

An exposition of facial expression recognition techniques

Somia Saeed¹ · M. Khalid Mahmood² · Yaser Daanial Khan¹

Received: 29 December 2015 / Accepted: 8 August 2016 / Published online: 29 August 2016
© The Natural Computing Applications Forum 2016

Abstract Automated facial expression recognition schemes have been a subject of interest ever since the inception of its idea. The initial efforts required supervised and controlled environments to get realistic and convincing experimental results. Various improved methods have been suggested to detect the facial expressions of a human. Some of the popularly used methods are automatic expression recognition system (AERS), graph-preserving sparse nonnegative matrix factorization (GSNMF) algorithm, two-phase test sample representation (TPTSR) technique and temporal template method. The current state-of-the-art techniques are able to detect facial expressions in difficult and obscured environments. Advancements in this technique have helped researchers to use video sequences as images to detect expressions. A single image is extracted from a video sequence, and elaborate techniques are applied to detect the expression. This article endeavors to discuss these intricate techniques and critically analyzes them. It helps the reader to understand the contemporary problems in facial recognition systems and how various researchers have employed different models to overcome these challenges. In this paper, the performance of various techniques such as AERS,

GSNMF algorithm, TPTSR, performance-based character animation, temporal template method, feature extractions using Gabor filter and image sequencing-based methods has been scrutinized in terms of their efficiency, accuracy and effectiveness. The efficiency and accuracy of the techniques have been compared using various benchmarks such as leave one out, cross-validation and receiver operating characteristics. Each technique bears its own advantages and disadvantages in terms of accuracy and efficiency. The highest accuracy rate is exhibited by the technique using canny edge detection algorithm and chamfer image method.

Keywords Facial expression · Expression recognition · Emotion recognition

1 Introduction

Facial expressions are the most prominent component of social communication among humans. Humans share more information through expressions as compared to verbal communication. Facial expression performs a significant role in understanding emotions of a human. The emotional factor of the facial expressions is not merely of reflexive nature, but it also imparts some communicative effect [14, 16]. The significance and potential applications of facial expressions in the communicative process have not been realized to its full potential as yet because effective facial expressions recognition techniques are very few. Facial expression is the process of drawing out a bunch of statistics regarding one's emotion and state of mind. Research on facial expressions has received considerable attention since past few years. Identification of facial expression contributes in making of various kinds of

✉ Yaser Daanial Khan
yaser.khan@umt.edu.pk

Somia Saeed
13021050002@umt.edu.pk

M. Khalid Mahmood
Khalid.math@pu.edu.pk

¹ School of Systems and Technology, University of Management and Technology, Lahore, Pakistan

² Department of Mathematics, University of Punjab, Lahore, Pakistan

communication applications. Numerous studies propose that in a normal human conversation almost 55 % of the communication data are transferred by facial expressions, 38 % by paralanguage and 7 % by linguistics [26]. This statistically analysis suggests that facial expression has a substantial role in the human communicative process. The child begins to exhibit its cognitive ability to detect facial expression from the very beginning of life. As a child grows, he starts learning different facial expressions before he learns how to speak. All the different looks and expressions are stored in the child's memory, as time passes his ability to recognize and mimic these expressions becomes stronger [5, 24, 34]. Children are exposed to an array of emotional aspects from birth, and evidence indicates that they imitate some facial expressions and gestures as early as few day after the beginning of life. A facial expression is formed as a result of changing positions in continuity by motion of one or more muscles below the facial skin. There are two cases of facial expressions: (1) voluntary—these facial expressions are often socially conditioned and follow a cortical route in the psyche. (2) Involuntarily—these facial expressions are viewed to be innate and follow a subcortical route in the psyche.

The evolutionary basis of facial expressions tracks back to Darwin. He proposed a theory titled “The expression of emotions in man and animal.” In this theory, he believed that humans have evolved their expressions from animals. He said that expressions are unlearned and inborn among humans. Darwin obtained significant results for his theory by studying various cultures and infants [11]. This theory had both supportive and critical reviews. After Darwin, Ekman continued the research work on facial expressions. He said that the facial expressions are not determined culturally but are of a universal nature. He was able to prove the first hypothesis proposed by Darwin that expressions were unlearned and they were independent of the culture. He conducted many experiments to learn the behavior of adults and children regarding facial expressions [10, 12, 22]. The work on facial expressions still continues. Many scientists are still working on the patterns and expressions molded by human expressions.

Facial expression recognition (FER) is being rapidly used by society for the past few years. It has become the replacement of pin for various locks. It is used in law enforcements for court records and in schools' surveillance cameras for child molesters. It is also used for security purposes at home. Recently, it has become beneficial for autistic children's learning and gaming [17].

The development of an intelligent FER system bears numerous challenges. In general terms, it is hard to determine facial expressions of a person as some people are expressive, and others are not; some have natural expressions and others are trained professionals. A lot of work has

been done in recording expressions of people of diverse ages and cultures [18, 23]. Their images and videos have been preserved in publicly available databases for further research work. Some of the famous databases maintained to preserve human expressions in digital form are: Japanese Female Facial Expression database (JAFFE) [7], Cohn–Kanade database [19, 22], Korea University Gesture database, ORL database of faces [31], MMI-Facial Expression, Genki's database and Face Recognition Technology (FERET) database.

2 Technical framework

FER systems are a type of computer system that automatically verifies or identifies a person's emotion, provided that an image or a video from a fixed source is available for recognition purposes. FER is done by many different ways; a typical method to do so is to select these facial features from an image and then compare it with a known facial database [4]. Expression recognition technology is being predominantly being used in a multitude of ways. Its uses range from photograph tagging on social networks to security and authentication and targeted advertising based on the mood of the user [32]. Typically, the working of a FER system has four main steps as shown in Fig. 1.

1. The facial area is captured from the full image.
2. In some cases, the facial attributes are localized/segmented.
3. These attributes nodes are further numerically transformed into a feature vector.
4. These features are acquired from specific databases and further compared or processed for classification purpose.

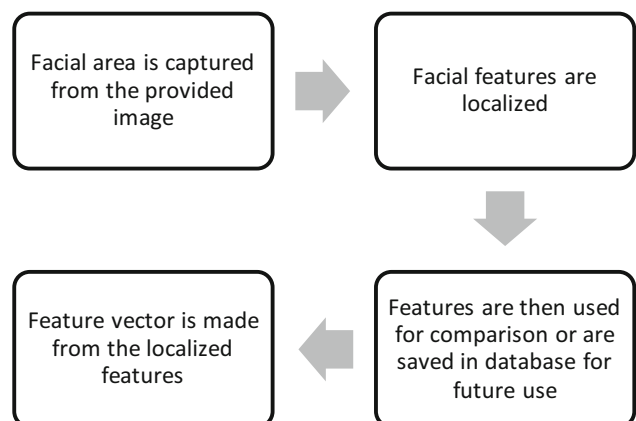


Fig. 1 Flowchart for working of a facial expression recognition system

This survey provides a comprehensive view of such systems found in text. Various mathematical and statistical models have been employed by different researchers for FER systems. Various models like fast Fourier transform, wavelet transform, principle component analysis (PCA) and local binary patterns (LBPs) have been used for feature extraction, while other models like support vector machine (SVM), neural networks, *k*-nearest neighbor and linear discriminant analysis (LDA) have been used for expression classification. This section discusses some state-of-the-art systems and explains how they have incorporated these models in order to obtain assiduous results.

2.1 Automatic expression recognition system (AERS)

Automatic expression recognition system (AERS) was being invented in the early 1990s. The work of Alirezaee et al. [2], Bartlett and Whitehill [3] studied AERS. Previously, such algorithms were tediously slow requiring time that extends over length manifolds as compared to the time span of a single facial expression. As time progressed, new optimal and efficient techniques started emerging, reducing the computational time [25]. Marian Stewart Bartlett gave the working of their system in three steps (1) localization—the process of narrow downing the region of face for detection purposes, (2) feature extraction—it sums out the most prominent features of the data and presents it in succinct form and (3) classification—the classification process is used to classify the different facial features into respective classes in the database as shown in Fig. 2. In Trujillo et al. [35], the authors proposed a system which used rule-based classifiers in which the mapping from feature values to facial expression is defined manually. Furthermore, some FER systems used machine learning-based classifiers, such as neural networks.

In such automated expression recognition system, the face of a person is first localized from the acquired image. The face is further subdivided into nodes to extract the features. Various types of features are used, i.e., geometric—these include the mouth, eyes, brows and cheeks; motion based—the features which can show some movement like the eyebrows, lips; and feature based—these are the types of features a face possess, e.g., size of eyes, nose shape (Fig. 3).

Firstly, the facial area is localized in an image by an entropy method to detect the expressions communicated.

