	299				235
Fe		728	4010		235		
Ha		251		4236	624		789
Ne			1475		4100		325
Sa		277		322		1325	
Su	234		509				1522

proposed GSDRS. For GDA, a kernel selection algorithm based on cross validation is used. In GDA, the training data is divided into two folds. A model based on first fold is created and then dimension reduction is done on second fold using different kernels. The kernel that produces high classification accuracy, while classifying the second fold using ELM-RBF is chosen as the best kernel for reducing the dimensions of testing data. This step takes considerable time. But in GSDRS, this step is excluded by using PGK as it reduces the time for dimension reduction, and its running time is more or less equivalent to PCA, LDA and LPP as in Fig. 9.

It can be seen from Fig. 10 that MRDNP has significantly less computation time because of the fewer number of steps when compared to MRDTP. Though MRDNP is inferior to MRDTP in terms of classification accuracy, it is still superior to the existing feature extraction techniques as shown in Table 2. Besides this, the dimension of the feature vectors created by MRDNP is low compared to MRDTP as shown in Table 3. This means that either MRDTP or MRDNP can be used for the classification of emotions depending on the application, whether it is computationally intensive or computation savvy.

Table 13 Comparison of classification accuracy of different dimension reduction techniques for seven-class emotion recognition

Datasets	MRDTP				MRDNP			
	PCA (%)	LDA (%)	LPP (%)	GSDRS (%)	PCA (%)	LDA (%)	LPP (%)	GSDRS (%)
JAFFE	93.4	94.1	93.8	94.3	91.1	92.1	92.3	92.9
CK+	97.5	97.2	96.8	98.4	94.5	94.2	94.8	98.1
MUG	93.3	95.1	94.9	95.7	94.3	94.1	96.9	94.6
SFEW	34.2	33.1	33.6	34.8	32.2	33.1	33.6	34.1
MMI	79.0	80.0	80.2	82.1	77.3	78.4	78.6	82.0
DISFA	65.3	68.2	68.9	70.7	67.8	68.5	68.6	70.3
DISFA& DISFA+	67.4	68.9	69.2	70.4	67.9	68.4	68.6	70.2

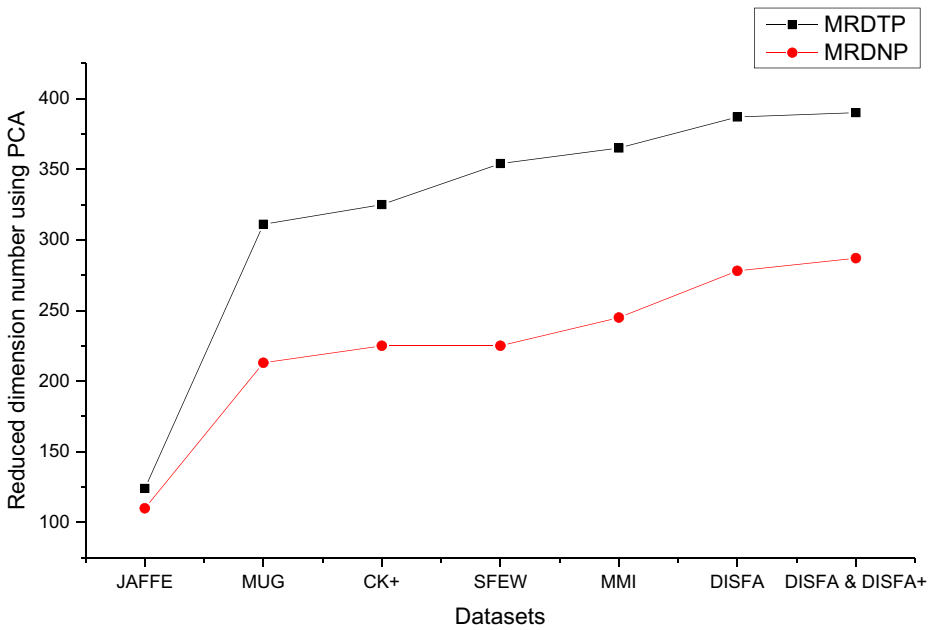


Fig. 8 Dimension of MRDTP and MRDNP feature vectors after using PCA for seven-class emotion recognition

While repeating the experiments using SVM classifier with RBF kernel (SVM-RBF) [15], the classification accuracy results obtained are more or less the same as the ELM-RBF, but the time consumed by SVM-RBF is more when compared to ELM-RBF as seen in Figs. 11, 12 and 13 respectively.

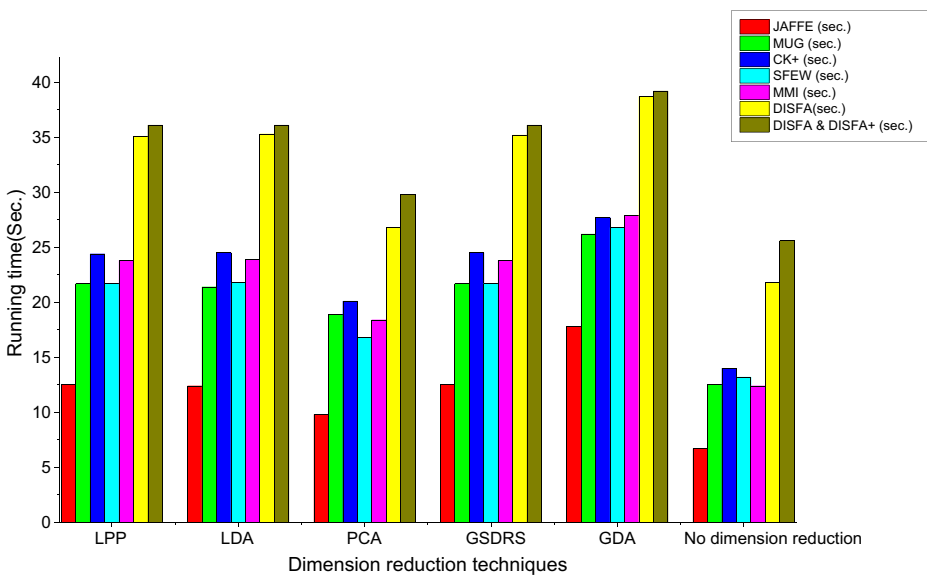


Fig. 9 Running time of the proposed approach without any dimension reduction and with several dimension reduction techniques

