Multimed Tools Appl (2016) 75:6663–6682
DOI 10.1007/s11042-015-2598-1

Effective semantic features for facial expressions recognition using SVM

Chen-Chiung Hsieh[1] · **Mei-Hua Hsih**[2] · **Meng-Kai Jiang**[1] · **Yun-Maw Cheng**[1] · **En-Hui Liang**[3]

Received: 21 August 2014 / Revised: 25 February 2015 / Accepted: 31 March 2015 /
Published online: 17 April 2015

Abstract Most traditional facial expression-recognition systems track facial components such as eyes, eyebrows, and mouth for feature extraction. Though some of these features can provide clues for expression recognition, other finer changes of the facial muscles can also be deployed for classifying various facial expressions. This study locates facial components by active shape model to extract seven dynamic face regions (frown, nose wrinkle, two nasolabial folds, two eyebrows, and mouth). Proposed semantic facial features could then be acquired using directional gradient operators like Gabor filters and Laplacian of Gaussian. A multi-class support vector machine (SVM) was trained to classify six facial expressions (neutral, happiness, surprise, anger, disgust, and fear). The popular Cohn–Kanade database was tested and the average recognition rate reached 94.7 %. Also, 20 persons were invited for on-line test and the recognition rate was about 93 % in a real-world environment. It demonstrated that the proposed semantic facial features could effectively represent changes between facial expressions. The time complexity could be lower than the other SVM based approaches due to the less number of deployed features.

Keywords Face detection · Active shape model (ASM) · Facial expression recognition · Facial texture · Support vector machine (SVM)

✉ Chen-Chiung Hsieh
cchsieh@ttu.edu.tw

[1] Department of Computer Science and Engineering, Tatung University, No. 40, Sec. 3, Jhongshan N. Rd., Taipei 104, Taiwan, Republic of China

[2] Department of Industrial Design, Tatung University, No. 40, Sec. 3, Jhongshan N. Rd., Taipei 104, Taiwan, Republic of China

[3] Department of Information Management, Tamkang University, No.151, Yingzhuan Rd. Tamsui Dist., New Taipei City 25137, Taiwan

1 Introduction

Interpersonal communication is conducted simultaneously at auditory, visual, and tactile levels. Auditory communication involves communicating through speaking, whereas visual communication involves communicating through writing, gestures, and facial expressions. People communicate their feelings by using these complementary channels. However, facial expression is the most crucial form of communication.

Facial expression is critical for people expressing their inner feelings, and they are a nonverbal form of communication. Due to the development of artificial intelligence on computer–user interaction, the ability of computers to detect emotional change and generate appropriate feedback can benefit the application and advancement of artificial intelligence. For example, robots that can communicate concepts with people are more than just tools, because they can also be used in medical aspect for monitoring changes of the facial expressions of patients during diagnosis and supplementary care procedures. For digital signage applications, in addition to counting the number of people and the length of time for which they stay, facial expression recognition systems can indicate additional clues regarding whether a viewer is a potential customer. This type of information is valuable for commercial advertisers.

Ekman and Friesen [14] made a pioneering facial expression recognition which provided a way for current expression recognition researches. The development of image processing and pattern recognition technologies [49] makes automatic facial expression recognition possible. In general, a facial expression recognition system could be divided into several modules [35]. First of all, face detection module [2, 20, 22, 40] is to find the face of the subject from captured image. There might be obstacles that make this detecting process harder due to changes in environments [46], lighting, or face orientations [52]. Once there is a face found, the feature extraction module detects whether there is an expression being displayed by tracking facial muscle changes or deformations of facial components like eyes and lips. Finally, the recognition module classifies the extracted features and a decision is made about the expression being displayed. Of course the system needs to learn from a database which is trained to detect expressions regardless of age, sex, and ethnicity.

The majority of work conducted in this field involves 2D imagery. Expression features can be divided into three categories [49]: deformation features, motion features, and statistical features. Deformation feature is used to measure the relative distance changes of feature points [24] or texture changes of feature patches [4] caused by the variety of expressions. Motion feature [34] is mainly used to track some feature points or patches from a sequence of images and calculate the movement distance and direction. Statistical features extracted by histogram, moment invariant, principal component analysis (PCA) [6], or linear discriminant analysis (LDA) [27], are used to describe the characteristics of expression images.

In order to deal with the problems caused by inherent pose and illumination variations, 3D and 4D (dynamic 3D) recordings [42] are increasingly used in expression analysis researches [32, 43]. Moreover, 3D video recordings could capture out-of-plane changes of the facial surface, or easy to see changes. Research in 3D facial expression analysis is still in its infant stage, with a large number of works expected in the near future as the current technological advances allow the easy and affordable acquisition of high quality 3D data. Still, there exist several issues such as computationally expensive correspondence algorithms and existing exaggerated 3D expressions databases.

Classification algorithms for facial expression recognition [49] are usually space-based methods like neural networks [25], support vector machine (SVM) [3, 9, 28, 31, 39, 41], k-

Nearest-Neighbor (k-NN) [31], AdaBoost method [47], etc. Time and space based methods are like hidden Markov model (HMM) [33], regression neural network, spatial and temporal motion energy templates method. To evaluate these classification algorithms is beyond this study. Here, we discuss the deployed feature vector and the efficiency of the classification algorithm for expression recognition while using the deployed features.

With regarding the issue about the size of feature vector, it ranges dramatically from the size of whole image to the size within hundreds or dozens. Of course, the less size of the feature vector, the lower complexity of the classification algorithm. The difference relies on the description ability of the deployed features. Thus, reduce the size of the feature vector while keep high recognition accuracy is one theme of this study. Most researches in facial expression recognition are focused on the feature extraction and recognition processes by assuming that the subject face is detected. The other theme of this study is to propose a real-time facial expression recognition system which integrates all the three modules discussed previously as a complete system.

The remainder of this paper is organized as follows. Section 2 introduces related studies on facial expression recognition, Section 3 discusses the proposed semantic facial features for expressions recognition, Section 4 describes the system architecture and implementation of the study, Section 5 presents the results and analyses of the study, and Section 6 concludes this study.

2 Overview of related work

Recently, facial expression recognition has been the main focus of studies on human–machine interfaces. Fasel and Luettin [15] indicated that deriving facial feature deformation and facial motion from facial images are crucial stages in analyzing facial expressions. Deformation extraction or feature extraction can be categorized as image-based [16] and model-based [23] approaches. There is not a conflict with the feature classifications mentioned in the previous section. But we change the view point to the adopted features and the extraction mechanisms.

