Cluster Comput (2018) 21:549–567
https://doi.org/10.1007/s10586-017-0935-z

Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery

Muhammad Sajjad[1] · Adnan Shah[1] · Zahoor Jan[1] · Syed Inayat Shah[2] · Sung Wook Baik[3] · Irfan Mehmood[4]

Received: 6 January 2017 / Revised: 5 March 2017 / Accepted: 28 March 2017 / Published online: 3 June 2017

Abstract Facial sentiment analysis has been an enthusiastic research area for the last two decades. A fair amount of work has been done by researchers in this field due to its utility in numerous applications such as facial expression driven knowledge discovery. However, developing an accurate and efficient facial expression recognition system is still a challenging problem. Although many efficient recognition systems have been introduced in the past, the recognition rate is not satisfactory in general due to inherent limitations including light, pose variations, noise, and occlusion. In this paper, a hybrid approach of facial expression based sentiment analysis has been presented combining local and global features. Feature extraction is performed fusing the histogram of oriented gradients (HOG) descriptor with the uniform local ternary pattern (U-LTP) descriptor. These features are extracted from the entire face image rather than from individual components of faces like eyes, nose, and mouth. The most suitable set of HOG parameters are selected after analyzing them experimentally along with the ULTP descriptor, boosting performance of the proposed technique over face images containing noise and occlusions. Face sentiments are analyzed classifying them into seven universal emotional expressions: Happy, Angry, Fear, Disgust, Sad, Surprise, and Neutral. Extracted features via HOG and ULTP are fused into a single feature vector and this feature vector is fed into a Multi-class Support Vector Machine classifier for emotion classification. Three types of experiments are conducted over three public facial image databases including JAFFE, MMI, and CK+ to evaluate the recognition rate of the proposed technique during experimental evaluation; recognition accuracy in percent, i.e., 95.71, 98.20, and 99.68 are achieved for JAFFE, MMI, and CK+, respectively.

Keywords Facial expression recognition · Sentiment based knowledge discovery · Histogram of oriented gradient · Uniform local ternary pattern · Support vector machine

1 Introduction

Facial expression is a very important communicative source in interpersonal relations. Facial expressions and other gestures convey non-verbal cues in face-to-face interactions. These cues may also complement speech helping the listener to elicit the intended meaning of spoken words. According to Mehrabian [1], in face to face communication among human beings, 7% of the message pertains to feelings and attitudes in the words spoken, 38% is in the way the words are said, and 55% is in facial expressions. Facial expressions are a universal and natural form of non-verbal communication. They help in showing emotions among human beings for better communication. Automatic face expression recognition systems have a variety of applications in several research and developmental areas such as lie detection, behavior analysis, surveillance systems, transportation, and robotics. The

✉ Irfan Mehmood
irfan@sejong.ac.kr; irfanmehmood@ieee.org;
irfan.memhood@live.com

[1] Digital Image Processing Laboratory, Department of Computer Science, Islamia College Peshawar, Peshawar, Pakistan

[2] Department of Mathematics, Islamia College Peshawar, Peshawar, Pakistan

[3] Digital Contents Research Institute, Department of Software, Sejong University, Seoul, South Korea

[4] Department of Computer Science and Engineering, Sejong University, Seoul, South Korea

importance of a robust expression recognition system is evident with the advances in robotics, especially humanoid robots. As robots start becoming a part of our living and work spaces, they need to become more intelligent in terms of understanding the moods and emotions of humans. A FER system will help in creating intelligent visual interfaces between humans and machines, facilitating human computer interaction (HCI). Moreover, many applications such as customer satisfaction studies for broadcast, video conferencing, user profiling, driver's mood detection systems, and virtual reality require efficient and powerful FER systems. Huge amounts of work has been done so far in the field of automatic FER systems due to its importance, hence, various techniques proposed for FER can be found in existing literature. A facial action coding system (FACS) has been incorporated in numerous research works to identify facial expressions. Ekman et al. [2] established a FACS capable of perceiving various emotions via the contraction and relaxation of different facial muscles both individually and simultaneously. FACS describes muscle movement as action units (AUs), where each unit is composed of letters and digits. Different combinations of AUs can label any facial expression to identify the mood of an individual. For example, the combination of AU4, AU5, AU7, and AU23 describes the angry mood of the underlying image.

Analyzing facial expressions directly from images can be categorized into two parts, i.e., geometry-based approaches and appearance-based approaches. In geometrical feature-based approaches, various facial points known as fiducial points are tracked and localized prior to the extraction process, and then these points are extracted as a feature vector representing facial expressions [3]. The active appearance model (AAM) and its variations are used by most of the geometric feature-based approaches. However, in the geometric-based approach [4], the accurate detection and tracking of a dense set of facial points is very difficult to adjust in many situations. For example, eye brow and mouth corner positions vary from mood to mood; hence, building a dynamic system for this problem is difficult. In appearance-based methods, dominant features based on facial appearance variations are extracted either from the whole face or from the region of interest (ROI) using different feature extraction filters. Both face and facial expression recognition systems incorporate various local and global universal features extracted through different feature extraction techniques from the whole face or from the ROI [5,6].

A majority of FER systems comprise of three key parts as shown in Fig. 1. In the pre-processing step, an image enhancement technique is used to enhance the quality of the input image by removing noise, various blurring effects, and degradation artifacts [7]. After the image enhancement step, the face and its components such as eyes, eye brows, cheeks, nose, and mouth are then detected in the input image.

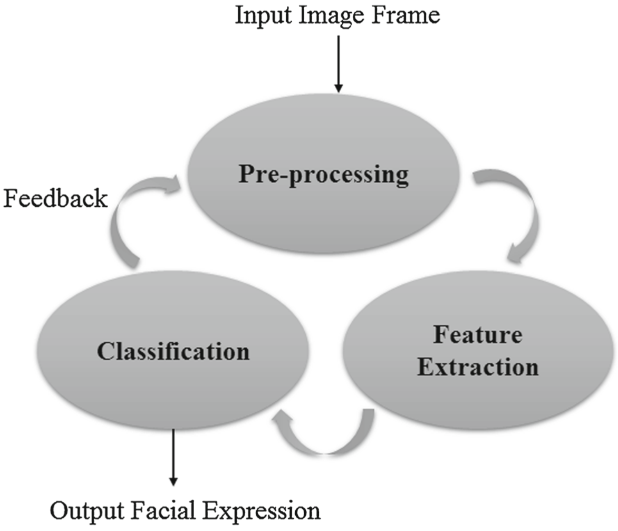


Fig. 1 General framework of facial expressions recognition systems

In most of the face recognition and FER systems, face detection and its sub-parts is usually done by existing state-of-the-art face detection algorithms such as the Viola-Jones object/face detection algorithm presented by Viola and Jones [8,9]. This algorithm comparatively provides more accuracy in in detecting faces and their components in real time over other state-of-the-art techniques. The ROI is then cropped and resized into specified dimensions, considering the availability of various resources such as computational, storage, and transmission resources. In the second step, i.e. feature extraction, raw pixel data is converted to a comprehensible representation of color, shape, or texture information. Various feature extraction methods such as the Local Binary Pattern (LBP) [4], Local Directional Number Pattern (LDN) [10], Gabor features, Scale-invariant feature transform (SIFT) [11], and Principal Component Analysis (PCA) [12] can be used for feature extraction from the ROI of the input image. In the third step, features are labeled according to the respective facial expression classes and then classification is performed over the labeled features. There is vast category of classifiers for face emotion recognition such as the Support Vector Machine (SVM) [13], K-Nearest Neighbor (KNN) [14], Logistic regression, Decision tree, and Naive Bayes.

The proposed FER framework considers both local and global descriptors to extract features from images containing human faces. For this purpose, both a histogram oriented gradient (HOG) descriptor in combination with a uniform-local ternary operator (U-LTP) are incorporated to extract features into a single feature vector. Instead of extracting features from individual components of faces like eyes, nose, and mouth as illustrated in the literature, we extract features using HOG and U-LTP from the whole cropped face, describing appearance, shape, and textural variations of the underlying face in the image. This is because in facial expression analysis, even a small part of a face can play an important role in expression recognition. Thus, extracting features from only individual components of the face results in losing a significant amount of information involved in facial expressions. The main contribution of the proposed work is the integration of features describing shape, appearance, and textural variations extracted via HOG and U-LTP descriptors respectively. This combination also induces the perception of local and global features as a single entity, which complements the weakness of the local and global features while improvising the generation of a stronger and more robust feature vector. The extracted feature vector is then fed to SVM for classification. Keeping in view the heterogeneity of human faces and variety in their expressions, a multiclass-SVM is adopted to produce a more accurate and robust FER system.

The main contributions of the proposed framework are as follows:

- A fully automated face expression recognition framework is presented which is robust to various real environment elements such as noise, illumination changes, and partial overlapping or occlusion.
- Histogram oriented gradient (HOG) descriptor in combination with uniform-local ternary operator (U-LTP) are presented to extract more robust features. This extracts comprehensible important features from the human face, thereby increasing face expression recognition accuracy.
- This combination of HOG and U-LTP induces the perception of local and global features as a single entity, which complements the weakness of the local and global features while improvising the generation of a stronger and more robust feature vector.

The rest of this paper is ordered as: Sect. 2 surveys related work. Section 3 depicts the proposed method. The experiments and results are discussed in Sect. 4, and finally, Sect. 5 concludes the paper.

2 Related work

This section discussed previous state-of-the-art works in the field of FER. Some of the extensive research efforts in this field have been made in the early 1990s, and since then a fair amount of research techniques have been developed due to the ever growing popularity of FER applications.