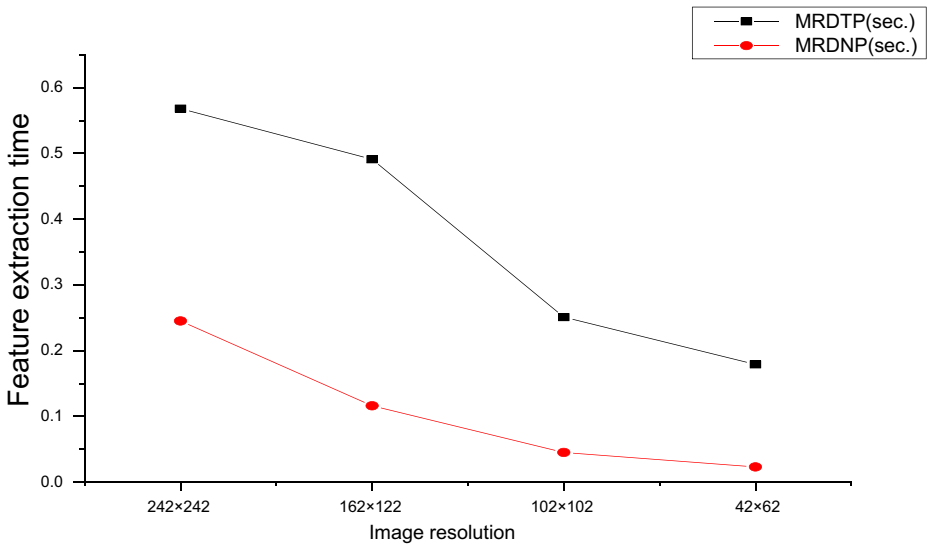


Fig. 10 Feature extraction time for a single image using MRDTP and MRDNP

To prove the robustness of MRDTP and MRDNP in the presence of noise, Gaussian white random noises of mean zero and different variance levels such as 0.0001, 0.0002, 0.0003, 0.004 are used to contaminate the images of datasets. The results of classification accuracy show the efficiency of the proposed patterns than the existing ones. The classification accuracy of the existing feature extraction techniques reduces significantly with increase in variance of Gaussian white random noise. But the proposed patterns are more robust to the random noise

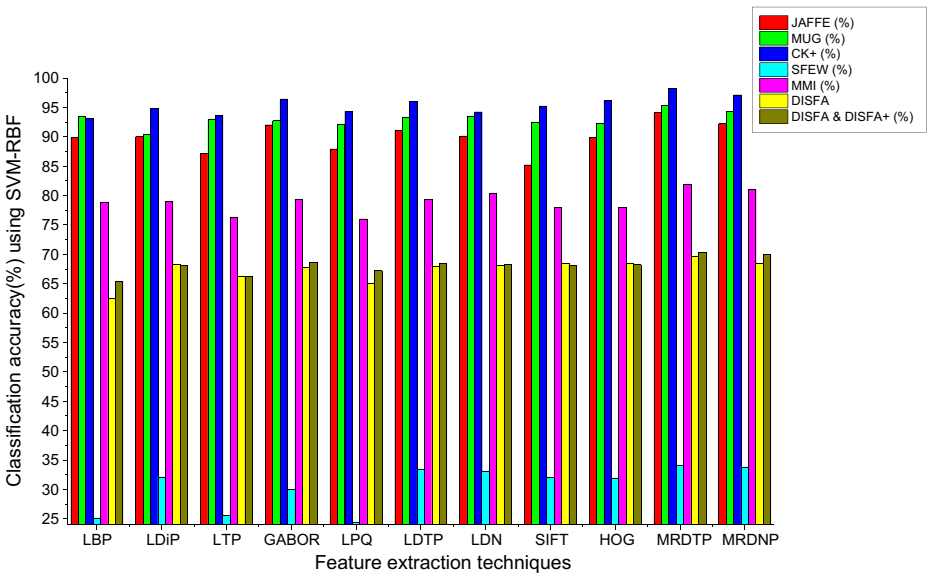


Fig. 11 Comparison of classification accuracies obtained using SVM-RBF without dimension reduction for seven-class emotion recognition

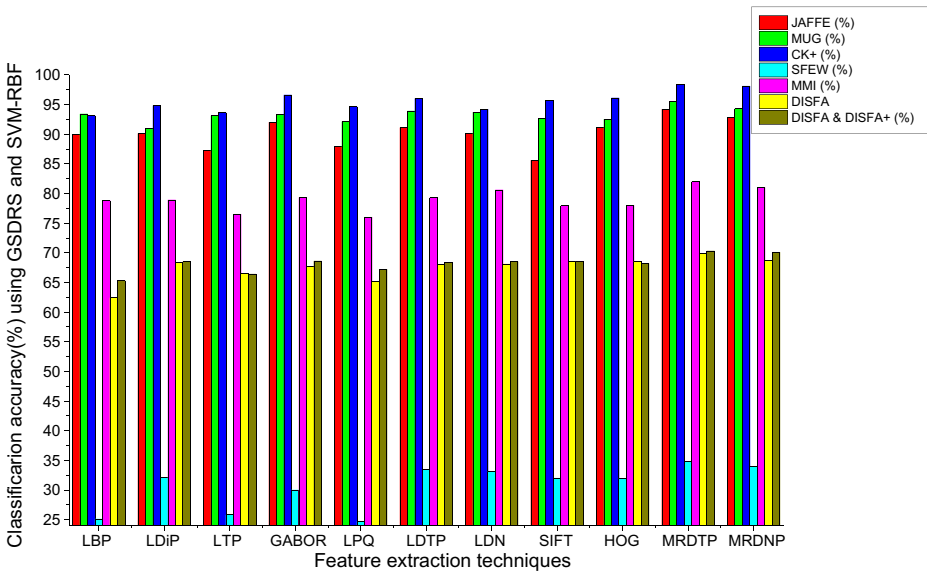


Fig. 12 Comparison of classification accuracies obtained using SVM-RBF with dimension reduction for seven-class emotion recognition

than other directional patterns in literature. It is seen in Figs. 14, 15, 16, 17, 18, 19 and 20 and is evident from the classification accuracy results.