Image-based approaches such as Gabor wavelets [48] and local binary patterns (LBP) [50, 51] involve processing facial images or specific facial regions to extract facial expression information without using additional facial expression knowledge. Thus, the size of feature vector is the whole original image or the utilized specific facial regions. LBPs and variations of LBPs as texture descriptors were even investigated for multi-view facial expression recognition [29]. Though PCA or LDA [53] could be used to reduce the dimensions of feature vector, the space complexity is still too large when compared with other methods. Conversely, model-based approaches mainly involve using facial models to represent facial structures, and therefore can determine the facial motion and deformation of facial features. However, the disadvantage of model-based approaches is that most feature points of facial models must be set manually, which involves relatively complex procedures, such as those involved in the active appearance model (AAM) [1] and point distribution model [19]. Face and landmark detection by using cascade of classifiers could be found in [7, 8]. Facial motion typically refers to optical flow [34, 36], motion modeling, and feature point tracking [45]. The more feature points used in the model, the more recognition accuracy. However, it must pay for higher space and time complexity.

Tang et al. [38] used AAM to extract 63 feature points in face and derive four effective features for expression recognition. The degree of openness of these facial components were

measured based on their height-to-width ratios. Facial expressions were determined by multiplying these ratios by their respective weights, and then summing the resulting values. They classified 4 basic expressions only and the recognition rate is nearly 88 %.

Ou et al. [30] defined 28 feature points proximal to facial components for measuring facial expression by using 40 Gabor filters comprising five frequency types in eight directions. Because of the high volume of the derived feature vectors, principal component analysis was used to reduce the data dimensions, and k-nearest neighbors was subsequently used to categorize the feature vectors into one of six expression types. However, facial feature extraction must be set manually and normalized.

Due to the small size of target set, SVM is quite suitable for facial expression recognition. Bartlett et al. [3] used a bank of Gabor features and selected 200 features per action unit by AdaBoost. Support vector machine was adopted as the classifier. However, the union of all features selected for each of the 20 action unit detectors resulted in a total of 4000 features.

Tsai et al. [39] employed angular radial transform (ART), discrete cosine transform (DCT), and Gabor filter (GF) on face image to extract facial features. The model adopts ART features with 35 coefficients, DCT features with 64 low-frequency characteristics, and GF features with 40 texture change elements. A SVM is trained to achieve high recognition rate.

Michel et al. [28] defined 22 feature points for automatic tracking. The motions of all the feature points from neutral to peak expression were measured as a feature vector. Chen et al. [9] used the feature points displacements and local texture differences between the normalized neutral and expressive face images for recognition. The combined feature vector contains a 42 dimensional geometric feature vector and a 21 dimensional texture feature vector. The average accuracy is 95 % using a SVM on the extended Cohn-Kanade database.

Valstar and Pantic [41] located 20 facial points based on Gabor-feature boosted classifiers. These points are tracked by particle filtering with factorized likelihoods and recognized by a combination of GentleBoost, SVM, and hidden Markov model. They attain an average AU recognition rate of 95.3 % on a benchmark set and 72 % when tested on spontaneous expressions. In addition, Saeed et al. [31] use eight facial points to achieve state-of-the-art recognition rate using a SVM. However, the expression recognition rate using geometrical features is adversely affected by the errors in the facial point localization, especially for the expressions with subtle facial deformations.

There have been several advances in the past few years in terms of feature extraction mechanisms that use SVM for expression classification. The contribution of this study is mainly to propose new facial clues and their corresponding facial features for recognition based on the description of action units from Ekman and Friesen [14]. This study analyzed the action units to recognize the following six facial expressions: 1) neutral, 2) happiness, 3) surprise, 4) anger, 5) disgust, and 6) fear as shown in Fig. 1. Based on the various facial muscles involving facial expressions, the relationships between expression and mood for effectively detecting facial variations are obtained. The essential feature vector includes both geometrical properties and facial textures of specific dynamic facial regions. They are of higher level and more semantic than traditional primitive facial points. Especially, the size of the feature vector is only six which speeds up the recognition process.

In this study, Adaboost [44] and active shape model (ASM) [10] are used to identify human faces and locate facial components, respectively, from camera captured images. Subsequently, Gabor filters [12] and Laplace of Gaussian (LoG) edge detection [37] are adopted to extract high level dynamic facial features, the proposed semantic facial features, from defined dynamic facial regions. The semantic features could be quantized from measured directional

Fig. 1 The six facial expression types: neutral, happiness, surprise, anger, disgust, and fear from left to right in this study

gradients and textures. Support vector machines (SVMs) [11] are then trained to classify the user facial expressions into one of the six types of expression.

The surveys on facial expression recognition [5, 33, 35, 42, 49] present multi-view of the advances in this field including the applications of automatic face expression recognizers, the characteristics of an ideal system, the databases that have been used and the advances made in terms of their standardization and a detailed summary of the state of the art. Compare with the state of the art, the proposed system has several advantages: 1) the proposed semantic facial features are new which include important geometrical properties and texture features of dynamic facial regions, 2) semantic features could be extracted from directional gradient operators like Gabor and Log filters, 3) multi-class non-linear SVM executes in less time by semantic features, and 4) integrated with Adaboost and ASM for real time on-line face detection.

3 Facial expressions analysis and proposed semantic facial features

According to research on facial expressions, the facial action coding system (FACS) [14], which is a standard method for describing facial expression, was used in this study. Facial expression in the FACS comprises variations of the upper face (i.e., the forehead, eyebrows, and eyes) and those of the lower face (i.e., the mouth and nose). These varying facial components, such as the eyebrows stretching upward or the eyes opening wide, are called action units, of which 44 have been identified. Figure 2 shows how these action units can be combined to describe various forms of facial expression.

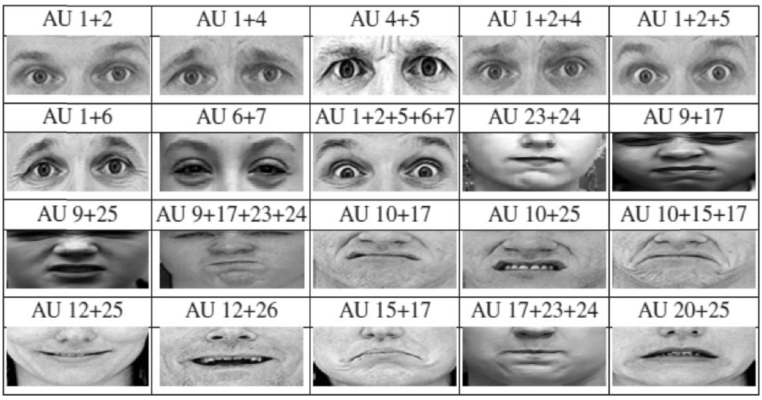

Fig. 2 Facial expressions represented as the combination of action units [14]

This study analyzed the action units to recognize the following six types of facial expressions: 1) neutral, 2) happiness, 3) surprise, 4) anger, 5) disgust, and 6) fear. Based on the various facial muscles involving facial expressions, the relationships between expression and mood for effectively detecting facial variations are summarized in Table 1. The facial clues and their corresponding facial features for recognition are proposed by us based on the description of action units from Ekman and Friesen [14].

