Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection

Muhammad Hameed Siddiqi, Rahman Ali, Adil Mehmood Khan, Eun Soo Kim, Gerard Junghyun Kim & Sungyoung Lee

ONLINE FIRST

Multimedia SYSTEMS

NUMBER 4

Volume 20 · July · 2014

🦌 Springer

🦌 Springer

REGULAR PAPER

Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection

**Muhammad Hameed Siddiqi · Rahman Ali ·
Adil Mehmood Khan · Eun Soo Kim ·
Gerard Junghyun Kim · Sungyoung Lee**

Abstract Knowledge about people's emotions can serve as an important context for automatic service delivery in context-aware systems. Hence, human facial expression recognition (FER) has emerged as an important research area over the last two decades. To accurately recognize expressions, FER systems require automatic face detection followed by the extraction of robust features from important facial parts. Furthermore, the process should be less susceptible to the presence of noise, such as different lighting conditions and variations in facial characteristics of subjects. Accordingly, this work implements a robust FER system, capable of providing high recognition accuracy even in the presence of aforementioned variations. The system uses an unsupervised technique based on active contour model for automatic face detection and extraction. In this model, a combination of two energy functions: Chan–Vese energy and Bhattacharyya distance functions are employed to minimize the dissimilarities within a face and maximize the distance between the face and the background. Next, noise reduction is achieved by means of wavelet decomposition, followed by the extraction of facial movement features using optical flow. These features reflect facial muscle movements which signify static, dynamic, geometric, and appearance characteristics of facial expressions. Post-feature extraction, feature selection, is performed using Stepwise Linear Discriminant Analysis, which is more robust in contrast to previously employed feature selection methods for FER. Finally, expressions are recognized using trained HMM(s). To show the robustness of the proposed system, unlike most of the previous works, which were evaluated using a single dataset, performance of the proposed system is assessed in a large-scale experimentation using five publicly available different datasets. The weighted average recognition rate across these datasets indicates the success of employing the proposed system for FER.

M. H. Siddiqi · R. Ali · S. Lee (✉)
Department of Computer Engineering, Kyung Hee University,
Suwon, Republic of Korea
e-mail: sylee@oslab.khu.ac.kr

A. M. Khan
Division of Information and Computer Engineering,
Ajou University, Suwon, Republic of Korea

E. S. Kim
Department of Electronic Engineering, Kwangwoon University,
Seoul, Republic of Korea

G. J. Kim
Department of Software Technology and Enterprize,
Korea University, Seoul, Republic of Korea

1 Introduction

Automatic facial expression recognition (FER) has emerged as an active area of research in the last two decades. It is because not only facial expressions are vital to social communications between human, they also play a significant role in many computing applications such as human computer interaction [3], robot control, and driver state surveillance [47].

An FER system consists of four basic modules: pre-processing, feature extraction, feature selection, and

recognition. It is the face that contains most of the expression-related information; therefore, in preprocessing module face is detected and extracted from a given image. Some common methods such as skin tone-based methods [40], and geometric-based methods [17] are used for face detection.

Feature extraction module deals with extracting the distinguishable features from each facial expression shape and quantizing it as a discrete symbol. Well-known methods like independent component analysis (ICA) [43], local feature analysis [26], local binary pattern [39], and local direction pattern [18] are commonly used for feature extraction. Next, feature selection is used to select a subset of relevant features from a large number of features extracted from the input data. The selected features are expected to contain better discriminatory power to help distinguish among different classes.

Lastly, in the recognition step, a classifier like support vector machine (SVM) [24], or hidden Markov model (HMM) [43] is first trained with training data and then used to generate the appropriate labels for the facial expressions contained in the incoming video data.

Though, a lot of work has already been done for automatic FER, a robust FER system is yet to be developed: a system capable of providing high recognition accuracy not just for one dataset but across different datasets, having expressions captured under different lighting conditions; using subjects of different gender, race and age. Developing such a robust FER system would require: (1) an accurate face detection and extraction method that can automatically detect and extract the faces from the expression frames; (2) a technique to get rid of noise prior to the feature extraction; (3) robust feature extraction and feature selection methods, such that the accuracy of these methods is not effected by the race and gender of subjects; (4) and finally, a good classifier to categorize the incoming expressions. Accordingly, in this paper, we have made the following contributions.

- An unsupervised automatic face detection model has been proposed, which is based on active contour (AC) model that automatically detects and extracts human faces from the expression frames. The proposed AC model is based on level set that is the combination of two energy functions: Chan–Vese (CV) [8] energy and Bhattacharyya distance [21] functions. The proposed AC model is robust because it not only minimizes the dissimilarities within the object (face) but it also maximizes the distance between the two regions such as face and the background.
- A robust feature extraction technique based on the facial movement features is proposed. The technique is based on symlet wavelet transform coupled with optical flow to get the facial movement features. The reason for using the wavelet transform is to diminish the noise before extracting the facial movement features.
- Even though the proposed feature extraction method extracts good features, there might still be some redundancy among these features. Therefore, a feature selection method called stepwise linear discriminant analysis (SWLDA) is applied to the selected feature space. SWLDA selects the most informative features taking the advantage of the forward selection model and removes irrelevant features by taking the advantage of the backward regression model. To the best of our knowledge, it is the first time that SWLDA is being utilized as a feature selection technique for FER systems.
- To show the robustness of the proposed system, large-scale experimentation is performed using multiple datasets.

The rest of the paper is organized as follows. Section 2 talks about related work and their limitations with regard to face detection and extraction, feature extraction, and feature selection. Section 3 presents an overview of the proposed approaches. Experimental setup for the proposed approaches is described in Sect. 4. Experimental results of the proposed methods are discussed in Sect. 5. Finally, the conclusion of the paper with future directions are provided in Sect. 6.

2 Literature review

2.1 Face detection and extraction

Essentially, face detection is the first step for a typical FER system, with the purpose of localizing and extracting the face region from the background. Some factors like illumination, pose, occlusion, and size of the image make it difficult for FER systems to accurately detect faces from expression video [29].

A variety of methods have been proposed in literature for face detection, including appearance-based and feature-based methods [45], geometric-based methods [17], knowledge-based methods [23], skin tone-based methods [40], and template-based methods [20]. However, each of these categories has its own limitations.

The performance of appearance-based methods is excellent in static environment; however, their performance degrades with the environmental change [36]. Therefore, feature-based methods were proposed to overcome the limitations of appearance-based approaches. These methods are more robust in illumination, pose and size of the image than the appearance-based methods.

However, a prior knowledge is required for these methods, i.e., at the time of implementation for these techniques, it is compulsory to decide randomly which intensity information will be important [7]. Due to these characteristics, these methods are known as heuristic techniques. Moreover, these methods have trouble in automatic feature detection [19].

In geometric-based approaches, parts of a face like mouth, eyes, and nose are positioned with their attributes and their reciprocal relationships. Face is recognized by calculating the distance, angles and areas between these parts of the face. These approaches are quite sufficient for small databases with stable lighting condition and stable viewpoint [11]. However, their performance degrades with the variation in lighting conditions and viewpoint [11].

Similarly, knowledge-based approaches are the rule-based techniques that convert human knowledge of what constitutes a typical face to capture the relationships between the facial features [46]. However, it is very hard for these approaches to build an appropriate set of rules. If the rules are too general then there could be several false positives, or there could be false negatives if the rules are in too detail [10]. Moreover, these approaches are incapable of finding faces in complex images [10].

Template-based approach is a simple process that has been widely employed to locate the human face in the input image. However, this method is very sensitive to pixel misalignment in sub-image areas and depends on facial component detection [49].

2.2 Feature extraction

According to the face descriptors, there are two types of features: global features and local features. In global features, features are extracted from the whole face, whereas in local features, features are extracted from some parts of the face.

Methods used for global feature extraction, also called holistic methods, include nearest features line-based subspace analysis [32], and independent component analysis (ICA) [43]. However, all these holistic methods do not know what exact facial features are the most important for FER. Moreover, performance of these approaches degrades in FER systems with the variation of illumination, pose, facial expression, occlusion and aging [36].

On the other hand, local feature-based methods compute local descriptors from different parts of a face and then integrate this information into one descriptor. These methods include local feature analysis (LFA) [26], Gabor features [15], non-negative matrix factorization (NMF) and local non-negative matrix Factorization (LNMF) [6] and local binary pattern (LBP) [48]. Among these methods, LBP achieved better performance. However, LBP does not

provide directional information of the facial frame. One recent study [18] employed local directional pattern (LDP) to solve the limitations of LBP. However, the performance of their method still degrades in non-monotonic illumination change, noise variation, change in pose, and expression conditions [36].