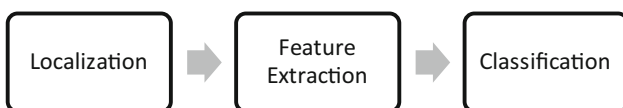


Fig. 2 Schematic of AERS technique

The entropy metric generate values based on the low- and high-level information corresponding to the facial location.

$$H = \sum_{i=1}^m p_i \log \frac{1}{p_i} \tag{1}$$

$$H = - \sum_{i=1}^m p_i \log p_i \tag{2}$$

where $e_1 \dots e_m$ are the events occurring with probabilities $p_1 \dots p_m$.

Various image filters and algorithm are then applied to filter all three features that are geometric, motion based and feature based, in their respective categories. After localization, feature extraction is performed. The primary objective behind feature extraction is to reduce a large amount of data into a succinct form such that information regarding obscure patterns is not lost. Such representation reduces the cost of resources that may be required to designate a large set of data. The major problem faced while performing analysis of complex and large data stems from the number of variables involved. Furthermore, color, texture and shape are the three main features of any image [33]. These features are also extracted to enable the system to distinguish among facial expression patterns.

2.1.1 Color feature extraction

Color is one of the most significant aspects of an image. Various color space or color models are used to depict the color depth. Color is one of the most immediate visual features perceived by the human sensory system [30]. Every color-based matching algorithm is performed in three steps (1) selection of a color space such as red, green and blue (RGB), (2) representation of color features and (3) matching algorithms. They are calculated as:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij} \tag{3}$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \tag{4}$$

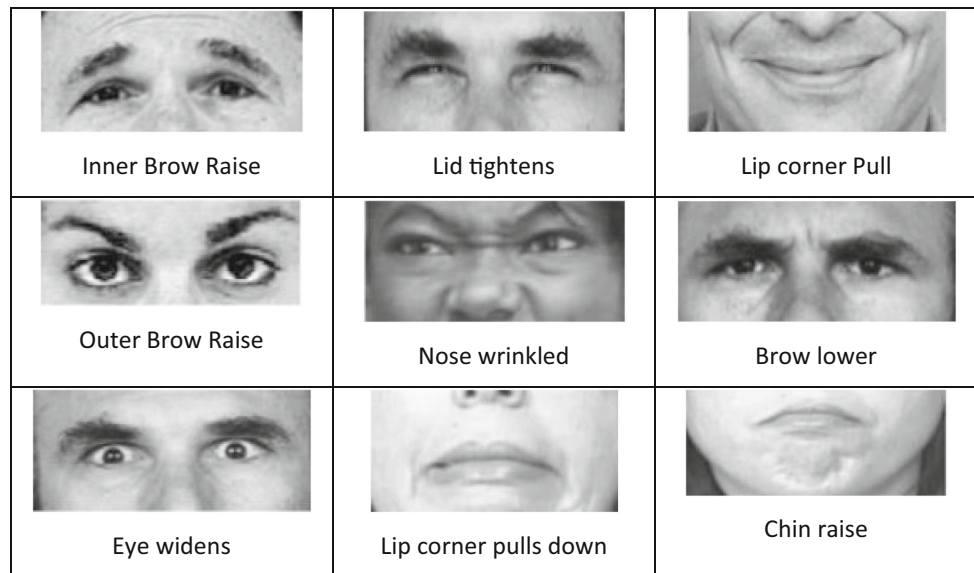
$$\gamma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \tag{5}$$

where f_{ij} is the color value of *i*th color component of *j*th image pixel and *N* is the total number of pixels in the image.

2.1.2 Texture feature extraction

Texture is a significant characteristic of a wide range of images. It is believed that human’s visual system uses texture for interpretation and recognition. Image texture is

Fig. 3 Role of various facial components in facial expressions [3]



quantized into a set of metrics computed using some image processing techniques. In this case, textures are detected using Gabor's filter. A bi-dimensional Gabor function $g(x, y)$ is given as:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W_x \right] \quad (6)$$

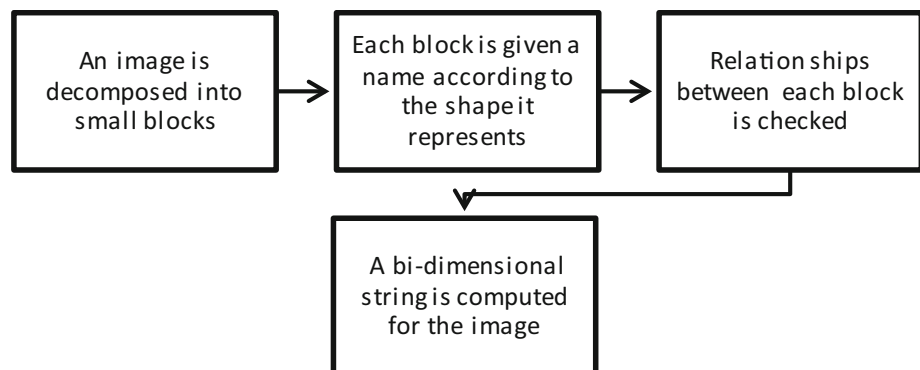
where σ_x and σ_y are the scaling parameters and W is the central frequency.

2.1.3 Shape feature extraction

Shape features are less developed than the color and texture features. The shape extraction works in a hierarchical flow as shown in Fig. 4.

After all the expressions are extracted, they are then mapped to the images from the databases. The images are then classified according to the represented expressions. These results are then stored in databases, for future detection.

Fig. 4 Shape feature extraction hierarchy



2.2 Facial image as a whole pattern

The previous technique uses some localized feature of the face in order to classify an expression. Some techniques use the changes in the whole facial image during an expression to identify it. In Kimura [20], the authors have used the idea that FER can be performed by extracting the discrepancy of an arbitrary expression as compared to a neutral while considering the facial area as a whole pattern. Kimura gave three types of expression classes, namely happiness, anger and surprise. The technique proposed by Kimura endeavors to identify any of the three expressions based on the overall features of the whole face. It works as follows:

1. It first normalizes the faces.
2. Then, a potential net is used to connect all nodes of the features to their neighbors. A potential net is a potential field using a physical elastic model made to explicitly define all feature nodes on the face.

3. The connection of all nodes removes any noise and gains a motion flow.
4. Lastly, when expressions are obtained, then it is smoothed by edge detection and Gaussian filters.
5. These features are then saved to databases for further research.

The flowchart for the working of system is given in Fig. 5.

Once the required image has been captured from the full image, normalization is performed to mark the features on the face. Normalization is done in four steps:

1. The center of right and left eye and mouth are selected and marked as E_r , E_l and M , respectively (Fig. 7).
2. Then, a line is drawn from left to right eye. The center point of this line is defined as O which is then connected to M (Fig. 7).
3. Affine transformed is performed on these lines to normalize their length.
4. Required facial components area is then selected to determine the facial expression.

Figure 6 depicts that E_r is the center of right eye, E_l is the center of left eye, O is the center of line drawn from points E_r and E_l and M is the center of the lips. Normalization is performed to eliminate any effect of location variation of the face in different images.

After required facial components are obtained, a potential net is drawn. It is a 2-dimensional mesh in which each node is joined to its neighboring node with flexible edges (Fig. 7). The basic purpose of using the potential net is to reduce noise, to track down motion of facial features and to track down the motion flow of featureless points of the face.

Now the deformation is performed. The motion of any node N can be determined by the following equation:

$$F_{\text{ext}} = m \frac{d^2 n_{i,j}}{dt} + \gamma \frac{dn_{i,j}}{dt} + F_{\text{spring}} \tag{7}$$

$$F_{\text{spring}} = k \sum_a^4 \left(\left| \frac{l_{i,j}^a}{l_{i,j}^a} \right| - l_0 \right) \frac{l_{i,j}^a}{\left| \frac{l_{i,j}^a}{l_{i,j}^a} \right|} \tag{8}$$

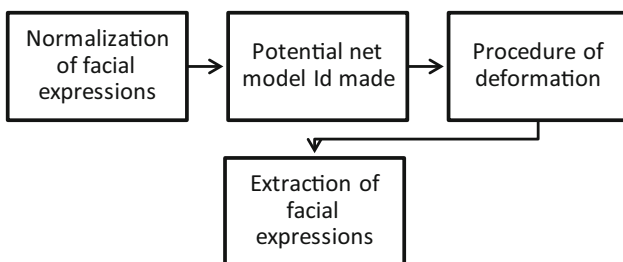


Fig. 5 Steps of facial expression recognition

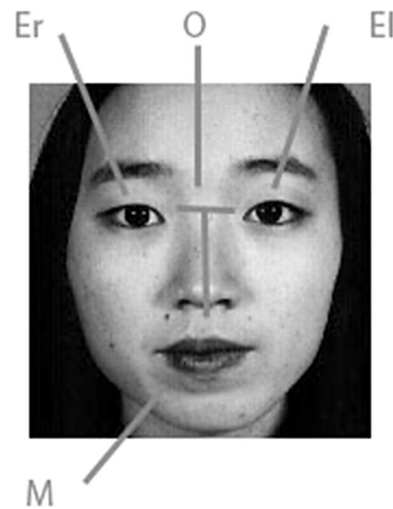


Fig. 6 Normalization of a facial image



Fig. 7 Potential net

where $n_{i,j}$ is the two-dimensional coordinates of the net, m is the mass of the node N , γ is the coefficient, l is the size of spring, k is elastic constant, l_0 is the distance between adjacent nodes, F_{spring} is an internal force and F_{ext} is an external force. Under stationary condition, we assume:

$$F_{\text{spring}} = F_{\text{ext}} \tag{9}$$

After the values are calculated, K - L expansion is applied for the extraction of features of face and their expressions. The expressions are then highlighted on a normalized face to see which expression is being proposed (Fig. 8). The expressions extracted are then saved in databases.

