Soft Comput (2018) 22:4389–4405
https://doi.org/10.1007/s00500-017-2634-3

METHODOLOGIES AND APPLICATION

Facial expression recognition using a combination of multiple facial features and support vector machine

Hung-Hsu Tsai[1] · **Yi-Cheng Chang**[1]

Published online: 19 May 2017

Abstract This paper presents a novel facial expression recognition (FER) technique based on support vector machine (SVM) for the FER. Here it is called the FERS technique. First, the FERS technique develops a face detection method that combines the Haar-like features method with the self-quotient image (SQI) filter. As a result, the FERS technique possesses better detection rate because the face detection method gets more accurate in locating face regions of an image. The main reason is that the SQI filter can overcome the insufficient light and shade light. Subsequently, three schemes, the angular radial transform (ART), the discrete cosine transform (DCT) and the Gabor filter (GF), are simultaneously employed in the design of the feature extraction for facial expression in the FERS technique. More specifically, they are employed in constructing a set of training patterns for the training of an SVM. The FERS technique then exploits the trained SVM to recognize the facial expression for a query face image. Finally, experimental results show that the recognition performance of the FERS technique can be better than that of other existing methods under consideration in the paper.

Keywords Angular radial transform · Facial expression recognition · Emotional expression recognition · Gabor filter · Self-quotient image filter · Support vector machine

Communicated by V. Loia.

✉ Hung-Hsu Tsai
thh@nfu.edu.tw

[1] Department of Information Management, National Formosa University, Huwei 632, Yulin, Taiwan, ROC

1 Introduction

Human facial expressions can be naturally used for the identification task of emotional tendencies, health qualities, and social interaction. Nowadays, the FER technologies are widely employed in the research fields such as human–computer interface, computer games, e-Learning, driver fatigue detection, and other emerging applications (Lee et al. 2008; Loh et al. 2006; Saeed et al. 2007).

Generally, facial expressions can be categorized into seven basic emotions, happy, sad, surprise, fear, disgust, anger, and neutral (Kim et al. 2008). Each has a corresponding specific facial expression. Consequently, it becomes a popular topic to develop computer algorithms to recognize facial expression. The FER technique commonly consists of three main components: face detection, facial expression feature extraction, and facial expression categorization. First, many approaches such as principal component analysis (PCA), artificial neural network (ANN), ART, and Haar features (HFs) have been applied in collecting a set of representation features for human face detection (Fang and Qiu 2003; Shih and Liu 2006). Some methods using the PCA to compute a linear projection that preserves the maximum variance of input data so as to reduce the data dimension for several research areas such as edge detection, face pattern, and face localization (Yang et al. 2002). Rowley et al. (1998) proposed a face detection method that devises a multiple neural network-based learning method that examines the existence of faces at each window. It detects face patterns and then makes the final decision from multiple detection results. A feature extraction method using ART is one of region-based shape descriptors. It derives a compact representation of face patterns (Fang and Qiu 2003). However, the method requires highly computational complexity. The scheme of Viola employs a boosted cascade of simple HFs rectangle features for the con-

struction of a fast face detection system (Viola and Jones 2001). The HFs method possesses some prominent characteristics, which performs multiple detection and robustness for size variation and orientation. The HFs rectangle features offer a satisfying performance in differentiating between face and non-face cases. However, solving multiple classification problems is one of principal limitations they suffer from (Kim et al. 2008).

Second, the feature-extraction methods of the FER techniques for facial expressions can be divided into two categories, image-based and feature-based methods. The former extracts features from face regions of an image. For example, Fellenz et al. (1999) proposed a feature-extraction method by employing Gabor wavelets in representing a whole face image by extracting GFs for the face image. In the method of Lajevardi and Hussain (2012), GFs, log GFs, local binary pattern (LBP) operators, and higher-order local autocorrelation methods are exploited to find out corresponding features for face images and then combines them to form a set of features. In the method of VenkataRamiReddy et al. (2014), local features and global features of each facial expression image are extracted via employing directional local binary patterns (DLBP) and DCT. The method then utilizes the PCA as a fusion scheme which fuses the local and global features of facial images. Subsequently, a radial basis function neural network is performed to be a classifier for classification of fusion features of facial expression images. Kharat and Dudul (2008) proposed a feature-extraction method via utilizing DCT, fast Fourier transform (FFT), and singular value decomposition (SVD) in the procedure of extracting entire facial features. Then, an SVM is trained by feeding the SVM with these features to construct an SVM-based classifier for human emotion recognition. The latter is to extract features components from a face image, such as nose, mouth, eye, and eyebrow. The work of Zhou et al. (2006) adopts selective feature extraction that selected feature components according to the results of rough classification. Chai et al. (2014) proposed a facial-feature-extraction method developed in the spatial image domain. It computes face image features via applying information entropy to spatial histograms of local patterns of face images. The work of Saeed et al. (2007) extracts the changes to eye shape for driver fatigue detection. The use of different features results in different contributions on the facial expression recognition. Meanwhile, a reasonable combination of the features mentioned above can improve the performance of FER systems. The work of Wei and Qingxuan (2016) utilizes facial key points of each image as feature points on emotion recognition based on facial expression. Here the classifier for emotion recognition is realized by an SVM with a weighted feature Gaussian kernel function. Al-Shabi et al. (2016) proposed a hybrid approach of combining multiple deep convolutional neural networks (CNNs) with scale invariant feature transform (SIFT) for feature extraction. The multiple deep CNNs extract CNN features from raw images. The SIFT relies on edges to locate the key points of facial images. The SIFT features of a facial image can be represented by a set of key-point descriptors. A CNN is performed to classify the combined feature by merging CNN and SIFT features. The CNN can be used to recognize facial image expression.

Third, facial expression categorization can be performed by a classifier with a decision procedure. The work of Loh et al. (2006) uses the back-propagation neural network to work for expression classification. The obtained results showed that the classification ability the scheme possesses is acceptable. However, it is examined by only using their own database instead of using current common databases. Kim et al. (2008) modified the HFs method by extending it to have more sub-categories in order to identify more precisely facial expressions. The work of Cheng et al. (2008) applies Burial Markov Model in the design of the facial expression recognition. The scheme inherits a shorting of Hidden Markov Model. Therefore, it cannot perform well due to the existence of some noises or the loss of features in facial expression images. The work of Kotsia et al. (2008) presents an FER method using an SVM for facial action unit classification from the shape information for the Cohn–Kanade database. The work of Sun et al. (2008) proposes a multi-class SVM to perform the feature classification. The SVM approach is robust to noisy data and adaptive to some misclassifications via training a classification model that can identify the outliers.

The previous methods merely adopt one kind of features such as texture or shape. Therefore, it motivates us to propose the FERS technique which can take the advantages of these two categories of features. That is, it combines texture features with shape features and then uses the combined features in the design of its feature extraction. Here, three kinds of features are utilized in the feature extraction of the FERS technique. First, a face image is processed by sequentially using an SQI filter and a Sobel filter to get a corresponding filtered image for the face image. The first kind of features is retrieved from the DCT version of the filtered image. The second kind of features is extracted from the ART version of the filtered image that is obtained by applying a Sobel filter to the original face image. The third kind of features is gained via computing the GF version of the face image. The first and the third features include image texture characteristics. The second features contain image shape characteristics. In order to promote the facial expression recognition rate, an assumption is that input images can be roughly categorized into eight classes, seven facial expressions and a non-face class. An image is classified to the non-face class if no face can be detected in the image. Finally, a trained SVM plays a role as a multi-classes classifier for these three kinds of features. The experimental results demonstrate that the recognition perfor-

mance of the FERS technique can be better than that of other existing methods under consideration in the paper.

The remainder of the paper is organized as follows. Section 2 reviews the SQI filter, HFs, ART, GF and SVM. Section 3 then describes the face detection, feature extraction and facial expression recognition algorithms used in the design of the FERS technique. Section 4 shows the experimental results. Finally, Sect. 5 draws conclusions.

2 Backgrounds

2.1 Self-quotient image

The SQI process is used for illumination normalization that removes the lighting effect by the variation of illumination, such as shadows and illumination direction changes (Kim et al. 2008). The SQI method, no need for alignment, provides an invariant representation of face images under different lighting conditions. The formal definitions for the SQI process can be specified in (1)–(4).

$$Q = \frac{I}{\hat{I}} = \frac{I}{g * I}, \tag{1}$$

$$g_i(x, y) = W(x, y) * G(x, y), \tag{2}$$

$$W(x, y) = \begin{cases} 0, & I(x, y) < \tau \\ 1, & \text{otherwise} \end{cases}, \tag{3}$$

$$\tau = \text{Mean}(I_\Omega) \tag{4}$$

where I and $\hat{I}$ represent the original image and its smoothed version, respectively. Moreover, g_i stands for the Gaussian kernel, W expresses the weight for the pixel (x, y), and G represents a Gaussian filter. Ω is the kernel size and τ is the mean of filtering region. The weight is set to 0 when the intensity is lower than τ. Otherwise, it is set to 1. Figure 1 shows the filtered images after performing the SQI process. The algorithm of the SQI process is described as follows.
Algorithm: the SQI process.