Zhang et al. [3] systematically represents facial visual signs at different levels of abstraction by combining the Dynamic Bayesian Networks (DBN) with Ekman's Facial Action Coding System (FACS) for modeling and understanding the temporal, dynamic, and stochastic behaviors of facial expressions in a series of images. Active and dynamic information was fused to increase robustness and efficiency in facial expression analysis. The recognition of facial expressions is done considering current as well as previous visual observations of the faces contained in the series of images. Lee et al. [15] detected faces via skin color using the YCbCr color space model. Then sub-components of the underlying face such as eyes and mouth are detected and considered as ROIs. Further, points are extracted from these ROIs by applying a Bezier curve. Recognition of facial expressions is performed measuring the Hausdorff distance with the Bezier curve between training and testing features. Al-Shabi et al. [16] presented a model integrating CNN and SIFT which was trained for facial expression recognition acquiring only a small sample dataset. Multiple deep neural networks and a hybrid CNN-SIFT classifier were combined to generate an effective classification model. Wang et al. [17] recognized expressions on the basis of Fuzzy Support Vector Machine (FSVM) and KNN classifiers, in which features are extracted from the static face of the input image using PCA and then the region is divided into different types. The extracted labeled feature vectors are passed to the classifier based on the combined characteristics of FSVM and KNN. Tong et al. [18] presented a new Local Gradient Code (LGC) algorithm dividing face images into several blocks and then separately calculating gradient binary coding for each block in the horizontal, vertical, and diagonal. They further compute the histogram of the LGC statistics and finally link the histogram for each block together to form a feature vector. The reduction in computational complexity is achieved by a new LGC operator based on the horizontal and diagonal gradient (LGC-HD) while preserving the main information contained in the texture of the face expressions. Luo et al. [19] used PCA to extract global features of the facial region contained in the input image. In this work, authors introduced a novel approach consisting of an integration of PCA with local binary pattern (LBP) coping the environmental sensitivity of the features extracting through PCA. The LBP extracts local features of the mouth region because it contributes most in expression recognition, counter accelerating the effect of global features in facial expression recognition. A hybrid feature-based approach is presented by Happy and Routary [20], in which two types of feature are extracted, i.e., pyramid of histogram of gradients (PHOG) and LBP describing shape and appearance of the underlying face respectively. Hybrid features are extracted localizing active patches of the face after landmark detection. These patches capture key changes of facial expressions showing different sentiments. Classification is performed over SVM after dimensionality reduction through Linear Discriminant Analysis (LDA). Carcagni et al. [21] evaluates the performance of HOG in FER by varying parameter values. For facial expression recognition, Sunil Kumar et al. [22] presented a method for the extraction of informative regions of a face. The authors' approach entails extracting discriminative features from the informative regions of a face. They proposed an informative region extraction model, which models the importance of facial regions based on the projection of expressive face images onto neural face images. Yulan Guo et al. presented an expression-invariant 3D face recognition based on feature and shape matching [23]. Each 3D face is first automatically detected from raw 3D data and normalized to achieve pose invariance. The 3D face is then represented by a set of keypoints and their associated local feature descriptors to achieve robustness to expression variations. During face recognition, a probe face is compared against each gallery face using both local feature matching and 3D point cloud registration. Enthused by the previous efforts, we present a novel robust FER framework in which local and global features are extracted considering human face texture and appearance properties. The extracted feature vector is then used in Multiclass-SVM for accurate expression detection and classification.

3 Proposed method

In this section, the processing steps involved in the proposed methodology are described in detail. In the first step, a database $\mathcal{D}$is selected and divided into training $\mathcal{D}_{\mathrm{TR}} = \left\{\mathcal{I}_1^T, \ldots, \mathcal{I}_{3N/4}^T\right\}$ and test sets, i.e., cross-validation test sets $\mathcal{D}_{\mathrm{C}} = \left\{\mathcal{I}_{(3N/4)+1}^C, \ldots, \mathcal{I}_N^T\right\}$ where N is the total number of images in $\mathcal{D}$. In the training phase, an image $\mathcal{I}_i^T$ is taken from $\mathcal{D}_{\mathrm{TR}}$ and the quality of this image is enhanced applying image enhancement techniques such as histogram equalization and median filters to obtain an image $\mathcal{I}_i^E$ free from noise and blurring artifacts. Further, a well-known Viola & Jones face detection technique [8] is used to detect the face region in the underlying image $\mathcal{I}_i^E$. After detection, the face region is cropped into an image $\mathcal{I}_i^F$. In the feature extraction step, HOG and U-LTP descriptors are used to extract features from the cropped image $\mathcal{I}_i^E$ into feature vectors $\mathcal{F}_i^{HOG}$ and $\mathcal{F}_i^{U-LTP}$ respectively. These features vectors $\mathcal{F}_i^{HOG}$ and $\mathcal{F}_i^{U-LTP}$ are then fused into a single feature vector $F_i^{(HOG+U-LTP)}$. The feature vector $\mathcal{F}_i^{(HOG+U-LTP)}$ is extracted for all $3N/4$images in $\mathcal{D}_{TR}$ and then these $3N/4$ feature vectors are labeled according to their corresponding label of expression $\mathcal{L}_J$ ={$\mathcal{L}_1$, ..., $\mathcal{L}_7$}, where $\mathcal{J}$ denotes seven standard facial expressions. The labeled feature vectors are fed into a multi-classifier SVM Ç to effectively train the underlying machine. In the testing phase, i.e., cross validation phase, an image $\mathcal{I}_i^C$ is taken from the datasets from $\mathcal{D}^C$, following the same procedure for image enhancement, detection, and cropping of the face region from the input image, and used in feature extraction as followed in the training phase. The $N/3$ extracted feature vectors $\overset{\frown}{\mathcal{F}}_i^{(HOG+U-LTP)}$: are cross-validated against their labels $\mathcal{L}_J$ over the trained classifier Ç. The cross-validation further optimizes the prediction ratio of the proposed framework. The overall framework of the proposed system is depicted in Fig. 2, comprising of three main stages. A description of the modle parameter of the proposed framework is given in Table 1.

3.1 Pre-processing of images

In the pre-processing step an input image $\mathcal{I}$ is further processed to enhance its quality. Initially, $\mathcal{I}$ may contain noise or other type of blurring elements that may reduce recognition accuracy. Therefore, to eliminate noise data and preserve important information, a median filter of size 3×3 is applied to the input image $\mathcal{I}$. It replaces each pixel in the neighborhood with the median value, which helps to get rid of those noises similar to salt and pepper noise without reducing the sharpness of the output image $\mathcal{I}$. Similarly, the recognition rate degrades when low resolution or low contrast images are used. For this purpose, a histogram equalization technique is applied to enhance image contrast and normalize the illumination effects. After filtering and histogram equalization, the face is detected in image $\mathcal{I}^E$. After face detection, it is cropped into image $\mathcal{I}^F$ and resized to 128×128 using our recent adaptive interpolation technique based on multi-kernel image super-resolution scheme [24]. Face detection, ROI extraction, and resizing of ROI is shown in Fig. 3.

3.2 Feature extraction

In the proposed technique, two types of feature descriptors based on appearance and shape information are used to extract dominant features from the face image. These features are then fused to form a 1-D feature vector.

3.2.1 Histogram of oriented gradients (HOG)

HOG, developed by Dalal and Triggs [25], is a popular global descriptor like SIFT [26] and SURF [27] which is used for object detection. The algorithm counts the edge's direction and visibility of a pixel, i.e., how many pixels in a localized slice of an image $\mathcal{I}^F$ have an edge passing it with a specific direction. In experimental evaluation, several variants of HOG descriptors are analyzed with different spatial associations, gradient computation, and different normalization techniques. HOG is a robust feature extraction technique that extracts features describing each pixel in the ROI of the underlying image.

The vital notion of the HOG descriptor is that local face appearance and shape features within an image can be described by the scattering of intensity gradients or edge

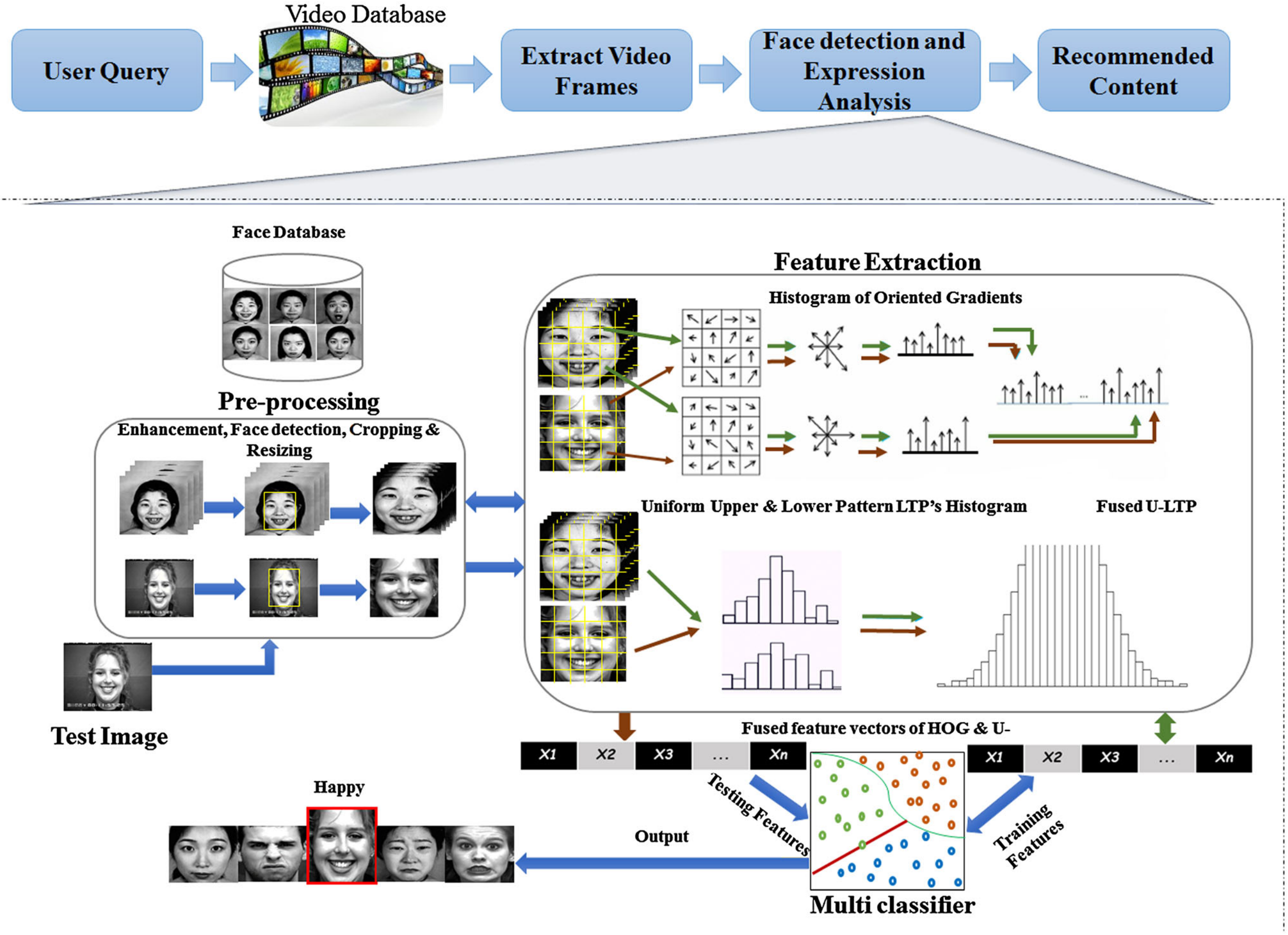


Fig. 2 Proposed facial expression analysis framework

Table 1 Description of the model parameters in the proposed work

$\mathcal{D}$: database of face images

$\mathcal{D}_{TR}$: Face image dataset for training

$\mathcal{D}_{C}$: Face image dataset for cross-validation

$\mathcal{I}_i^T$: Training image

$\mathcal{I}_i^E$: Enhanced image

$\mathcal{I}_i^F$: Image with cropped face

$\mathcal{F}_i^{HOG}$: HOG based feature vector extracted from $\mathcal{I}_i^F$

$\mathcal{F}_i^{U-LTP}$: U-LTP based feature vector extracted from $\mathcal{I}_i^F \mathbf{I}_{FACE}$