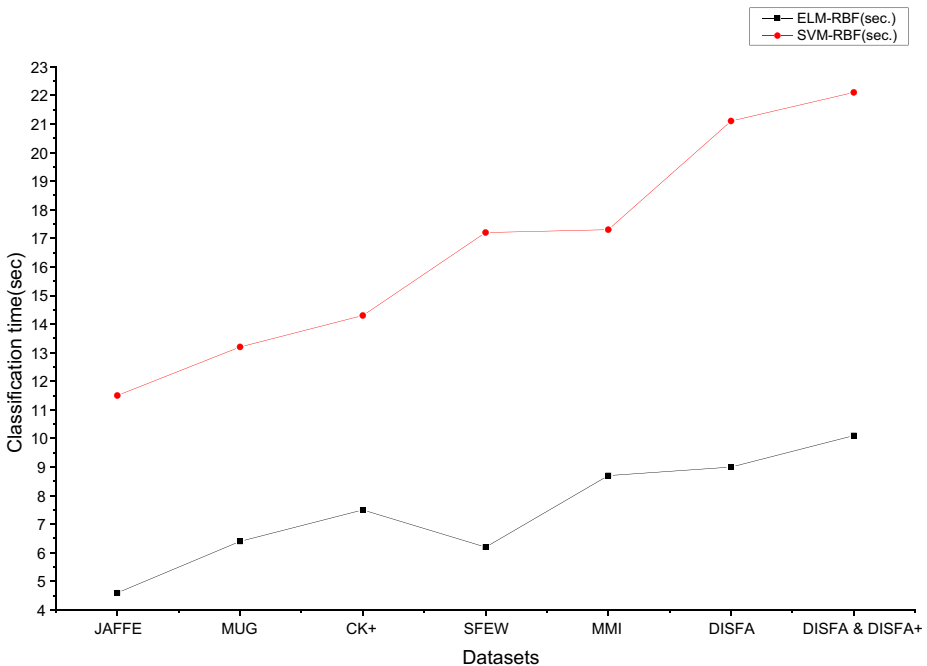


Fig. 13 Classification time obtained using different classifiers while using features of reduced dimension six

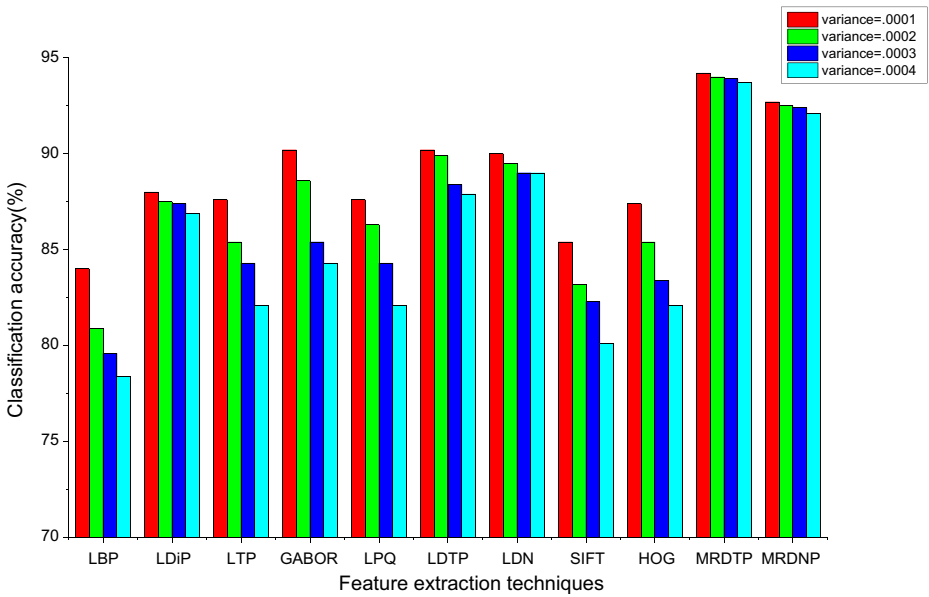


Fig. 14 Experimental results on JAFFE images with added Gaussian white noise of different variances for seven-class emotion recognition

While using other feature extraction techniques in the proposed approach and applying GSDRS before classification using RBF-ELM, all the feature descriptors perform better and

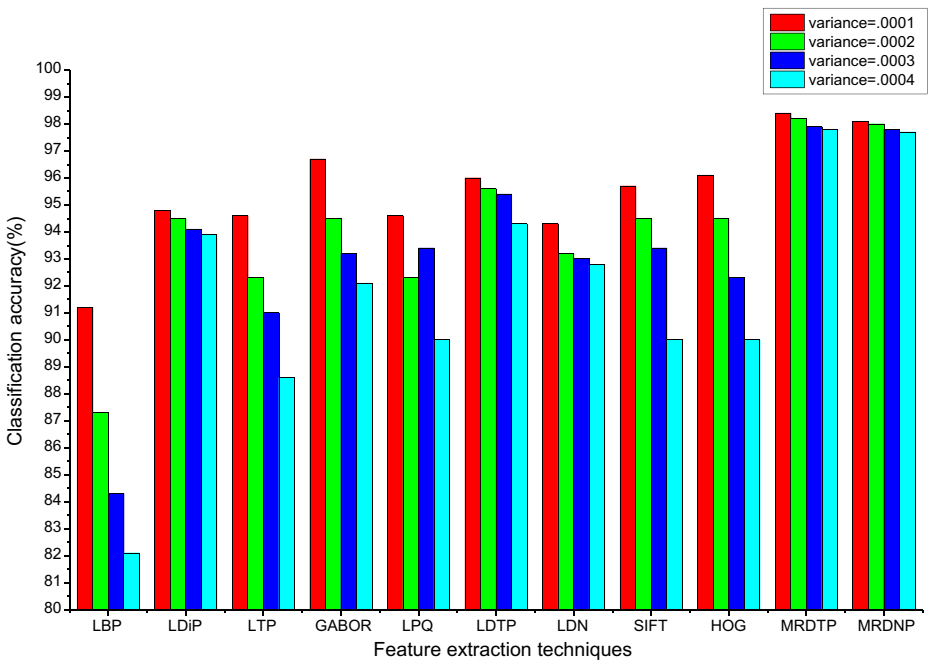


Fig. 15 Experimental results on CK+ images with added Gaussian white noise of different variances for seven-class emotion recognition