2.3 Feature selection

Feature selection module is used for selecting subset of relevant features from a large number of features extracted from the input data. The selected features are expected to to help distinguish one class from the others with a better accuracy. Feature selection module also helps reducing the dimensions of the feature space by selecting only the distinguishable features.

Please see Fig. 1 to appreciate the use of a feature selection technique in an FER system. Figure 1 shows 3D feature plots for six expression classes. It can be seen that the feature values for the six classes are highly merged, which can result in a high misclassification rate. It is because the use of inappropriate coefficients results in high within-class differences and low between-class differences. Therefore, a method is required to address the aforementioned problem and to reduce the dimension space and to increase the class separability. This idea is employed by various feature selection methods, mainly principal component analysis (PCA) [5], linear discriminant analysis (LDA) [30], kernel discriminant analysis (KDA) [31], and generalized discriminant analysis (GDA) [4].

PCA has poor discriminating power [12]. LDA-based methods suffer from limitations that their optimality criteria are not directly associated to the classification capability of the achieved feature representation [27]. Similarly, KDA does not have the capability to provide better
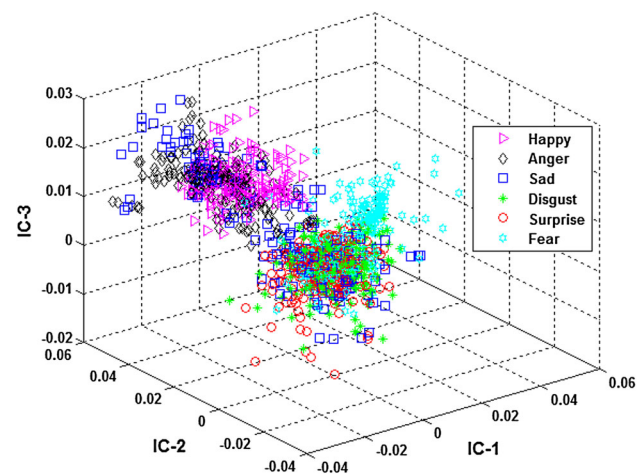

Fig. 1 3D-feature plot for six types of facial expressions

performance in the case if the face images of the same subjects are scattered rather than dispersed as clusters [33]. Likewise, the solution of GDA might not be stable and perhaps is not optimal in terms of the discriminant ability if there is small sample-sized training data [50].

3 Materials and methods

The overall architecture of the proposed FER system is shown in Fig. 2, whereas the description of different parts of the system is given in the following subsections.

3.1 Proposed face detection approach

The AC model is a well-known technique in the field of image segmentation. It is a deformable spline influenced by constraint and image forces that pull it towards object contours. It tries to move into a position where its energy is minimized. The AC model tries to improve by imposing desirable properties such as continuity and smoothness to the contour of the object, which means that the AC approach adds a certain degree of prior knowledge for dealing with the problem of finding the object contour.

Recently, Chan–Vese proposed in [8] a novel form of AC for object segmentation based on level set framework. Its energy function is defined by.

$$F(C) = \int_{\text{inside}(C)} |I(x) - c_{\text{in}}|^2 \mathrm{d}x + \int_{\text{outside}(C)} |I(x) - c_{\text{out}}|^2 \mathrm{d}x \tag{1}$$

where c_{in} and c_{out} are, respectively, the average intensities inside and outside of the curve C. Compared to the other AC models, the CV AC model can detect the faces more
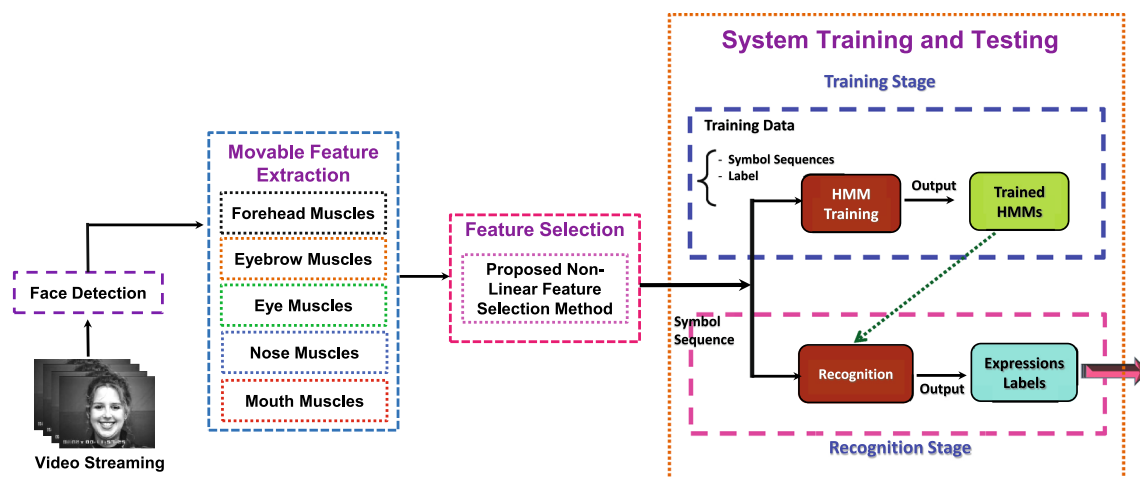
exactly since it does not need to smooth the initial facial image (via the edge function $g|\nabla I_\sigma|^2$), even if it is very noisy. In other words, this model is more robust to noise. The CV AC model does not use the edge information but utilizes the difference between the regions inside and outside of the curve, making itself one of the most robust and thus widely used techniques for image segmentation, especially, in the area of face detection. However, the convergence of CV AC model depends on the homogeneity of the segmented faces. When the inhomogeneity becomes large, the CV AC model provides unsatisfactory results. Moreover, the global minimum of the above energy function does not always guarantee the desirable results. The unsatisfactory result of the CV AC in this case is due to the fact that it tries to minimize the dissimilarity within each segment but does not take into account the distance between different segments.

The proposed methodology in this work is to incorporate an evolving term based on the Bhattacharyya distance to the CV energy function that minimizes the dissimilarities within the object (face) and maximizes the distance between the two regions (face and background). The proposed energy function is:

$$E(C) = \beta F(C) + (1 - \beta)B(C) \tag{2}$$

where $\beta \in [0, 1]$. Note that the value of $B(C)$ is always within the interval $[0, 1]$ whereas $F(C)$ is calculated based on the integral over the facial image plane.

The intuition behind the proposed model (in $E(C)$) is that we seek for a curve which is regular (the first two terms) and partitions the facial image into regions such that the differences within each region are minimized (the $F(C)$ term) like reducing environmental effects and the distance between the two regions (i.e., face and background) is maximized (the term $B(C)$ term).

Fig. 2 Architectural diagram for the proposed FER system

For the level set formulation, let us define ϕ as the level set function, $I : \Omega \to Z \subset R^n$ as a certain image feature such as intensity, color, texture, or a combination thereof, and $H(\bullet)$ and $\delta_0(\bullet)$ as the Heaviside and the Dirac function, respectively. The energy function can then be rewritten as

$$E(\phi) = \gamma \int_{\Omega} |\nabla H(\phi(x))| \mathrm{d}x + \eta \int_{\Omega} H(-\phi(x))$$
$$+ \beta \left[\begin{array}{c} \int_{\Omega} |I(x) - c_{\mathrm{in}}|^2 H(-\phi(x)) \\ + \int_{\Omega} |I(x) - c_{\mathrm{out}}|^2 H(\phi(x)) \end{array} \right] \qquad (3)$$
$$+ (1 - \beta) \int_{Z} \sqrt{p_{\mathrm{in}}(z) p_{\mathrm{out}}(z)} \mathrm{d}z$$

where

$$p_{\mathrm{in}}(z) = \frac{\int_{\Omega} \delta_0(z - I(x)) H(-\phi(x)) \mathrm{d}x}{\int_{\Omega} H(-\phi(x)) \mathrm{d}x}$$
$$p_{\mathrm{out}}(z) = \frac{\int_{\Omega} \delta_0(z - I(x)) H(\phi(x)) \mathrm{d}x}{\int_{\Omega} H(\phi(x)) \mathrm{d}x} \qquad (4)$$

where γ and η are non-negative constants, and $p_{\mathrm{in}}(z)$ and $p_{\mathrm{out}}(z)$ are given in (4) are the local fitting functions [16] that depend on the level set function ϕ and need to be updated in each contour evaluation. Differentiating w.r.t $\phi(x)$, then, the evaluation flow associated with minimizing the energy function in (3) is given as

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E}{\partial \phi}$$
$$= \delta_0(\phi) \left\{ \begin{array}{c} \gamma k + \eta + \beta[(I - c_{\mathrm{in}})^2 + (I - c_{\mathrm{out}})^2] \\ \\ -(1-\beta) \left[\begin{array}{c} \frac{B}{2} \left(\frac{1}{A_{\mathrm{in}}} - \frac{1}{A_{\mathrm{out}}} \right) \\ \delta_0(z-1) \\ + \frac{1}{2} \int_z \left(\frac{1}{A_{\mathrm{out}}} \sqrt{\frac{p_{\mathrm{in}}}{p_{\mathrm{out}}}} \\ - \frac{1}{A_{\mathrm{in}}} \sqrt{\frac{p_{\mathrm{out}}}{p_{\mathrm{in}}}} \right) \mathrm{d}z \end{array} \right] \end{array} \right\}$$
$$(5)$$

where A_{in} and A_{out} are, respectively, the areas inside and outside the curve C. Thus, the proposed AC model overcame the limitation of conventional CV AC model in the area of face detection.