Step 1: Using (1)–(4) to produce three filtered images, $Q_{3\times3}$, $Q_{9\times9}$, and $Q_{15\times15}$, employing three different mask sizes.
Step 2: The final version of the filtered image $\hat{Q}$ can be computed by $\hat{Q} = \log(Q_{3\times3}) + \log(Q_{9\times9}) + \log(Q_{15\times15})$.
Step 3: Normalizing pixel values of $\hat{Q}$ between 0 and 255.

2.2 Haar-like features

A set of HFs can be regarded as the digital image feature that has been powerfully employed for face detection (Chen et al. 2008; Wong and Cho 2009; Yang et al. 2009). The HFs method encodes differences in average intensities between two or more rectangular regions. It can extract textures without depending on absolute intensities (Mita et al. 2005). Figure 2 displays four types of rectangular features for the enclosing detection window. The value of each rectangular feature is obtained by subtracting the sum of the pixels within white rectangles from the sum of the pixels within black rectangles. Here each feature performs a classifier. In order to enhance classification robustness, each weak classifier is cascaded into a strong classifier. The AdaBoost procedure selects features for gaining the best combination of classifiers. A weak classifier $h_j(x)$ in (5) contains a feature $f_j(x)$, a threshold θ_j, and the polarity term p_j indicating the direction of the inequality sign. Figure 2 illustrates an example for the calculation of HFs.

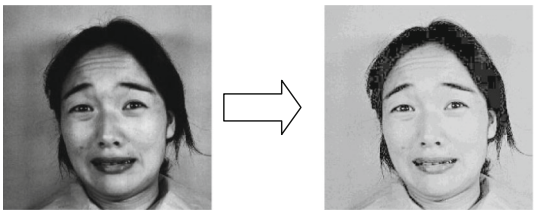

Fig. 1 Illumination normalization using the SQI process

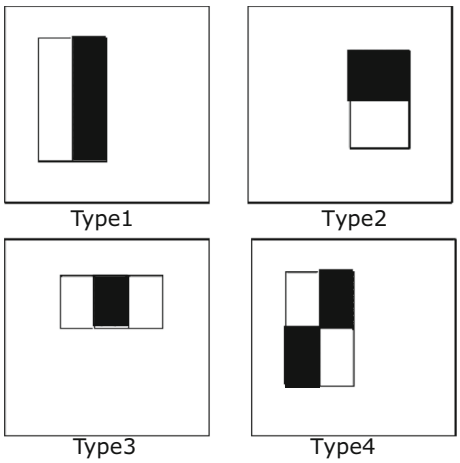

Fig. 2 Four types of rectangular features of HFs

$$h_j(x) = \begin{cases} 1, & \text{if } p_j f_j(x) < p_j \theta_j, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

2.3 Angular radial transform

ART is adopted in MPEG-7, and it is regarded as a regional based shape descriptor for the binary version of an images. The principle design for ART is to adopt a unit disk in the polar coordinate that is separable along the angular and the radial directions in a complex 2D transformation. The ART

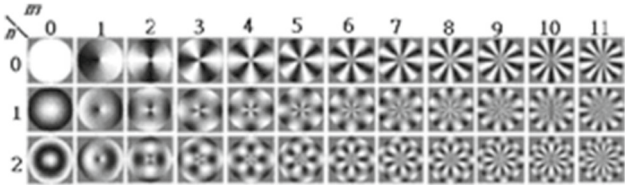

Fig. 3 ART basis function changes for 35 circumstances ($m = 12$ and $n = 3$)

features can be used to describe the distribution of pixels in the shape (Fang and Qiu 2003). The ART feature is defined as ARTM $[m][n]$ in (6) where m and n represent the feature changes to angle and radius, respectively. The process of normalizing the scales for ART features is to divide each ART coefficients by the magnitude of ARTM [0][0]. The region-based shape descriptor defined in MPEG-7 uses 12 angular and 3 radial functions (i.e., totally, 35 ART features where $m = 12$ and $n = 3$) to express the shape characteristics of images. A set of the ART features for the ith image is denoted by $F_{\mathrm{ART}}(i) = \{F_{\mathrm{ART}}(i, j) | j = 1, 2, \ldots, \mathrm{NF}_{\mathrm{ART}}\}$, where $\mathrm{NF}_{\mathrm{ART}} = 35$ represents the number of the ART features. In short, after the ART algorithm is applied to a binary image, 35 ART features can be produced for the image. Figure 3 draws the real part of some ART basis of change situations. Here the darker color indicates stronger basis, and, oppositely, light color indicates weaker basis.

$$F_{\mathrm{ART}}(i, j) = \mathrm{ARTM}\ [m]\,[n] = \frac{\mathrm{ARTM}\,[m]\,[n]}{\mathrm{ARTM}\ [0][0]} \tag{6}$$

2.4 Gabor filter

GF can effectively express the texture features (Tsai et al. 2014). It captures the salient visual properties and has greatly successful results in face recognition (Liu and Wang 2006). Figure 4 displays the GF kernels containing the real part and the imaginary part. GF kernels are similar to the receptive field profiles in cortical simple cells, which are characterized as localized, orientation selective, and frequency selective. An image is processed by the kernel element and, then, to produce its corresponding frequency images, which are further employed to compute to obtain Gabor features for the image. Figure 5a presents an original image. Figure 5b, c depicts the corresponding real-part and imaginary-part images of the Gabor version of the image in Fig. 5a, respectively.

Figure 6 exhibits a result image by combining the real part with the imaginary part calculated using (7) where $\mathrm{GF}(x, y)_{R,S}$, $\mathrm{real}(x, y)_{R,S}$, and $\mathrm{img}(x, y)_{R,S}$ denote the combined image, real part image, and imaginary part image, respectively. Moreover, R and S express orientation factor and scale factor for the GF kernels, respectively. In the experiment of the paper, each base element is with size 40×40 pixels, therefore, $n_x = n_y = 40$. The mean $\mu_{R,S}$ and standard deviation $\sigma_{R,S}$ for each base element are specified in (8) and (9), respectively. Consequently, the GF feature of the ith image in an image dataset can be expressed by $F_{\mathrm{GF}}(i) = \{F_{\mathrm{GF}}(i, j) | j = 1, 2, \ldots, \mathrm{NF}_{\mathrm{GF}}\}$ where $\mathrm{NF}_{\mathrm{GF}}$ represents the number of Gabor features. In the experiment of the paper, NF_{GF} is set to 80. The formula is given in (10). In the paper, n_R and n_S are set to 7 and 4, respectively.

$$\mathrm{GF}(x, y)_{R,S} = \sqrt{\mathrm{real}(x, y)_{R,S}{}^2 + \mathrm{img}(x, y)_{R,S}{}^2} \tag{7}$$

$$\mu_{R,S} = \frac{\sum_{x=1}^{n_x} \sum_{y=1}^{n_y} |\mathrm{GF}(x, y)_{R,S}|}{n_x \times n_y} \tag{8}$$

$$\sigma_{R,S} = \sqrt{\frac{\sum_{x=1}^{n_x} \sum_{y=1}^{n_y} (|\mathrm{GF}(x, y)_{R,S}| - \mu_{R,S})^2}{n_x \times n_y}} \tag{9}$$

$$F_{GF}(i, j) = \left\{(\mu^i_{R,S}, \sigma^i_{R,S}) | R = 0, 1, \ldots, n_R, \; S = 0, 1, \ldots, n_S\right\} \tag{10}$$

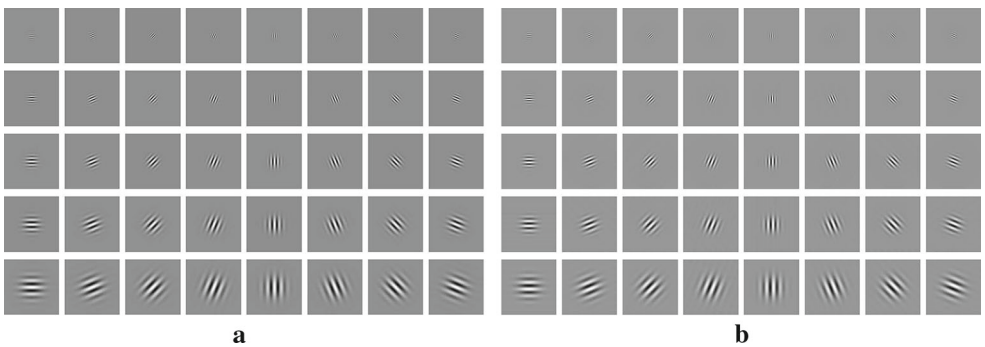

Fig. 4 Forty Gabor kernels are used to represent an image. (**a**) and (**b**) represent the real part and the imaginary part of the Gabor version of an image