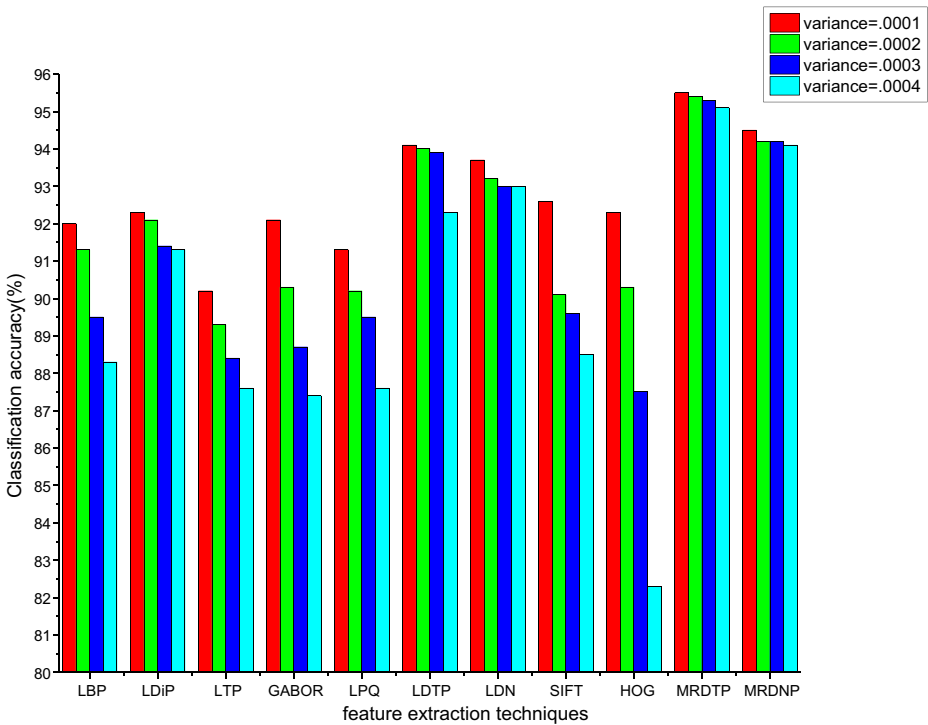


Fig. 16 Experimental results on MUG images with added Gaussian white noise of different variances for seven-class emotion recognition

achieve better classification accuracy as in Fig. 21 than the ones achieved without using GSDRS in Table 4. This ensures the efficiency of GSDRS on other local coding methods too for emotion recognition.

4.3 Discussion

4.3.1 The steps that improve classification accuracy in MRDTP and MRDNP

MRDTP and MRDNP encode the structural information well. The noise restraining step using DOG filter in MRDTP helps in eliminating the random noise, thereby improving the classification accuracy. The efficiency of the MRDNP is also high in recognizing the noisy images because it encodes the structural information of the entire neighborhood based on the direction number of only the strong edges. Because of the noise resistant property of MRDNP and MRDTP, no other filtering is performed in the preprocessing stage of the proposed work. Also, the method of feature vector calculation using grids helps to remove the illumination artifacts due to monotonic grayscale transformations. Due to these factors, both MRDTP and MRDNP help in achieving good facial emotion recognition.

4.3.2 Dimension reduction and kernel parameters

PGK is used in the GSDRS dimension reduction system. The smoothness of the kernel is more dependable on σ than the ω parameter. Anyway, two parameters are necessary to create a

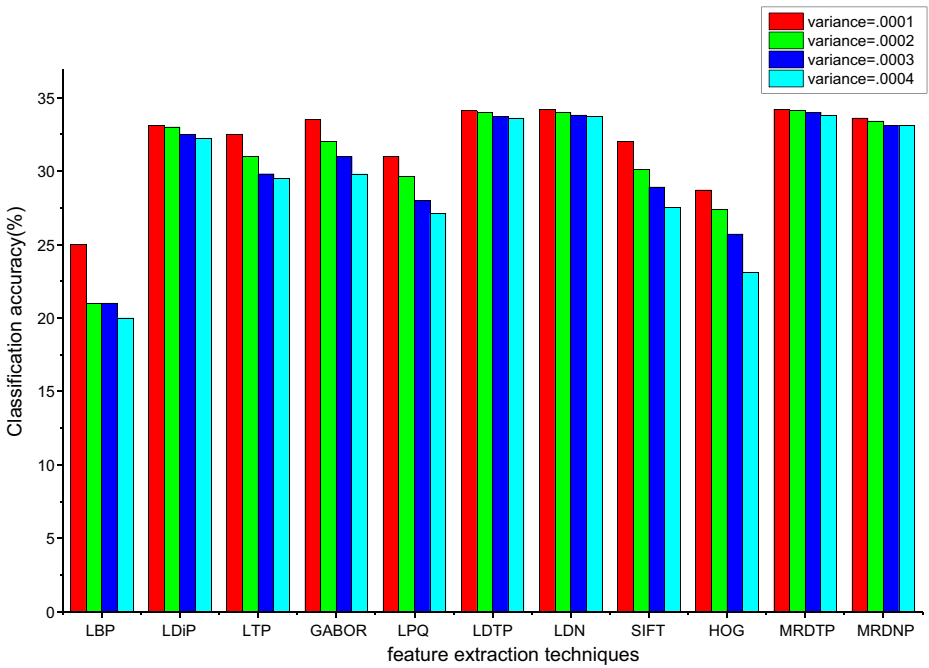


Fig. 17 Experimental results on SFEW images with added Gaussian white noise white noise of different variances for seven-class emotion recognition

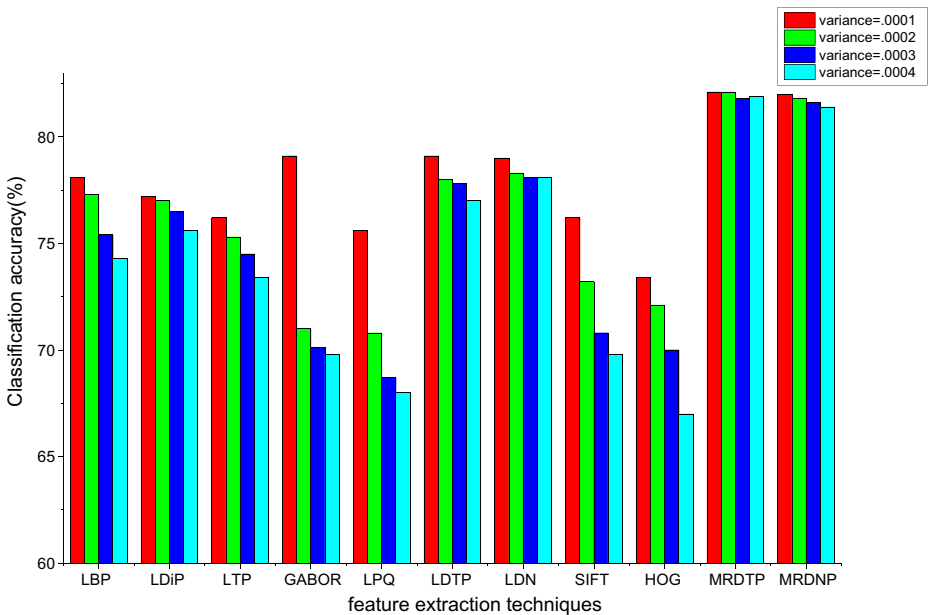


Fig. 18 Experimental results on MMI images with added Gaussian white noise white noise of different variances for seven-class emotion recognition