3.2 Feature extraction

Noise reduction via wavelet transform In real-life scenarios, some environmental parameters (such as lighting effects) may produce some noise in the expression frames that could reduce the recognition rate. The proposed

method employs symlet wavelet to reduce such noise. Facial frames are converted to gray scale prior to applying this step.

The wavelet decomposition could be interpreted as signal decomposition into a set of independent feature vector. Each vector consists of sub-vectors like

$$V_0^{2D} = V_0^{2D-1}, V_0^{2D-2}, V_0^{2D-3}, \ldots, V_0^{2D-n} \qquad (6)$$

where V represents the 2D feature vector. If we have an expression frame X in the decomposition process, and it breaks up into the orthogonal sub images corresponding to different visualization. The following equation shows one level of decomposition.

$$X = A_1 + D_1 \qquad (7)$$

where X indicates the decomposed image and A_1 and D_1 are called approximation and detail coefficient vectors, respectively. If a facial frame is decomposed up to multiple levels, then Eq. 7 can be written as

$$X = A_j + D_j + D_{j-1} + D_{j-2} + \cdots + D_2 + D_1 \qquad (8)$$

where j represents the level of decomposition. The detail coefficients mostly consist of noise, so for feature extraction only the approximation coefficients are used. In the proposed algorithm, each facial frame is decomposed up to two levels, i.e., the value of $j = 2$, because by exceeding the value of $j = 2$, the facial frame looses significant information, due to which the informative coefficients cannot be detected properly, which may cause misclassification. The detail coefficients further consist of three sub-coefficients, so Eq. 8 can be written as

$$X = A_2 + D_2 + D_1$$
$$= A_2 + [(D_{\mathrm{h}})_2 + (D_{\mathrm{v}})_2 + (D_{\mathrm{d}})_2] + [(D_{\mathrm{h}})_1 \qquad (9)$$
$$+ (D_{\mathrm{v}})_1 + (D_{\mathrm{d}})_1]$$

where D_{h}, D_{v} and D_{d} are known as horizontal, vertical and diagonal coefficients, respectively. Note that at each decomposition step, approximation and detail coefficient vectors are obtained by passing the signal through a low-pass filter and high-pass filter, respectively.

In a specified time window and frequency bandwidth wavelet transform, the frequency is estimated. The signal (i.e., facial frame) is analyzed using the wavelet transform [42].

$$C(a_i, b_j) = \frac{1}{\sqrt{a_i}} \int_{-\infty}^{\infty} y(t) \Psi_{f.e}^* \left(\frac{t - b_j}{a_i} \right) \mathrm{d}t \qquad (10)$$

where a_i is the scale of the wavelet between lower and upper frequency bounds to get high decision for frequency estimation, and b_j is the position of the wavelet from the start to the end of the time window with the specified signal

sampling period, t is the time, the wavelet function $\Psi_{f.e}$ is used for frequency estimation, and $C(a_i, b_i)$ are the wavelet coefficients with the specified scale and position parameters. Finally, the scale is converted to the mode frequency, f_m for each facial frame:

$$f_m = \frac{f_a(\Psi_{f.e})}{a_m(\Psi_{f.e}).\Delta} \tag{11}$$

where $f_a(\Psi_{f.e})$ is the average frequency of the wavelet function, and Δ is the signal sampling period. Next, the facial movement feature have been extracted by exploiting the optical flow [2].

Feature extraction via optical flow The motion information of the facial pixels is found by employing the optical flow to generate the feature vector for each expression frame. To find the optical flow of the two expression frames, first, a kernel is made for partial derivative of Gaussian (like g_x and g_y) with respect to x and y such as

$$g_x(i,j) = -\frac{j-k-1}{2\pi\Sigma^3}\exp\left(-\frac{(i-k-1)^2 + (j-k-1)^2}{2\Sigma^2}\right) \tag{12}$$

$$g_y(i,j) = -\frac{j-k-1}{2\pi\Sigma^3}\exp\left(-\frac{(i-k-1)^2 + (j-k-1)^2}{2\Sigma^2}\right) \tag{13}$$

where $k = \frac{N-1}{2}$ and N is the size of the kernel, x and y derivatives are computed for both frames. After that the images are smoothed by building a Gaussian kernel such as

$$\text{kernel}(i,j) = -\frac{1}{2\pi\sigma^2}\exp\left(-\frac{(i-k-1)^2 + (j-k-1)^2}{2\sigma^2}\right) \tag{14}$$

$$A = \begin{bmatrix} \Sigma I_x, I_x & \Sigma I_x, I_y \\ \Sigma I_x, I_y & \Sigma I_y, I_y \end{bmatrix} \quad B = \begin{bmatrix} \Sigma I_x, I_t \\ \Sigma I_y, I_t \end{bmatrix} \tag{15}$$

The resultant image is given as

$$R = A^{-1}(-B) \tag{16}$$

where R is the resultant image that has the pixels' motion information. For instance, such pixel motion information for the two consecutive expression frames are shown in Fig. 3. The average feature vector is obtained by taking the average of the whole pixels' motion information for all the facial frames in a video clip which is given below:

$$f_{\text{ave}} = \frac{R_1 + R_2 + R_3 + \cdots + R_K}{N} \tag{17}$$

where f_{ave} indicates the average feature vector of all the expressions frames that a single expression video have, R_1 R_2 R_3 $\cdots$ R_K are the motion information for each expression frame, K is the last frame of the expression video, and N represents the whole number of frames in each expression video.

3.3 Stepwise linear discriminant analysis (SWLDA)

In the this step, the most informative features are selected using SWLDA, which maximizes the ratio of between-class variance to within-class variance in any particular data set, thereby guaranteeing maximal separability. SWLDA is based on two processes: forward and backward regression. These forward and backward selection techniques enable SWLDA to effectively reduce the dimensions of the feature space.

In the forward step, the most correlated features are selected based on partial F test values from the feature space. On the other hand, in the backward step, the least significant values are removed from the regression model, that is, lower F test values. In both processes, the F test values are calculated on the basis of defined class labels. The advantage of this method is that it is very efficient in seeking the localized features.