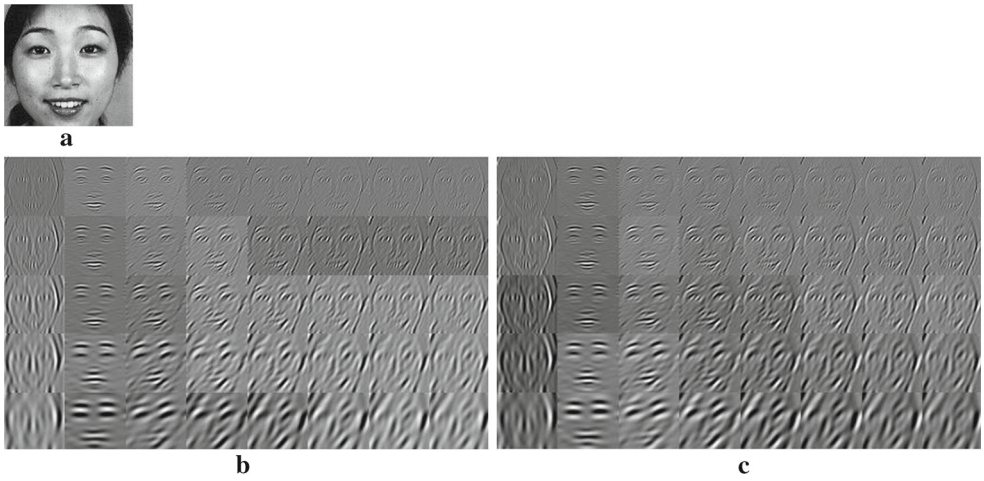

Fig. 5 A face image using GF transformation. **a** is an original image; **b**, **c** are the real part and the imaginary part of (**a**)

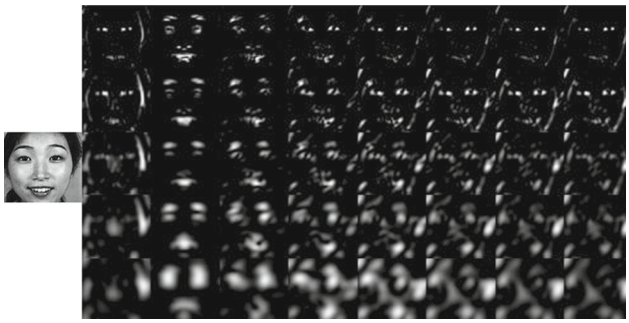

Fig. 6 A result image with Gabor version of Fig. 5a, which is produced via combining the real part with the imaginary part in Fig. 5b, c, respectively

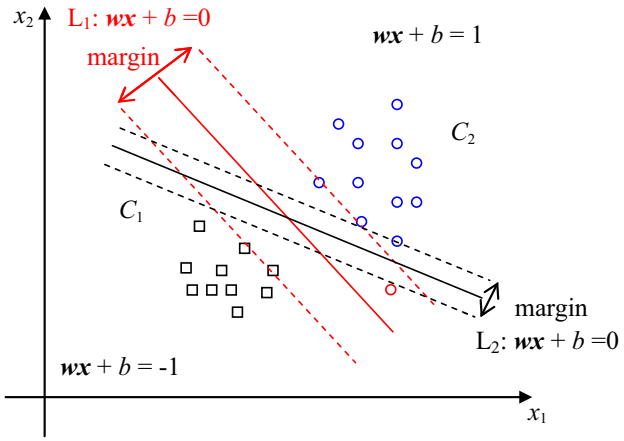

Fig. 7 The hyperplanes separate the two classes C_1 and C_2 with the margins

2.5 Support vector machine

An SVM maps the input space to a higher dimensional feature space and constructs a hyperplane to separate class members from non-members. It is a useful classification technique based on statistical learning theory (Kharat and Dudul 2008; Tsai et al. 2013). Figure 7 shows the hyperplanes separating the two classes C_1 and C_2 with the margins. Many hyperplanes can separate the data. However, only one of them can achieve a maximal margin in order to avoid the possible misclassifications while determining which type of the new data belongs to. Let $\boldsymbol{S} = \{(\boldsymbol{x}_i, d_i) \mid i = 1, 2, \ldots, N\}$ be a set of N training patterns, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \in R^n$ denotes an input vector in the input space and $d_i \in \{-1, 1\}$ represents the label class of $\boldsymbol{x}_i$. Assume $\boldsymbol{S}$ be linear separable. A linear SVM seeks an optimal separating hyperplane with maximum margin so that the linear SVM has good generalization. More specially, the SVM searches a pair $(\boldsymbol{w}, b)$ for an optimal hyperplane by maximizing the margin $\frac{2}{\|w\|}$ between $\boldsymbol{wx} + b = 1$ and $\boldsymbol{wx} + b = -1$ where the vector w points perpendicular to the separating hyperplane and $b \in R$.

Assume the pair $(\boldsymbol{w_o}, b_o)$ is an optimal solution for a corresponding separating hyperplane for $\boldsymbol{S}$. The linear SVM can be performed by the decision function f defined as

$$f(\boldsymbol{x}) = \mathrm{sgn}(\boldsymbol{w}_o \boldsymbol{x}^t + b_o) \tag{11}$$

where sgn(·) and t stand for the sign function and the matrix transpose operator.

While classifying a set of linearly non-separable data, several slack variables, ξ_i, are added into the derivation of a linear non-separable SVM which introduces the following optimization problem (Burges 1998).

$$\text{Minimize } \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i$$
$$\text{Subject to } d_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, N. \tag{12}$$

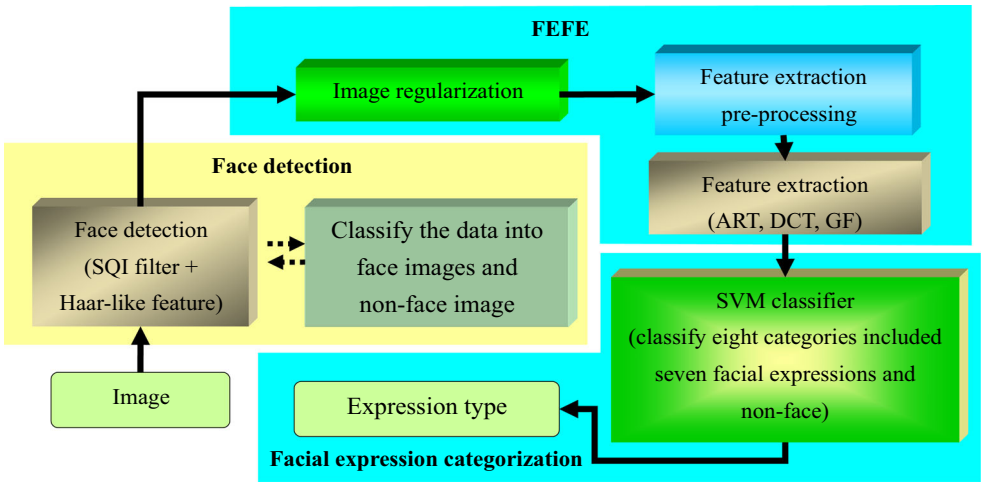

Fig. 8 The block diagram of the FERS technique

Here C denotes the penalty for vectors incorrectly classified or inside the margin, which also plays a role as the regularization parameter for the SVM classifier.

Let the pair $(\bar{w}, \bar{b})$ be computed using an optimization solution $\bar{\boldsymbol{\alpha}} = (\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_N)$. Then, the decision function f for a nonlinear SVM can be rewritten as

$$f(x) = \mathrm{sgn}(\bar{w}^t x + \bar{b}) = \mathrm{sgn}\left(\sum_{i=1}^{N} \bar{\alpha}_i d_i K(\boldsymbol{x}_i, \boldsymbol{x}) + \bar{b}\right) \quad (13)$$

where $K(\cdot, \cdot)$ is a kernel function that allows the SVM algorithm to fit the maximum-margin hyperplane in a transformed feature space. In the paper, the Gaussian radial basis function (GRBF) is taken to realize the kernel functions which can be specified by $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(\frac{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\gamma^2})$ where γ is taken as denotes the spread of the GRBF and $\|\cdot\|$ denotes L_2 norm for vector operations.

3 The FERS technique

Figure 8 depicts the block diagram of the FERS technique. A set of HFs is used for face detection, and the SQI filter is utilized to reduce the illumination affect. The Facial Expression Feature Extraction (FEFE) includes three processes: image regularization, preprocessing for feature extraction, and feature extraction to be exploited in getting ART, DCT, and GF features. In the paper, the facial emotions have eight classes: the seven classes of basic facial expressions (anger, disgust, fear, happiness, sadness, surprise and neutral) and a non-face class. Sequentially, an SVM classifier is trained using three sets of features. Then, the trained SVM can be utilized in the prediction of the corresponding expression class an input face image belongs to.