2.3 Graph-preserving sparse nonnegative matrix factorization (GSNMF)

Graph-preserving sparse nonnegative matrix factorization (GSNMF) is a derivative of the original nonnegative matrix



Fig. 8 Normalized face depicting the expression and highlighted face

factorization (NMF) process. This process uses both graph-preserving and sparse properties. Its sparse properties are most beneficial for dimension reduction. A dimension reduction technique is the process of decreasing the number of unplanned variables. In Zhi et al. [39], the authors have proposed a GSNMF method. GSNMF can use in both unsupervised and supervised manner for dimension reduction. Supervised method requires manual intervention for labeling input data, while the unsupervised version works on unlabeled data. It has two components, i.e., feature selection and feature extraction. The supervised dimension reduction projects data in a linear format, and it uses different labels to choose projection. It projects data into lower dimensions which improves the classification. On the other hand, unsupervised dimension reduction uses the transform function to project the lower dimensions. It is often used before the supervised way of dimension reduction, and they can also be linked together. In GSNMF technique, sparse demonstration of the images is acquired by minimizing the first-degree normalization of the base images. The neighbors of the reduced base image are then saved by keeping the graph construction in the mapped space. After that, decomposition of the images is performed which converts high-dimensional expressions to low-dimensional subspace. The observed facial expressions were disgust, anger, happiness, fear, surprise and sadness as illustrated in Fig. 9.

Experimental results performed by Ruicong Zhi, Markus Flierl and Qiuqi Ruan on facial images from Cohn–Kanade database show that GSNMF overtakes other NMF and other renowned FER approaches, like eigenfaces, Fisher faces and Laplacian faces. The local arrangement of the test samples and class label information to predict the emotion is used which is helpful for organizing various facial features. The databases used are Cohn–Kanade, the JAFFE and the GENKI.

The NMF algorithm is used for finding nonnegative decompositions of the acquired data matrix. It characterizes a facial expression as a linear grouping of various

imageries. The extracted image contains the basic vectors of eyes, nose and mouth.

In GSNMF algorithm, NMF is extended to include a sparseness constraint. The neighboring samples are saved by minimizing the graph-preserving standard in the drawn space. First of all, the facial image is converted into a matrix, and then, l^0 -norm operation is applied to place zeros at non-feature place. By applying this norm, a cost function is formed:

$$\text{GSNMF}(S||\text{TA}) = G(S||\text{TA}) + \sum_{k,j} \omega_{k,j} \quad (10)$$

such that: $\omega_{k,j} \geq 0, \quad h_{k,j} \geq 0, \quad \forall k, i, j$

where W is the basis images, H is the elements of the coefficient matrix and X is the original image. By this function, the features are localized and are matched in GENKI database system to determine which expression is represented by the original image. In this technique, three different occlusions are compared, i.e., eye, nose and mouth occlusion (Fig. 10).

2.4 Two-phase test sample representation method

Two-phase test sample representation (TPTSR) method is a powerful algorithm for face detection. It is a two-step process which identifies the facial expression from a facial image. In Xu et al. [38], the authors have concentrated on a TPTSR technique for facial expression identification. The initial stage of the method is used to present the test model of an image as a linear amalgamation of all the training set. Every training model is used to determine N which is the nearest neighbors for the test model. The second stage uses the gained test model and the nearest neighbors and performs sorting of the results. The principal component analysis PCA [8] and the LDA are two examples of transform techniques discussed by Xu et al. The authors show that these transformation methods usually employ the integral band of training samples to obtain transformation axes and then map each training and test instance onto the transformation axes to create a depiction of the tester. A large part of experiments illustrate a respectable execution of this technique. Databases used for this method were the FERET database, ORL database and AR database. Some of the test samples from respective database are given further (Fig. 11).

In this procedure, initially it is presumed that there are O classes and n number of samples. The first five images of each test subject are used as the training samples, while the rest are used as the test instances. In the beginning stage of the two-phase test sample representation (TPTSR), it uses all of the training data to understand every test model and associates the result to identify the N nearest neighbors of the test instance by consolidating on the training data.

Fig. 9 Images from Cohn–Kanade database of three different people, showing seven types of emotions [19]

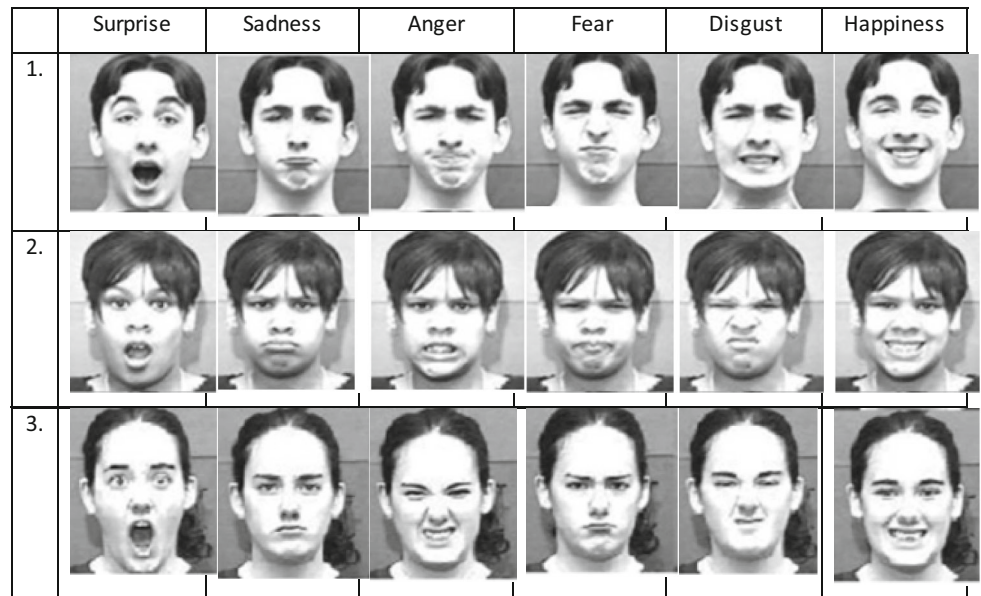


Fig. 10 Comparison of eyes, nose and mouth occlusion

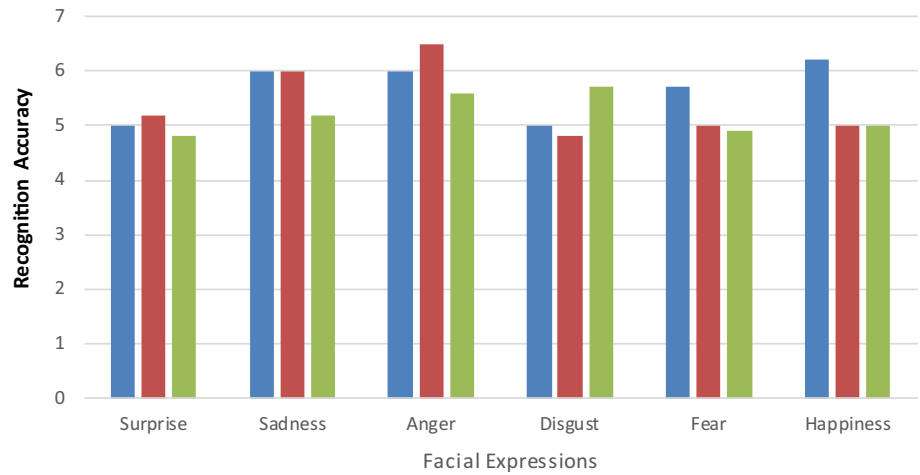


Fig. 11 Some face images of a subject from the AR database [38]



$$y = a_1x_1 + \dots + a_nx_n \quad (11)$$

where y is the test instance and a_i are the coefficients. By rewriting the statement, we get

$$y = XA \quad (12)$$

where $A = [a_1 \dots a_n]$, $X = [x_1 \dots x_n]^t$. This equation indicates that training samples make their own contribution to test the sample. Also

$$y = c_1\tilde{x}_1 + \dots + c_N\tilde{x}_N \quad (13)$$

where \tilde{x}_i ($i = 1, 2, \dots, N$) are the identified instances of nearest neighbors and c_i ($i = 1, 2, \dots, N$) are the coefficients. By rewriting the equation, we find:

$$y = \tilde{X}C \quad (14)$$

where $C = [c_1 \dots c_N]$, $\tilde{X} = \tilde{x}_1 \dots \tilde{x}_N$

If \tilde{X} is not a singular matrix, C is computed by:

$$C = (\tilde{X})^{-1}y \quad (15)$$

Otherwise:

$$C = (\tilde{X}^T\tilde{X} + \gamma I)^{-1}\tilde{X}^Ty \quad (16)$$

where γ is the positive constant, I is the identity matrix. After C is obtained, $\tilde{X}C$ is referred as the representational result of our technique.