As shown in the last column of Table 1, the semantic features found to be descriptive are 1) distance between the eyes and eyebrows, 2) mouth width, 3) frown lines between the eyebrows, 4) bunny lines between the nose and eyes, 5) left nasolabial folds, and 6) right nasolabial folds. These features include both geometrical properties (1 and 2) and facial textures (3 to 6) about dynamic facial regions. They are of higher level and more semantic than traditional primitive facial features while describing facial expressions.

Regarding to the changes between facial components, Euclidean distance was adopted for the measurement. Firstly, the distance between the eyes and eyebrows (*BtoE*) is defined as the distance between the lines formed by connecting the inner corners of the eyebrows (*EyeBrowL* and *EyeBrowR*) and eyes (*EyeL* and *EyeR*) which are designated as in Fig. 3. The eyebrows typically move upward and downward when expressing surprise or fear. The distance between the center of line (*EyebrowL* and *EyebrowR*) as well as the center of line (*EyeL* and *EyeR*) is calculated as the distance between the eyebrows and eyes. Secondly, the width of the mouth to that of the face, as depicted in Fig. 4, is used for normalization and expressed in Eq. (1).

$$Feature_{mouth} = \frac{Mouth_{width}}{Face_{width}} \tag{1}$$

Table 1 Facial clues and features for expression recognition

Mood	Ekman Description	Our observed facial clues	Proposed semantic features
Happy	cheeks are raised, lip corners are pulled obliquely	mouth stretched wide and upward slightly; nasolabial folds appear on both sides between nose and mouth; cheeks lifted.	ratio of mouth to face width; textures on both sides of the nasolabial folds
Angry	nostrils raised, mouth compressed, furrowed brow, eyes wide open	inner eyebrows extruded downward; two eyebrows closed; frown lines appear between the eyebrows; mouth closed.	distance between eyebrows and eyes; distance between eyebrows; frown lines between eyebrows;
Surprised	eyebrows raised, mouth opened, eyes opened	inner eyebrows up; chin downward; mouth opened big.	distance between eyebrows and eyes; ratio of mouth width to face width
Scared	eyes opened, mouth opened, eyebrows raised, lip stretch	inner eyebrows up; mouth corners stretched downward.	distance between eyebrows and eyes; ratio of mouth width to face width
Disgust	upper lip raised, nose wrinkle	nose squeezed up; bunny lines appear between nose and the inner corners of eyes; eyebrows slightly downward; mouth slightly recessed; upper lip lifted up.	bunny lines appear between nose and eyes; distance between eyes and eyebrows; width ratio of mouth and face;

Fig. 3 Inner feature points of eyes and the distance between eyes and eyebrows

As shown in Fig. 5, the ROIs used for expression recognition are the regions around the eyebrows, nose wrinkles, and nasolabial folds on both cheeks. These ROIs are adaptable which could be defined by the feature points extracted by ASM. Regarding the measurement of texture features, Eq. (2) is applied to calculate the ratio of the total edge pixels count to the area of the texture regions of interest (ROIs), where the ROIs are the regions after processed with edge operators. Together with the two relative changes, there are totally six features defined for recognizing facial expressions. The whole system and functional components including feature extraction and recognition module are described in more detail in the following section.

$$Feature_{texture} = \frac{\sum_{i=1}^{ROI_{width}} \sum_{j=1}^{ROI_{Height}} Pixel_{i,j}}{ROI_{Area}}, \quad Pixel \in Texture\ Area \tag{2}$$

4 System architecture

According to the previous discussions, most researches [3, 9, 28, 30, 31, 38, 39, 41] focus on the classification module of facial expressions without considering face detection module. This may be due to the challenge caused by complex environments, lighting, or face orientations. Here, we design the system flowchart, as shown in Fig. 6, which could conquer the mentioned problems. For each input image, Adaboost [44] which could eliminate the effects caused by environments and lighting is used to detect the presence of a human face. When a face is detected, the ASM [10] is used to extract the facial shape and facial components even under some facial deformations. The ASM feature points are then used to calculate the adaptable ROIs for semantic facial feature extraction. Setting the ROIs could eliminate many unnecessary regions and filter out excessive noises. Afterward, geometric properties are calculated from the developed feature points and texture features are calculated by deploying multidirectional Gabor filter [12] and LoG [37]. The formed semantic feature vector is used for facial

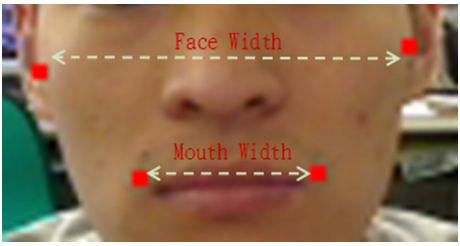

Fig. 4 Mouth width and facial width

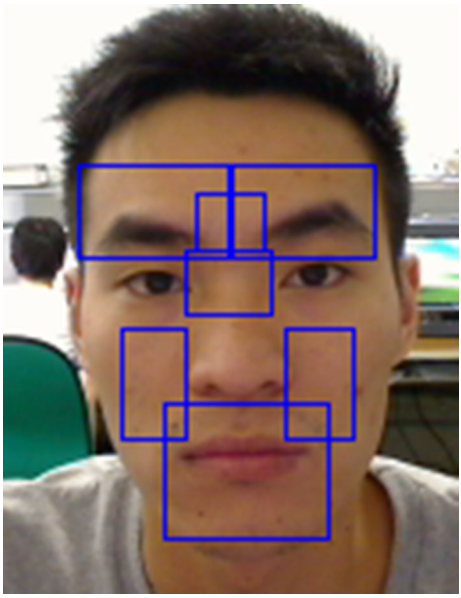

Fig. 5 The adaptable ROIs used for feature extraction

expression recognition by a SVM [11] from a neutral expression to any other expression. Each of the system modules is detailed in the following subsections.