3.1 The proposed face detection scheme

Figure 9 exhibits the procedure of the proposed face detection method in the paper. A set of HFs is used in the face detection method, and OpenCV is also applied to implement the face detection (OpenCV, 2008). The face detection method in OpenCV is modified here for strengthening the human face detection for a wide range of light of an illumination environment. The procedure of the proposed face detection method is briefly described as follows. First, color images are converted into their grayscale version. A median filter is then used to remove noises generated by the video camera. Second, the high-pass filter and the SQI filter are applied in face detection so as to promote the performance of the next phase while suffering from a variety of different light and shade of the light environment. To accelerate the calculation of HFs, the length and width are divided by three for resizing images instead of being divided by 2 in OpenCV. Third, it adopts a Gaussian filter with size 5×5 to remove noises from images due to the SQI filter introducing more noises. Finally, the HFs are employed in face detection in accordance with the scaled proportion to identify the face portion of the original images.

Furthermore, Fig.10 displays a test result with and without using an SQI filter, respectively. The result shows that the FERS technique can be applied at night or in the extreme lighting condition while using the SQI filter. In the general situation, there are some miss images in the experiment. Therefore, it is necessary to add a non-face class in facial expression categories. In the classification process, the FERS technique can judges whether an image has face regions or not. Here the FERS technique is also investigated under the video camera environment with web camera. In an appropriate illumination environment, 18 volunteers join the experiment. Each has three images.

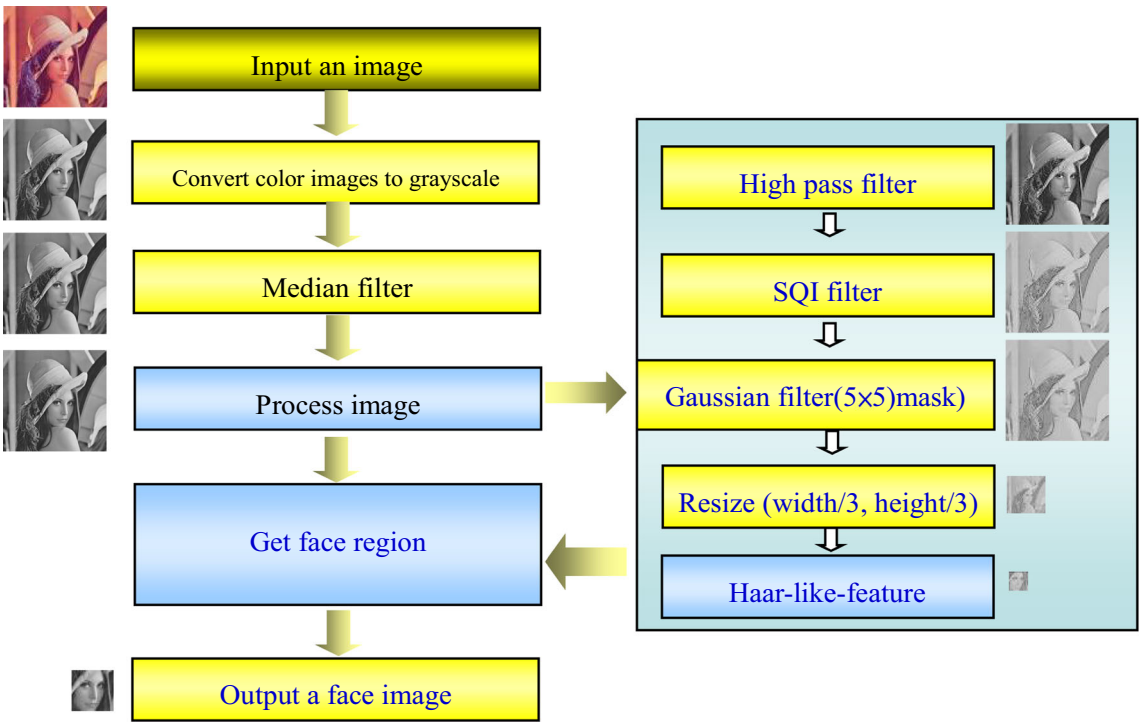

Fig. 9 The procedure of the proposed face detection method used in the FERS technique

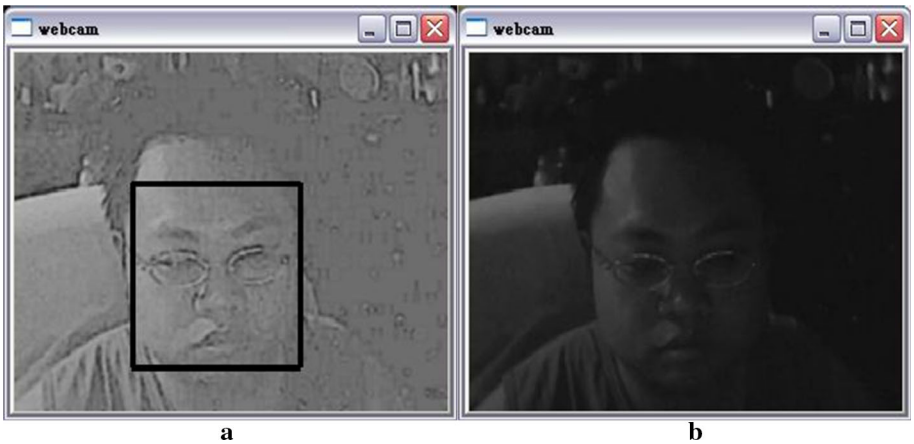

Fig. 10 The results of using the proposed face detection method with and without involving an SQI filter, which are displayed in (**a**) and (**b**), respectively

Totally, here 54 images are collected in the experiment. Figure 11 exhibits the results of using the proposed face detection method in the normal illumination situation for 18 volunteers.

3.2 Feature extraction

The aim of performing the component, image regularization, in Fig. 8 is to reserve face emotion regions of an image as largest as it has. The proposed method merely leaves the detected facial image but drops non-face portions of the image. Figure 12 displays a result of performing the image regularization. Before extracting DCT, GF and ART features, it needs a feature-extraction preprocessing in Fig. 8, which converts a color image to a binary version for ART procedure and to a gray-level version for DCT and GF schemes. Subsequently, a set of ART features is obtained by feeding a binary image into the ART procedure. Two features sets for DCT and GF versions are extracted by inputting a gray-level image into DCT and GF algorithms, respectively. Figure 13 draws the block diagram of the feature extraction for three kinds of features used in the FERS technique. The extracted features contain explicit information for various facial features such as the human face shape features, simple texture features, and complex texture features. According to each feature, different feature types are combined to form several new combined features. The paper investigates the best combination of them.

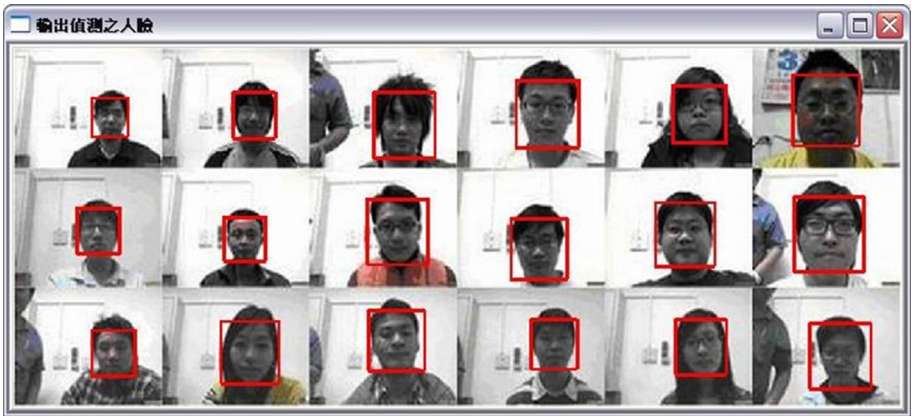

Fig. 11 The results of using the proposed face detection method in the normal illumination situation for 18 volunteers

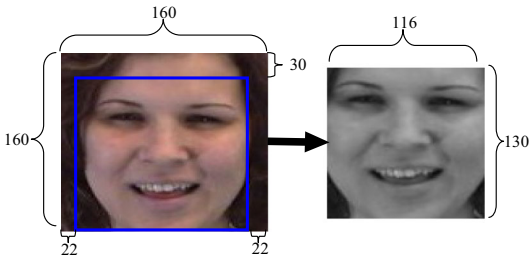

Fig. 12 A result of using the image regularization method in the FERS technique

Figure 14 displays the results obtained on the completion of feature extraction. Figure 14a, b shows an original image and the result of applying an SQI filter to the original image, respectively. Figure 14c depicts the result of applying a Sobel filter to the original image. Figure 14d represents the result of applying the Sobel filter to Fig. 14b. Figure 14e, f illustrates the binary versions of the filtered images as shown in Fig. 14c, d, respectively. Figure 14b–d depicts the images used for the feature extraction to get the DCT and GF features, which can represent the image texture information. Figure 14e–f exhibits the images used for the feature extraction to compute ART features that can retain the image shape information. The paper exploits several filters in enhancing image features, which can remove the unnecessary information such as using SQI filters for removing the interferences of light and using Sobel filters for enhancing image edges. As a result, the feature extraction of the FERS technique can sharpen facial expression and facial-line changes.