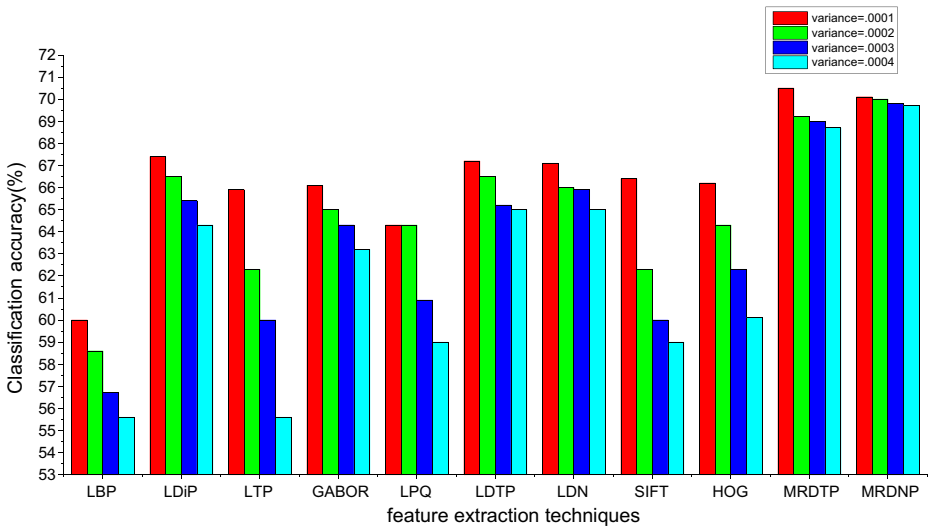


Fig. 19 Experimental results on DISFA images with added Gaussian white noise white noise of different variances for seven-class emotion recognition

space mapping that is similar to linear, polynomial, RBF kernel. GSDRS selects the best discriminative features and also brings together the features of the images that belong to same class together, while reducing the feature vector dimension.

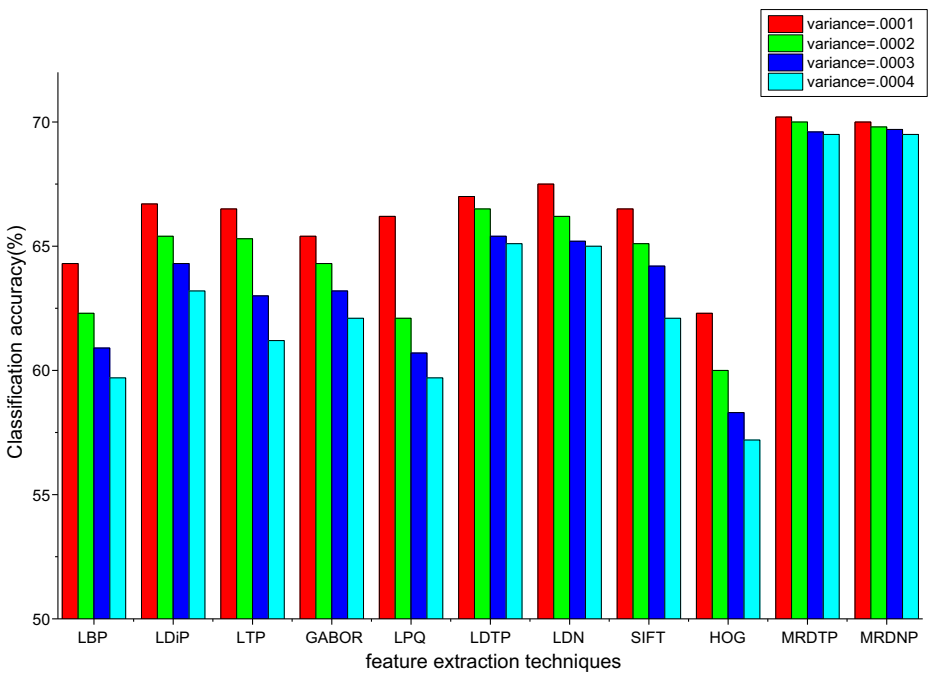


Fig. 20 Experimental results on combined DISFA & DISFA+ images with added Gaussian white noise white noise of different variances for seven-class emotion recognition

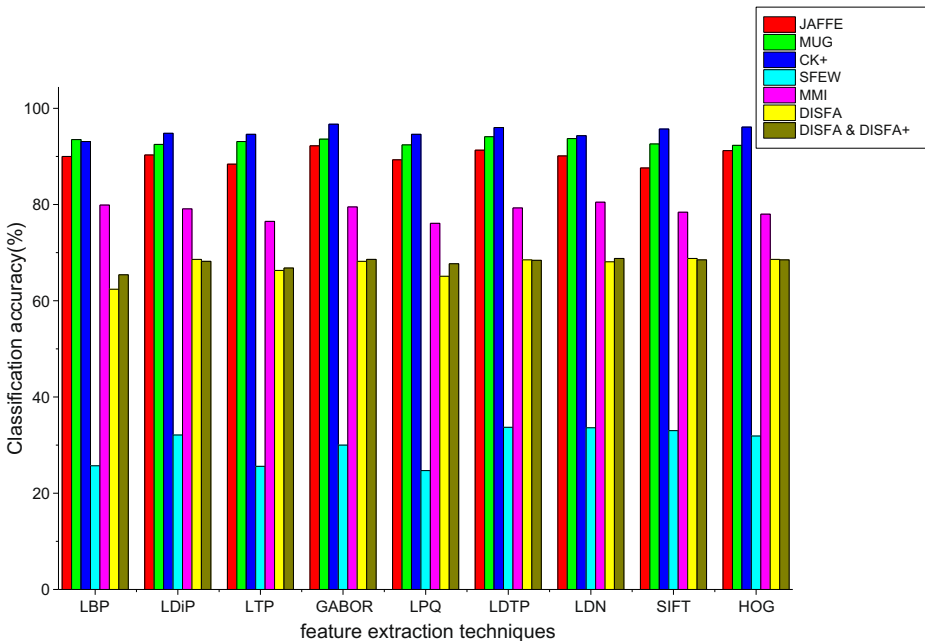


Fig. 21 Classification accuracies obtained after substituting different feature descriptors in proposed approach and applying GSDRS for seven-class emotion recognition