4.1 Face detection

Accurate facial expression recognition depends on robust face detection. The facial detection method proposed by Viola and Jones [44] and extended by Lienhart and

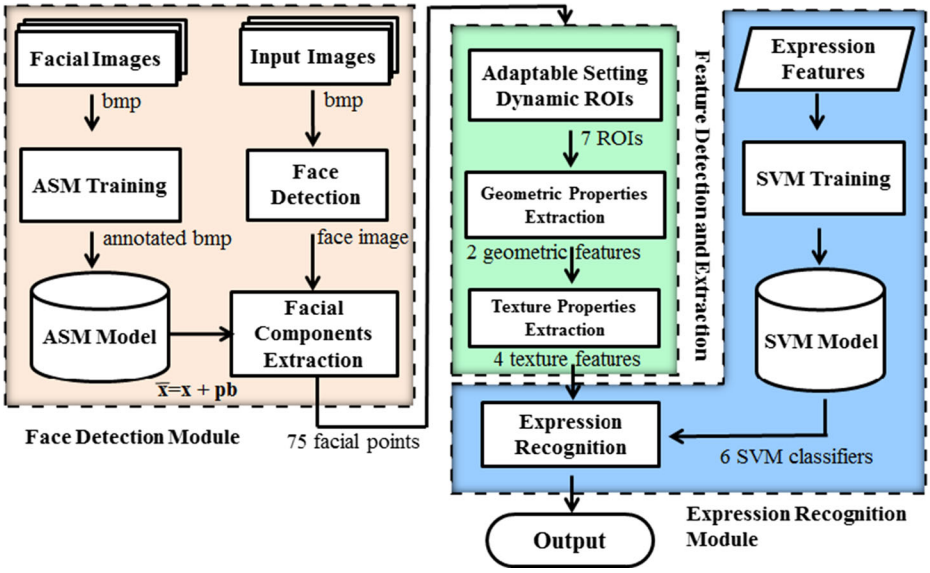

Fig. 6 System architecture

Maydt [26] is adopted as one of the key components in this study. The characteristic of their methods involves using the black–white Haar-like patterns for identifying eyes on a face, regardless of skin color. However, false alarms occur when eye-like patterns are detected. In this study, the ASM [10] is used to position the 75 defined facial feature points (Fig. 7). False alarms are filtered out if the amount of deformation is greater than a given threshold.

To build a deformable shape model for various human faces, the ASM is used to detect and extract facial components. The main advantage of using the ASM is that various targets can be detected if enough training samples are given. The ASM is described briefly as follows. Positive samples are used to train the ASM. Figure 3 shows the 75 landmarks used in this study, including feature points of the face shape, eyebrows, eyes, nose, and mouth. These feature points are selected manually during training because they involve corner points, high curvatures, or junction points. Interpolated points are inserted at equal intervals between two consecutive feature points.

The ASM is built based on the mean shape as shown by Eq. (3) after completing all of the training processes. From all the training data and mean shape, a covariance matrix S of $2n \times 2n$ could be calculated. By singular value decomposition of the covariance matrix, Eigen-system consists of Eigenvalue $\boldsymbol{\lambda}(\lambda_1, \lambda_2, \ldots, \lambda_{2n})$ and Eigenvector $\boldsymbol{P}(P_1, P_2, \ldots, P_{2n})$ was obtained.

$$\mathbf{x} = \overline{\mathbf{x}} + P\mathbf{b}, \tag{3}$$

where $\mathbf{b}$ is a variable for the feature vector. The ASM varies according to the changes in $\mathbf{b}$, as shown in Eq. (4), where $m = 2 \sim 3$. The range of scope $\boldsymbol{b}$ is verified from a set of known shapes $\boldsymbol{x}$, as expressed in Eq. (5), where P must be of a square matrix to ensure that the transpose matrix of P exists.

$$-m\sqrt{\lambda_i} \le b_i \le m\sqrt{\lambda_i} \tag{4}$$

$$\mathbf{b} = P^T(\mathbf{x} - \overline{\mathbf{x}}) \tag{5}$$

Before the first change, $\boldsymbol{b}$ was set to zero to obtain the target shape that is equal to the mean shape. The shape of the ASM could then be changed by modifying $\mathbf{b}$. Furthermore, shape and position parameters could be adjusted to change the rotation angle and scaling factor for a

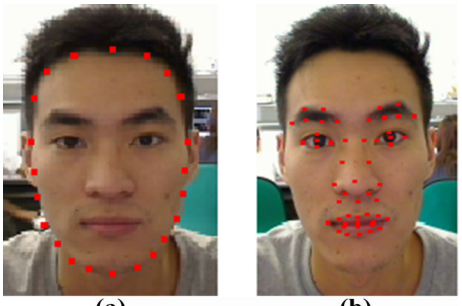

Fig. 7 The 75 defined ASM feature points. **a** Face shape consists of 22 feature points. **b** Facial components consist of 53 feature points

deformable ASM, as expressed in Eqs. (6) and (7), where s is the scaling factor, θ is the rotation angle, and (T_x, T_y) is the translational offset.

$$\mathbf{x} = A(p) \cdot (\overline{\mathbf{x}} + \mathbf{Pb}) \tag{6}$$

$$A(p) = \begin{bmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix} \tag{7}$$

4.2 Semantic facial features extraction

To extract the semantic facial features accurately, seven adaptable ROIs (Fig. 5) are calculated according to the ASM feature points. These ROIs are scalable, relative to the size of a detected face, thereby saving processing time and reducing noises. Among extracted facial components and shapes, eyebrows and mouth are particularly different when people exhibit various facial expressions. The relative changes of eyebrows and mouth are measured from a neutral position to any given facial expression.

Regarding wrinkles, the nasolabial folds, nose wrinkles, and frown wrinkles are the most prominent. This study employ directional Gabor filter to detect eyebrows and nasolabial folds on both cheeks, and a LoG operator is used to detect the edges of the mouth, nose wrinkles, and frown wrinkles. Dynamic facial texture could thus be measured. The semantic features including some geometric properties and dynamic facial textures differ from those surveyed in Section 2. Due to the high level descriptive ability of the features, we call them semantic features.

Based on the directional changes of dynamic facial textures, the Gabor filter is used to detect 0°, 45°, and 135° edges of the previous set of ROIs. Because both the mouth and eyebrows are horizontal, the 0° Gabor filter is used. The dynamic facial textures on both cheeks, such as those of nasolabial folds, are mostly diagonal; therefore, the 45° and 135° Gabor filters are used. The other parameters (γ, σ), wave length and scale at orthogonal directions, are set as (1, 6.28) when the distance between user and camera is around 30~50 cm. Figure 8 shows the detection results of the Gabor filters with various orientations.

Because the Gabor filters are less effective in detecting the upper and lower jaw movements, this study employs LoG edge detection to extract the mouth shape (Fig. 9a), nose wrinkles (Fig. 9b), and frown lines (Fig. 9c). After the 8-directional connected components [17] are identified, the eyebrows and mouth which have larger area are identified as the larger components. Noises such as scars, moles, facial hair, and acne around the mouth area could be removed effectively.