SWLDA works as follows: In the beginning, there is no predictor in the model. Based on the significance test, i.e., partial F test (the t-test), predictor is either entered or removed from the model in each iteration. Two predictors Alpha-to enter and Alpha-to remove are defined for significance level test. Alpha-to enter $a_e = 0.25$ and Alpha-to-remove $a_\gamma = 0.30$ are set as threshold parameters. These values are chosen based on various experiments. These threshold parameters show the significance level of the predictors which are entered or removed from the model, respectively. The algorithm stops when there are no more predictors to enter or remove from the stepwise model. For more details on SWLDA, please refer to a previous study [25].
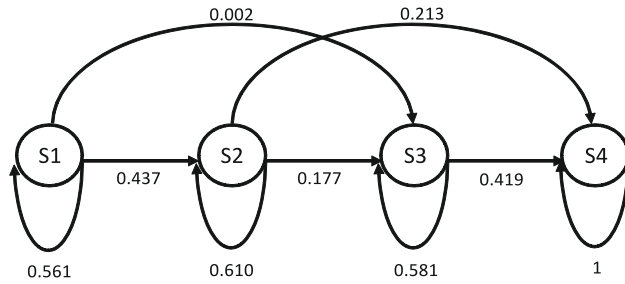
3.4 Recognition via hidden Markov model (HMM)

Commonly, the function of HMM is to provide a statistical model λ for a set of observation sequences. These observations are called "frames" in FER domains. Suppose there are sequence of observations of length T that are denoted by $O_1, O_2, \ldots, O_T$, and an HMM also consists of particular sequences of states, S, whose lengths range from 1 to N ($S = S_1, S_2, \ldots, S_N$), where N is the number of states in the model, and the time t for each state is denoted by $Q = q_1, q_2, \ldots, q_N$. The states are connected by arcs, as shown in Fig. 4, and each time

Fig. 3 Two consecutive expression frames and its corresponding optical flow (pixel movement information)

Fig. 4 After training, the arrangement of HMM and transition probabilities for happy facial expression

that a state j is entered, an observation is generated according to the multivariate Gaussian distribution $b_j(O_t)$ with the mean value μ_j and covariance matrix V_j correlated with that state. The arcs also have transition probabilities correlated with them such that probability a_{ij} is the resultant transition probability from state i to state j. The initial model probability for the state j is Π_j. An HMM can be defined by this set of parameters, such as $\lambda = A, B, \Pi$, where A indicates the probability of the state transition such that $A = a_{ij}$, $a_{ij} = \text{Prob}(q_{t+1} = S_j | q_t = S_i)$, $1 \leq i, j \leq N$, where B represents the probability of observations such that $B = b_j(O_t)$, $b_j = \text{Prob}(O_t | q_t = S_j) 1 \leq j \leq N$, and the initial state probability is indicated by Π such that $\Pi = \Pi_j$, $\Pi_j = \text{Prob}(q_1 = S_1)$. All the equations are based on the work by [34] and make use of the initial state probability distribution. In the training step, for a given model λ, the multiplication of each transition probability by each output probability at each step t provides the joint likelihood of a state sequence Q and the corresponding observation O that is calculated as:

$$P(O, Q|\lambda) = \pi_{q1} b_{q1}(O_1) \left[\prod_{t=2}^{T} a_{qt-1,qt} b_{qt}(O_t) \right] \quad (18)$$

During testing, the appropriate HMM can be determined by mean of likelihood estimation for the sequence observations O calculated based on the trained λ as

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \quad (19)$$

The maximum likelihood for the observations provided by the trained HMM indicates the recognized label. For more details on HMM, please refer to [37].

4 System validation

4.1 Datasets used

As mentioned before, the motivation behind this work is to build an accurate and robust FER system whose accuracy is not affected by noise and by the race and gender of subjects in a given image. Therefore, for an effective validation, the proposed system was tested against five different datasets: Cohn–Kanade [22], JAFFE [28], natural visible and infrared facial expression (USTC-NVIE) dataset [44], Yale [13], and FEI face dataset [41].

Some of the subjects in USTC-NVIE, Yale B and FEI datasets have worn glasses, whereas the subjects in Cohn–Kanade and JAFFE datasets are free of glasses. Similarly, most of expressions in JAFFE are captured from subjects with their hair tied. This exposes all the sensitive regions of the face from where we can easily extract the facial movement features. In other datasets, subjects have hair on their foreheads and eyebrows, which could reduce the recognition rate. Likewise, the facial features, such as the eyes of the subjects in the JAFFE dataset, are totally different from the eyes of the subjects of the other datasets [38]. Moreover, in each dataset, expressions were captured under different lighting conditions, which could also cause misclassification. Furthermore, in some datasets like in Cohn–Kanade, some subjects have a darker skin color. Also, some expressions in USTC-NVIE and FEI datasets are taken from different angles, which could also degrade the accuracy of FER systems.

To summarize, it is for these different parameters and characteristics that the proposed FER system is tested and

validated on five publicly available standard datasets. Detailed description on each of these datasets is as follows:

- *Cohn–Kanade Dataset*
  In this facial expressions' dataset 100 subjects (university students) performed basic six expressions. The age range of the subjects were from 18 to 30 years and most of them were female. We employed those expressions for which the camera was fixed in front of the subjects. By the given instructions, the subjects performed a series of 23 facial displays. Six expressions were based on descriptions of prototypic emotions such as happy, anger, sad, surprise, disgust, and fear. To utilize these 6 expressions from this dataset, we employed total 450 image sequences from 100 subjects, and each of them was considered as one of the 6 basic universal expressions. The original size of each facial frame was 640×480 or 640×490 pixel with 8-bit precision for grayscale values. For recognition purpose, twelve expression frames were taken from each expression sequence, which resulted in total 5,400 expression images.

- *Japanese female facial expressions (JAFFE) Dataset*
  The expressions in this dataset were posed by 10 different subjects (Japanese female). Each image has been rated on 6 expressions' adjectives by 60 Japanese subjects. Most of the expression frames were taken from the frontal view of the camera with tied hair to expose all the sensitive regions of the face. In the whole dataset, there were total 213 facial frames, which consisted of seven expressions including neutral. Therefore, we selected only 195 expression frames for 6 facial expressions performed by ten different Japanese female as subjects. The original size of each facial frame was 256×256 pixel.

- *USTC-NVIE dataset*
  In this dataset, an infrared thermal and a visible camera was used to collect the two types of expressions; such as the spontaneous and the posed-based expressions, but we utilized only posed-based expressions in which some subjects worn glasses and some were free of glasses. For posed-based expressions, the subjects were asked to perform a series of expressions with illumination from three different directions. There were 108 subjects (university students) performed posed-based expressions, and the age range was from 17 to 31 years. The size of each facial frame was 640×480 or 704×490 pixels. The total 1,027 expressions frames were utilized from this datasets.

- *Yale B face dataset*
  In this dataset, there are a total 5,760 facial frames taken in single light source performed by 10 distinct subjects each seen under 576 viewing conditions

(9 poses $\times$ 64 illumination conditions). For every subject, when capturing a particular pose, the ambient illumination was also captured.

- *FEI Face Dataset*
  It is a Brazilian face database in which 200 subjects (university students and staff) performed six expressions. Each expression sequence contains 14 frames. The age range of the subjects was from 19 to 40 years old with distinct appearance, hairstyle, and adorns. In this dataset, the number of male and female subjects are exactly the same and equal to 100. All the expressions are colorful and were taken against a white homogenous background in an upright frontal position with profile rotation of up to about 180°. Scale might vary about 10 and the original size of each image is 640×480 pixels. For recognition purpose, all the expression frames were taken from each expression sequence, which resulted in total 2,800 expression images.

4.2 Experimental setup

For the experiments, six basic universal facial expressions were selected: happy, anger, sad, surprise, disgust, and fear. The size of each input image (expression frame) was 60×60 pixels based on ground truth positions of the eye and mouth. The images were converted to a zero-mean vector of size $1 \times 3,600$ for feature extraction. All the experiments were performed in Matlab using an Intel® Pentium® Dual-Core™ (2.5 GHz) with a RAM capacity of 3 GB. For a thorough validation, seven experiments were performed as follows.

- In the first experiment, performance of the proposed face detection and extraction technique was analyzed.
- In the second experiment, performance of the proposed FER system (including all the components: face detection, feature extraction, feature selection, and classification) was validated using the five datasets. For each dataset, a 10-fold cross-validation scheme (based on subjects) was used. In other words, out of 10 subjects data from a single subject was used as the validation data, whereas data for the remaining 9 subjects were used as the training data. This process was repeated 10 times with data from each subject used exactly once as the validation data.
- Third experiment was similar to the second experiment, with one major difference: cross-validation scheme was applied based on datasets. In other words, from the five datasets four were used as validation datasets, whereas one dataset was used as the training data. This process was repeated five times, with data from each dataset used exactly once as the training data.

- In the fourth experiment, performance of the proposed feature extraction method was compared against existing feature extraction methods using all the five datasets.
- As for the fifth experiment, performance of the SWLDA method was compared against other well-known feature selection methods, again using all the five datasets.
- Finally, in the sixth experiment, the performance of proposed FER system was compared against previous state-of-the-art FER systems.

5 Experimental results

5.1 First experiment

It should be noted that in the proposed FER system, active contour evolution in a certain frame is performed independently of the other frames. It means that the face detection is performed on frame-by-frame bases. In any given frame, the only information utilized from the previous frame is the final contour obtained in the previous frame. This information is used to determine the initial position of the active contour in the current frame. First, an ellipse with major axis along y-axis of length 20 and minor axis along x-axis of length 20 is selected as the initial contour. In this experiment, the initial shape was the same for all frames, and only the center location varied. In each video sequence, the first frame is segmented using manual initialization such that the initial contour is closer to the face.