Some test samples have been taken from the ORL database to apply this technique to get facial expressions. The initial row depicts the original test data. Subsequent rows show the outcome obtained using the TPTSR and its globalized version, respectively (Fig. 12).

2.5 Facial expression recognition and analysis: detection

In Valstar et al. [36], authors have placed the initial experiment on facial expression recognition and analysis (FERA). It consists of two subchallenges, namely emotion recognition and action unit detection. This delineates the data to be used for the challenge and the challenge protocol. In summation, it describes a standard system that uses SVMs, LBP and PCA to either observe the activation of action units (AU) per frame or classify emotions in a video. Subsequently, these techniques are used to either perceive the initiation of AUs per frame or recognize sensations in a video sequence. The results show that the data imports some level of difficulty, but with new techniques discovered it is not impossible to detect emotions from sequencing videos.

An overview of baseline approach is given in Fig. 13.

The main working of the system starts at the stage of feature extraction (Fig. 14). LBP is a potent way of texture depiction. It yields a 3×3 neighborhood of every pixel along with a key value.

The operator for the general case is represented by $LBP_{P,R}^U$, where P is the neighbor set and R is the radius of the circle. Superscript U shows the role of even rules. Later using the LBP operator as an icon, a histogram of the labeled image $f(x, y)$ is defined as:

$$H_i = \sum_{x,y} I(f(x, y) = i) \quad (17)$$

where $i = 0, 1, \dots, n - 1$ and n is the possible labels produced by LBP operator

Fig. 12 Samples from the ORL database for a single subject and corresponding output [38]



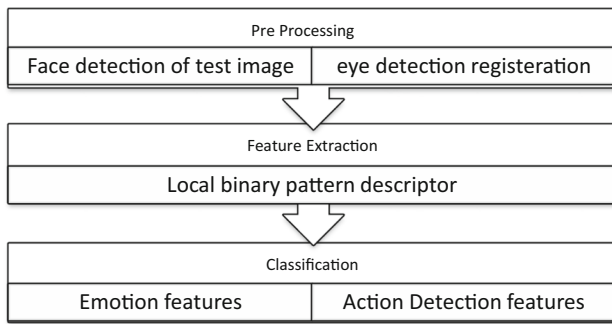


Fig. 13 Overview of baseline approach

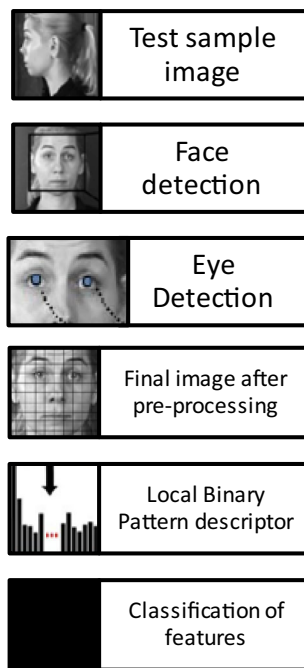


Fig. 14 Working of baseline approach

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

A distinct SVM classifier was made for each AU individualized. The set of AU has two subsets which are:

1. Upper face AUs: $G_u\{AU 1, AU 2, AU 4, AU 6, AU 7\}$
2. Lower face AUs: $G_l\{AU10, AU 12, AU 15, AU 17, AU 18, AU 25, AU 26\}$

Each video sequence is divided into sets of AU classifiers. These sets work on multiple frame units. The central figure of every set with a discrete AU combination is chosen. If an image has multiple similar AU sets, then the frame is trained from the first existence of this grouping. Finally, the features are standardized to lie in a scope of $[-1, 1]$.

Subsequently, another filter is employed to find the case of emotion, i.e., P , where $P = \{\text{anger, sadness, relief, joy, fear}\}$. To resolve the label Y of a video of m frames, the emotion with the most number of frames is searched:

$$Y = \arg \max_e \sum_{j=1}^m y_{e,j} \quad (19)$$

2.6 Performance-based character animation: avatar creation

In Weise et al. [37], the authors have formed performance-based character animation system that enables a user to govern the facial expressions of an animated avatar in real time. The user actions are recorded in a natural surroundings utilizing commercially available non-intrusive three-dimensional (3D) sensor. The latest developments in gaming technology, for example Nintendo Wii and the Xbox Kinetic, track the motion of a person for real-time communication. The objective is to use these technical approaches and create affordable facial animation system. The system should be able to detect motion of a user and transform it into an avatar. The main support of this system is that it combines 3D geometry and two-dimensional (2D) textures. They are then blended together to get a final image. The method is very successful in its working as it tracks compound facial expressions even from noisy inputs. Mapping of acquired depth maps is done with the extracted animation features. These 2D and 3D combinations produce more realistic results as compared to other techniques. The systems run as illustrated in Fig. 15.

The primary contribution of this technique is a method that combines 3D and 2D textures in such a manner that they can map various expressions on an avatar. This technique can get expressions, even if the input is real noisy.

In this method, a particular user is selected who requires an avatar for gaming purposes. Various expressions of the user are recorded through kinetic systems. These systems capture color images in 2D and depth image in 3D. This is carried out at 30 frames/s. These expressions are then converted to 2D/3D images. These images are then mapped to the avatar. Diverse tracking algorithms are put to work to track down the facial aspects and identify the expressions made by the performing user.

To integrate tracking and animation into one optimization blend, shape weights are used to represent facial expressions. These facial expressions are then directly animated by using commercial animation tools.

To track the expressions by using blend shape weights, following equation is used to calculate the final expressions for the avatar:

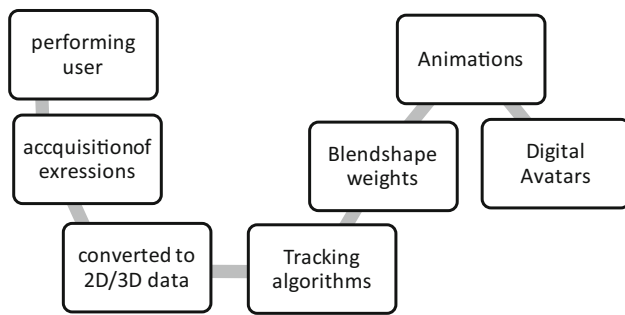


Fig. 15 Overview of processing pipeline

$$t_s^* = \frac{\sum_{t=0}^u w_t t_{s-t}}{\sum_{t=0}^u w_t} \tag{20}$$

where t_{s-t} denotes the trajectory at frame $s - t$. The weights w_s are defined as:

$$w_s = e^{-s \cdot H \cdot \max_{v \in [v,u]} ||t_s - t_{s-v}||} \tag{21}$$

where H is a constant used for noise reduction and k is the window size.

The avatars are a mixture of texture features and geometric features [4] of the input data. Both the features are extracted separately and are then combined to get the final look for the avatar.

2.7 Temporal template method

In Ahad et al. [1], the authors have described the motion history image (MHI) method which is a survey-based temporal template method. This method uses the action movements of the image to capture the emotions. The MHI method is a simple and understandable template matching approach. In this approach, the image sequence is first changed to static shape pattern and is then compared to already stored action samples during recognition. These approaches can be implemented easily, and very less computational load is required for them.

This is a new way to detect the expressions represented by any human. In this technique, the image of a person is first localized from the data. The nodes are then marked down as points to capture the important facial organs that are taking part in the making of the facial expression. These images are then converted to a static shape pattern. The MHI approach uses prestored action prototypes to compare to the results gained from the imaging data (Fig. 16).

MHI works as a temporal template. Each part of the pixel has some function of the motion specified at that particular pixel. These connected pixel meshes are then compared to stored models in databases. The MHI is calculated using the following function:

$$H_\tau(a, b, c) = \begin{cases} \tau & \text{if } \varphi(a, b, c) = 1 \\ \max(0, H_\tau(a, b, c - 1) - \delta) & \text{otherwise} \end{cases} \tag{22}$$

where a, b, c shows the position and time, $\varphi(a, b, c)$ represents object’s presence in the image, τ decides the temporal movement and δ is the decay parameter. This function is applied to each image of the video sequence, and expressions are detected.