3.3 Facial expression categorization

On the complete of feature extraction, an SVM is utilized in the FERS technique for the facial expression classification. A set of the images are randomly partitioned into two exclusive sets including a set of training images and the other of testing images. The set of training images is employed to specify the parameters in the SVM for the facial expression classifier and the testing set is then used to verify the performance of the facial expression classifier. In the experiment of the paper, Libsvm is exploited to perform the SVM for multi-classes classifier (Chang and Lin 2011). Let TRAFERS denote the algorithm for training the FERS technique is briefly described as follows.

Algorithm TRAFERS

Step 1 : Input a set of training images.
Step 2 : Categorize these images. Let $I_{i,j,c}$ express the ith examined image, I_i, having the jth features and belonging to the cth class.
Step 3 : Perform face detection for each image I_i.

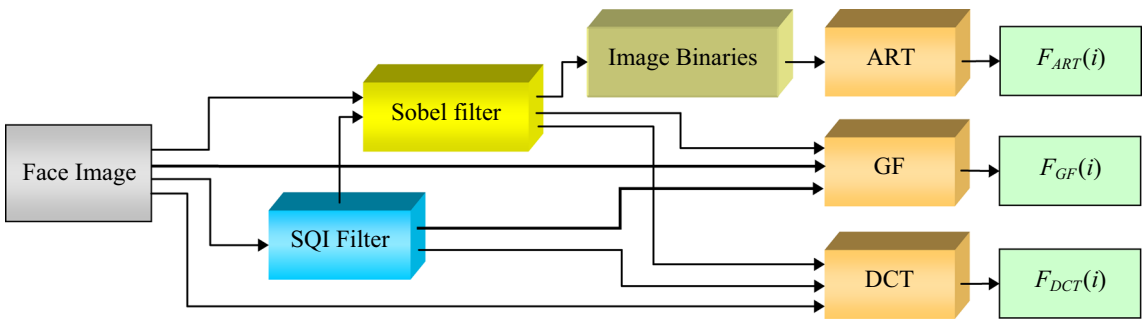

Fig. 13 The block diagram of feature extraction for three kinds features used in the FERS technique

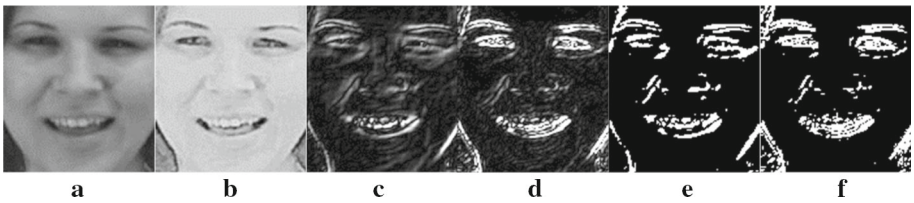

Fig. 14 The obtained images by performing the feature-extraction method used in the FERS technique. **a** An original image; **b**, **c** are the images of applying an SQL filter and a Sobel filter to (**a**), respectively; **d** is the image of applying a Sobel filter to (**b**); **e**, **f** are the binary versions of (**c**) and (**d**), respectively

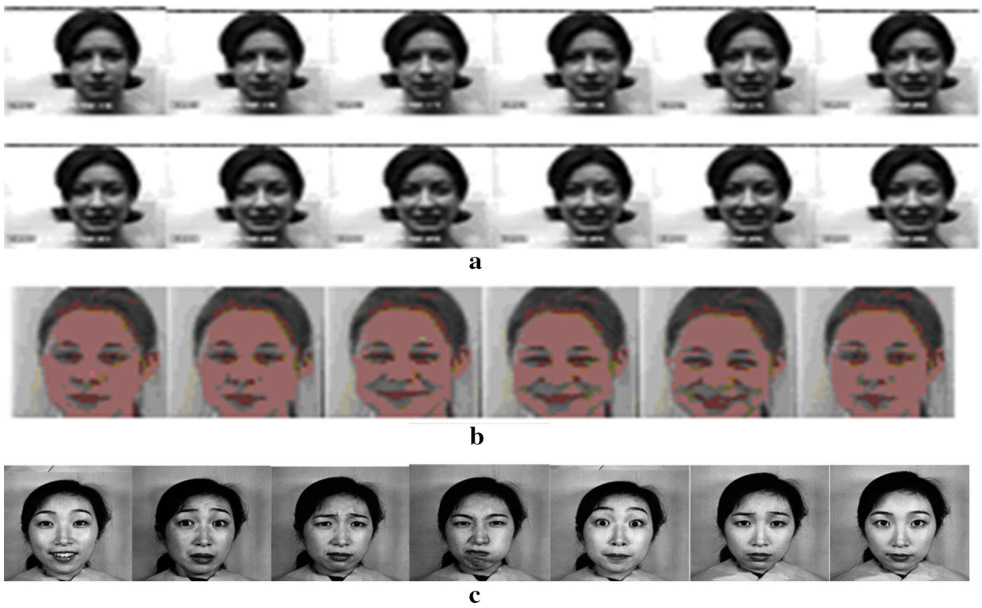

Fig. 15 **a–c** Shows some facial expression images in CKFI, FG-NET, and JAFFE databases, respectively

Step 3.1: Identify non-face images among these examined images.
Step 3.2: Output human facial image for each image.

Step 4 : Regularize face-region images.
Step 5 : Perform preprocessing before feature extraction.
Step 6 : Extraction of ART, DCT, and GF features for the ith image I_i and, then, get a set, F_i, of its features, $F_i = \bigcup_{k\in\{\mathrm{ART,\ DCT,\ GF}\}} F_k(i)$. The block diagram of feature-extraction algorithm is shown in Fig. 13.
Step 7 : Train a multi-class SVM as the facial expression classifier used in the FERS technique.
Step 8 : Output the trained SVM.

4 Experimental results

4.1 Face image databases

In the experiment, the paper employs three common facial expression databases to assess the performance of the proposed method. Figure 15 displays some examples of facial images. Figure 15a–c shows some images in the Cohn–Kanade face (CKFI) image database (Kanade et al. 2000), the FG-NET database (Wallhoff 2005), and the Japanese female facial expression (JAFFE) database (JAFFE 2008), respectively. Each image in JAFFE database has been categorized by facial expression definitions. However, the images in CKFI and FG-NET databases are sequential images without classification. All images in the two databases are also classified manually during the experiment. Accordingly, the examined image database consists of 3624 images here, which includes 688 happiness images, 270 sadness images, 802 neutral images, 582 surprise images, 280 anger images, 544 disgust images, 373 fear images, and 85 non-face images. Henceforth, the examined face image database is called FID.

4.2 Performance measurement

A quantitative index, *accuracy*, for the performance measurement can be defined by (14). Four cases are described as follows. First, Hits: true positive (*tp*), means that the actual

Table 1 The experimental results of face detection

Method	Total	*Tp*	*fn*	Miss
HFs	3539	3539	532	0
OpenCV	3539	3539	71	0
FERS	3539	3539	85	0

condition is present and the test result is positive. Second, False alarm: false positive (*fp*), indicates that the actual condition is absent but detected. The case is also called as type I error. Third, Misses: false negative (*fn*), stands for that actual condition is present but not detected. It is also called as type II error. Fourth, Correct rejections: true negative (*tn*), denotes that actual condition is absent and test result is negative. Hits and correct rejections are good, but false alarms and misses are harmful to overall performance.

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{14}$$

Another popular quantitative index, *F-measure*, for the performance measurement can be specified by (15),

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \tag{15}$$

where

$$Precision = \frac{tp}{tp + fp} \quad \text{and} \quad Recall = \frac{tp}{tp + fn}.$$

A classifier has a high *F-measure* if it possesses high *precision* and *recall* rates.

4.3 Face detection performance

Table 1 presents the comparison results for face detection performances of the HFs method, the OpenCV method, and the FERS technique. Their face detection rates reach 100% because the facial images in the database are rather simple (a single human face in each image). However, the *fn* values

Table 2 The results for regular illumination environments for face detection

Method	Total	*tp*	*fn*	Zoom in and zoom out face	Miss
HFs	54	53	10	0	1
OpenCV	54	54	4	0	0
FERS	54	54	1	0	0

of using the OpenCV method and the FRES technique have better than that of using the HFs method without performing image preprocessing. Here non-face images (images without face regions) may be extracted because noises may make the texture changes which are similar to the face features. Consequently, image filters as shown in Fig. 9 are used to remove the noises so as to reduce the *fn* rate. Figure 16 depicts some examples for the set of non-face images including partial face regions.