$\mathcal{F}_i^{(HOG+U-LTP)}$: Denoting fusion of feature vectors $\mathcal{F}_i^{HOG}$and $\mathcal{F}_i^{U-LTP}$

$\mathcal{L}_J$: Vector containing facial expressions

$\widehat{\mathcal{F}}_i^{(HOG+U-LTP)}$: Feature vector extracted for each image containing in $\mathcal{D}_C$

Ç: SVM based classifier

directions. The implementation criteria of HOG can be achieved by dividing an image into small spatial portions referred to as cells. To compute the gradient orientations for each pixel within a cell, x and y derivatives are computed for each pixel in a cell. Masks for calculating x and y derivatives are shown in Eq. 1. The x and y derivatives are calculated by convolving masks as in Eq. 2 over the whole image. Then gradient magnitude and angle are computed for each pixel in the ROI of the image $\mathcal{I}^F$ as in Eqs. 3 and 4 respectively, and are quantized into a 9 bins histogram forming a feature vector for the underlying cell in the ROI. To improve accuracy while keeping in view lightening variations and shadow problems, the preprocessed image $\mathcal{I}^F$ is divided into blocks with each block containing a fixed number of cells. The histograms computed for each cell in the block are concatenated to form a single vector presenting the corresponding block. Further, the resultant vector is normalized through a contrast normalization technique based on its energy (that is, regularized

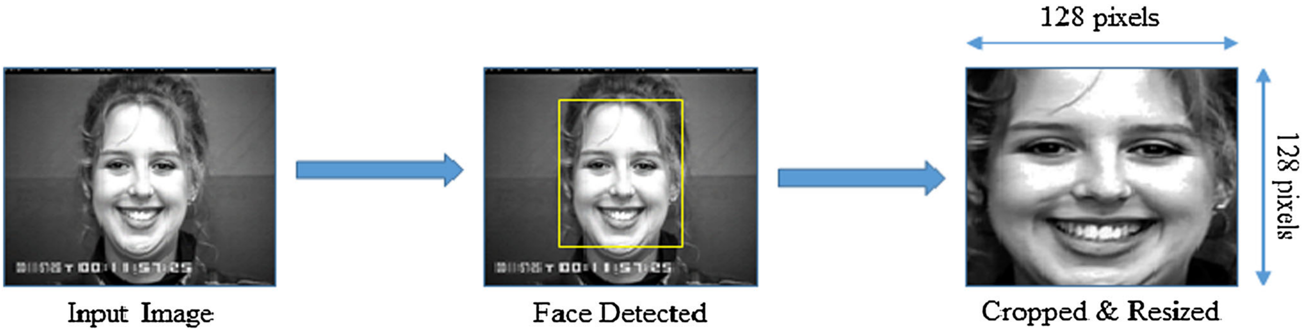


Fig. 3 Face detection, cropping the ROI, and then resizing through our recent adaptive interpolation technique based on a Multi-kernel Image Super-resolution scheme [24]

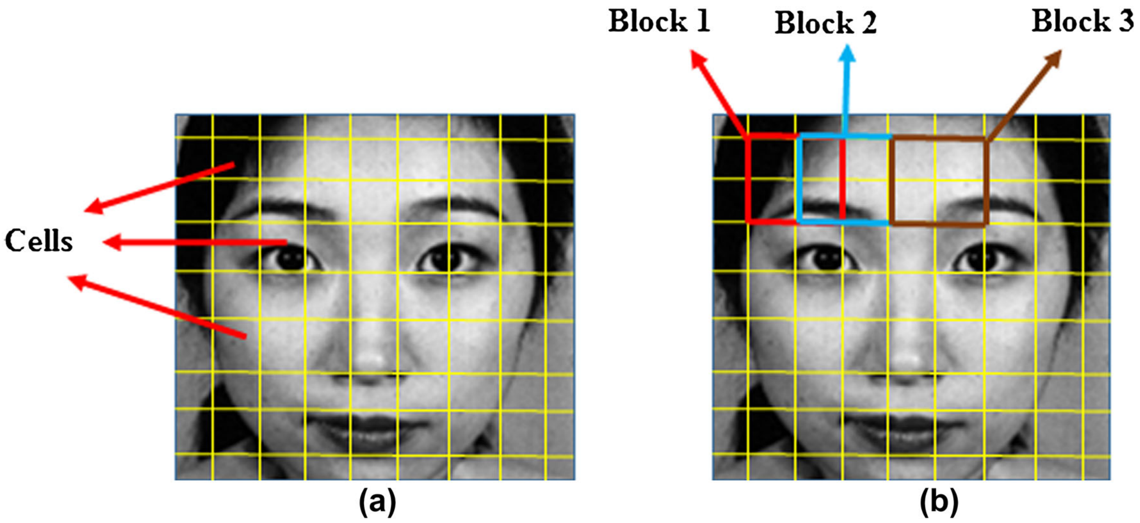


Fig. 4 **a** Dividing an image into cells **b** Each block contains 2 × 2 number of cells, with 50% overlapping with its neighboring block

L1 or L2 norm). Figure 4 describes the procedure of image division into blocks and blocks into cells. The same normalization procedure is repeated for all the blocks in image $\mathcal{I}^F$. Finally, a histogram generated from all blocks in image $\mathcal{I}^F$ are combined into a global feature vector.

In gradient computation, the centered horizontal and vertical gradients of each pixel are computed. Dalal and Trigg et al. [25] analyzed different kinds of masks for gradient calculation including one-dimensional point derivatives (namely un-centered [−1, 1], centered [−1, 0, 1], and cubic [1, −8, 0, 8, −1]). According to Dalal and Trigg [25], the most convenient and effective way to apply a one-dimensional centered discrete derivative mask in both horizontal and vertical directions is

$$D_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \quad and \quad D_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{1}$$

The convolution operation in each direction is given by

$$\mathcal{I}_x^F = \mathcal{I}^F \otimes D_x \quad and \quad \mathcal{I}_y^F = \mathcal{I}^F \otimes D_y \tag{2}$$

where $\mathcal{I}^F$ is a preprocessed image containing the ROI, and $\mathcal{I}_x^F$ and $\mathcal{I}_y^F$ are the x and y derivatives of an image $\mathcal{I}^F$. HOG features descriptor is composed of gradient magnitude |M| and edge direction θ, where the magnitude and orientation components of the gradients are calculated as

$$|M| = \sqrt{(\mathcal{I}_x^F)^2 + (\mathcal{I}_Y^F)^2} \tag{3}$$

$$\theta = \arctan \frac{\mathcal{I}_Y^F}{\mathcal{I}_x^F} \tag{4}$$

Before gradient magnitude |M| and edge direction θ computation, image $\mathcal{I}^F$ is divided into $\mathcal{K}$ blocks, where each block is further divided into a specified number of cells of equal size. The histogram of sum of magnitudes of a pixel is then stored in orientation bins over a radial or rectangular cell. These frequencies are either equally spread over 0 – 180° (un-signed gradients) or 0–360° (signed gradients) and usually quantized into 9 bins, each having 20°.

Finally, a nine-bin histogram computed for each cell of the blocks in image $\mathcal{I}^F$ are fused into a single vector. There are n-cells in each block. In addition, the histogram of gradient orientations is normalized to suppress lighting variations, improving the overall performance of the proposed technique. Dalal et al. [25] reported two types of blocks, that is, rectangular blocks (R-HOG) and circular blocks (C-HOG). R-HOG blocks are composed of square grids denoted by three parameters: the number of cells per block, size of the cells, and number of bins in the histogram for each cell of the block. The block normalization is computed dividing the resultant 9 × n-sized vector by the formula given in Eq.5, where C_i^h denotes the histogram for each cell in the block.

$$\sqrt{\left(C_1^h\right)^2 + \left(C_2^h\right)^2 \ldots + \left(C_n^h\right)^2} \quad (5)$$

In experimental evaluation, Sec. 4, different sized cells (e.g., 8 × 8, 12 × 12, and 16 × 16), different sized blocks (e.g., 2 × 2), and size of overlapping blocks with different orientation bins are evaluated as shown in Fig 4. The accuracy and effectiveness of cell size 16 ×16, block size 2 × 2, blocks with fifty percent overlapping, and nine bins in orientation histograms are proved experimentally. The following factors play a vital role in the input value selection of these parameters.

- To capture spatial information at a larger-scale, a suitable size for the cell is chosen. It is proved experimentally that a cell with proper size improves the effectiveness of the HOG descriptor.
- Block size also plays an important role in locally controlling the impact of illumination variation while extracting features.
- The size of the overlapping area of the two underlying blocks can increase or decrease the length of the feature vector extracted, which in turn affects the overall performance of the HOG descriptor.