2.8 Image sequencing

In Raducanu [9], the authors have discussed the challenging problem of analyzing image sequence for the purpose of FER. This scheme uses a temporal classifier to detect key frames. It is a view-based technique and is totally texture independent. Key frames are detected in the video. The temporal recognition scheme upsurges whenever a key frame is detected. The suggested system has two benefits. Firstly, the CPU time is significantly reduced as only very few key frames are observed for the recognition of the emotion. Secondly, since a key frame and its neighboring frames are describing the structure of the emotions proposed, the performance of the recognition system will be stimulated. The learning phase is simple as compared to other techniques.

The flow of the proposed technique is illustrated in Fig. 17.

Initially, key frames are detected from a video sequence. Whenever a key frame is found, a temporal pattern recognition algorithm is called forth. Two targets are computed from this realization

1. The L_1 norm τ_{a1}^a and
2. The temporal derivative

$$D_t = \frac{\partial \tau_{a1}}{\partial t} = \sum_{i=1}^6 \frac{\partial \tau_{a(i)}}{\partial t} \tag{23}$$

where D is the positive local maximum and τ_{a1}^a is a pre-defined threshold.

The system has three levels to get the required feature from the face.

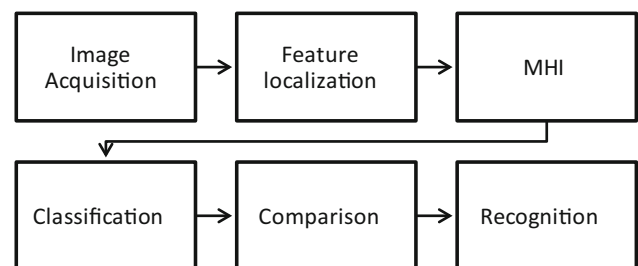
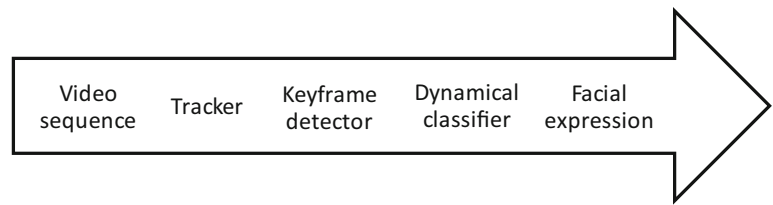


Fig. 16 Steps to MHI approach

Fig. 17 Process of FER using keyframes



1. Tracking level.
2. Key frame detection level.
3. Recognition level.

A tracker is used throughout the system processing which is used to track down the image sequence in the video. The tracker is timed manually to get the key frames from the video. When the tracker covers a selected amount of time, the user key frame detection is invoked. This invoked key frame then selects the current segment for recognition.

A confusion matrix is used to represent the result for key frames. Every key frame detects the level of feature represented separately. A sample confusion matrix is given in Table 1.

2.9 Facial expression recognition in cinematic series

In Moore et al. [28], the authors propose another innovative technique method for FER within videos. They have covered six basic facial expressions classifiers. Extensive experiments illustrate that two types of feature used for recognition which are geometric and appearance-based features. Geometric features are formed using the position and shape of various facial components. Variations in the appearance of the face during an expression are captured by appearance-based features. They are typically extracted by convolving specially designed image filters with the expression or its subareas. Geometric features exhibit sensitivity to noise and frequently call for dependable and precise facial feature exposure and tracking. In this technique, the features are drawn out, and it is changed to a binary map which consists of non-feature and feature pixels. Ultimately, an image is formed in which every pixel value represents the distance to the nearest feature pixel and is called chamfer of an icon. Strong classifiers are used for temporal boosting; lastly, the expression is classified

using a database. Cohn–Kanade database has been used to verify the results of this technique.

The classifier marks images to extract features. The canny edge procedure is applied to make edge maps for icons. Classifier banks are made, and temporal boosters learn the ideal subclass of features from this bank (Fig. 18).

The canny edge detection procedure boosts the borders. Image is smoothened to get rid of the interference. Sobel operator is used to obtain spatial gradient measurements. The source image is then compared along the direction of the image gradient. In case the image does not lie within the local maxima obtained by Sobel operator, it is rendered as zero. Chamfer matching is applied to measure the edge power alongside that feature in an icon. Every image within the training set goes through edge detection using canny edge detector. This creates an edge map given as F . Distance transform is applied to produce a chamfer image. The result contains pixel values r , and these values are bound by a proportional relation to the distance to its adjoining edge point in F (Figs. 19, 20):

$$DT_F(r) = \min_{e \in F} ||r - e||_2 \tag{24}$$

A chamfer total is calculated for every outline piece S , where $S = \{s\}$:

$$d_{\text{cham}}^{(S)}(DT_F) = \frac{1}{N} \sum_{s \in S} DT_F(s) \tag{25}$$

where N is border points in S . The mean distance among feature S and the chamfer image DTF determines the overall chamfer score. The function $d_{\text{cham}}^{(S)}(DT_F)$ is an effective classifier which is effectively used to classify and differentiate between different expressions. The technique connotes that the contours around the cheek and the edge of the mouth are used to classify the joy expression. However, mostly in case of negative expression, different areas of the

Table 1 Confusion matrix with 60 key frame strokes for a single unseen person

	Surprise (14)	Sadness (14)	Joy (14)	Disgust (14)	Anger (14)
Surprise	14	0	0	0	0
Sadness	0	14	0	10	0
Joy	0	0	14	0	0
Disgust	0	0	0	4	4
Anger	0	0	0	0	10

Fig. 18 Overview of FER system

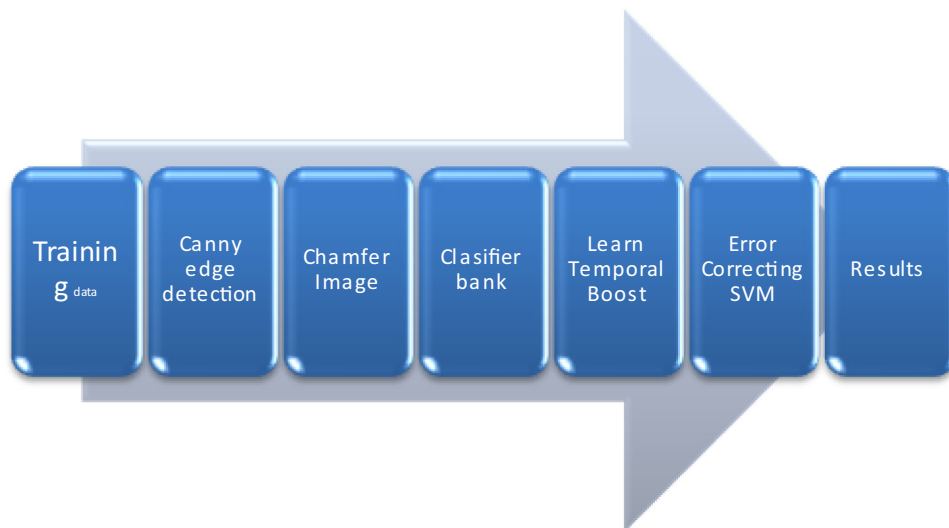


Fig. 19 Comparison of common facial expressions classifiers [28]



Fig. 20 Test samples of four individuals from ORL database



cheek and mouth contribute differently to the identification process (Fig. 19).

The temporal opening s starts at 0 and is prolonged as long as the complete classification error for the existing weights declines. The temporal boost procedure tries to distinct training instances by choosing the best weak classifier $h_j(x)$.

A weak classifier therefore contains a feature p_{ij} , a threshold s_j and a parity p_j which represents the path of the inequality sign.

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j < p_j s_j \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

where

$$f_j = d_{\text{cham}}^{(T)}(DT_E) \tag{27}$$

The effect of this technique is tested on Cohn–Kanade database. It is presented in a confusion matrix as given in Table 2.

2.9.1 Combination of Gabor filter, KPCA and SVM

In Meshgini et al. [27], the authors present unique face recognition method based on the Kernel principle component analysis (KPCA), SVM and Gabor filter bank.

Initially, a Gabor filter bank with eight orientations and five frequencies is applied to every face image in order to extract features minimizing distortions mainly caused by variations of facial expression, pose and illumination. The output of the filter bank is subjected to feature reduction techniques of KPCA to decrease the size of feature vectors. SVM is trained and subsequently used to classify the expressions. Furthermore, these expressions are compared to various databases to check the accuracy of the results (Fig. 22).

Experimental subjects are chosen and images are taken. The expressions gathered for experimental purposes exhibit variations in similar facial expressions such as glasses/no glasses, smiling/no smiling, and facial details such as open/closed eyes. All the images are taken against a dark homogeneous background with the subjects in an upright, frontal position with tolerance for some side movements (Fig. 20). The ORL face database is used for expression matching. This technique exhibits a maximum recognition rate of 98.5 %, which is higher than the other related algorithms applied to the ORL database.