Then from the second frame, the position of the initial contour's center in the current frame is the mean value of the points along the final contour in the previous frame. For example, suppose that along the final contour of frame $n(n \geq 1)$, there are M points $(x_i^{(n)}, y_i^{(n)}), i = 1 \cdots M$. Then, the center $(c_x^{(n+1)}, c_y^{(n+1)})$ of the initial contour in the frame $(n + 1)$ is calculated as

$$(c_x^{(n+1)}) = \frac{1}{M} \sum_{i=1}^{M} x_i^{(n)}; \quad (c_y^{(n+1)}) = \frac{1}{M} \sum_{i=1}^{M} y_i^{(n)}$$

Now that we have explained how the contours are calculated, please see the results of the first experiment in Fig. 5, which compares the performance of the AC model with that of the CV AC model. It is obvious that the employed AC model showed better performance than the CV AC model (as shown in Fig. 5b).

5.2 Second experiment

As mentioned earlier, the purpose of this experiment was to analyze the performance of the proposed FER system for all five datasets. The recognition results for this experiment are shown in Table 1–Fig. 6 (using Cohn–Kanade dataset), Table 2–Fig. 7 (using JAFFE dataset), Table 3–Fig. 8 (using USTC-NVIE dataset), Table 4–Fig. 9 (using Yale B face dataset), and Table 5–Fig. 10 (using FEI dataset).It is clear from Tables 1, 2, 3, 4, and 5 that the proposed FER consistently achieved a high recognition rate when applied to these datasets separately: 99.33 % for the Cohn–Kanade dataset, 96.50 % for JAFFE dataset, 99.17 % for USTC-NVIE dataset, 99.33 % for Yale B face dataset, and 99.50 % for FEI face dataset. Please note that the system achieved comparatively lower recognition rates for JAFFE dataset. This is because most of the expressions in JAFFE dataset are quite similar with other expressions [36], which can result in similar pixel motion information. Nevertheless, the results are quite satisfactory for all the datasets.
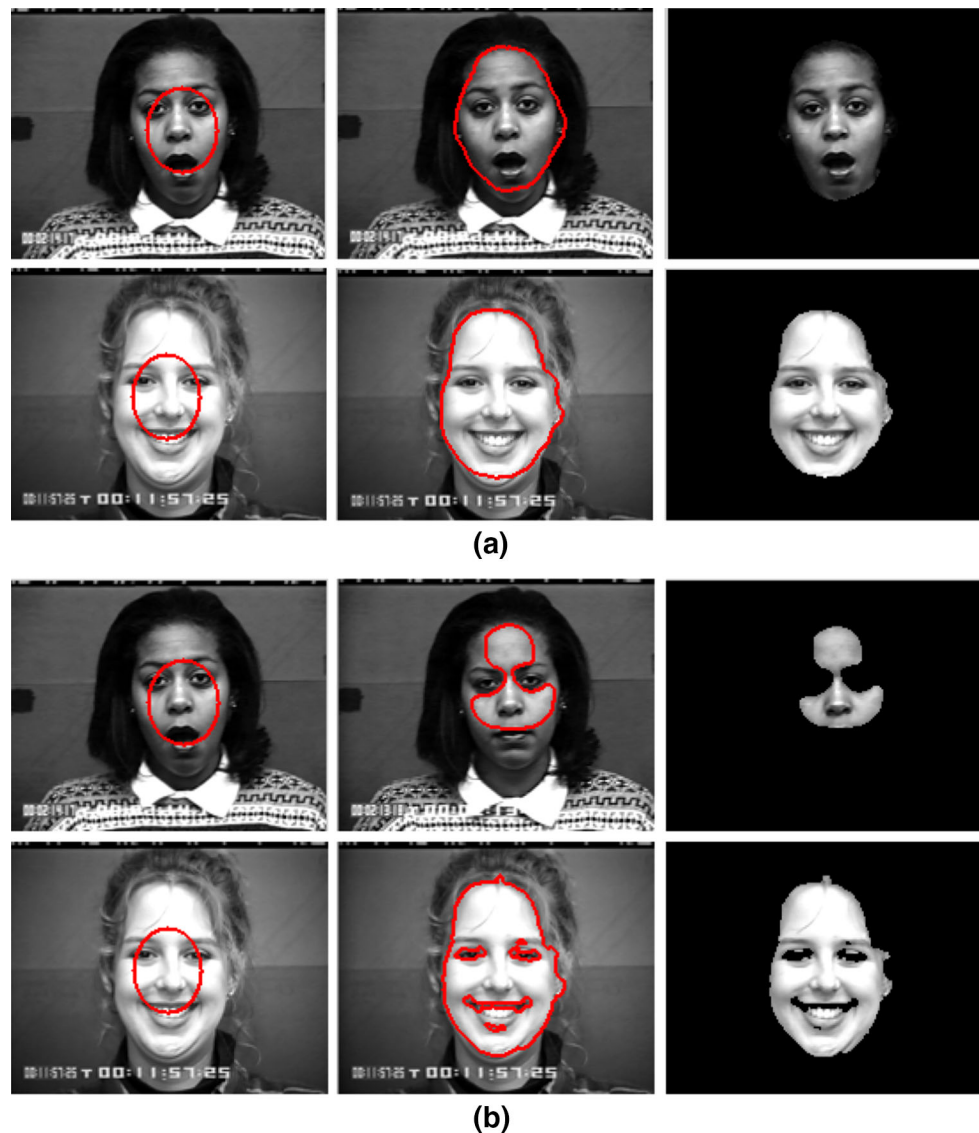
5.3 Third experiment

As mentioned earlier, in this experiment, the FER system was validated using a cross-validation scheme based on datasets. The overall results are shown in Fig. 11.

It is clear from Fig. 11(a), (c), and (d) that the system achieved high recognition rates when it was trained on the Cohn–Kanade, Yale B, and USTC-NVIE datasets. However, its accuracy of classification was low when the system was trained on the JAFFE and FEI datasets (shown in Fig. 11(b) and (e)). This may be because these datasets have different facial features; for instance, some of the subjects in the Yale B face dataset have worn glasses, while subjects in the Cohn–Kanade and JAFFE datasets did not wear glasses. Furthermore, eye features in the JAFFE dataset are very different from those in the Cohn–Kanade and Yale B face datasets. Similarly, some expressions in FEI datasets were taken using different angles of the camera, while other datasets have the expressions only from frontal view of the camera. Nevertheless, the results are very encouraging and this suggests that the proposed FER system is robust, i.e., the system not only achieved a high recognition rate on one dataset, but also provided good recognition rates when used across multiple datasets.

5.4 Fourth experiment

In the fourth experiment, performance of the feature extraction module of the proposed FER system was compared against four existing feature extraction methods. This experiment was performed on all the five datasets. For each dataset, the experiment was repeated five times. Each time a 10-fold cross-validation scheme (based on subjects) was used, and feature extraction module was changed, whereas all the other parts of the system remained the same. The overall results are shown in Fig. 12.

Fig. 5 **a** Sample results of the proposed AC model for which $\gamma = 0.5$ and $\beta = 0.2$, while **b** indicates the sample results of CV AC model using $\gamma = 0.5$ and $\beta = 1.0$. *Left image* shows the initial contour, *middle image* represents the final contour, and *last image* presents the extracted face

(a)

(b)

It can be seen from Fig. 12 that the proposed feature extraction method (symlet wavelet transform with optical flow) outperformed the existing well-known feature extraction methods in all cases. This is because, the proposed idea of noise reduction using symlet wavelet transform relies on the operation of the wavelet coefficients using a filter that takes into account the local regularity of the coefficients in the transform domain. Similarly, the estimation of the threshold is not required for this method. The initial probabilities have been assigned to show how noisy the coefficients are [9]. Also, the symlet wavelet has the capability to support the characteristics of orthogonal, biorthogonal, and reverse biorthogonal of grayscale images, that is why it provides better enhanced results. Moreover, we measure the statistic dependency of wavelet coefficients for all the facial frames of gray scale. Joint probability of a grayscale frame is computed by collecting

geometrically aligned frames of the expression for each wavelet coefficient.