3.2.2 Uniform local ternary pattern

The local ternary pattern (LTP) is the generalization of the local binary pattern (LBP). LTP features are more effective than the local binary pattern (LBP) because LTP features are more robust to noise. The LTP is the extended version of the LBP as described by Tan and Triggs [28], where the gray-level of the centered pixel is compared with the gray-levels of the pixels present in its neighborhood vicinity while keeping in view a threshold value $\pm\boldsymbol{\tau}+ z_c$, where $\boldsymbol{\tau}$ is a user specified threshold and z_c indicates the gray-level at the center. The neighboring pixel values are quantized to 0, 1, and −1 if they lie between $\pm\boldsymbol{\tau} + z_c$ or above $\boldsymbol{\tau}+ z_c$ and below $-\boldsymbol{\tau}+ z_c$ respectively.

$$g(u,\ z_c,\tau) = \begin{cases} 1 & u \geq z_c + \tau \\ 0 & |u - z_c| < \tau \\ -1 & u \leq z_c - \tau \end{cases} \quad (6)$$

where $g(u, z_c, \boldsymbol{\tau})$ is a three valued function, u denotes the pixel value in eight neighborhoods of the center pixel C. The dimensionality of the LTP histogram is large enough resulting in the histogram having larger dimensions [29]. Thus, to decrease feature dimensions, a coding technique is proposed by Tan and Trigg [28] that divides the LTP code in to its two LBP; upper (positive) LBP and lower (negative) LBP as shown in Fig. 5.

The idea of a uniform local binary pattern discussed in [30] is incorporated in this work due to its resiliency to distortion. It has been proved experimentally by Ojala et al. [30] that out of 256-bit patterns, 58 are uniform. They further added that approximately ninety percent (90%) of all observed pixels in the neighborhood vicinity of the origin pixels are redundant. This redundancy increases the dimensions of the resultant feature vector. The non-uniform patterns can be assigned to a single bin without considerably reducing the resultant histogram and without losing meaningful information. Thus, combining the patterns which have more than two shifts into a single bin forms an LBP operator that is represented by $\mathcal{LBP}_{\mathcal{PR}}^{\mathcal{U}2}$ with less than 2^{P} bins, i.e., the number of bins for the original LBP is 256 but for uniform LBP it is 59 bins. The subscript ($\mathcal{PR}$) represents neighborhood pixels and the superscript $\mathcal{U}2$ stands for uniform binary patterns which label all the remaining patterns with a single label. Thus, instead of using two 256 bins of histogram for the LTP operator that results in a feature vector of size 512, we have used two 59 bins of histogram of the LTP that results in a feature vector of size 118. The LTP is more resilient to noise as it encodes small pixel differences into a separate state, which is why the LTP is embedded with HOG to help the proposed technique to be robust to noise to greater extent. Finally, extracted HOG and U-LTP feature vectors are fused into a single feature vector and are labeled for facial expression recognition in the FER system.

3.3 Expression classification based on a multi-classifier

Researchers have proposed various methods for the evaluation of different classifiers in expression classification problems in recent times. The common problem in facial expression recognition is intensity variations while naming expressions. An effective and accurate facial expression classification approach must be able to classify each facial expression out of seven classes under these variation constraints. Cohn et al. [31] evaluated static classifiers e.g. Naïve Bayes and Tree Augmented Naïve Bayes [32], and dynamic classifiers like the single and multi-level Hidden Markov Model [33]. It is proved experimentally that the Support

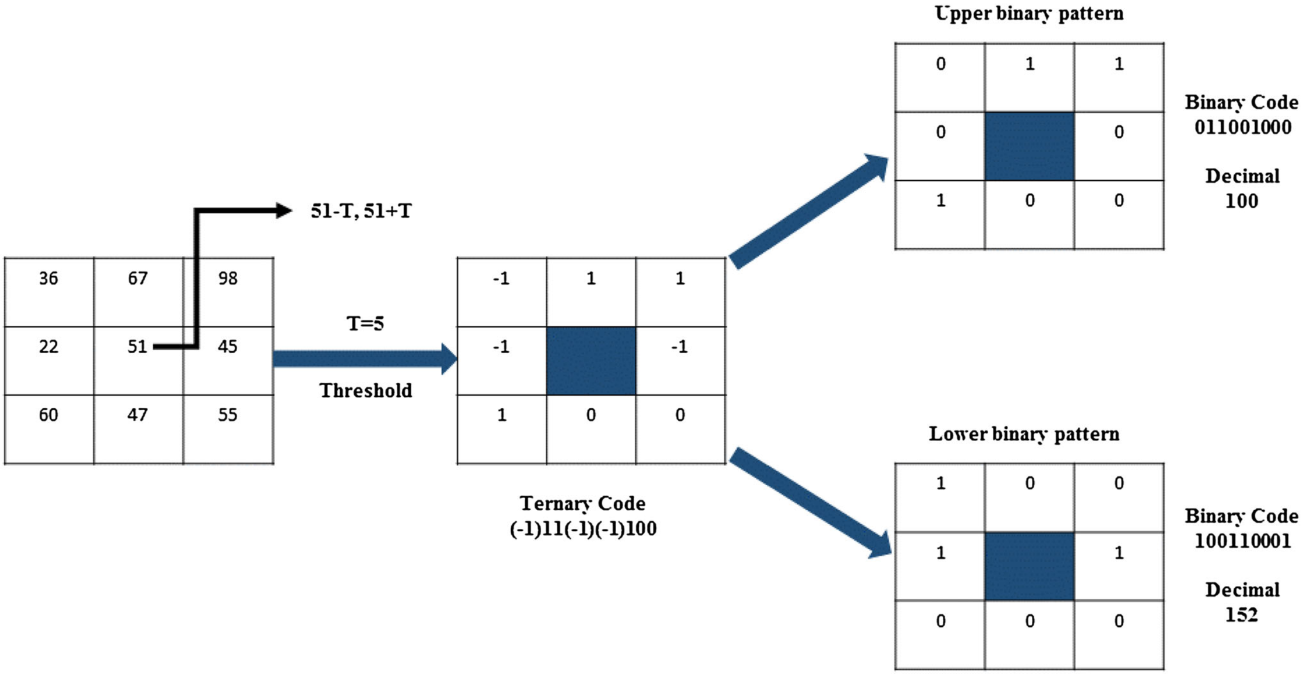


Fig. 5 Splitting an LTP code into upper and lower LBP codes

Vector Machine (SVM) plays an effective role in terms of multi-classification in the presence of various constraints [34–36].

The proposed work adopted an SVM as a classifier for facial expression classification and recognition, which first maps the training data of two classes into a higher dimensional space, then constructs an optimal separating hyper plane with a fine margin between the data of two classes. Since our problem is based on more than two classes, a multiclass SVM for multi classification is used for the proposed work while considering the hyper plane with the mentioned attributes. Multi classification is the problem of classifying features into more than two classes. Although the SVM was originally designed for binary classification it can be tuned into multinomial classification through a variety of methods. However, two methods are most commonly used for multi classification problems by reducing it to multiple binary classification problems, i.e., one versus rest (or one versus all) and one versus one. In the proposed study, the one versus rest strategy is employed. An SVM with a linear kernel is employed in the proposed work due to the large amount of training features. The linear kernel Ψ of the SVM classifier can be described as:

$$\Psi\left(\widehat{\mathcal{F}}_i^{(HOG+U-LTP)}, \Phi_i\right) = 1/\left(1 + e^{\mathcal{F}_I(HOG+U+LTP)^T \mathcal{L}_{\mathrm{j}}}\right) \tag{7}$$

Given the labeled training sample $(\widehat{\mathcal{F}}_i^{(HOG+U-LTP)}, \mathbf{L}_i)$ where i =1, 2, 3 ….,m,$\widehat{\mathcal{F}}_i^{(HOG+U-LTP)} \in \Re^{n+1}$ and $\mathcal{L}_J \in [1, 2, \ldots, 7]$. The classification can be described as:

$$\text{Ç}(\widehat{\mathcal{F}}_i^{(HOG+U-LTP)}) = sign\left(\sum_{i=1}^{n} \alpha_i L_J \Psi\left(\widehat{\mathcal{F}}_i^{(HOG+U-LTP)}, \Phi_i\right) + b\right) \tag{8}$$

where α_i are Lagrange multipliers of dual optimization problem, Ψ is a kernel function, and b is the bias of the hyperplane. The detailed algorithm of the proposed method is given in Table 2

4 Experiments and results

In this section, the proposed facial expression recognition method has been evaluated on various standard databases. Furthermore, a detailed comparative analysis of the proposed technique is performed with state-of-the-art FER techniques, performing both quantitative and qualitative assessments. Three databases are used, and each database in the proposed system is randomly divided into training and testing sets. Experiments are conducted by varying the number of training and testing images. The platform used for all simulations in the proposed method is MATLAB R2016a running on a