The system works in a linear method. An input is taken as a facial image. Gabor filter is applied to the face to highlight the features used by the person. Downsampling is performed on the image, and then, it is transformed. KPCA algorithm is applied to reduce extra features of the face leaving behind the expression owned features only. The expressed results are then detected by multi-class SVM, and final results are tabulated (Fig. 21).

Human facial features are firstly extracted using Gabor wavelets. A 2D Gabor filter is a Gaussian kernel function represented by:

$$\Psi_{\omega,\theta}(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \exp(j\omega x') \quad (28)$$

$$x' = x \cos \theta + y \sin \theta, \quad y' = x \sin \theta + y \cos \theta \quad (29)$$

where (x, y) is the pixel position, ω is the central angular frequency, σ represents the sharpness of Gaussian filter and θ is the anticlockwise rotation of Gaussian filters. The acquired image is convolved with Gabor filters to obtain following equation:

$$G_{m,n}(x,y) = I(x,y) \times \Psi_{\omega_m,\theta_m}(x,y) \quad (30)$$

PCA is a technique used for reducing a large set of pixels to a succinct set. It is used to detect the face space required to extract the facial expressions. After the Gabor filter, KPCA is applied to the subject image. It is applied in the form of kernel matrix which is as follows:

$$K_{ij} = \Phi(x_i) \cdot \Phi(x_j) \quad (31)$$

where $\Phi(x_{i-j})$ represents a data matrix in the feature space.

After the face space is fixed, SVM is used to minimize the number of misclassified training models. The decision function that is used to minimize models comparison is given as:

$$f(x) = \text{sgn}\left(\sum_{x_i \in S} \alpha_i y_i K(x_i, x) + b\right) \quad (32)$$

where f is the decision function and S is the set of support vector. After calculating the result, they are saved in databases for further comparison.

2.10 Gabor filter in combination with a neural network classifier

In Kumbhar et al. [21], the authors have discussed the application of feature extractions using Gabor filter and neural network classifier. This technique recognizes seven types' of facial expressions from distinct images of humans. FER has various applications in diverse phases of everyday life. The importance of these applications has not yet been comprehended due to lack of effective expression recognition methods. A fusion of statistical methods for expression recognition system has been proposed. Initially, the facial images are extracted from a database, after that feature points are captured. Later, various feature extraction filters are applied to them such as Gabor filters. Furthermore, PCA is applied over the extracted features to reduce the feature vector length of the image. The acquired image contains all the required image features (Fig. 22). These features are used by the classifier which gives the best result of the FER. The Japanese Female Expression (JAFEE) database is used for the input purpose for this technique.

The system's working executes as follows:

The process begins with the acquisition of the image. During the image preprocessing step, image scaling,

Table 2 Confusion matrix for fivefold cross-validation on Cohn–Kanade database

	Joy (100)	Surprise (100)	Fear (100)	Sad (100)	Disgust (100)	Anger (100)
Joy (100)	94	0	3	7	1	0
Surprise (100)	0	95	0	2	4	0
Fear (100)	5	0	75	2	4	6
Sad (100)	0	0	6	85	0	8
Disgust (100)	0	0	3	0	91	6
Anger (100)	0	0	0	9	15	76

Fig. 21 Process flow of proposed algorithm

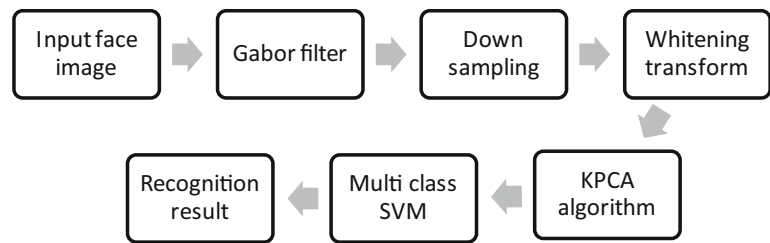
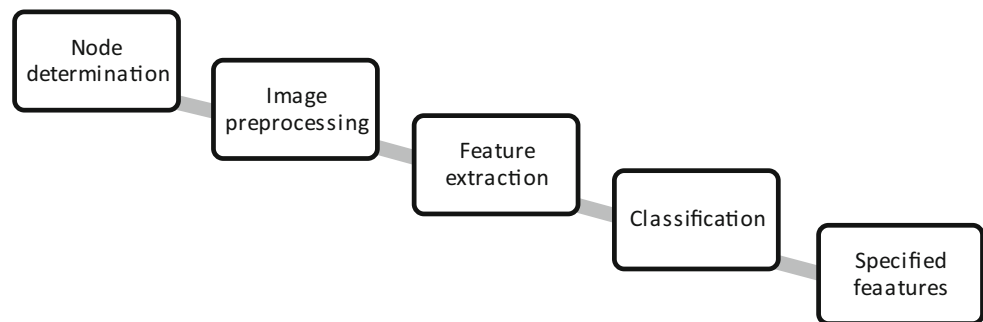


Fig. 22 Process flow of FER using a Gabor filter



illumination intensity, contrast, correction and other enhancement operations are applied. The image is adjusted to the required setting, and feature extraction is done. In this step, a set of feature vector is used to map features in the image. After this, a 2D Gabor filter is applied to smooth the image and remove noise. Next PCA is used to lower the face space dimension. It reduces the amount of datasets and reduces the complexity of image process. Finally, the classifier is applied to compare the facial expressions.

2.11 Eigenface method

In Chakrabarti and Dutta [6], the authors have introduced a modification of eigenfaces approach for FER. In this method, the eigenface technique is used for expression recognition and not for identification of the person. It starts with the human vision as a standard reference point and then computes the expressions contained by the image of a test face. Six standard expressions anger, disgust, fear, happy, sad or surprise are detectable by this method. In this technique, a projection of test image to each eigenspace is made, and then, the closest matching eigenspace is selected and the class of the corresponding eigenspace is the class of the input image. The current focus is on FER from still images using a modified recognition method. The JAFEE database is used for the base of images to detect the selected expressions using the eigenface method.

In this proposed technique, PCA is used along with the snap sort method to reduce dimensionality. The images are captured from a standard database of images. These images are free from fringe space of the face, which makes it easier to calculate the expression measure and its intensity. They



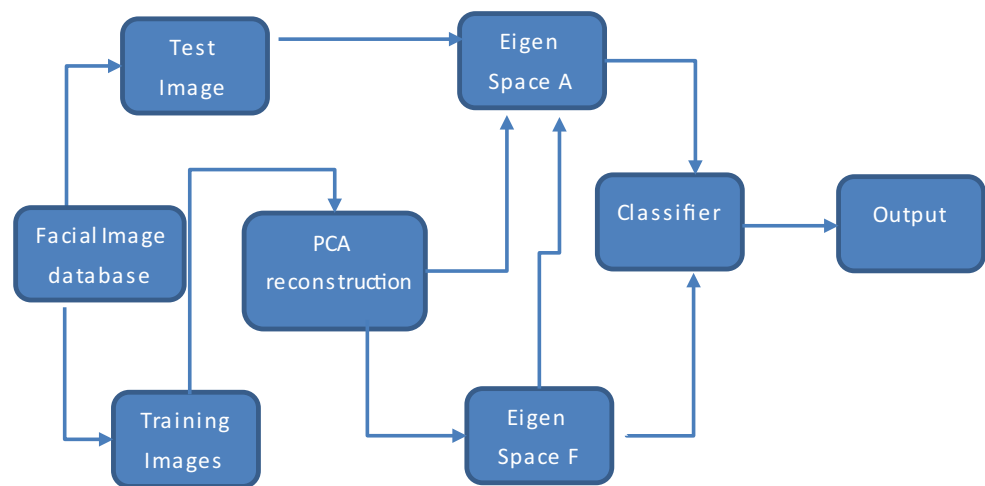
Fig. 23 Universal facial expressions

are then categorized into six classes based on six universal facial expressions proposed by humans (Fig. 23).

Eigenspace is calculated by each image separately. After calculating it, similarity is measured. And the image is placed in the class which has the most common traits. The eigenspace method is carried out as given in Fig. 24.

2.12 Databases

Database is an essential part of any computing system. It is used to store significant and expensive information. This stored information is later retrieved by other users for future use such as comparison and analysis of data in terms of accuracy and consistency. It is a matter of mutual concern that a common database is used by all the researchers.