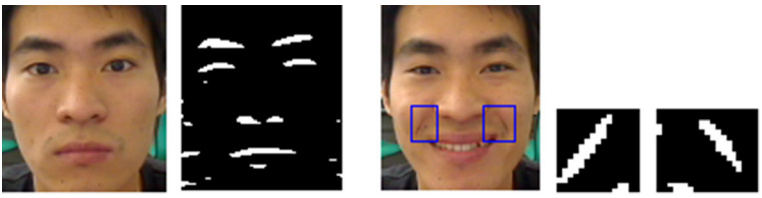

Fig. 8 Resulting images after Gabor filter using phase angle (**a**) 0° on facial components, and (**b**) 45° and 135° on both cheeks

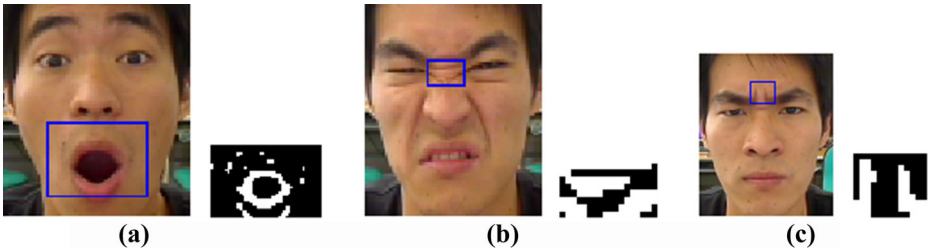

Fig. 9 LoG edge detection on dynamic ROIs. **a** Mouth ROI. **b** Nose wrinkle ROI. **c** Nasolabial ROI

4.3 Facial expression classification

A feature vector extracted from the previous module is classified by the SVM into one of the following six facial expression types (Fig. 1): 1) neutral, 2) happiness, 3) surprise, 4) anger, 5) disgust, and 6) fear. SVM is adopted as the recognition engine because of its efficiency and nonlinear classification capability, which could effectively reduce the probability of classification errors.

Cortes and Vapnik [11] developed an SVM as a machine learning system based on statistical theory. SVMs can be used to solve nonlinear and high-dimensional problems in finite samples. Supervised learning is the machine learning task of inferring a function from labeled training data. In a linear division environment, SVMs can use hyper-planes directly for classification. However, most problems arise from nonlinear division environments. To solve this problem of data classification, Cortes and Vapnik proposed using a kernel function to transform primary data at lower dimensions for forming a higher-dimension feature space to identify a linear hyperplane in that higher dimension. Thus, data points that cannot be classified using linear functions can be categorized using a hyperplane in a high-dimensional feature space. Equation (8) expresses the function for classifying data transformed in higher dimension:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i k(x_i, x) + b\right), \tag{8}$$

$$0 \le \alpha_i \le C, \quad i = 1, \ldots, n \tag{9}$$

where α_i is a Lagrange multiplier and $k(x_i, x)$ represents the kernel function of the conversion to a high dimension, C is a parameter used to model the soft margin to choose a hyperplane that splits the examples as cleanly as possible, b is the offset derived from the support vectors, and n is the number of training samples. The radial basis function expressed in Eq. (10) was selected as the kernel function for its classification capability involving nonlinear and high dimensional data. Subsequently, the derived feature vector is identified using this nonlinear SVM.

$$k(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \quad \gamma > 0 \tag{10}$$

Intuitively, the parameter γ defines how far the influence of a single training example reaches and parameter C trades off misclassification of training example against simplicity of the decision surface. Therefore, the parameters (C, γ) require adjustment. It is not known

beforehand which *(C, γ)* are good for a given problem. Consequently, a grid-search on *(C, γ)* using cross-validation is usually adopted to find the best parameters so that the classifier can accurately predict unknown data. Trying exponentially growing sequences of *(C, γ)* is often a practical method to identify good parameters (for example, $C \in \{2^{-5}, 2^{-3}, \ldots, 2^{13}, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{-13}, \ldots, 2^{1}, 2^{3}\}$. To avoid doing a complete grid-search which is time-consuming, we adopt the technique proposed by Hsu et al. [18] to search using a coarse grid first. After identifying a better region on the grid, a finer grid search on that region can be conducted. From the experimental results, $(C, \gamma)=(3, 0.0625)$ is found to have the best performance.

The dominant approach for forming a multiclass SVM is to reduce the single multiclass problem into multiple binary classification problems [13]. They distinguish (i) between one of the labels and the rest (one-versus-all) or (ii) between every pair of classes (one-versus-one). Classification of new instances for the one-versus-all case is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class. For the one-versus-one approach, classification is done by a max-wins voting strategy, in which every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally the class with the most votes determines the instance classification. For simplicity, classification of instances is done by a winner-takes-all strategy. Therefore, we have totally six SVMs for expression recognition.

5 Experimental results and analysis

The proposed facial expression recognition system was implemented on a personal computer (Intel® Core™2 Q6600, 2 GB RAM). A Logitech portable QuickCam Pro 9900 was deployed as the input device to capture 320×240-pixel images. For portability, the software was developed in C programming language using MS Visual Studio 6.0 with image processing library OpenCV 1.1 installed under environment of Microsoft Windows XP. The goal of this study is to design a real-time robust human–machine interface with facial expression recognition capabilities. The left most image in Fig. 10 shows the user interface in which the user could start or stop the system. The recognition result is indicated by the corresponding expressional graphic icon. A text box is also used to display a textual description of the facial expression recognition for verification and log.

The facial expressions of 20 persons (18–26 years old) were tested. Each person was required to perform six expressions 50 times. For the training data, 10 images per expression per person were used (totally 20×6×10=1200 feature vectors), and the remaining images were used for testing (total number of test images, 20×6×40=4800). Table 2 lists the recognition results. The recognition rate for neutral expressions was 100 %, and the average expression recognition rate reached 93.08 %. Figures 10 and 11 show some of the successful results and also the recognition errors, respectively. Figure 11a shows that most of the erroneous cases involve eyebrows that were occluded by hair thereby inhibit the feature point extraction. As shown in Fig. 11b, the “angry” expression was not identified because of variations in the eyebrows, and the frown lines were too minor even for human observation. The comparatively low rates of successfully recognizing the “disgust” (Fig. 11c) and “fear” (Fig. 11d) expressions were because these expressions were psychological emotions that differ among people.