Table 2 shows face detection results under regular illumination environments for the cases of zoom in and zoom out. The results in Table 2 demonstrate that the FERS technique is superior to others methods under consideration here. Figure 17 exhibits face images for the cases of zoom in and zoom out.

4.4 Facial expression recognition performance

4.4.1 Evaluation for classifiers and filters

In the experiment, there are many combinations (models) for several filters and three kinds of features, ART, DCT and GF. Table 3 lists these combinations used in the experiment. Each combination can produce its corresponding set of features. The set is employed to train an SVM, and then the trained SVM is applied to perform the facial expression classification. Table 3 presents the compared results for investigating them. For example, the first model just adopts the ART features with 35 coefficients. The second model first applies an SOI filter to a face image and then

Fig. 16 Some images occur in the case of false negative during face detection

Fig. 17 Face images are detected in the cases for zoom in and zoom out

Table 3 The comparison results of using different combined features for two training cases in selecting training patterns

Model no.	Feature extraction	Verification	
		Tenfold cross-validation (average accuracy for testing)	Training patterns (66%) testing patterns (34%) (average accuracy for testing)
1	ART	97.21%	96.02%
	(35 features)	($C = 34, \gamma = 8$)	($C = 32, \gamma = 8$)
2	SQI filter+ART	94.95%	93.18%
	(35 features)	($C = 34, \gamma = 8$)	($C = 34, \gamma = 8$)
3	DCT	98.39%	95.77%
	(64 features)	($C = 32, \gamma = 0.125$)	($C = 32, \gamma = 0.5$)
4	SQI +DCT	97.18%	94.15%
	(64 features)	($C = 32, \gamma = 0.125$)	($C = 32, \gamma = 0.5$)
5	Sobel+DCT	98.26%	96.34%
	(64 features)	($C = 32, \gamma = 0.125$)	($C = 32, \gamma = 0.5$)
6	SQI+Sobel+DCT	97.84%	96.65%
	(64 features)	($C = 32, \gamma = 0.125$)	($C = 32, \gamma = 0.5$)
7	GF	97.07%	95.12%
	(80 features)	($C = 8, \gamma = 0.5$)	($C = 8, \gamma = 0.5$)
8	SQI+GF	91.41%	88.23%
	(80 features)	($C = 8, \gamma = 0.5$)	($C = 8, \gamma = 0.5$)
9	Sobel+GF	94.42%	90.42%
	(80 features)	($C = 8, \gamma = 0.5$)	($C = 8, \gamma = 0.5$)
10	SQI+Sobel+GF	88.35%	86.28%
	(80 features)	($C = 8, \gamma = 0.5$)	($C = 8, \gamma = 0.5$)
11	SQI+Sobel+DCT+ART	97.95%	96.59%
	(99 features)	($C = 32, \gamma = 0.125$)	($C = 32, \gamma = 0.125$)
12	SQI+Sobel+DCT+GF	98.59%	96.91%
	(144 features)	($C = 8, \gamma = 0.125$)	($C = 8, \gamma = 0.125$)
13	ART+GF	97.70%	95.29%
	(116 features)	($C = 32, \gamma = 0.125$)	($C = 32, \gamma = 0.125$)
14	SQI+Sobel+DCT+ART+GF	98.59%	97.10%
	(179 features)	($C = 8, \gamma = 0.125$)	($C = 8, \gamma = 0.125$)

extracts ART features from the filtered image. The third model employs DCT features with 64 low-frequency coefficients (characteristics). The sixth model first performs an SOI filter and a Sobel filter sequentially and then extract DCT features from the processed image. The seventh model uses GF features computed by first using 40 texture change elements with eight directions and five strengths, and then calculating the mean and standard deviation of the 40 ele-

ments for each texture change element. Totally, there are 80 GF features in the experiment. The difference between the sixth model and the tenth model is in utilizing different features. The twelfth model differs from the fourteenth model at the use of features. The former adopts DCT and GF features, and the latter employs a combination of three kinds of features.

K-fold cross-validation is employed for the assessment of the applicability of learning algorithms. First, data size is divided into equal K partitions where each partition is mutually exclusive. For example, the size of the data set A is divided into equal K partitions, namely $A = A_1 \cup A_2 \cup \cdots \cup A_K$ and then repeated several times. Let At, $t \in \{1, 2, \ldots, K\}$, denote a set of testing data and the remaining data are used for training data. After repeated for K times, the average accuracy can be calculated. Table 3 presents compared results for these eight models while verifying tenfold cross-validation and randomly selected data as training and testing data sets. While taking the GRBF as the kernel function in an SVM, C and γ represent the regularization and the kernel width (spread) of the SVM parameters, respectively. Here, empirical results show that the case, taking 66% of a whole set of patterns as training patterns and the remaining patterns as testing patterns, is better for the trade-off consideration to computation complexity and accuracy. Note that, although utilizing more training patterns in the training of an SVM can get better classification accuracy, the generalization ability of the SVM commonly reduces greatly (Haykin 1995).

4.4.2 *Combinations of filters and features*

To observe experimental results, it is necessary to involve an SQI filter under the various light environments because the image signal is vulnerable for varied strengths of lights. Figure 18 exhibits a comparison of using and without using an SQI filter followed by a Sobel filter. Figure 18b illustrates the result that an SQI filter can remove many image details. Figure 18c depicts more image details of the images still existing while only applying a Sobel filter to the original image. Accordingly, an image filtered by an SQI filter is inappropriate to be an input of a Sobel filter or an ART filter. The main purpose of using the Sobel filter is to enhance the line features. Moreover, the ART filter can preserve image details. More line features and image details of an image can form a unique set of features of the image for image recognition. Consequently, the unique feature set can help a classifier to get better performance of the discrimination process.

Here, some image-processing filters and three kinds of features are combined to form a combined feature to represent an image face. First, a face image, I, is filtered by an SQI filter to produce I_{SQI}, and then I_{SQI} is processed by the Sobel filter to get $I_{\mathrm{SQI,Sobel}}$. The first kind of features is extracted from DCT version of the filtered image $I_{\mathrm{SQI,Sobel}}$. The second kind of features is extracted from the ART version of the filtered image I_{Sobel}. Note that I is processed by the Sobel filter to produce I_{Sobel}. The third kind of features is extracted from the GF version of I. The first and the third kinds of features include texture characteristics. The second kind of features contains shape characteristics. Combining these three kinds of features to construct a new feature used in the FERS technique, therefore, it can get a better accuracy rate due to their complementary effects. Note that "SQI+Sobel+DCT+ART+GF" represents the new feature.

Fig. 18 **a** The original image; **b** the result image is obtained via first applying an SQI filter to (**a**) and then applying Sobel filter to the image processed by the SQI filter; **c** the result image is gained via applying Sobel filter to (**a**)

Table 4 presents a confusion matrix for the recognition performance of the FRES method for a case of the testing patterns with 34%. Here the non-face class can be detected correctly. An observation is that the prediction results of facial expressions for "happy" and "surprise" have higher accuracy rate. The reason is that the shapes of happy and surprise facial expressions changed greatly. These changes in texture and shape are rather unique for other expressions. However, it gets lower accuracy rate for "fear" and "disgust" expressions due to little changes in between their shapes as well as similar facial shape and texture changes between them.

Table 4 Using the FRES method in predictions for the testing patterns with 34% of FID

Predicted								FRES	
Anger	Disgust	Fear	Happy	Neutral	Non-face	Sadness	Surprise		
92	2	3	0	1	0	0	0	Anger	correct
0	174	3	0	4	0	1	3	Disgust	
0	1	114	2	0	0	3	0	Fears	
0	0	1	234	1	0	0	0	Happy	
2	4	0	0	263	0	0	0	Neutral	
0	0	0	0	0	27	0	0	Non-face	
1	2	0	0	0	0	85	0	Sadness	
0	0	0	0	1	0	0	208	Surprise	