Fig. 24 Flow of eigenface technique

Such standardization will make testing, benchmarking and comparison of the new with the existing more credible and authentic. The FERET database has been accepted as a standard for testing face recognition systems [13]. The existence of a large dataset is necessary for FER as it can have all variation of expression and also a variety of posed or spontaneous expressions [29]. Cohn–Kanade database also known as the CMU-Pittsburgh AU-coded database is a fairly widespread database and has been broadly used by the face expression recognition community. It is a posed expression database containing 500 image sequences from 100 subjects. The age group of the people is between 18 and 30 years [19]. Some of the other databases that contain only posed expressions are the AR face database which has a dataset of 126 peoples, the CMU Pose with 68 individual’s dataset (gross) and the JAFFE database which has images of 10 Japanese female models expressing 6 different expressions [7].

3 Comparative study

FER is not perfect and struggles to perform under certain conditions. Among the different biometric techniques, FER has reliability issues. Along with many benefits, it has few challenges too. Some automated expression recognition systems have problems in detection of the expressions when a person is speaking continuously. People such as professional actors can fake an expression very well, so sometimes it is difficult to distinguish between real and fake expressions. Certain techniques are very fast and accurate, but at the same time they are supervised so it requires continuous manual intervention throughout the experiment. Similarly, some techniques require continuous uninterrupted user support during preprocessing. Another factor that seriously effects the performance of the system

is the variance in individual facial expressions. A smile larger or smaller than usual can make the system less effective. Besides the few benchmark expressions, there are a number of other expressions that can be recognized. Segmenting and classifying spontaneous expressions is even more perplexing. None of the methods yield the same result for all of the classes; for example, anger is often confused with disgust. Furthermore, illustrated results demonstrate that the recognition percentages are different for varying expressions. Factors that cause variation in similar expressions such as cultures, age groups and other social aspects also effect the accuracy of the system. A face expression system that provides consistent results must be immune to such factors.

The work of Brunelli and Poggio [4] describes the working of AERS. They have used various databases to compare the facial expressions. Their working shows that this technique is 89 % accurate and the system yields 80 % efficiency. The study of Carlson [5] used FER technique using normalization of face and potential net model. This technique is fundamentally built on the degree of facial expression that can be extracted from a human face. The results have been tested on ORL database. It yields 85 % result on accuracy and 78 % result in efficiency of the process. The work of Deng et al. [8] proposed a novel GSNMF algorithm for FER. This technique exhibits 90 % of accuracy for mouth, nose and eye occlusion on Cohn–Kanade database, whereas with JAFFE database it yields 93.45 % accurate results. In Tian [33], the authors have concentrated on a TPTSR method for facial expression identification. This technique has performed experimental results on FERET, ORL and AR database. The accuracy of with the mentioned databases is 89, 90 and 87 %, respectively. The work of Chakrabarti and Dutta [6] described the initial experiment on facial FERA. This technique uses each frame as a single unit, thus providing more accuracy.

The system has an accuracy rate of 91 %, but the efficiency is 70 % each frame needs to be checked. A system for performance-based character animation has been presented by Weise [37]. This system takes a human input, studies its features and different movements and finally produces an avatar. This technique is prone to noise so its accuracy is less as compared to other FER techniques. With 79 % accuracy and 76 % efficiency, this technique is used only for gaming purposes because the avatar's expressions can be detected even with the noise. The work of Gross has described the MHI approach, and 2D and 3D approaches have been used. Although these approaches are expensive, they provide reasonable results. The accuracy on this particular approach is 89 %, whereas the efficiency is 72 %. A lot of future research work is required in this direction as it is new and is prone to errors. Fridlund [15] has discussed the challenging problem of image sequence in FER. This technique uses key frame method to detect the facial expression. It is a robust technique because it uses a video frame to detect the expressions of a person. It produces more accurate results because it judges the expression along with other bodily motions. This makes the expression more accurate. The accuracy of this technique is 93 %, but due to single key frame detection it is a bit slower in its efficiency which is 80 %.

The work of Kimura [20] proposed a facial expression classifier which uses a canny edge detection technique to verify the prominent facial expressions from an input image. This technique is applied to video sequences. It yields 86.1 % clear recognition results. The results have been tested on Cohn–Kanade database for confirmation. The most accurate result is gained for surprise achieving over 95 % accuracy. A method has been presented by Fasel and Luetin [13] which combines Gabor filter for feature extraction, KPCA for feature reduction and SVM for classification of the extracted features. This is one of the best and the most accurate techniques proposed which yields 98.5 % accuracy when the ORL database is used for comparison. It works fast and precisely yielding accurate measurement of the extracted feature. An application of Gabor filter-based feature extraction is described by Pantic and Rothkrantz [29]. It uses Gabor filter along with feed-forward neural network as the classifier. Work on seven facial expressions was performed with still pictures of human faces. The algorithm is 70 % accurate when tested on JAFFE database. PCA along with snap sort technique is used in the work of Bartlett and Whitehill [3]. It is a modified method of eigenface approach. The algorithm was applied to the JAFFE database to check the accuracy. About 60 images were chosen randomly for the training set. The result of the test is given in Tables 3 and 4.

The comparison of all techniques is summarized in Table 5:

Table 3 Accuracy of the system for each class

Expression class	Accuracy (%)
Anger	95
Disgust	80
Fear	90
Happy	80
Sad	75
Surprise	85
Overall	84.16

Table 4 Overall accuracy of each system

Author	Accuracy (%)
Bartlett and Whitehill	89
Kimura	85
Zhi	93.45
Xu	90
Valstar	91
Weise	79
Ahad	89
Raducanu	93
Moore	95
Meshgini	86.1
Kumbhar	70
Chakrabarti and Dutta	94

Table 5 Accuracy and efficiency of FER techniques

Author name	Cross-validation (%)	Leave one out	Time
Bartlett	88	89.67	0.0367
Kimura	85.98	86	0.0345
Zhi	93	93.88	0.0221
Xu	89.56	90.34	0.0456
Valstar	91	91.66	0.037
Weise	78	79.09	0.0541
Ahad	88.67	90	0.0343
Raducanu	92	93.56	0.0430
Moore	95.23	95.12	0.0252
Meshgini	85.41	86.34	0.0534
Kumbhar	70	71.45	0.0311
Chakrabarti and Dutta	93.81	95.12	0.0452

This bar chart in Fig. 25 exhibits the accuracy of all the discussed techniques. With 95 % accuracy, canny edge detection algorithm using chamfer image has the highest rate which is also depicted in Fig. 26. Similarly, the ROC graph given in Fig. 26 illustrates and compares the overall accuracy of each system discussed.

Fig. 25 Comparative analysis of the accuracy of each system

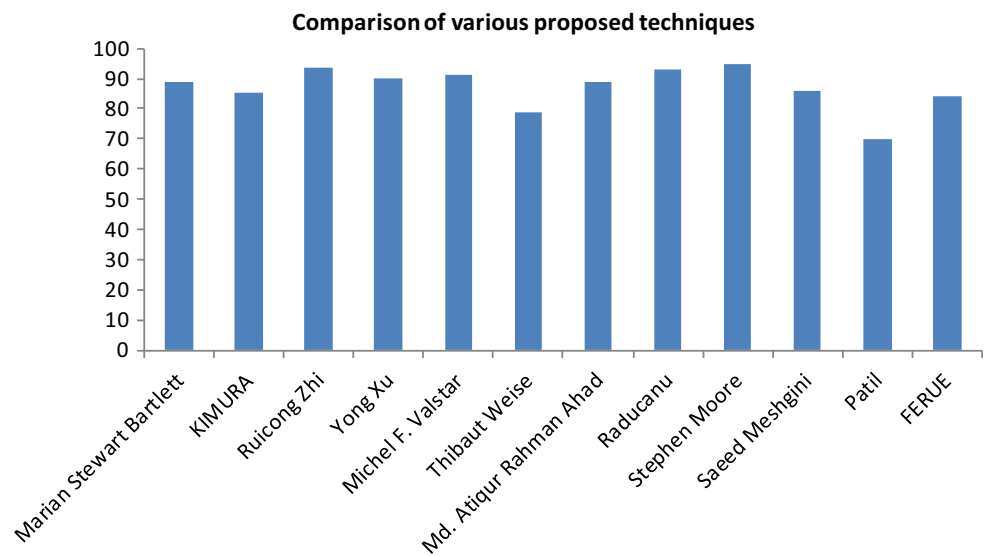
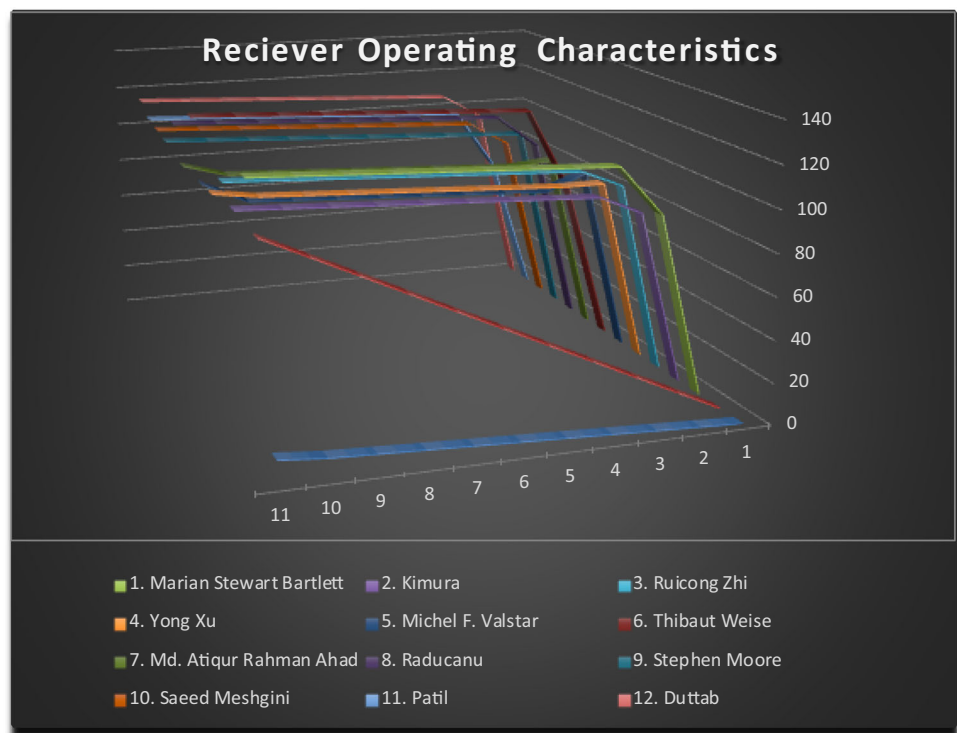