Table 2 Proposed Algorithm

```
I.    Given a database of face images $Đ$.
II.   Dividing the database $Đ$ into training $Đ_{TR}$ and testing $Đ_{C}$
III.  Conduct Training
        a. For $i=1$ to size of $(Đ_{TR}) = 3N/4$, where $N$ is the total number of images in $Đ$
        b. Select an image $\mathcal{I}_i^T$ from $Đ_{TR}$
        c. Preprocess the selected $\mathcal{I}_i^T$ applying histogram equalization technique and median filter:
           $\mathcal{I}_i^E \leftarrow \text{Preprocessing}(\mathcal{I}_i^T)$
        d. Detect the face region through the Viola & Jones algorithm: $\mathcal{I}_i^F \leftarrow \text{ViloaJones}(\mathcal{I}_i^E)$
        e. Crop the face region $\mathcal{I}_i^F \leftarrow \text{Cropped}(\mathcal{I}_i^F)$
        f. Extract HOG features into feature vector $\mathcal{F}_i^{HOG} : \mathcal{F}_i^{HOG} \leftarrow HOG(\mathcal{F}_i^F)$
        g. Extract U-LTP features into feature vector $\mathcal{F}_i^{U-LTP} : \mathcal{F}_i^{U-LTP} \leftarrow U-LTP(\mathcal{F}_i^F)$
        h. Fuse feature vectors $\mathcal{F}_i^{HOG}$ and $\mathcal{F}_i^{U-LTP}$ into 1-D feature vector $\mathcal{F}_i^{HOG+U-LTP}$
        i. END
IV.   Label all $3N/4$ feature vectors $\mathcal{F}_i^{HOG+U-LTP}$ extracted in the previous step according to their corresponding expression labels in $\mathcal{L}_J$, where $\mathcal{J}=\{1,\ldots,7\}$ or $\mathcal{J}=$ {Happy, Angry, Sad, Surprise, Disgust, Fear, Neutral}: $\hat{\mathcal{F}}_i^{HOG+U-LTP} \leftarrow \text{Label}(\mathcal{F}_i^{HOG+U-LTP}, \mathcal{L}_J)$
V.    Train the classifier Ç by passing all the feature vectors with their corresponding labels $\hat{\mathcal{F}}_i^{HOG+U-LTP}$.
VI.   Conduct cross-validation test
        a. For $i=1$ to size of $(Đ_C) = (N-3N/4)$
        b. Repeat the sub-steps from b to h in step III for the images in ĐC
        c. END
VII.  Pass the resultant feature vector $\hat{\mathcal{F}}_i^{(HOG+U-LTP)}$ of each test image in $Đ_C$ to the classifier Ç for cross-validation and updates the classifier Ç accordingly.
VIII. The classifier Ç checks the similarity between $\mathcal{F}_i^{(HOG+U-LTP)}$ and $\hat{\mathcal{F}}_i^{(HOG+U-LTP)}$ by the following formula:
        $Ç(\hat{\mathcal{F}}_i^{(HOG+U-LTP)}) = sign\left(\sum_{i=1}^{n} \alpha_i \mathcal{L}_J \Psi\left(\hat{\mathcal{F}}_i^{(HOG+U-LTP)}, \Phi_i\right) + b\right)$
IX.   Predict the $\mathcal{L}_J$ with a proper category $\mathcal{J}==$ {Happy, Angry, Sad, Surprise, Disgust, Fear, Neutral}
```

PC with a CPU speed of 2.70 GHz equipped with 4.00 GB of RAM and a Windows 8.1 64 bit-version operating system. The detailed explanation of each of the three databases (that is, JAFFE, MMI, and CK+) used in the experiment are described in the subsequent sections.

4.1 Japanese female facial expression database (JAFFE)

This database [38] consists of 213 images of 10 Japanese females, and for each female seven basic facial expressions (angry, sad, happy, disgust, surprise and fear with neutral) are recorded. 3∼4 images were captured for each user for each expression out of seven. All 213 images have been used in experiments. Table 3, illustrates the number of images per expression. Some sample images from the JAFFE database are shown in Fig 6 (a).

4.2 Extended cohn kanade database (CK+)

The CK database [31] is the first version of CK+. It is composed of a series of images containing both posed and non-posed expressions, taken from 123 subjects. Each subject is taught to show seven basic expressions. Fig 6 (b), illustrates some sample images from the CK+ database. The beginning of each subject's image for each expression is from neutral to the peak of an expression. From each video five frames are captured which shows a particular expression. A different number of images were used for experiments. A total of 630 images from CK+ (90 images for each out of seven expressions) have been used in this study.

Table 3 Number of images from each expression from the three different databases

	No. of images		
Expression name	MMI	JAFFE	CK+
Neutral	36	30	90
Happy	39	32	90
Angry	45	30	90
Fear	41	32	90
Disgust	39	29	90
Sad	34	30	90
Surprise	39	30	90
Total	**273**	**213**	**630**

4.3 MMI

The MMI [37] database contains more than 20 subjects of both genders (44% females). Their ages range from 19 to 62 years, belonging to different countries (from Europe, Asia, South America, etc.). Both males and females were taught to display 79 sequences of emotion, six of which are basic expressions (sad, angry, fear, surprise, disgust, and happy with neutral). It comprises a sequence of videos including both posed and spontaneous expressions. A total number of 273 frames are extracted from different videos. Fig 6 (c) shows some sample images from the MMI database. Table 3 illustrates the number of images per expression.

4.4 Experiments on the JAFFE database

Experiments on the JAFFE database are performed using K-fold splitting. In the first phase, 70 (36%) images are used as a training set and 143 (64%) images are used for testing purposes. In the second phase, 64% of the images are used

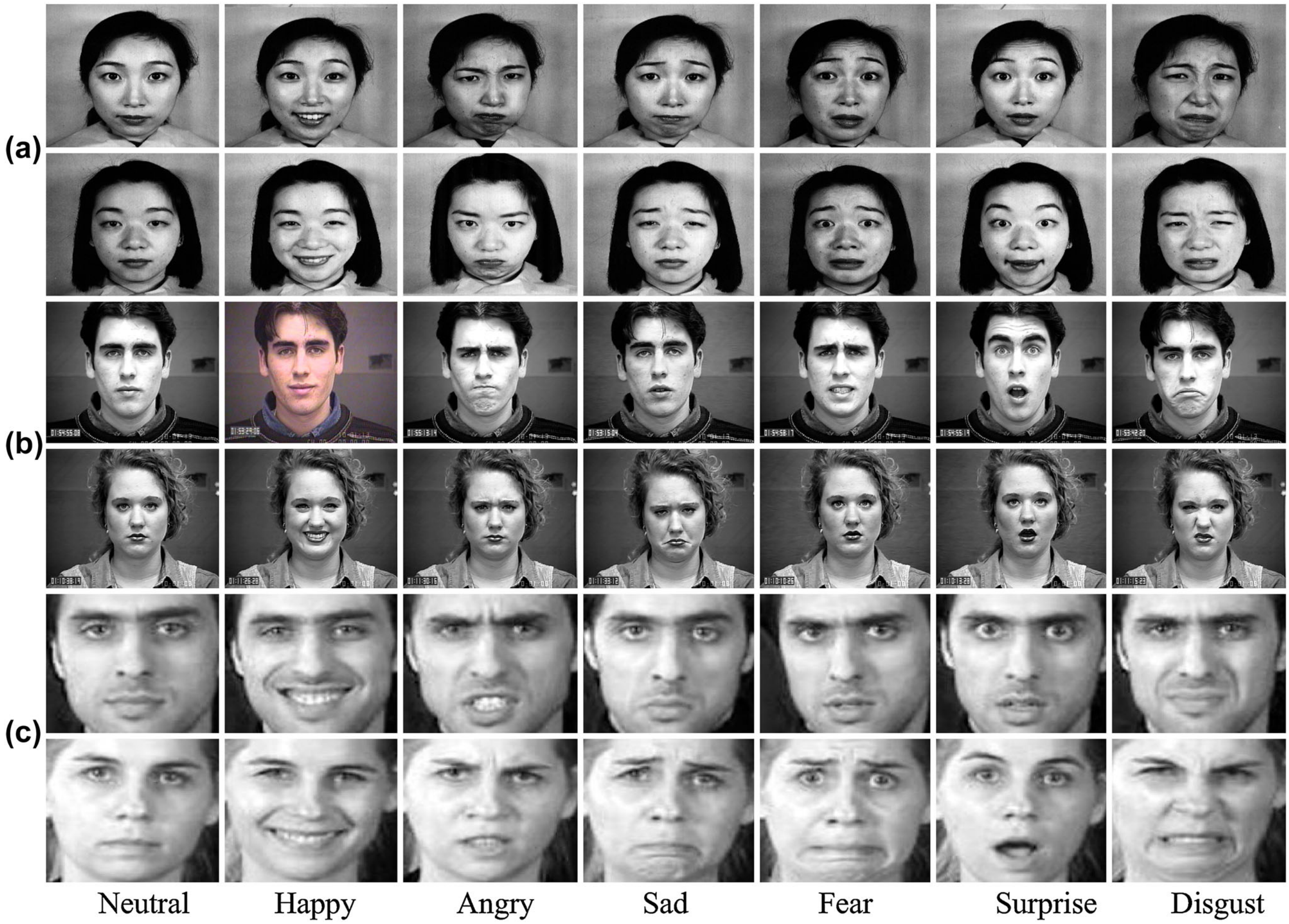


Fig. 6 Sample images from three standard datasets: JAFFE, CK+, and MMI, databases shown in group **a**–**c**, respectively. *1st column* shows Neutral images, *2nd* shows happy, *3rd* shows Angry, *4th* shows Sad, *5th* shows Fear, *6th* shows Surprise, and *7th column* shows Disgust expressions from all of three databases

Table 4 Confusion matrix of the proposed method on JAFFE [training (36%) and testing (64%)]

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**89.999**	0	0	0	0	9.999	0
Happy	0	**89.999**	0	0	0	9.999	0
Angry	4.999	0	**89.999**	0	4.999	0	0
Fear	0	0	0	**74.999**	4.999	15.999	4.999
Disgust	0	0	0	0	**99.999**	0	0
Sad	0	0	0	4.999	4.999	**89.999**	0
Surprise	0	4.999	0	4.999	0	0	**89.999**
Average recognition accuracy	**89.284%**						

Table 5 Confusion matrix of the proposed method on JAFFE [training (64%) and testing (36%)]

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**99.999**	0	0	0	0	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**99.999**	0	0	0	0
Fear	0	0	0	**89.999**	9.999	0	0
Disgust	0	0	0	0	**99.999**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	9.999	0	0	0	**0**	**89.999**
Average recognition accuracy	**97.141%**						

Table 6 Confusion matrix of the proposed method on JAFFE (10-fold cross validation)

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**99.999**	0	0	0	0	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**86.666**	0	9.999	3.333	0
Fear	0	0	0	**96.774**	0	0	3.222
Disgust	0	0	10.344	0	**89.655**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	3.333	0	4.999	0	0	**96.666**
Average recognition accuracy	**95.714%**						

Bold figures in Tables 4, 5, and 6 indicate the average accuracy for each expression

for training and remaining 36% for testing purposes. In the third phase, 10-fold cross validation is adopted, randomly dividing the database into 10 equal segments in terms of varying expressions. 9 out of 10 segments are trained each time and the remaining 10% of images are used for testing.

In the first phase of experiments, an overall correct recognition rate of 89.28% is achieved. Due to a lower number of images (36%) in the training set the recognition rate is comparatively low. Table 4 shows the confusion matrix of the said experiment, where bold digits shows average accuracy of neutral, happy, angry, and sad and surprise expressions. In case of disgust expressions a highest accuracy of 99.99% is achieved and a lowest accuracy rate of 74.99% is achieved for fear expressions. In the second phase of the experiment an average recognition rate of 97.14% is achieved as shown in Table 5. The third phase of experiments is the most well-known experiment done by many researchers. The confusion matrix of the given experiment is shown in Table 6. The correct recognition accuracy of each expression is computed for neutral, happy, and sad is 99.99%, and for fear and surprise it is almost the same at 96.77 and 96.66% respectively. For angry and disgust 86.666, and 89.655% are achieved. It is observed that the accuracy for angry and disgust is low compared to other expressions, due to similarities for both of these expressions.