To test the capability of the trained SVMs, another experiment was conducted using the same trained SVMs on 20 different unseen individuals during the training. By testing 20 images per expression per person, Table 3 shows the classification rates which are a little lower than those in

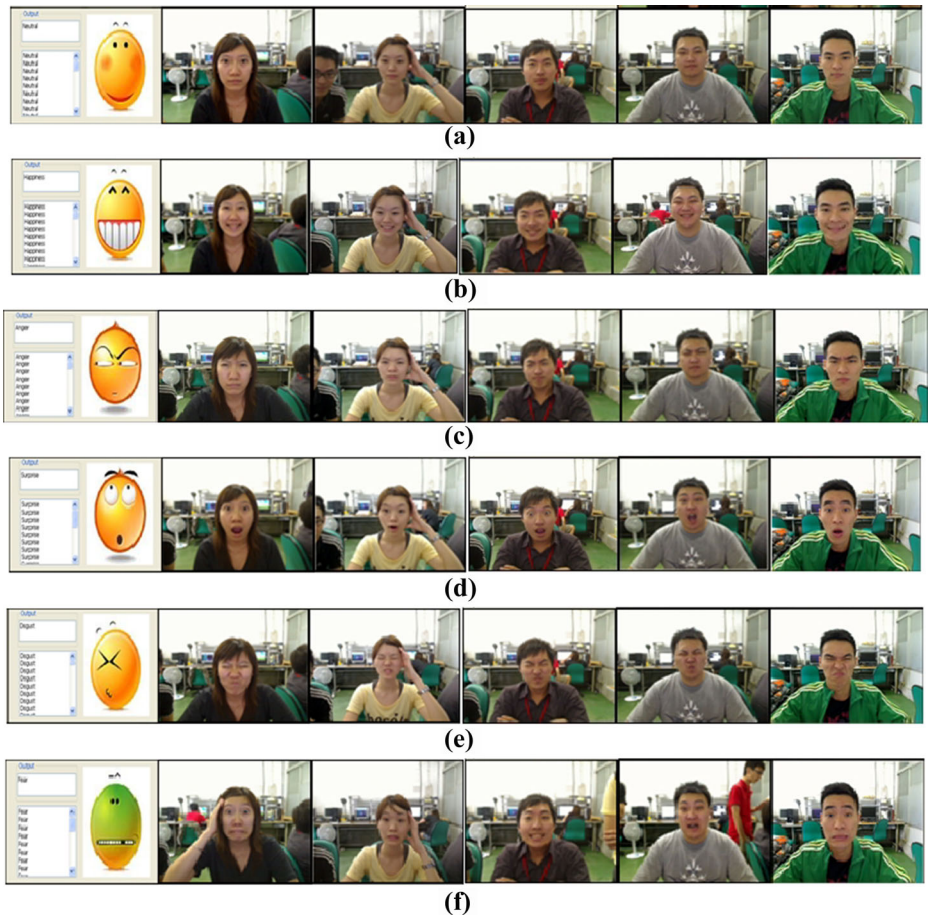

Fig. 10 Some successful results. **a** Neutral. **b** Happy. **c** Angry. **d** Surprised. **e** Disgusted. **f** Fear

Table 2. This may be due to the displayed facial expressions were not obvious or exaggerated. With regarding this issue, some finer expression changes of individuals may not be easily detected. However, one way to improve the accuracy could seek to higher resolution camera.

Table 2 Classification rates on self-collected data

Out In	N	H	S	A	D	F	RR (%)	ARR (%)
N	800	0	0	0	0	0	100	92.95
H	26	754	0	12	0	8	94.3	
S	0	0	769	0	0	31	96.1	
A	41	0	0	724	35	0	90.5	
D	45	0	0	51	704	0	88.0	
F	38	31	0	21	0	710	88.8	

N neutral, *H* happy, *S* surprised, *A* angry, *D* disgusted, *F* afraid, *RR* recognition rate, *AAR* average recognition rate

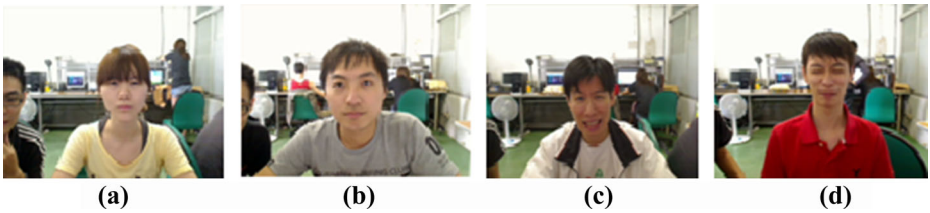

Fig. 11 Some recognition errors. **a** Eyebrows could not be identified. **b** Angry face of small variations from neutral. **c** Disgust face. **d** Scared face

This study employed the Cohn–Kanade AU-coded facial expression database [21] to verify the system performance. This database is typically used in facial expression recognition research. The database comprises 486 image sequences of 97 university students (18–30 years old), including expressions such as neutrality, happiness, surprise, anger, disgust, fear, and sadness. Each sequence begins with a neutral expression and ends with the target facial expression. Each trainee was required to perform single or compound actions. After filtering unavailable sequences and sad expressions, we retained 341 sequences as in Table 4 for the experiments. Among these, 20 images per facial expression were used for training data (20×6 feature vectors), and the remaining 221 images were used for test data. Table 5 shows examples of some of these images, and Table 6 shows the recognition results, in which the recognition rate was 100 % for neutral expression and the average recognition rate reached 94.7 %. Some errors occurred because the lighting on the face was not uniform, thereby causing incorrect image extraction. Regarding the "angry" facial expressions, some people performed minor movements with their eyebrows, and no frown lines were observed. The errors for the "afraid" facial expressions were because the eyebrows did not move.

Not only the semantic features developed are different from those proposed in previous studies but also the size of feature vector is quite smaller. Table 7 gives the comparisons based on the adopted features, size of feature vector, recognition mechanism, and recognition rates of some recent papers using SVM on the same CK database. With regarding the issue about real time processing, most surveyed approaches could be executed promptly with results. However, the size of feature vector surveyed in Table 6 ranges from 20 to 200. Due to the high level descriptive ability of proposed semantic features, ours is only 6. This alleviates the complexity of the kernel function of SVM for feature transformation and speeds up the recognition process while maintaining good recognition accuracy.