Fig. 26 ROC graph depicting the overall performance of each system



To check the performance of the different techniques, cross-validation and leave-one-out strategies were used. In cross-validation, the facial expressions were randomly divided into different sections. Training and testing was done by repeating the procedure ten times (tenfold cross-validation). The average result of all procedures was then taken and is given in the above table. In leave-one-out strategy, one image was used as a test sample and the rest were used to train. The technique proposed by Stephen Moore has given the best recognition rates in the least amount of time used.

4 Conclusion

In the 1960s, the scientist began working on various softwares for FER. It has come a long way since then. FER is based on the ability to distinguish among facial expressions and also evaluate the various features of the expression. It is another form of interacting with people and recognizing them. Several methods are employed to observe the facial expression projected by a mortal. This report has reviewed the previously used techniques and algorithms. Working of each algorithm has been explained in detail. Each

technique has its own advantages and disadvantages in terms of accuracy and efficiency. Some systems offer very precise solutions, but are less effective and vice versa. The highest accuracy rate is provided by the technique using canny edge detection algorithms and chamfer image method. In the future, two or more classifiers or algorithms can be utilized together to find more accurate and resourceful solutions.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding this publication.

References

- Ahad MAR, Tan JK, Kim H, Ishikawa S (2012) Motion history image: its variants and applications. *Mach Vis Appl* 23(2):255–281. doi:10.1007/s00138-010-0298-4
- Alirezaee S et al (2006) An efficient algorithm for face localization. *Int J Inf Technol* 12(7):30–36
- Bartlett MS, Whitehill J (2010) Automated facial expression measurement: recent applications to basic research in human behavior, learning, and education. In: Calder A, Rhodes G, Haxby JV (eds) *Handbook of face perception*, vol 7. Oxford University Press, San Diego
- Brunelli R, Poggio T (1993) Face recognition: features vs templates. *IEEE Trans Pattern Anal Mach Intell* 15(10):1042–1052
- Carlson NR (1986) *Physiology of behavior*. Allyn & Bacon, Boston
- Chakrabarti D, Dutta D (2013) Facial expression recognition using eigenspaces. In: *International conference on computational intelligence: modeling techniques and applications*, India
- Dailey MN, Joyce C, Lyons MJ, Kamachi M, Ishi H, Gyoba J, Cottrell GW (2010) Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion* 10(6):874–893
- Deng HB, Jin LW, Zhen LX, Huang JC (2005) A new facial expression recognition method based on local Gabor filter bank and PCA plus LDA. *Int J Inf Technol* 86–96
- Dornaika F, Raducanu B (2009) Facial expression recognition for HCI applications. *Encycl Artif Intell* 2(2009):625–631
- Ekman EL (1993) Facial expression and emotion. *Encycl Neurosci (Suppl. 3)*:51–52
- Ekman P (2006) Darwin and facial expression: a century of research in review. *Ishk*, San Jose, CA
- Ekman P, Rolls ET, Perrett DI, Ellis HD (1992) Facial expressions of emotion: an old controversy and new findings [and discussion]. *Philos Trans R Soc B Biol Sci* 335(1273):63–69
- Fasel B, Luetttin J (2003) Automatic facial expression analysis: a survey. *Pattern Recognit* 259–275
- Fogel A, Thelen E (1987) Development of early expressive and communicative action: reinterpreting the evidence from a dynamic systems perspective. *Dev Psychol* 747–761
- Fridlund AJ (1994) *Human facial expression: an evolutionary view*. Academic Press, San Deigo
- Frith C (2009) Role of facial expressions in social interactions. *Comput Emot Man Mach* 3453–3458
- Jack RE, Garrod OGB, Yu H, Caldara R, Schyns PG (2012) Facial expressions of emotion are not culturally universal. *Proc Natl Acad Sci* 109(19):7241–7244
- Kanade T, Yamada A (2003) Multi-subregion based probabilistic approach toward poseinvariant face recognition. In: *Proceedings of the 2003 IEEE international symposium on computational intelligence in robotics and automation*, 2003, vol 2. IEEE, pp 954–959
- Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. In: *Proceedings of the fourth IEEE international conference on automatic face and gesture recognition*. IEEE
- Kimura MY (2010) *Facial expression recognition and its degree estimation*. The Kansai Electric Power, Osaka
- Kumbhar M, Jadhav A, Patil M (2012) Facial expression recognition based on image feature. *Int J Comput Commun Eng* 1(2):117
- Lucey PC (2010) The extended Cohn–Kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: *IEEE Computer Society conference on computer vision*. IEEE
- Lyons MJ, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with Gabor wavelets. In: *Proceedings of the third IEEE international conference on automatic face and gesture recognition*, IEEE Computer Society, Nara, Japan, 14–16 April 1998, pp 200–205
- Malatesta CZ, Culver C, Tesman JR, Shepard B, Fogel A, Reimers M, Zivin G (1989) The development of emotion expression during the first two years of life. *Monogr Soc Res Child Dev* 1–136
- Matsumoto D, Ekman P (2008) Facial expression analysis. *Scholarpedia* 3(5):4237
- Mehrabian A (1981) *Silent messages: implicit communication of emotions and attitudes*. Wadsworth, Belmont, CA
- Meshgini S, Aghagolzadeh A, Seyedarabi H (2012) Face recognition using Gabor filter bank, kernel principle component analysis and support vector machine. *Int J Comput Theory Eng* 4(5):767
- Moore S, Ong EJ, Bowden R (2010) Facial expression recognition using spatiotemporal boosted discriminatory classifiers. *Int Conf Image Anal Recognit* 1(1):405–414
- Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state art. *IEEE Trans Pattern Analysis and Mach Intell* 22(12):1424–1445
- Pitas KS (1996) Face localization and facial feature extraction based on shape and color information. In: *International conference on image processing*, Lausanne, pp 483–486
- Samaria FS (1994) Parameterisation of a stochastic model for human face identification. In: *Proceedings of the second IEEE workshop on applications of computer vision*. IEEE, pp 138–142
- Shergill GS et al (2008) Computerized sales assistants: the application of computer technology to measure consumer interest—a conceptual framework. *J Electron Commer Res* 9(2):176
- Tian D (2013) A review on image feature extraction and representation techniques. *Int J Multimed Ubiquitous Eng* 385–396
- Trevarthen C, Aitken CT (2001) Infant intersubjectivity: research, theory, and clinical applications. *J Child Psychol Psychiatry* 3–48
- Trujillo L, Olague G, Hammoud R, Hernandez B (2005) Automatic feature localization in thermal images for facial expression recognition. In: *IEEE Computer Society conference on computer vision and pattern recognition (CVPR '05)-workshops*, pp 14–25. IEEE
- Valstar MF, Jiang B, Mehu M, Pantic M, Scherer K (2011) The first facial expression recognition and analysis challenge. In: *9th IEEE international conference on face and gesture recognition*, Santa Barbara, CA, USA
- Weise T, Bouaziz S, Li H, Pauly M (2011) Realtime performance-based facial animation. *ACM Trans Graph* 30(4):9

38. Xu Y et al (2011) A two-phase test sample sparse representation method for use with face recognition. *IEEE Trans Circuits Syst Video Technol* 21(9):1255–1262
39. Zhi R et al (2011) Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition. *IEEE Trans Syst Man Cybern B* 41(1):38–52