4.5 Experiments on the MMI database

The same experimental criteria discussed in Sect. 4.4 is also used on the MMI database. In the first phase, the database is divided so that 42% of the images are selected for the training set and 58% were selected for testing. In the second

Table 7 Confusion matrix of the proposed method on MMI [training (42%) and testing (58%)]

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**99.999**	0	0	0	0	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**88.888**	0	11.111	0	
Fear	0	0	0	**99.999**	0	0	0
Disgust	0	0	0	0	**99.999**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	0	0	0	0	0	**99.999**
Average recognition accuracy		**98.41%**					

Table 8 Confusion matrix of the proposed method on MMI (Training (58%) and Testing (42%))

Expression Name	Normal	Happy	Angry	Fear	Disgust	Sad	Surprise
Normal	**99.999**	0	0	0	0	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**96.296**	0	0	0	3.703
Fear	0	0	0	**99.999**	0	0	0
Disgust	0	0	0	0	**99.999**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	0	0	0	0	0	**99.999**
Average recognition accuracy		**99.47%**					

Table 9 Confusion matrix of the proposed method on MMI (10-Fold Cross Validation)

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	**Surprise**
Neutral	**97.222**	0	2.777	0	0	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**97.777**	0	2.777	0	0
Fear	0	0	2.439	**97.560**	0	0	0
Disgust	0	0	0	2.564	**97.435**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	0	0	2.564	0	0	**97.435**
Average recognition accuracy		**98.203%**					

Bold figures in Tables 7, 8, and 9 indicate the average accuracy for each expression

phase, 58% of the images are used for training and 42% of the images for testing (training: 3 images from each expression, testing: 2 images). The selection criteria is on a random basis. In the third phase 10-fold cross validation is performed for experiments. Tables 7, 8, and 9 describe the confusion matrix for all three phases on the MMI database. The results in bold figures show that the proposed FER system successfully recognizes each face expression with high recognition rates. However, a minor error rate is observed in the case of angry expressions which is caused by highly similar features to disgust and angry expressions.

4.6 Experiments on the CK+ database

The proposed three phase's experimental methods have also been applied to the CK+ database. 40% of the images are used for training and the remaining 60% for testing in the first phase of the experiment. In the second phase, 60% of the images were used for training and 40% for testing. 10-fold cross validation criteria is employed in the third phase. Like for MMI, CK+ also shows high correct recognition rates for all of the three phase experiments, but the results are better than MMI. This shows that the proposed system works best for the CK+ database. The confusion matrix of the first, second, and third phases are shown in Tables 10, 11, and 12 respectively.

4.7 Robustness to noise

In the real environment, noise is the main factor which degrades the quality of images, resulting in the performance seediness of various computer vision and pattern recognition

Table 10 Confusion matrix of the proposed method on CK+ (Training (40%) Testing (60%))

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**98.148**	0	0	0	1.851	0	0
Happy	0	**96.296**	0	3.703	0	0	0
Angry	0	0	**94.444**	3.703	0	0	0
Fear	0	0	0	**99.999**	0	0	0
Disgust	0	0	0	0	**99.999**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	0	0	0	0	0	**99.999**
Average recognition accuracy		**98.412**					

Table 11 Confusion matrix of the proposed method on CK+ (Training (60%) Testing (40%))

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**99.999**	0	0	0	0	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**98.888**	1.111	0	0	0
Fear	0	0	0	**99.999**	0	0	0
Disgust	0	0	0	1.851	**98.148**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	0	0	0	0	0	**99.999**
Average recognition accuracy		**99.575%**					

Table 12 Confusion matrix of the proposed method on CK+ (10-Fold Cross Validation)

Expression name	Neutral	Happy	Angry	Fear	Disgust	Sad	Surprise
Neutral	**98.888**	0	0	0	1.111	0	0
Happy	0	**99.999**	0	0	0	0	0
Angry	0	0	**98.888**	0	1.111	0	0
Fear	0	0	0	**99.999**	0	0	0
Disgust	0	0	0	0	**99.999**	0	0
Sad	0	0	0	0	0	**99.999**	0
Surprise	0	0	0	0	0	0	**99.999**
Average recognition accuracy		**99.681%**					

Bold figures in Tables 10, 11, and 12 indicate the average accuracy for each expression

systems. A practical FER system must be able to perform well despite the presence of noise in images. This section aims to examine the robustness of the proposed technique in the presence of noise. In this context, salt and pepper noise of different levels were added randomly to the test images of size [128 128]. Fig.7 shows the example images under different levels of noise. Noise up to level 0.05 was added because in real environment this the average noise level observed in images. The robustness of the proposed FER system is evaluated on the three databases, in which 60% of data are used for training and 40% of data are used as testing sets. It is noticed that the proposed system is fairly robust against salt and pepper noise as shown in Fig 8. Different variations occur in the recognition rate while changing noise densities. It is observed that with the increase in noise density, the recognition rate decreases. It can be seen that noise weakens the recognition rate of the JAFFE and MMI databases further compared to CK+ when the density of noise is 0.01. The recognition rate of all the databases decreases gradually as the noise density increases to 0.02 but the recognition rate for the JAFFE database is better to some extent compared to MMI and CK+. It can be noticed that the rate decreases more severely for CK+ compared to JAFFE and MMI when noise density is increased.

4.8 Robustness to occlusions

The presence of occlusion also affects image quality and degrades the performance of facial expression recognition systems. In this section, random sized blocks were added to the test images to check robustness to occlusion. The block sizes range from 15×15 to 55×55. These blocks are added randomly on facial images as shown in Fig 9. Experiments are performed on 64% of the training data and 36% of the testing

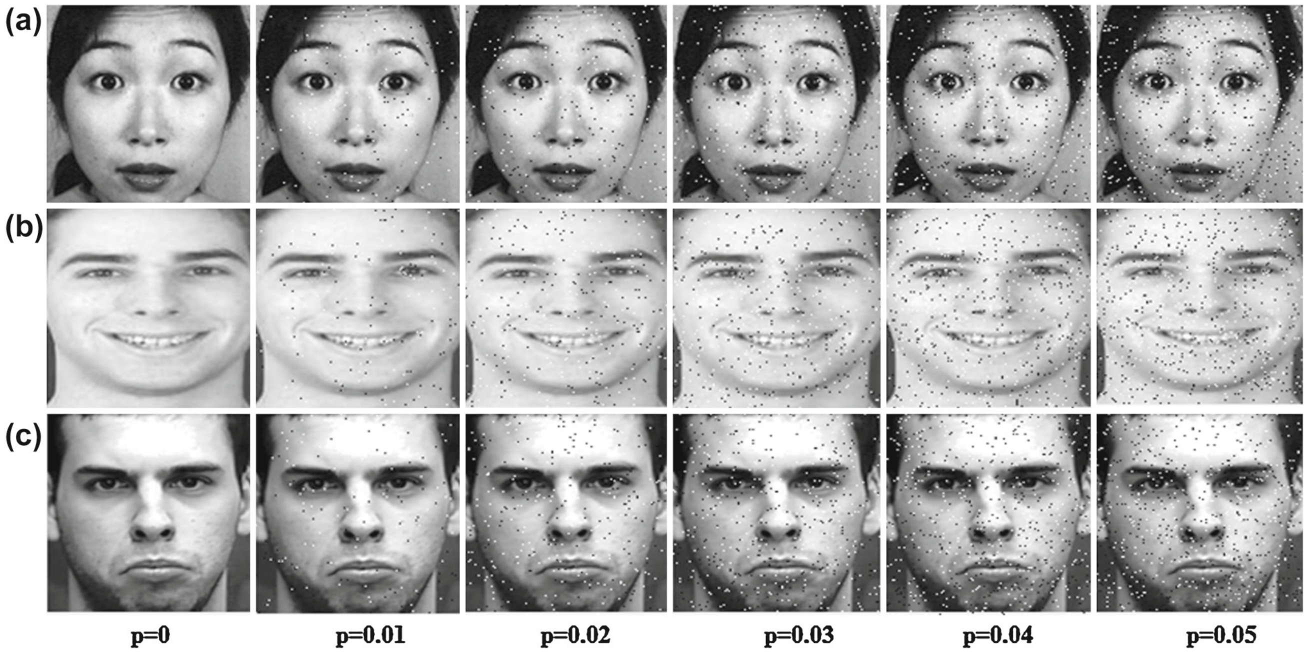
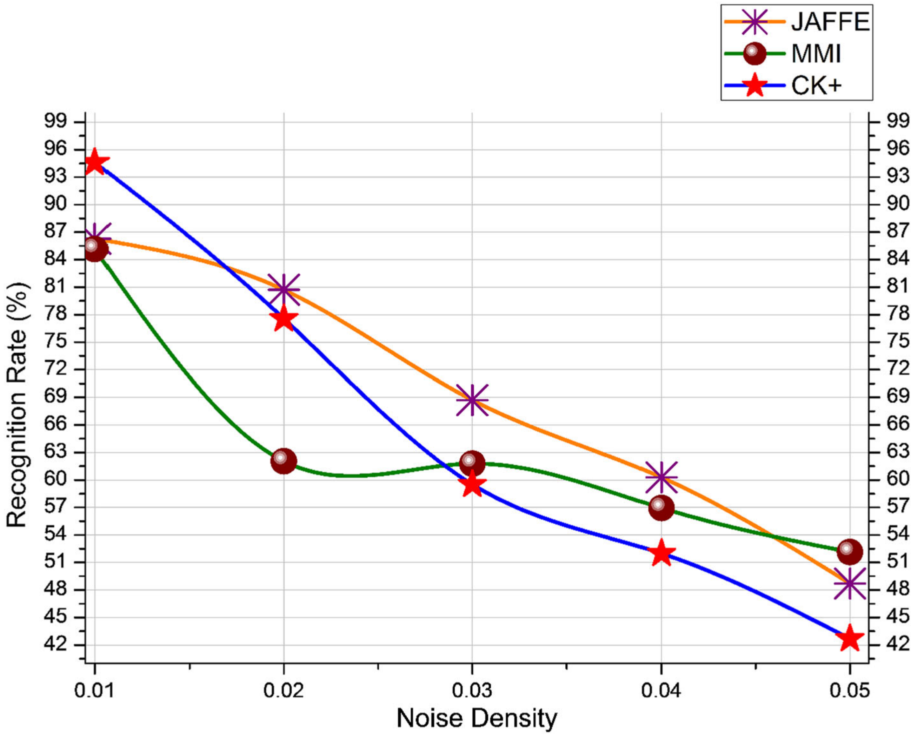


Fig. 7 Example images from three databases with salt and pepper noise added are shown in group **a**–**c** respectively, where p is the power of noise

Fig. 8 Recognition accuracy of three databases in the presence of salt & pepper noise



data from each of the three databases. The resultant average recognition rate is shown in Fig 10. Recognition accuracy of both CK+ and MMI databases are excellent compared to JAFFE. The recognition rate for both CK+ and MMI is almost the same i.e. 97.89 and 95.73 when block sizes of 15 × 15 and 25 × 25 were added, which again shows the better recognition accuracy of the proposed system. The recognition rate for the JAFFE database is almost equal for block sizes of 15 × 15 and 25 × 25. The overall accuracy of the proposed system for the CK+ database is very fair for the aforesaid block sizes compared to JAFFE and MMI.