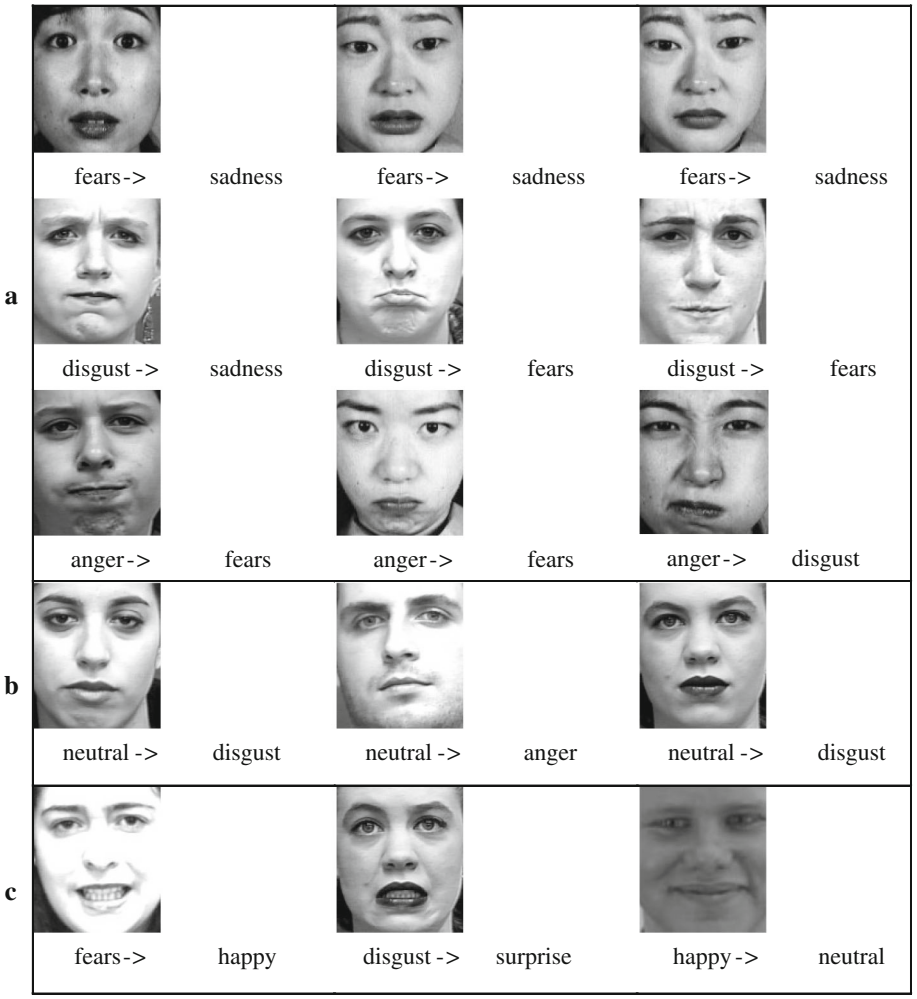

Fig. 19 Examples for misclassifications of using the FRES technique

Figure 19 illustrates some misclassifications. For example, fears → sadness as shown in Fig. 19a means actual condition is "fears" but the test result shows sadness. Figure 19b shows some "neutral" images are misclassified to the "disgust" classification or the "anger" one. Another finding as shown in Fig. 19c is that "fears" image is misclassified as "happiness" because they are analogous for texture and shape.

4.4.3 Classification performance

The results in Table 5 indicate that the SVM method is more suitable for facial expression recognition than the other classification methods evaluated in this paper. The MLP stands for the multilayer perceptrons where I, H, and O stand for the input layer, hidden layer, and output layer of the MLP, respectively (Haykin 1995; Tsai and Cheng 2005). The MLP

Table 5 Different classifiers with the average accuracy rate for 179 features

Method	SQI+Sobel+DCT+ART+GF (179 features)	
	Tenfold cross-validation	Training patterns (66%) testing patterns (34%) (average accuracy for testing)
SVM	98.59%	97.10%
	($C = 8, \gamma = 0.125$)	($C = 8, \gamma = 0.125$)
K-NN ($K = 3$)	95.11%	91.15%
MLP	95.25%	93.01%
	($I = 179, H = 93, O = 8$)	($I = 179, H = 93, O = 8$)
DT	78.06%	70.29%
BN	89.12%	85.30%

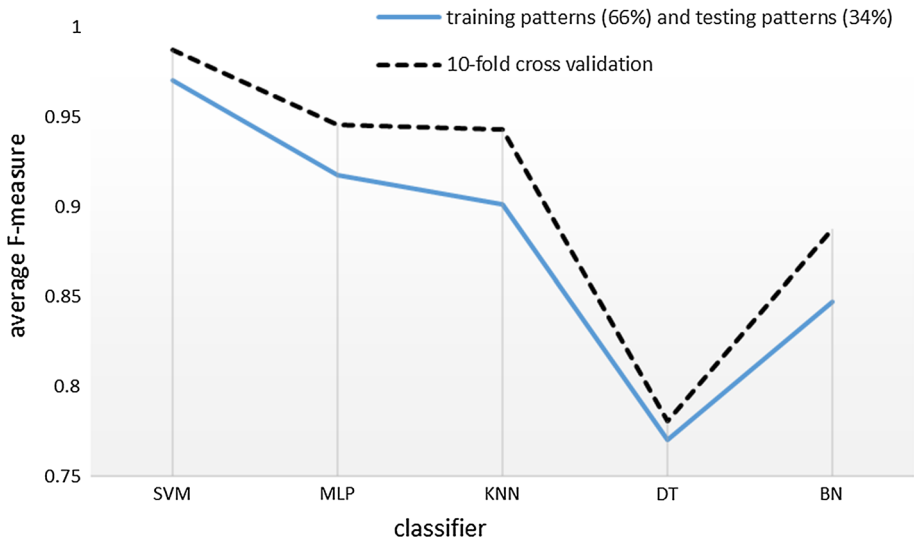

Fig. 20 A comparison for these classifiers shown in Table 5 in terms of average *F-measure*

is one model of artificial neural networks, which has 17391 weights ($179 \times 93 + 93 \times 8$), and these weights are estimated via the back-propagation algorithm. K-NN indicates the K-nearest neighbor method where K is set to 3 in the experiment. DT denotes the decision tree method. BN represents the Bayesian network method.

Observing the experiment results, the recognition accuracy of using an SVM is better than that of using other methods for two training cases. The results in Table 5 reveal that average accuracy rates of using K-NN and MLP are more than 90%. Note that K-NN is one of example-based methods. If the patterns in the training set are similar to the input data, then the input data can be easily predicted which class they belong to. The prediction is easily failure if there are no patterns similar to the input data. This method is very susceptible to the characteristics of training patterns. The experimental results in Table 5 also reveal that the average accuracy of using the SVM is better than that of using existing classification methods under consideration here. Although a few of papers are found in discussion for facial expression recognition involving two methods, DT and BN, the results also show that these two methods may be inappropriately applied to the facial expression identification problem because they are more suitable to find out the changes in properties of rules while solving the rule-changed problems. Figure 20 displays another comparison measure using *F-measure* for these classifiers shown in Table 5 for two training cases, tenfold cross-validation and training patterns (66%) as well as testing patterns (34%). Additionally, Tables 6, 7, 8 and 9 present confusion matrices for the recognition performance of K-NN, MLP, BN and DT methods, respectively, in predictions for the testing patterns with 34% of FID. To compare these confusion matrices in Table 4 with Tables 6, 7, 8 and 9, the FRES technique outperforms these methods mentioned above.

Table 10 lists the comparison results of the FERS techniques with existing methods for the examined images in CKFI, FG-net and JAFFE databases. The experimental results demonstrate that the FERS technique is superior to existing methods under consideration in the paper. The work of Chen et al. (2008) mixed two characteristics of face images to gain a higher accuracy. In the design of the FERS techniques, it combines various types of features and applies appropriate preprocessing methods (for instances,

Table 6 Using the K-NN method in predictions for the testing patterns with 34% of FID

Predicted								K-NN	
Anger	Disgust	Fear	Happy	Neutral	Non-face	Sadness	Surprise		
85	4	0	0	7	0	2	0	Anger	Correct
0	175	1	0	4	0	2	3	Disgust	
2	6	104	3	5	0	0	0	Fears	
4	5	0	218	8	0	1	0	Happy	
12	7	0	3	245	0	2	0	Neutral	
0	1	0	1	2	21	0	2	Non-face	
6	3	0	0	2	0	77	0	Sadness	
0	2	0	0	8	0	1	198	Surprise	

Table 7 Using the MLP method in predictions for the testing patterns with 34% of FID

Predicted								MLP	
Anger	Disgust	Fear	Happy	Neutral	Non-face	Sadness	Surprise		
77	8	3	0	6	0	4	0	Anger	Correct
0	168	4	0	2	0	3	8	Disgust	
0	2	102	3	2	2	2	7	Fears	
0	0	0	236	0	0	0	0	Happy	
2	5	0	0	261	0	1	0	Neutral	
0	0	0	0	0	27	0	0	Non-face	
4	2	3	0	1	0	77	1	Sadness	
0	4	2	0	1	0	4	198	Surprise	

Table 8 Using the BN method in predictions for the testing patterns with 34% of FID

Predicted								BN	
Anger	Disgust	Fear	Happy	Neutral	Non-face	Sadness	Surprise		
68	13	3	1	9	0	2	2	Anger	Correct
4	155	1	4	15	0	1	5	Disgust	
1	5	94	9	6	0	3	2	Fears	
0	0	3	230	1	0	1	1	Happy	
1	38	3	0	217	0	8	2	Neutral	
0	1	0	0	3	23	0	0	Non-face	
2	6	3	0	7	0	67	3	Sadness	
0	2	3	4	3	0	0	197	Surprise	