4.3.3 Comparison with other existing techniques

In Table 14 the classification accuracy of the existing other approaches in literature like SURF, LBP, LDTP, ELM space mapping, and Local Fisher Discriminant Analysis (LFDA) are specified with their experimental setups. LFDA has similar drawbacks as LDA, as not having S_w^{-1} in the presence of less number of training data. The proposed approach cannot be directly comparable with the results obtained by other existing approaches in literature as the experimental setups differ. Some of the approaches use deep learning methods for classification. But the deep learning methods are slower than ELM. JAFFE dataset is classified by Deep Belief Network of 300 hidden nodes by using 27.46 s [29] whereas the proposed approach uses 4.6 s while classifying using ELM-RBF. Even though no preprocessing techniques are used, MRDTP and MRDNP achieve good results for various datasets. The GSDRS system also achieves high recognition rate even under low dimension because of the selection of only the highly discriminative features with the help of the discrimination analysis done using training samples. MRDTP and MRDNP are very compact and low in complexity when compared to other existing techniques. The dimension reduction technique GSDRS also consumes less amount of time than other dimension reduction methods.

4.3.4 Time complexity

The computation complexity of LDTP is $O(MNKK + 21MN + 4PQ)$ where K is the size of Kirsch masks used, M and N are the number of rows and columns in the image, and P, Q are the sizes of the matrix grid from image used for subtraction. For LDN, the computation complexity is $O(MNKK + 9MN)$. The proposed MRDTP using code image has computation

Table 14 Performance comparison of different approaches

Author	Technique	No. of images	No. of classes	Dataset	Accuracy (%)
Iosifidis et al. (2015) [32]	ELM space mapping of input data	210 images	7	JAFFE	87.4(cross validation)
Hao and Tian (2017) [29]	Double Weber Local Descriptor with Deep Belief network	213 images	7	JAFFE	92.66(not specified)
Rivera et al. (2015) [52]	Local Directional Texture Pattern (LDTP) with PCA and SVM	213 images	7	JAFFE	89.2(cross validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	213 images	7	JAFFE	94.3(cross validation)
Iosifidis et al. (2015) [32]	ELM space mapping of input data	245 images	7	CK	92.9(cross validation)
Rivera et al.(2015) [52]	LDTP and SVM	1632 images	7	CK	68.69(cross validation)
Shan et al., (2009) [54]	LBP and SVM	1280 images	7	CK	95.7(cross-validation)
Yang and Bhanu (2012) [68]	Emotion avatar image with SVM classifier	316 sequences of images	7	CK+	88.9(cross-validation)
Xie and Hu (2017) [67]	Feature Redundancy-Reduced Convolutional Neural Network (FRCNN)	309 sequences of images	6	CK+	82.6 (leave-one-subject-out cross-validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	1281 images	7	CK+	92.06(cross validation)
Zia and Jaffar (2015) [74]	SURF descriptor with Nearest Neighbor (NN) classifier	1281 images	7	CK+	98.4(cross validation)
		Not specified	7	MUG	98.1(cross validation)
Agarwal et al. (2016) [2]	Entropy and correlation analysis with SVM	324 sequences of images	7	MUG	73.43(70–80% Images for training, remaining images for testing)
Rahulathavan et al. (2013) [48]	Local Fisher Discriminant Analysis in encrypted domain with NN classifier	1022 images	7	MUG	78.57(cross validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	567 images	7	MUG	95.24(leave-one subject out)
Wen et al. (2017) [65]	Probability-based fusion with Deep neural network	1238 images	7	MUG	95.7(cross validation)
Eleftheriadis et al. (2015) [24]	Discriminative shared Gaussian processes with NN classifier	700 images	7	SFEW	94.6(cross validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	1355 images	7	SFEW	34.09 (two independent folds)
					24.7 (two independent folds)
					34.8 (two independent folds)
					34.1 (two independent folds)

Table 14 (continued)

Author	Technique	No. of images	No. of classes	Dataset	Accuracy (%)
Kim et al. (2017) [37]	Deep generative-contrastive networks	312 sequences of images	6	MMI	81.53(cross validation)
Zhang et al. (2017) [72]	Spatial-Temporal Networks	205 sequences of images	6	MMI	81.18(cross validation)
Zhao et al. (2017) [73]	Local Binary Pattern and Gabor Multi-orientation Fusion with SVM classifier	203 sequences of images	6	MMI	71.92(cross validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	1050 images	7	MMI	82.1(cross validation)
Mollahosseini et al.(2016) [45]	MRDNP + GSDRS + ELM-RBF	1050 images	7	MMI	82.0(cross validation)
	Deep neural networks	89,000 images	7	DISFA	69.8%(cross validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	22,224 images	7	DISFA	70.7(cross validation)
	MRDNP + GSDRS + ELM-RBF	22,224 images	7	DISFA	70.3(cross validation)
Proposed method	MRDTP + GSDRS + ELM-RBF	31,524 images	7	DISFA & DISFA+	70.4 (cross validation)
	MRDNP + GSDRS + ELM-RBF	31,524 images	7	DISFA & DISFA+	70.2(cross validation)

complexity of $O(MNKK + MN + 8MNP^2)$, where $(2P + 1) \times (2P + 1)$ is the size of the Gaussian kernel. The proposed MRDNP using direction has a very low computation complexity of $O(MNKK + MN)$ when compared to the existing LDTP and LDN. For GSDRS, the computational complexity is $O(m^2)$, where m denotes the feature vector dimension. The existing GDA has $O(m + 2n^2)$, where m, n represent the feature vector dimensions. In GDA, an additional step is involved to find the optimum kernel among the three conventional kernels (Linear, Poly ($d = 1$) and a Gaussian RBF).