Table 3 Classification rates on unseen individuals

Out In	N	H	S	A	D	F	RR (%)	ARR (%)
N	385	5	3	7	0	0	96.25	91.0
H	20	368	0	8	0	4	92.0	
S	5	0	370	4	2	19	92.5	
A	23	0	0	355	22	0	88.75	
D	16	0	0	28	356	0	89.0	
F	21	16	0	13	0	350	87.5	

N neutral, *H* happy, *S* surprised, A angry, *D* disgusted, *F* afraid, *RR* recognition rate, *AAR* average recognition rate

Table 4 Cohn-Kanade AU-coded facial database

Types	N	H	S	A	D	F	Subtotal
Training	20	20	20	20	20	20	120
Testing	63	55	51	15	16	21	221
Subtotal	83	75	71	35	36	41	341

Due to the different software/hardware issues, the speed comparison is not easily done. We choose to do complexity analysis of SVM algorithm. To keep the computational load reasonable, the transformation kernel function used by SVM schemes are designed to ensure that dot products or differences (Eq. (10)) be computed easily in terms of the variables in the original space. Thus, the time complexity of SVM in Eq. (8) is $O(n \times m)$, where n is the number of training samples and m is the size of adopted feature vector. By reduction the size of the feature vector to k, the speed up factor is m/k. Note that the complexity referred here is for classification and the complexity of the training is different. Also, the classification time depends on the number of support vectors (with non-zero Lagrange multipliers), and this is much smaller than n. On the other hand, our accuracy is not the highest though, it still comparable to the top ones. Note that the recognition rate is not absolute because the selected test set from CK database may not be totally the same.

As to error analysis, false positive faces would occur by Adaboost based face detection. In this situation, the ASM was used to position the 75 defined facial feature points. False alarms are filtered out if the amount of deformation is greater than a given threshold. Still, from the experimental results as shown in Fig. 11, most of the erroneous cases involve eyebrows that were occluded by hair thereby inhibit the feature point extraction. In this situation, the trained SVM could only tolerate the errors to a certain degree.

6 Conclusion

Previous studies have adopted entire facial images [6, 23, 48, 50, 51] or have manually set feature points [19, 29, 34, 38] for classifying expressions. Although these methods have generated satisfactory results, they are neither rapid nor automatic. Most state of the art adopt

Table 5 Part of the facial expressions (640×490) from Cohn-Kanade database

N	H	S	A	D	F
					
					

Table 6 Classification rates on public Cohn-Kanade database

Out / In	N	H	S	A	D	F	Total	RR (%)	ARR (%)
N	63	0	0	0	0	0	63	100	94.7
H	3	52	0	0	0	0	55	94.5	
S	0	0	49	2	0	0	51	96.1	
A	1	0	0	14	0	0	15	93.3	
D	0	0	0	1	15	0	16	93.8	
F	0	1	0	1	0	19	21	90.5	

geometric features of facial components (include displacement vectors) and texture characteristic of facial components detected by edge operators like Gabor and Canny separately. Though some works integrate both kinds of features but they did on different facial regions from ours and by different classification mechanisms.

In this study, we extract semantic features including relative changes of facial components and dynamic facial textures of frown lines, nose wrinkles, and nasolabial folds for facial expression recognition. For clarity, the semantic features are 1) distance between the eyes and eyebrows, 2) mouth width, 3) frown lines between the eyebrows, 4) bunny lines between the nose and eyes, 5) left nasolabial folds, and 6) right nasolabial folds. These features include both geometrical properties (1 and 2) and facial textures (4 to 6) about dynamic facial regions. They are of higher level and more semantic than traditional primitive facial points while describing facial expressions. The recognition targets are six types of facial expression including neutral, happy, surprised, angry, disgusted, and scared.

To achieve the goals covered in [26, 33, 35, 42, 49] to be a complete robust system in real-time, this study integrates all the three modules [35] together such that subjects in images could be detected and recognized concurrently. Firstly, face detection is conducted using the robust Adaboost which is proved to be less sensitive to the lighting and illumination changes. Next, the ASM is trained to identify the human face and calibrates facial components. Subsequently,

Table 7 Comparisons with current state of the art using SVM on CK database

Recent works on CK database	Features	Feature vector size	Classification Mechanisms	Kernel fun. & and parameter selection	Average recognition rate (%)
Bartlett et al. [3]	Gabor wavelets	200	Linear SVM & AdaBoost	Polynomial	90.9 %
Tsai et al. [39]	SQI+Sobel+DCT +ART+GF	139	Nonlinear SVM	10-fold cross-validation	97.15 %
Michel & Kaliouby [28]	Facial points	22	Nonlinear SVM	RBF	87.9 %
Chen et al. [9]	Facial points and local textures	63	ASM+Nonlinear SVM	RBF	95 %
Valstar & Pantic [41]	Facial points	20	SVM & HMM	RBF	91.7
Saeed et al. [31]	Geometrical features	8	Nonlinear SVM	RBF	83.01 %
Proposed method	Semantic features	6	ASM+Nonlinear SVM	RBF and coarse to fine selection	94.7 %

Gabor filters and LoG edge detection are used to extract the semantic features from defined dynamic adaptable facial ROIs. A one-versus all multi-class non-liner SVM is then used to classify facial expressions into one of the six types of expression. From experimental results, the average classification accuracy of facial expressions was 93.08 % for the test of 20 persons from on-line video sequences. A Cohn–Kanade AU-coded facial expression database was also used to verify the system, and the average recognition rate was 94.7 %, thereby demonstrating the feasibility of the proposed system.

For all the surveyed papers, the experiments were all conducted on a standard test set and some self-recorded set. From the test set, some expressions seem exaggerated. This is because the subject is asked to perform an expression which tends to be unnatural and sometimes over exaggerated. The situation is more common in 3D data as discussed in [42] for the depth information is not fine enough. Even some experiments are done on natural data set; the problem still exists to need a standard test set that is recorded under natural environments. This is listed as one of the future work since the issue tends to capture natural facial expressions by means of someway like telling a joke is beyond our scope.