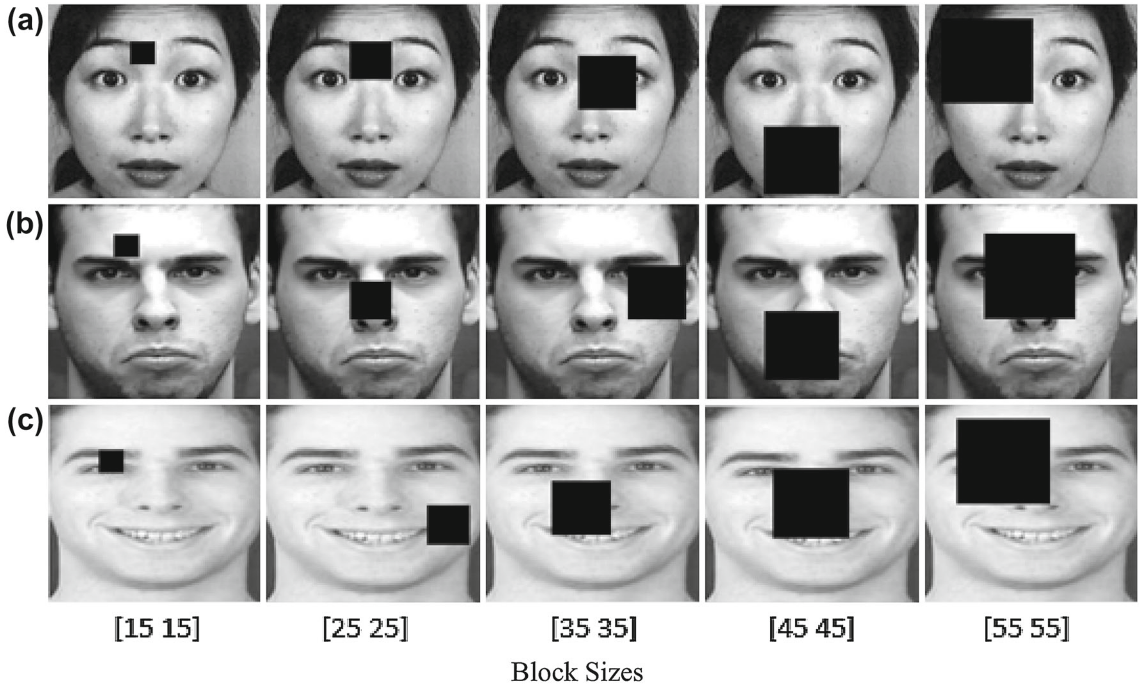
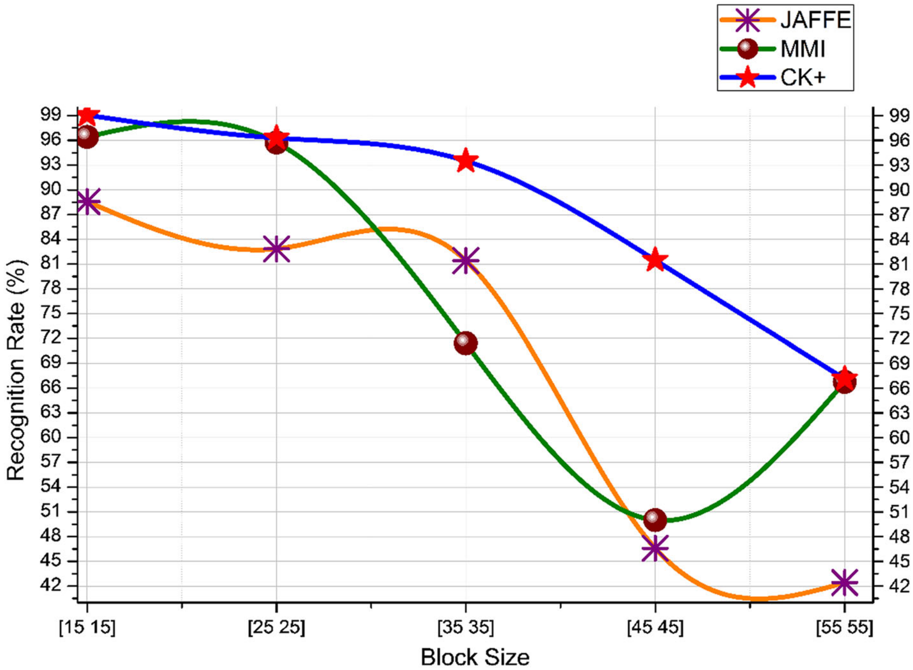


Fig. 9 Sample images with block occlusion from **a** JAFFE **b** MMI and **c** CK+ databases

Fig. 10 Recognition accuracy of three databases in the presence of block occlusions



4.9 Evaluation of HOG parameters

In this section different parameters of HOG are evaluated to test the high recognition rate of the proposed method and to select the optimal value for the HOG parameters to perceive the most significant information for the facial expressions. The HOG descriptor is categorized by its two key parameters: cell size and orientation bins. Cell size represents the dimension of the pixel's block involved in the single histogram calculation. Large cell size can be used to capture large scale spatial information but increasing cell size may result in losing small scale details of an image. Orientation histogram bins is specified as positive scalar. Increasing cell size to capture better orientation details also increases the size of the feature vector resulting in increased processing time. Therefore, by taking processing time and recognition

Fig. 11 Recognition accuracy of the JAFFE database with varying parameters of HOG

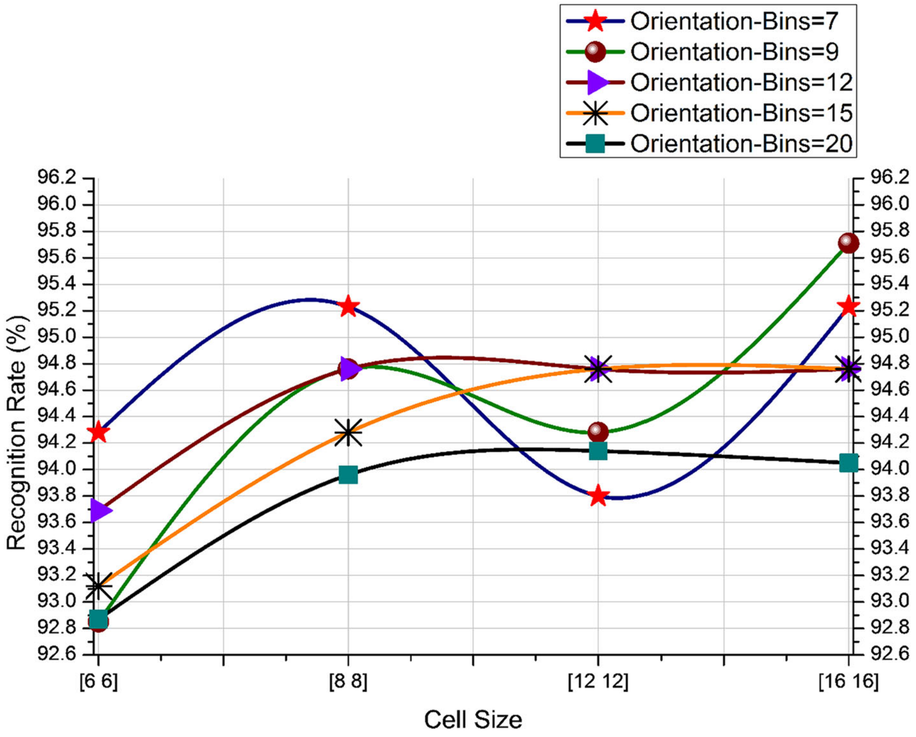


Table 13 Comparison of the proposed FER method with other state-of-the-art techniques

Ref	Database	Feature extraction	Dimensionality Reduction	Classifier	Avg. recognition Rate (%)
[39]	JAFFE	LPQ + es—LBP-s	Cr-LPP	SVM	94.88
[21]	CK+	HOG	N/A	SVM	98.8
[40]	JAFFE	Radial Encoding of Local Gabor features	PCA	KNN	89.67
	CK				91.51
[41]	CK+	HOG	N/A	SVM	95
[42]	JAFFE	Patch-base Gabor	N/A	SVM	92.93
	CK+				94.48
[4]	JAFFE	LBP	PCA	SVM	81.0
	MMI				86.9
[13]	JAFFE	LGC – HD	N/A		90
Proposed	MMI	HOG + U-LTP	N/A	SVM	98.20
	JAFFE				95.71
	CK+				99.68

rate of the proposed system into consideration, several experiments were performed in which different sizes of cells with different orientation bins were evaluated. These experiments are performed on the JAFFE database by taking 64% of the data for training and 36% of the data for testing. The results of these experiments are illustrated in Fig 11. It can be seen in the graph that an optimal parameter has been found, in which the cell size of 16×16 and 9 orientation bins is selected. This increases the correct recognition accuracy of the proposed system as well as decreasing processing time due to a smaller feature vector.