5 Conclusion and future enhancement

Two novel feature extraction techniques, namely MRDTP and MRDNP are proposed in this paper for extracting the features pertaining to emotions from face. Both the patterns are very simple, compact, and robust against noise. Edges detected using MRDTP are better than the Gabor edge detection techniques because of the noise restraining process. The major difference from other existing methods is that only the maximum response-based information is utilized instead of all the available information. This eliminates the inclusion of erroneous information that degrades the overall performance. Both the MRDTP and MRDNP achieve good classification results even at a reduced dimension of six, using the proposed GSDRS for seven-class emotion recognition. The classification accuracies obtained are above 92% for JAFFE, CK+, MUG datasets, above 70% for MMI, DISFA datasets and more than 34% for SFEW dataset. The accuracies achieved by the proposed techniques are better than the results stated in literature for facial emotion recognition. The proposed patterns perform better in the images disturbed with Gaussian white random noise too. This proposed work can also be extended on an audio and video-based emotion recognition application so that the information captured from audio also helps us to enhance the classification accuracy.

References

1. Abdulrahman M, Gwadabe TR, Abdu FJ, Eleyan A (2014) Gabor wavelet transform based facial expression recognition using PCA and LBP. In 2014 22nd Signal Processing and Communications Applications Conference (SIU), IEEE, pp 2265–2268
2. Agarwal S, Santra B, Mukherjee DP (2016) Anubhav: recognizing emotions through facial expression. *Vis Comput*. <https://doi.org/10.1007/s00371-016-1323-z>
3. Ahmed F, Hossain E (2013) Automated facial expression recognition using gradient-based ternary texture patterns. *Chin J Eng*. <https://doi.org/10.1155/2013/831747>
4. Ahmed F, Kabir MH (2012) Directional ternary pattern (DTP) for facial expression recognition. In *IEEE International Conference on Consumer Electronics*, pp 265–266
5. Aifanti, N, Papachristou C, Delopoulos A (2010) The MUG facial expression database. In *Proc. 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Desenzano, Italy, April 12–14
6. Anisetti M, Bellandi V (2009) Emotional state inference using face related features. In *New directions in intelligent interactive multimedia systems and services-2*. Springer Berlin Heidelberg, pp 401–411
7. Asthana A, Zafeiriou S, Cheng S, Pantic M (2014) Incremental face alignment in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 1859–1866
8. Bartlett MS, Littlewort G, Fasel I, Movellan JR (2003) Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. In *Computer Vision and Pattern Recognition Workshop, CVPRW'03*, pp 53–53
9. Baudat G, Anouar F (2000) Generalized discriminant analysis using a kernel approach. *Neural Comput* 12(10):2385–2404

10. Berretti S, Amor BB, Daoudi M, Del Bimbo A (2011) 3D facial expression recognition using SIFT descriptors of automatically detected key points. *Vis Comput* 27(11):1021–1036
11. Bhat FA, Wani MA (2016) Elastic bunch graph matching based face recognition under varying lighting, pose, and expression conditions. *IAES International Journal of Artificial Intelligence (IJ-AI)* 3(4):177–182
12. Bourbakis N, Esposito A, Kavradi D (2011) Extracting and associating meta-features for understanding people's emotional behaviour: Face and speech. *Cogn Comput* 3(3):436–448
13. Calder AJ, Burton AM, Miller P, Young AW, Akamatsu S (2001) A principal component analysis of facial expressions. *Vis Res* 41(9):1179–1208
14. Castillo JA, Rivera AR, Chae O (2012) Facial expression recognition based on local sign directional pattern. In 19th IEEE International Conference on Image Processing, pp 2613–2616
15. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3):27
16. 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, pp 884–888
17. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn*:20273–20229
18. Dhall A, Goecke R, Lucey S, Gedeon T (2011) Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In Proc. Int. Conf Comput. Vis. Workshops, pp 2106–2112
19. Dhall, A, Asthana A, Goecke R, Gedeon T (2011) Emotion recognition using PHOG and LPQ features. In Automatic Face & Gesture Recognition and Workshops (FG 2011), IEEE International Conference, pp 878–883
20. Dhall A, Goecke R, Lucey S, Gedeon T (2012) Collecting Large, Richly Annotated Facial Expression Databases from Movies. *IEEE MultiMedia* 19:34–41
21. Dhall A, Goecke R, Joshi J, Sikka K, Gedeon T (2014) Emotion recognition in the wild challenge 2014: baseline, data and protocol, *ACM ICMI* 2014
22. Ekman P (2004) Emotional and conversational nonverbal signals. In Language, knowledge, and representation, Springer Netherlands, pp 39–50
23. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. *J Pers Soc Psychol* 17(2):124
24. Eleftheriadis S, Rudovic O, Pantic M (2015) Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition. *IEEE Trans Image Process* 24(1):189–204
25. Ghimire D, Lee J, Li ZN, Jeong S (2016) Recognition of facial expressions based on salient geometric features and support vector machines. *Multimedia Tools and Applications* 15:1–26
26. Gupta SK (1998) Peak decomposition using Pearson type VII function. *J Appl Crystallogr* 31(3):474–476
27. Haghghat M, Zonouz S, Abdel-Mottaleb M (2015) CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Syst Appl* 42(21):7905–7916
28. Hamester D, Barros P, Wermter S (2015) Face expression recognition with a 2-channel Convolutional Neural Network. In 2015 International Joint Conference on Neural Networks (IJCNN), pp 1–8
29. Hao XL, Tian M (2017) Deep belief network based on double weber local descriptor in micro-expression recognition. In Advanced Multimedia and Ubiquitous Engineering May 22. Springer, Singapore, pp 419–425
30. Huang GB, Siew CK (2005) Extreme learning machine with randomly assigned RBF kernels. *Int J Inf Technol* 11(1):16–24
31. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. Part B: *IEEE Transactions on Systems, Man, and Cybernetics* 42(2):513–529
32. Iosifidis A, Tefas A, Pitas I (2015) On the kernel extreme learning machine classifier. *Pattern Recogn Lett* 54:11–17
33. Jabid T, Kabir MH, Chae O (2010) Robust facial expression recognition based on local directional pattern. *ETRI J* 32(5):784–794
34. Jolliffe I (2002) Principal component analysis. Wiley, New York
35. Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. Proceedings of Fourth IEEE International Conference in Automatic Face and Gesture Recognition, pp 46–53
36. Kim Y, Lee H, Provost EM (2013) Deep learning for robust feature generation in audiovisual emotion recognition. In 2013 I.E. International Conference on Acoustics, Speech and Signal Processing, pp 3687–3691
37. Kim Y, Yoo B, Kwak Y, Choi C, Kim J (2017) Deep generative-contrastive networks for facial expression recognition, arXiv preprint arXiv:1703.07140
38. Kirsch RA (1971) Computer determination of the constituent structure of biological images. *Comput Biomed Res* 4(3):315–328
39. Kotsia I, Pitas I (2007) Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans Image Process* 16(1):172–187
40. 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. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 94–101
41. Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with gabor wavelets. Third IEEE International Conference on Automatic Face and Gesture Recognition, pp 200–205