Furthermore, the motion of the pixels in important parts of the face could help in recognizing expressions. Therefore, once, the noise has been diminished from the expression frames, then the optical flow has been employed to extract the frequency-based features from the expression frames. High recognition results show that optical flow is capable of extracting prominent features from the enhanced expression frames with the aid of locality in frequency, orientation and in space as well.
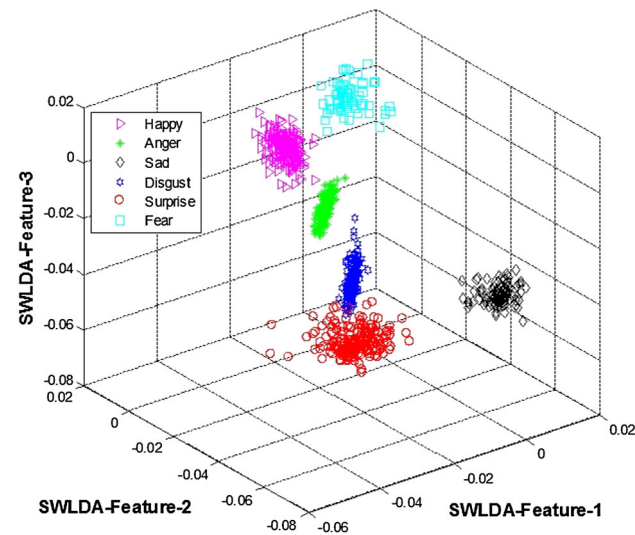
5.5 Fifth experiment

Similarly, in the fifth experiment, performance of SWLDA was compared against four well-known feature selection methods. The overall settings for this experiment were

Table 1 Confusion matrix of the proposed approaches using Cohn–Kanade dataset of facial expressions (unit %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	99	0	0	1	0	0
Sad	0	100	0	0	0	0
Anger	0	0	99	0	0	1
Disgust	0	0	0	100	0	0
Surprise	0	0	0	0	100	0
Fear	0	1	0	1	0	98
Average	99.33					

Table 2 Confusion matrix of the proposed approaches using JAFFE dataset of facial expressions (unit %)

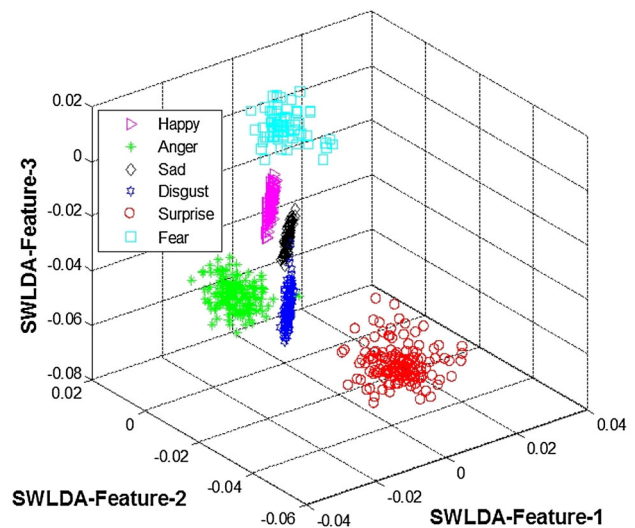| Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	96	1	0	0	1	2
Sad	1	97	0	2	0	0
Anger	1	2	95	0	0	2
Disgust	0	1	0	97	2	0
Surprise	0	2	0	0	98	0
Fear	0	1	1	2	0	96
Average	96.50					

Fig. 6 3D-feature plot of the proposed approaches for six facial expressions. It is indicated that the proposed approaches provided best classification rate on Cohn–Kanade dataset
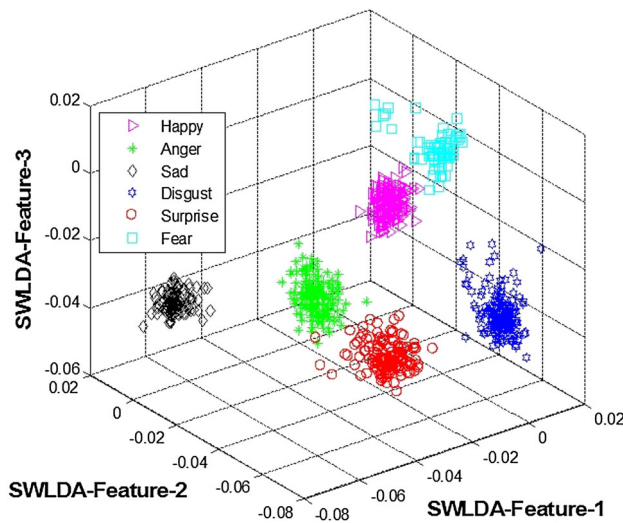

Fig. 7 3D-feature plot of the proposed approach for six facial expressions. It is indicated that the proposed approaches provided better classification rate on JAFFE dataset

similar to those of the fourth experiment with one major difference: this time the feature selection module was changed for each case whereas all the other parts of the system remained the same. The overall results are shown in Fig. 13.

It can be seen from Fig. 13 that SWLDA outperformed the other feature selection methods in all cases. This is because SWLDA is a robust feature selection technique that addresses the limitations of previous such techniques, especially PCA, lDA, KDA, and GDA. As shown earlier in Fig. 1, due to the similarity among the expressions that results in high within-class variance and low between-class variance, the features for the six classes are highly merged. This can result later in a high misclassification rate. Thanks to its forward and backward regression models, SWLDA solves this problem by not only providing dimension reduction, but also by increasing the low between-class variance to increase the class separation before the features are fed to a classifier.

Table 3 Confusion matrix of the proposed approaches using USTC-NVIE dataset of facial expressions (unit %)

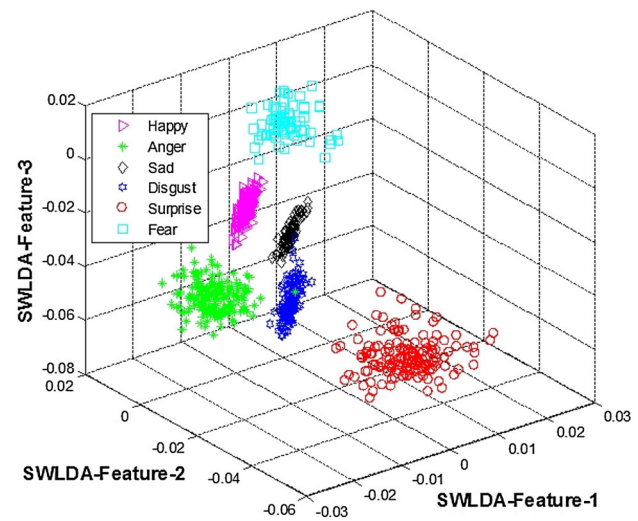| Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	100	0	0	0	0	0
Sad	1	98	0	0	1	0
Anger	0	0	100	0	0	0
Disgust	0	0	0	98	0	2
Surprise	0	1	0	0	99	0
Fear	0	0	0	0	0	100
Average	99.17					

5.6 Sixth experiment

Finally, in the last experiment, the performance of the proposed approaches was compared with some state-of-the-art methods [1, 14, 18, 35, 36] on all the datasets, i.e., Cohn–Kanade, JAFFE, USTC-NVIE, Yale B, and FEI

Fig. 8 3D-feature plot of the proposed approaches for six facial expressions. It is indicated that the proposed approaches provided best classification rate on USTC-NVIE dataset

Fig. 9 3D-feature plot of the proposed approaches for six facial expressions. It is indicated that the proposed approaches provided best classification rate on Yale B face dataset

Table 4 Confusion matrix of the proposed approaches using Yale B face dataset of facial expressions (unit %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	99	0	1	0	0	0
Sad	0	100	0	0	0	0
Anger	0	0	100	0	0	0
Disgust	0	1	0	98	1	0
Surprise	0	0	0	0	100	0
Fear	0	0	0	1	0	99
Average	99.33					

Table 5 Confusion matrix of the proposed approaches using FEI dataset of facial expressions (unit %)
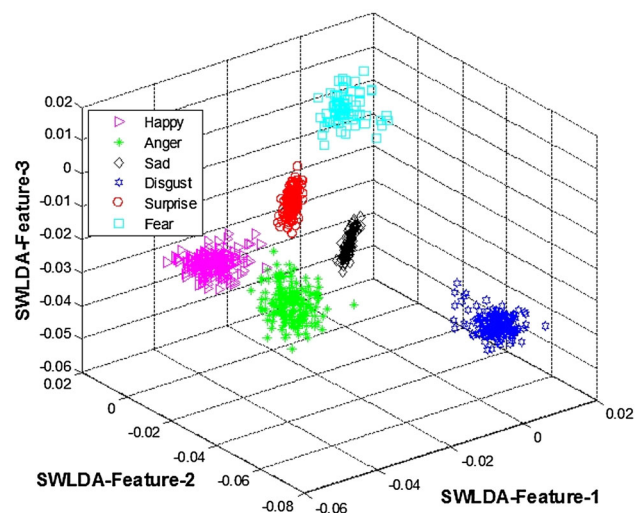
Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	100	0	0	0	0	0
Sad	0	100	0	0	0	0
Anger	1	1	98	0	0	0
Disgust	0	0	0	99	0	1
Surprise	0	0	0	0	100	0
Fear	0	0	0	0	0	100
Average	99.50					

datasets of facial expressions. All these methods were implemented by us using the instructions provided in their respective papers. For each dataset, 10-fold cross-validation scheme (based on subjects) was applied as described in Sect. 4. The average recognition rate for each method along with the proposed FER system are presented in Fig. 14.