Table 9 Using the DT method in predictions for the testing patterns with 34% of FID

Predicted								DT	
Anger	Disgust	Fear	Happy	Neutral	Non-face	Sadness	Surprise		
58	5	4	5	16	0	8	2	Anger	Correct
11	123	8	10	11	0	17	5	Disgust	
10	14	61	11	10	0	5	9	Fears	
12	12	8	189	10	0	2	3	Happy	
8	11	9	9	211	0	13	8	Neutral	
0	0	3	1	2	21	0	0	Non-face	
4	8	4	5	11	0	48	8	Sadness	
3	12	5	12	23	0	11	143	Surprise	

Table 10 The comparison results of the FERS techniques and other FER methods

Method	Database			Facial emotion classification	Training patterns selection	Average accuracy for training patterns	Average accuracy for testing patterns
	CKFI	FG-net	JAFFE				
Loh et al. (2006)				6			83.75%
Sun et al. (2008)			•	4	Tenfold		97.28%
Kim et al. (2008)			•	7			92.20%
Kharat and Dudul (2008)			•	6	Tenfold	100%	94.29%
Chen et al. (2008)	•			7			93.10%
Yang et al. (2009)	•			7			95.00%
Kotsia et al. (2008)	•			7	Fivefold		92.30%
Zhou et al. (2006)	•			6			90.00%
Lee et al. (2008)	•			6			92.85%
Cheng et al. (2008)	•			6			90.40%
Liu and Wang (2006)	•			7			92.57%
FRES	•			7	Fivefold		99.54% ($C = 8, \gamma = 0.125$)
FRES	•			7	Tenfold		99.54% ($C = 8, \gamma = 0.125$)
FRES	•			7		100%	99.03% ($C = 8, \gamma = 0.125$)
FRES		•		8	Tenfold		98.59%
FRES			•	8		99.95%	97.10%

an SQI filter, a Sobel filter, and image regulation) in getting these features. This results in that the FERS techniques can effectively improve the facial expression recognition accuracy of traditional schemes for these three examined databases. Additionally, using SVMs, which possess high generalization ability, also promotes the accuracy of the FERS techniques.

5 Conclusions

The paper has presented a novel approach, called the FERS technique, based on an SVM for facial expression recognition. It first utilizes HFs combining with SQI filters for improving face detection accuracy under various light source environments. Moreover, the face region of an image can be correctly located so as to remove the non-face regions of the image such as different hairstyles. Furthermore, the Sobel filter is also exploited to strengthen the texture features for edge detection. The FERS technique extracts three kinds of features based on DCT, GF, and ART features accompanying with appropriate preprocessing methods such as an SQI and a Sobel filter. One of main contributions of the paper is to find out a new feature of face images for facial expression recognition. The new feature is selected from these combinations of blending preprocessing methods with these three features, DCT, GF, and ART. The new feature mentioned above forms the input vector of each pattern (including training and testing patterns). Finally, the trained SVM plays a role as a corresponding multi-classes classifier that can predict facial expressions. The experimental results showed that the classification performance of the FERS technique is definitely better than other existing schemes under consideration here.

Acknowledgements Authors would like to thank Ministry of Science and Technology of Taiwan, ROC, for financially supporting this research under Contract Nos. NSC 98-2511-S-150-002 and MOST 104-2221-E-150-010.

Compliance with ethical standards

Conflict of interest Author Prof. Tsai has received research grants from Company Taiwan. OR if no conflict exists: Author Prof. Tsai declares that he has no conflict of interest. Author Mr. Chang declares that he has no conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Al-Shabi M, Cheah WP, Connie T (2016) Facial expression recognition using a hybrid CNN–SIFT aggregator. https://arxiv.org/pdf/1608.02833

Burges CJC (1998) A Tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2:121–167

Chai Z, Vazquez MH, He R, Sun Z, Tan T (2014) Explore semantic pixel sets based local patterns with information entropy for face recognition. EURASIP J Image Video Process 26:1–11

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:1–27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chen HY, Huang CL, Fu CM (2008) Hybrid-boost learning for multi-pose face detection and facial expression recognition. Pattern Recogn 41:1173–1185

Cheng K, Chen Y, Zhan Y (2008) A new approach for facial expression recognition based on Burial Markov model. In: Proceedings of the fifth international conference on fuzzy systems and knowledge discovery, pp 3–9

Fang J, Qiu G (2003) Human face detection using angular radial transform and support vector machines. In: Proceedings of international conference on image processing, pp 669–672

Fellenz WA, Taylor JG, Tsapatsoulis N, Kollias S (1999) Comparing template-based, feature-based and supervised classification of facial expressions from static images. In: Proceedings of circuits systems, communications and computers, pp 5331–5336

Haykin S (1995) Neural networks: a comprehensive foundation. Macmillan College Publishing Company, New York

JAFFE (2008) The Japanese female facial expression (JAFFE) database. http://www.kasrl.org/jaffe.html

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, pp 46–53

Kharat GU, Dudul SV (2008) Human emotion recognition system using optimally designed SVM with different facial feature extraction techniques. WSEAS Trans Comput 7:650–659

Kim DH, Jung SU, Chung MJ (2008) Extension of cascaded simple feature based face detection to facial expression recognition. Pattern Recogn Lett 29:1621–1631

Kotsia I, Zafeiriou S, Pitas I (2008) Texture and shape information fusion for facial expression and facial action unit recognition. Pattern Recogn 41:833–851

Lajevardi SM, Hussain ZM (2012) Automatic facial expression recognition: feature extraction and selection. SIViP 6:159–169

Lee JJ, Uddin MZ, Kim TS (2008) Spatiotemporal human facial expression recognition using fisher independent component analysis and Hidden Markov Model. In: Proceedings of the 30th annual international IEEE EMBS conference, Vancouver, British Columbia, pp 2546–2549

Liu W, Wang Z (2006) Facial expression recognition based on fusion of multiple Gabor features. In: Proceedings of the 18th international conference on pattern recognition, pp 536–539

Loh MP, Wong YP, Wong CO (2006) Facial expression recognition for e-Learning systems using Gabor wavelet & neural network. In: Proceedings of the sixth international conference on advanced learning technologies, pp 523–525

Mita T, Kaneko T, Hori O (2005) Joint haar-like features for face detection. In: Proceedings of the tenth IEEE international conference on computer vision, pp 1619–1626

OpenCV (2015) http://opencv.org/

Rowley HA, Baluja S, Kanade T (1998) Neural network-based face detection. IEEE Trans Pattern Anal Mach Intell 20:23–38

Saeed I, Wang A, Senaratne R, Halgamuge S (2007) Using the active appearance model to detect driver fatigue. In: Proceedings of the third international conference on information and automation for sustainability, pp 124–128

Shih P, Liu C (2006) Face detection using discriminating feature analysis and support vector machine. Pattern Recogn 39:260–276

Sun X, Xu H, Zhao C, Yang J (2008) Facial expression recognition based on histogram sequence of local Gabor binary patterns. In: Proceedings of IEEE conference on cybernetics and intelligent systems, pp 158–163

Tsai HH, Lai YS, Lo SC (2013) A zero-watermark scheme with geometrical invariants using SVM and PSO against geometrical attacks for image protection. J Syst Softw 86:335–348

Tsai HH, Chang BM, Liu SH (2014) Rotation invariant image retrieval using PSO and SVR. Appl Soft Comput 17:127–139

Tsai HH, Cheng JS (2005) Adaptive signal-dependent audio watermarking based on human auditory system and neural networks. Appl Intell 23:191–206

VenkataRamiReddy Ch, Kishore KVK, Bhattacharyya D, Kim TH (2014) Multi-feature fusion based facial expression classification using DLBP and DCT. Int J Softw Eng Appl 8:55–68

Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, pp 511–518

Wallhoff F (2005) Facial expressions and emotion database. http://www.mmk.ei.tum.de/~waf/fgnet/

Wei W, Jia Q (2016) Weighted feature Gaussian kernel SVM for emotion recognition. Comput Intell Neurosci 2016:1–7

Wong JJ, Cho SY (2009) A local experts organization model with application to face emotion recognition. Expert Syst Appl 36:804–819

Yang MH, Kriegman D, Ahuja N (2002) Detecting faces in images: a survey. IEEE Trans Pattern Anal Mach Intell 24:34–58

Yang P, Liu Q, Metaxas DN (2009) Boosting encoded dynamic features for facial expression recognition. Pattern Recogn Lett 30:132–139

Zhou G, Zhan Y, Zhang J (2006) Facial expression recognition based on selective feature extraction. In: Proceedings of the sixth international conference on intelligent systems design and applications, pp 412–417