42. Mavadati SM, Mahoor MH, Bartlett K, Trinh P (2012) Automatic detection of non-posed facial action units. In *Image Processing (ICIP), 19th IEEE International Conference on* 2012 Sep 30 IEEE, pp 1817–1820
43. Mavadati SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF (2013) Disfa: A spontaneous facial action intensity database. *IEEE Trans Affect Comput* 4(2):151–160
44. Mavadati M, Sanger P, Mahoor MH (2016) Extended DISFA Dataset: Investigating Posed and Spontaneous Facial Expressions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp 1–8
45. Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. In *Applications of Computer Vision (WACV), 2016 I.E. Winter Conference*, pp 1–10
46. Ojansivu V, Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*, pp 236–243
47. Pantic M, Valstar M, Rademaker R, Maat L (2005) Web-based database for facial expression analysis. In *Multimedia and Expo, ICME 2005. IEEE International Conference on* Jul 6 IEEE, pp 5
48. Rahulmathavan Y, Phan RC, Chambers JA, Parish DJ (2013) Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. *IEEE Trans Affect Comput* 4(1):83–92
49. Ramirez Rivera A, Rojas Castillo J, Chae O (2013) Local directional number pattern for face analysis: Face and expression recognition. *IEEE Trans Image Process* 22(5):1740–1752
50. Rivera AR, Castillo JA, Chae O (2012) Recognition of face expressions using local principal texture pattern. In *19th IEEE International Conference on Image Processing*, pp 2609–2612
51. Rivera AR, Rojas J, Chae O (2012) Local gaussian directional pattern for face recognition. In *Pattern Recognition (ICPR), 21st International Conference*, pp 1000–1003
52. Rivera AR, Castillo JR, Chae O (2015) Local directional texture pattern image descriptor. *Pattern Recogn Lett* 51:94–100
53. Shan C, Gong S, McOwan PW (2005) Appearance manifold of facial expression. In *Computer Vision in Human-Computer Interaction*, pp 221–230
54. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis Comput* 27(6):803–816
55. Suja P, Tripathi S, Deepthy J (2014) Emotion recognition from facial expressions using frequency domain techniques. In *Advances in signal processing and intelligent recognition systems*, pp 299–310
56. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process* 19(6):1635–1650
57. Tao J, Tan T (2005) Affective computing: A review. In *Affective computing and intelligent interaction*, pp 981–995
58. Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3(1):71–86
59. Üstün B, Melssen WJ, Buydens LMC (2006) Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. *Chemom Intell Lab Syst* 81(1):29–40
60. Valstar M, Pantic M (2010) Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect* May 21, p 65
61. Valstar MF, Pantic M (2012) Fully automatic recognition of the temporal phases of facial actions. *IEEE Transactions on Systems, Man, and Cybernetics* 42(1):28–43
62. Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vis* 57(2):137–154
63. Vo A, Ly NQ (2015) Facial expression recognition using pyramid local phase quantization descriptor. In *Knowledge and Systems Engineering*, Springer International Publishing, pp 105–115
64. Wang H, Huang H, Makedon F (2014) Emotion detection via discriminant laplacian embedding. *Univ Access Inf Soc* 13(1):23–31
65. Wen G, Hou Z, Li H, Li D, Jiang L, Xun E (2017) Ensemble of deep neural networks with probability-based fusion for facial expression recognition, cognitive computation, pp 1–4
66. Wu T, Bartlett MS, Movellan JR (2010) Facial expression recognition using gabor motion energy filters. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp 42–47
67. Xie S, Hu H (2017) Facial expression recognition with FRR-CNN. *Electron Lett* 53(4):235–237
68. Yang S, Bhanu B (2012) Understanding discrete facial expressions in video using an emotion avatar image. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42(4):980–992
69. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 26(1):131–137
70. Yang J, Frangi AF, Yang JY, Zhang D, Jin Z (2005) KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. *IEEE Trans Pattern Anal Mach Intell* 27(2):230–244
71. Zhang B, Zhang L, Zhang D, Shen L (2010) Directional binary code with application to PolyU near-infrared face database. *Pattern Recogn Lett* 31(14):2337–2344
72. Zhang K, Huang Y, Du Y, Wang L (2017) Facial expression recognition based on deep evolutionary spatial-temporal networks. *IEEE Trans Image Process* 26(9):4193–4203

73. Zhao L, Wang Z, Zhang G (2017) Facial expression recognition from video sequences based on spatial-temporal motion local binary pattern and gabor multiorientation fusion histogram. *Math Probl Eng*. <https://doi.org/10.1155/2017/7206041>
74. Zia MS, Jaffar MA (2015) An adaptive training based on classification system for patterns in facial expressions using SURF descriptor templates. *Multimedia Tools and Applications* 74(11):3881–3899



A. Sherly Alphonse has received her B.E degree from Manonmaniam Sundaranar University, India and M.E degree from Anna University, India. She has five years of teaching experience in various prestigious institutions. She is currently pursuing her full time Ph.D. at Anna University, Chennai, India. Her research interest includes image processing and facial expression analysis.



Dr. Deje received her B.E. and M.E. degrees in Computer Science and Engineering from Manonmaniam Sundaranar University, Tirunelveli, India, in 2003 and 2005 respectively. Later, she was with the Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India, as a Junior Research Fellow under the UGC Research Grant. She completed her Ph.D. in Computer Science and Engineering in 2011. She has been with the Department of Computer Science and Engineering, Regional Campus, Anna University: Tirunelveli Region as an Assistant Professor since 2010 and as the Head of the Department from 2011 to 2015. She is a member in IEEE, IET and ISTE. Her research interests include image and signal processing, watermarking, information hiding and multimedia security.