It can be seen from Fig. 14 that the proposed FER system outperformed the existing state-of-the-art methods. Thus, the proposed system shows significant potential in its ability to accurately and robustly recognize human facial expressions using video data.

6 Conclusion

Automatic facial expression recognition (FER) has received a lot of attention from the research community in

Fig. 10 3D-feature plot of the proposed approaches for six facial expressions. It is indicated that the proposed approaches provided best classification rate on FEI dataset
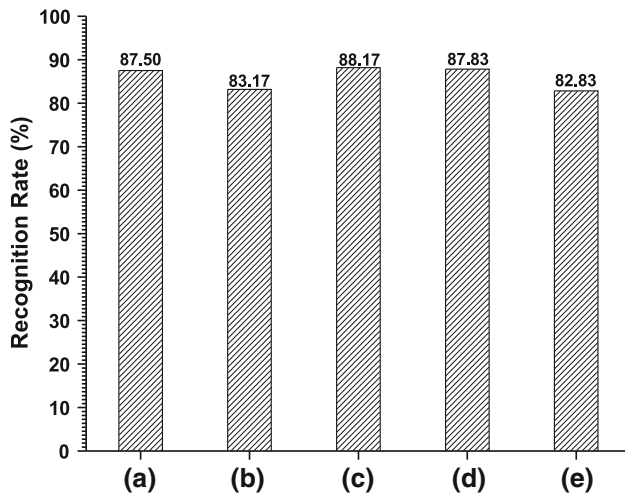
the past two decades. Several FER systems have been proposed; however, automatic face detection and recognizing human facial expressions accurately are still major concerns for such systems. Several factors can reduce the accuracy of an FER system. Two such factors are the lack of robust features and high similarity among different facial expressions that can result in low between-class variance in the feature space.

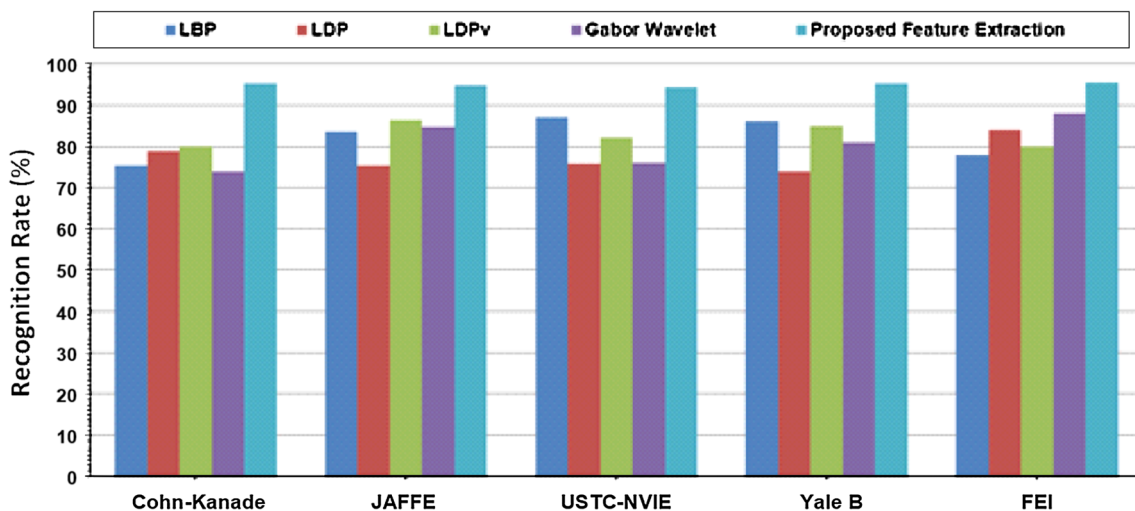This research proposes an accurate and robust FER system, capable of providing a high recognition accuracy

even when images are captured under different lighting conditions and subjects' facial features vary. In the proposed system, the recognition starts with automatic detection of human faces. For this an active contour model is employed, which is based on level set that is a combination of two energy functions: (1) Chan–Vese energy function, helps minimizing the dissimilarities within an object (in our case it is the human face), and (2) Bhattacharyya distance function, which helps maximizing the distance between the face and the background. Our experimental results show that this scheme can correctly detect human face from the expression videos.

Facial features are very sensitive to noise and illumination. Furthermore, due to high similarity among different facial expressions, features from different classes can merge with each other in the feature space, making it very hard to distinguish among different facial expressions. These problems are solved using three strategies in the feature extraction module: Firstly, noise is reduced by applying wavelet decomposition to the facial frames. Secondly, pixel movement is captured using optical flow. Lastly, SWLDA is employed as a feature selection method to select only the most relevant features. SWLDA not only increases the class separation, it also helps reducing the number of dimensions of the feature space. When compared against existing feature extraction and selection methods, the employed schemes provided much better results.
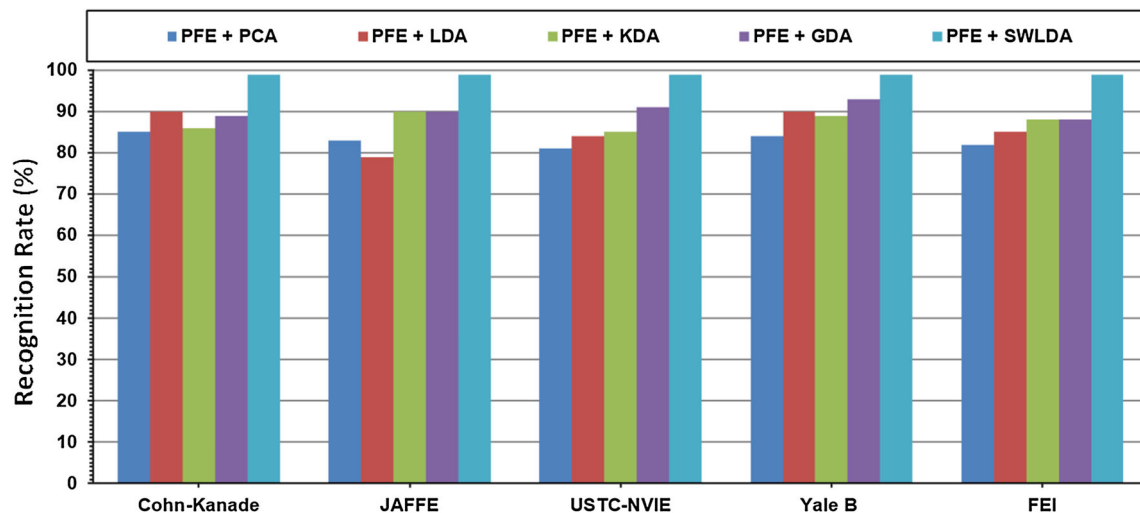
It should be noted that though it is shown that the proposed face detection and extraction method can accurately extract human faces from the expression frames, it requires manual placement of the initial contour near the human face in the initial frame. The placement of the contour in the subsequent frames is done automatically. Thus the proposed face detection and extraction method might not
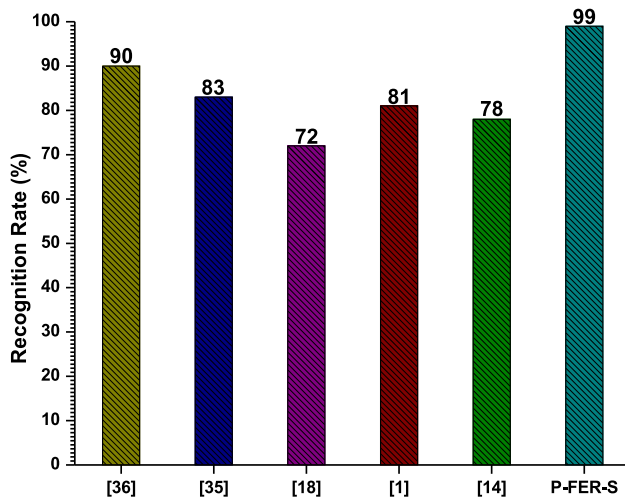