4.10 Comparisons with state-of-the-art methods

In this section, the proposed method is compared with the existing facial expression recognition algorithms (including the methods of [4], [13], [21], [39], and [40–42]). The resultant recognition accuracy and efficiency of the proposed system is compared with the reported results of the benchmarked approaches. These approaches are selected because they produced state-of-the-art performance using a similar testing strategy on the same databases. It is observed that the proposed method outperforms other existing systems using same databases as shown in Table 13 (using 10-fold cross validation). The correct recognition rate of the proposed system is 95.719% on JAFFE and 98.203% on the MMI database and 99.681% on the CK+ database using a 10-fold cross-validation strategy. Table 13 shows the performance comparison among the proposed system and the existing systems using the same JAFFE, MMI, and CK+ databases.

5 Conclusion

In this paper a facial emotion recognition based sentiment knowledge discovery framework is presented. A Histogram of Oriented Gradients descriptor with ULTP is extracted for robust classification of human face emotions. Experiments show that the proposed technique attains the highest recognition accuracy against the state-of-the-art techniques in 10-fold cross validation. It is also validated that the proposed FER system recognizes facial expressions in various challenging situations such as lighting effects and the presence of occlusions or noise. It has been shown by the experiments that the proposed method performs fairly well in the presence of various elements, although noise can badly affect the recognition accuracy; however, this aspect will be focused on in more detail in future work. Moreover, the presented framework will be extended from static images to FER in videos, providing solutions to various video-centric challenges in relation to facial emotion analysis.

Acknowledgements This research was supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under National program for Excellence in SW (2015-0-00938) supervised by the IITP (Institute for Information & communications Technology Promotion).

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interest.

References

1. Mehrabian, A.: Nonverbal Communication. Transaction Publishers, Los Angeles (1972)
2. Ekman, P., Friesen, W.V., Hager, J.C.: A technique for the measurement of facial action. In: Facial action coding system (FACS), p. 22. Palo Alto, Consulting (1978)
3. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 699–714 (2005)
4. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
5. Song, K.T., Chien, S.C.: Facial expression recognition based on mixture of basic expressions and intensities. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE (2012)
6. Dhall, A., et al.: Emotion recognition using PHOG and LPQ features. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE (2011)
7. Apte, S.: Facial Emotion Identification
8. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
9. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, IEEE (2001)
10. Rivera, A.R., Castillo, J.R., Chae, O.O.: Local directional number pattern for face analysis: face and expression recognition. IEEE Trans. Image Process. **22**(5), 1740–1752 (2013)
11. Jeyashree, T., et al.: An efficient algorithm for face and expression recognition. Int. J. Sci. Technol. **2**(3), 172 (2014)
12. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE (1991)
13. Hsu, C.W., Chang, C.C., Lin, C.J.: A practical guide to support vector classification (2003)
14. Kozma, L.: k Nearest Neighbors algorithm (kNN). Helsinki University of Technology (2008)
15. Lee, Y.H., Han, W., Kim, Y.: Emotional recognition from facial expression analysis using bezier curve fitting. In: 2013 16th International Conference on Network-Based Information Systems, IEEE (2013)
16. Al-Shabi, M., Cheah, W.P., Connie, T.: Facial expression recognition using a hybrid CNN-SIFT aggregator. arXiv preprint arXiv:1608.02833 (2016)
17. Wang, X.H., Liu, A., Zhang, S.Q.: New facial expression recognition based on FSVM and KNN. Optik-Int. J. Light Electron Opt. **126**(21), 3132–3134 (2015)
18. Tong, Y., Chen, R., Cheng, Y.: Facial expression recognition algorithm using LGC based on horizontal and diagonal prior principle. Optik-Int. J. Light Electron Opt. **125**(16), 4186–4189 (2014)
19. Luo, Y., Wu, C.M., Zhang, Y.: Facial expression recognition based on fusion feature of PCA and LBP with SVM. Optik-Int. J. Light Electron Opt. **124**(17), 2767–2770 (2013)
20. Happy, S., Routray, A.: Robust facial expression classification using shape and appearance features. In: Advances in Pattern Recognition (ICAPR), 2015 Eighth International Conference on, IEEE (2015)
21. Carcagnì, P., et al.: Facial expression recognition and histograms of oriented gradients: a comprehensive study. SpringerPlus **4**(1), 1 (2015)
22. Kumar, S., Bhuyan, M., Chakraborty, B.K.: Extraction of informative regions of a face for facial expression recognition. IET Comput. Vis. **10**(6), 567–576 (2016)
23. Guo, Y., et al.: EI3D: expression-invariant 3D face recognition based on feature and shape matching. Pattern Recognit. Lett. **83**, 403–412 (2016)
24. Sajjad, M., Ejaz, N., Baik, S.W.: Multi-kernel based adaptive interpolation for image super-resolution. Multimed. Tools Appl. **72**(3), 2063–2085 (2014)
25. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE (2005)
26. Geng, C., Jiang, X.: SIFT features for face recognition. In: Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on, IEEE (2009)
27. Dreuw, P., et al. SURF-Face: face recognition under viewpoint consistency constraints. In: BMVC (2009)

28. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1650 (2010)
29. Ren, J., Jiang, X., Yuan, J.: Relaxed local ternary pattern for face recognition. In: ICIP (2013)
30. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
31. Lucey, P., et al.: The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE (2010)
32. Cohen, I., et al.: Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, IEEE (2003)
33. Cohen, I., et al.: Facial expression recognition from video sequences: temporal and static modeling. Comput. Vis. Image Underst. **91**(1), 160–187 (2003)
34. Bartlett, M.S., et al.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE (2005)
35. Valstar, M., Pantic, M.: Fully automatic facial action unit detection and temporal analysis. In: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), IEEE (2006)
36. Valstar, M.F., Patras, I., Pantic, M.: Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, IEEE (2005)
37. Pantic, M., et al.: Web-based database for facial expression analysis. In: 2005 IEEE international conference on multimedia and Expo, IEEE (2005)
38. Lyons, M., et al.: Coding facial expressions with gabor wavelets. In: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, IEEE (1998)
39. Chao, W.L., Ding, J.J., Liu, J.Z.: Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection. Signal Process. **117**, 1–10 (2015)
40. Gu, W., et al.: Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recognit. **45**(1), 80–91 (2012)
41. Donia, M.M., Youssif, A.A., Hashad, A.: Spontaneous facial expression recognition based on histogram of oriented gradients descriptor. Comput. Inf. Sci. **7**(3), 31 (2014)
42. Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. IEEE Trans. Affect. Comput. **2**(4), 219–229 (2011)

Muhammad Sajjad received his Master degree from Department of Computer Science, College of Signals, National University of Sciences and Technology, Rawalpindi, Pakistan. He received his Ph.D. degree in Digital Contents from Sejong University, Seoul, Republic of Korea. He is now working as an assistant professor at Department of Computer Science, Islamia College Peshawar, Pakistan. He is also head of "Digital Image Processing Laboratory (DIP Lab)" at Islamia College Peshawar, Pakistan., where students are working on research projects such social data analysis, image super-resolution and reconstruction, medical image analysis, multi-modal data mining and summarization, prioritization, image/video quality assessment, and image/video retrieval. His primary research interests include computer vision, image understanding, pattern recognition, and robot vision and multimedia applications, with current emphasis on raspberry-pi and deep learning-based bioinformatics, video scene understanding, activity analysis, and real-time tracking.

Adnan Shah received his BS degree in computer science from Islamia College Peshawar, Pakistan. He is an active member of Digital Image Processing Lab, Department of Computer Science, Islamia College Peshawar. His research interests include digital image retrieval and recognition, pattern recognition and computer vision.

Zahoor Jan is currently holding the rank of an associate professor in computer science at Islamia College Peshawar, Pakistan. He received his M.S. and Ph.D. degree from FAST University Islamabad in 2007 and 2011 respectively. He is also chairman of Department of Computer Science at Islamia College Peshawar, Pakistan. His areas of interests include image processing, machine learning, computer vision, artificial intelligence and medical image processing, biologically inspired ideas like genetic algorithms and artificial neural networks, and their soft-computing applications, biometrics, solving image/video restoration problems using combination of classifiers using genetic programming, optimization of shaping functions in digital watermarking and image fusion.

Syed Inayat Ali Shah received his M.Phil degree from Quaid-e-Azam University Islamabad, Pakistan in 1990 and Ph.D. degree from Saga University Japan in 2002. He is currently working as Professor of Mathematics and Dean Faculty of Physical & Numerical Science, Islamia College Peshawar, Pakistan. His research interest includes fuzzy theory, computing, number theory, modeling & simulation and image processing.

Sung Wook Baik received the B.S. degree in computer science from Seoul National University, Seoul, Korea, in 1987, the M.S. degree in computer science from Northern Illinois University, Dekalb, in 1992, and the Ph.D. degree in information technology engineering from George Mason University, Fairfax, VA, in 1999. He worked at Datamat Systems Research Inc. as a senior scientist of the Intelligent Systems Group from 1997 to 2002. In 2002, he joined the faculty of the College of Electronics and Information Engineering, Sejong University, Seoul, Korea, where he is currently a Full Professor and Dean of Digital Contents. He is also the head of Intelligent Media Laboratory (IM Lab) at Sejong University. His research interests include computer vision, multimedia, pattern recognition, machine learning, data mining, virtual reality, and computer games.

Irfan Mehmood received B.S. degree in Computer Science in 2010 from National University of Computer and Emerging Sciences, Pakistan. From 2010 to 2011, he served as a Software Engineer in Talented Earth Organization, where he provided various services such a mobile application development. At the end of 2011, he joined Intelligent Media Laboratory (IM Lab) as a research associate while enrolling as a Ph.D. student in Sejong University, Seoul, South Korea. In IM Lab, he worked on various projects related to video summarization and prioritization, image super-resolution, image quality assessment, and mixed reality. In 2016, he joined Sejong University South Korea and serving as an assistant professor in computer science and engineering department. He is also head of Visual Analytics Lab, where students are working on research projects such social data mining and information retrieval, steganography, digital watermarking, and medical imaging analysis etc. under his supervision. His current research activities include social data mining, multi-modal medical data analysis, information summarization and multi-modal surveillance frameworks.