Fig. 11 Recognition rate of the proposed approaches (*a*) training on Cohn–Kanade dataset and testing on JAFFE, USTC-NVIE, Yale B, and FEI datasets, (*b*) training on JAFFE dataset and testing on Cohn–Kanade, USTC-NVIE, Yale B, and FEI datasets, (*c*) training on USTC-NVIE dataset and testing on Cohn–Kanade, JAFFE, Yale B, and FEI datasets, (*d*) training on Yale B face dataset and testing on Cohn–Kanade, JAFFE, USTC-NVIE, and FEI datasets, and (*e*) training on FEI face face dataset and testing on Cohn–Kanade, JAFFE, USTC-NVIE and Yale B face datasets (unit %)

Fig. 12 Comparison of the proposed feature extraction method against some recent existing feature extraction methods using all the five datasets (unit %)

Fig. 13 Comparison of SWLDA method against some well-known existing feature selection methods coupled with the proposed feature extraction (PFE) method using all the five datasets (unit %)

Fig. 14 Comparison results of the proposed FER system (P-FER-S) with some existing state-of-the-art methods (unit %)

work if the initial contour is far from the face. This is one of the limitations of this work. Similarly, if the expressions in the datasets posses very high similarity (such as in JAFFE dataset), then the performance of the proposed FER system decreases. Furthermore, it should also be noted that the experiments were performed in laboratory. The performance of the system is not yet investigated in real-time for outside laboratory settings. There exist several factors in real-life environment, such as background clutter, image rotation and, blur, that can decrease the performance of the proposed FER system. Therefore, further study is needed to tackle these issues and maintain the same high recognition rate in real-life situations.

References

1. Ahsan, T., Jabid, T., Chong, U.P., et al.: Facial expression recognition using local transitional pattern on Gabor filtered facial images. IETE Tech. Rev. **30**(1), 47–52 (2013)
2. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. Int. J. Comput. Vis. **12**(1), 43–77 (1994)
3. Bartlett, M.S., Littlewort, G., Fasel, I., Movellan, J.R.: Real time face detection and facial expression recognition: development and applications to human computer interaction. In: Conference on Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03, vol. 5, pp. 53–53. IEEE (2003)
4. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Comput. **12**(10), 2385–2404 (2000)
5. Boutsidis, C., Mahoney, M.W., Drineas, P.: Unsupervised feature selection for principal components analysis. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 61–69. ACM (2008)
6. Buciu, I., Pitas, I.: Application of non-negative and local non negative matrix factorization to facial expression recognition. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, vol. 1, pp. 288–291. IEEE (2004)
7. Cendrillon, R., Lovell, B.C.: Real-time face recognition using eigenfaces. In: Proceedings-SPIE the International Society for Optical Engineering, number 1, pp. 269–276. International Society for Optical Engineering (1999, 2000)
8. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
9. Chandra, D.V.S.: Image enhancement and noise reduction using wavelet transform. In: Proceedings of the 40th Midwest Symposium on Circuits and Systems, 1997, vol. 2, pp. 989–992. IEEE (1997)
10. de Carrera, P.F.: Face recognition algorithms. PhD thesis, Universidad del País Vasco (2010)

11. Duc, N.M., Minh, B.Q.: Your face is not your password face authentication bypassing Lenovo-Asus-Toshiba. Black Hat Briefings (2009)

12. Feng, G.-C., Yuen, P.C., Dai, D.-Q.: Human face recognition using PCA on wavelet subband. J. Electron. Imaging **9**(2), 226–233 (2000)

13. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 643–660 (2001)

14. Ghimire, D., Lee, J.: Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors **13**(6), 7714–7734 (2013)

15. Gu, W., Xiang, C., Venkatesh, Y.V., Huang, D., Lin, H.: Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recognit. **45**(1), 80–91 (2012)

16. He, L., Wee, W.G., Zheng, S., Wang, L.: A level set model without initial contour. In: 2009 Workshop on Applications of Computer Vision (WACV), pp. 1–6. IEEE (2009)

17. Hjelmås, E., Low, B.K.: Face detection: a survey. Comput. Vis. Image Underst. **83**(3), 236–274 (2001)

18. Jabid, T., Kabir, Md.H., Chae, O.: Robust facial expression recognition based on local directional pattern. ETRI J. **32**(5), 784–794 (2010)

19. Jafri, R., Arabnia, H.R.: A survey of face recognition techniques. J. Inf. Process. Syst. **5**(2), 41–68 (2009)

20. Jin, Z., Lou, Z., Yang, J., Sun, Q.: Face detection using template matching and skin-color information. Neurocomputing **70**(4), 794–800 (2007)

21. Kailath, T.: The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. Technol. **15**(1), 52–60 (1967)

22. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 46–53. IEEE (2000)

23. Kotropoulos, C., Pitas, I.: Rule-based face detection in frontal views. In: 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997. ICASSP-97, vol. 4, pp. 2537–2540. IEEE (1997)

24. Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans. Image Process. **16**(1), 172–187 (2007)

25. Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: Toward enhanced p300 speller performance. J. Neurosci. Methods **167**(1), 15–21 (2008)

26. Li, S.Z., Hou, X.W., Zhang, H.J., Cheng, Q.S.: Learning spatially localized, parts-based representation. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001, vol. 1, pp. I207–I212. IEEE (2001)

27. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using LDA-based algorithms. IEEE Trans. Neural Netw. **14**(1), 195–200 (2003)

28. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 200–205. IEEE (1998)

29. Mahmood, M.T.: Face detection by image discriminating. M.Sc. thesis (unpublished), Blekinge Institute of Technology (2006)

30. Mika, S.: Kernel fisher discriminants. PhD thesis, Universitäts-bibliothek (2002)

31. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, pp. 41–48. IEEE (1999)

32. Pang, Y., Yuan, Y., Li, X.: Iterative subspace analysis based on feature line distance. IEEE Trans. Image Process. **18**(4), 903–907 (2009)

33. Park, S.W., Savvides, M.: A multifactor extension of linear discriminant analysis for face recognition under varying pose and illumination. EURASIP J. Adv. Signal Process. **2010**, 6 (2010)

34. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989)

35. Rahulamathavan, Y., Phan, R.C.-W., Chambers, J.A., Parish, D.J.: Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. IEEE Trans. Affect. Comput. **4**(1), 83–92 (2013)

36. Rivera, A.R., Castillo, J.R., Chae, O.: Local directional number pattern for face analysis: face and expression recognition. IEEE Trans. Image Process. **22**(5), 1740–1752 (2013)

37. Samaria, F.S.: Face recognition using hidden Markov models. PhD thesis, University of Cambridge (1994)

38. Shamir, L.: Evaluation of face datasets as tools for assessing the performance of face recognition methods. Int. J. Comput. Vis. **79**(3), 225–230 (2008)

39. 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)

40. Singh, S.K., Chauhan, D.S., Vatsa, M., Singh, R.: A robust skin color based face detection algorithm. Tamkang J. Sci. Eng. **6**(4), 227–234 (2003)

41. Thomaz, C.E., Giraldi, G.A.: A new ranking method for principal components analysis and its application to face image analysis. Image Vis. Comput. **28**(6), 902–913 (2010)

42. Turunen, J.: A wavelet-based method for estimating damping in power systems. PhD thesis, Aalto University (2011)

43. Uddin, Md.Z., Lee, J.J., Kim, T.-S.: An enhanced independent component-based human facial expression recognition from video. IEEE Trans. Consum. Electron. **55**(4), 2216–2224 (2009)

44. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., Wang, X.: A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Trans. Multimed. **12**(7), 682–691 (2010)

45. Yang, J., Zhang, D., Frangi, A.F., Yang, J.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **26**(1), 131–137 (2004)

46. Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)

47. Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. IEEE Trans. Affect. Comput. **2**(4), 219–229 (2011)

48. Zhang, S., Zhao, X., Lei, B.: Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans. Signal Process. **8**(1), 21–31 (2012)

49. Zhang, X., Gao, Y.: Face recognition across pose: a review. Pattern Recognit. **42**(11), 2876–2896 (2009)

50. Zheng, W., Zhao, L., Zou, C.: A modified algorithm for generalized discriminant analysis. Neural Comput. **16**(6), 1283–1297 (2004)