References

1. Abboud B, Davoine F, Dang M (2004) Facial expression recognition and synthesis based on an appearance model. Signal Process Image Commun 19:723–740
2. Amjad A, Griffiths A, Patwary MN (2012) Multiple face detection algorithm using colour skin modelling. Image Process IET 6(8):1093–1101
3. Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J (2006) Fully automatic facial action recognition in spontaneous behavior, 7th International Conference on Automatic Face and Gesture Recognition. 223–230, doi:10.1109/FGR.2006.55
4. Bashyal S, Venayagamoorthy GK (2008) Recognition of facial expressions using Gabor wavelets and learning vector quantization. Eng Appl Artif Intell 21:1056–1064
5. Bettadapura V (2012) Face expression recognition and analysis: the state of the art, Technical Report, arXiv: 1203.6722
6. Calder AJ, Burton AM, Miller P, Young AW, Akamatsu S (2001) A principal component analysis of facial expressions. Vis Res 41(9):1179–1208. doi:10.1016/S0042-6989(01)00002-5, ISSN 0042–6989
7. Castrillón-Santana M, Déniz-Suárez O, Hernández-Sosa D, Lorenzo J (2011) A comparison of face and facial feature detectors based on the Viola-Jones general object detection framework. Mach Vis Appl 22(3):481–494
8. Cevikalp H, Triggs B, Franc V (2013) Face and landmark detection by using cascade of classifiers, IEEE FG
9. Chen J, Chen D, Gong Y, Yu M, Zhang K, Wang L (2012) Facial expression recognition using geometric and appearance features, Proceedings of the 4th International Conference on Internet Multimedia Computing and Service (ICIMCS' 12), ACM, New York, NY, USA, 29–33, doi:10.1145/2382336.2382345
10. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models—their training and application. Comput Vis Image Underst 61:38–59
11. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
12. Daugman JG (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am 2(7):1160–1169
13. Duan KB, Keerthi SS (2005) Multiple classifier systems. LNCS 3541:278–285. doi:10.1007/11494683_28
14. Ekman P, Friesen WV (1978) The facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto
15. Fasel B, Luettin J (2003) Automatic facial expression analysis: a survey. Pattern Recogn 36:259–275
16. Fellenz WA, Taylor JG, Tsapatsoulis N, Kollias S (1999) Comparing template-based, feature-base and supervised classification of facial expressions from static images. Proc. Circuits Syst Commun Comput, pp 5331–5336
17. Gonzalez R, Woods R (1992) Digital image processing, Addison-Wesley Publishing Company, Chap. 2
18. Hsu CW, Chang CC, Lin CJ (2004) A practical guide to support vector classification, Technical Report, Department of Computer Science and Information Engineering, National Taiwan University
19. Huang C, Huang Y (1997) Facial expression recognition using model-based feature extraction and action parameters classification. Image Represent 8(3):278–290

20. Jain V, Learned-Miller E (2011) Online domain adaptation of a pretrained cascade of classifiers. IEEE Conf Comp Vis Pattern Recognit (CVPR), 577–584, 20–25
21. Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis, Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46–53
22. Kawulok M, Szymanek J (2012) Precise multi-level face detector for advanced analysis of facial images. Image Process IET 6(2):95–103
23. Kobayashi H, Hara F (1997) Facial interaction between animated 3D face robot and human beings. Proc Int Conf Syst Man Cybern, pp 3732–3737
24. Lanitis A, Taylor C, Cootes TF (2002) Automatic interpretation and coding of face images using flexible models. IEEE Trans Pattern Anal Mach Intell, 743–756
25. Lee HC, Wu CY, Lin TM (2013) Facial expression recognition using image processing techniques and neural networks, Advances in Intelligent Systems and Applications - Volume 2. Smart Innov Syst Technol Vol 21(2013):259–267
26. Lienhart R, Maydt J (2002) An extended set of Haar-like features for rapid object detection. Image Process, 900–903
27. Ma R, Wang J (2005) Automatic facial expression recognition using linear and nonlinear holistic spatial analysis. Affect Comput Intell Interaction Lect Notes Comput Sci 3784:144–151
28. Michel P, Kaliouby RE (2003) Real time facial expression recognition in video using support vector machines, Proceedings of the 5th international conference on Multimodal interfaces, 258–264
29. Moore S, Bowden R (2011) Local binary patterns for multi-view facial expression recognition. Comput Vis Image Underst 115(4):541–558
30. Ou J, Bai XB, Pei Y, Ma L, Liu W (2010) Automatic facial expression recognition using Gabor filter and expression analysis. Int Conf Comput Model Simul 2:215–218
31. Saeed A, Al-Hamadi A, Niese R, Elzobi M (2014) Frame-based facial expression recognition using geometrical features, Advances in Human-Computer Interaction, Article ID 408953, 13 pages, doi:10.1155/2014/408953
32. Sandbacha G, Zafeirioua S, Pantica M, Yin L (2012) Static and dynamic 3D facial expression recognition: a comprehensive survey. Image Vis Comput 30(10):683–697
33. Schmidt M, Schels M, Schwenker F (2010) A hidden Markov model based approach for facial expression recognition in image sequences. Artif Neural Netw Pattern Recognit Lect Notes Comput Sci 5998:149–160
34. Shin G, Chun J (2008) Spatio-temporal facial expression recognition using optical flow and HMM, Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. Stud Comp Intell 149:27–38
35. Sumathi CP, Santhanam T, Mahadevi M (2012) Automatic facial expression analysis: a survey. Int J Comput Sci Eng Survey (IJCSES) 3(6):47–59. doi:10.5121/ijcses.2012.3604
36. Surendran N, Xie S (2009) Automated facial expression recognition—an integrated approach with optical flow analysis and support vector machines. Int J Intell Syst Technol Appl 7:316–346
37. Tabbone S (1994) Detecting junctions using properties of the Laplacian of Gaussian detector. Pattern Recogn 1:52–56
38. Tang F, Deng B (2007) Facial expression recognition using AAM and local facial features. Int Conf Nat Comput, pp 632–635
39. Tsai HH, Lai YS, Zhang YC (2010) Using SVM to design facial expression recognition for shape and texture features. Int Conf Mach Learn Cybern (ICMLC) 5:2697–2704. doi:10.1109/ICMLC.2010.5580938
40. Tsao WK, Lee AJT, Liu H, Chang HW, Lin HH (2010) A data mining approach to face detection. Pattern Recogn 43(3):1039–1049
41. Valstar MF, Pantic M (2012) Fully automatic recognition of the temporal phases of facial actions. IEEE Trans Syst Man Cybern Part B: Cybern 42(1):28–43. doi:10.1109/TSMCB.2011.2163710
42. Vezzetti E, Marcolin F (2012) 3D human face description: landmarks measures and geometrical features. Image Vis Comput 30(10):698–712
43. Vezzetti E, Marcolin F (2014) 3D Landmarking in multiexpression face analysis: a preliminary study on eyebrows and mouth. Aesthetic Plastic Surg, ISSN 0364-216X
44. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. Comput Vis Pattern Recognit, 511–518
45. Vukadinovic D, Pantic M (2005) Fully automatic facial feature point detection using Gabor feature based boosted classifiers. IEEE Int Conf Syst, 1692–1698
46. Wan C, Tian Y, Chen H, Wang X (2011) Rapid face detection algorithm of color images under complex background, in Proc. 8th Int. Symp. on Neural Networks, LNCS 6676: 356–363
47. Wang Y, Ai H, Wu B, Huang C (2004) Real time facial expression recognition with AdaBoost, Pattern Recognition, Proceedings of the 17th International Conference on ICPR, 3: 926–929
48. Wu T, Bartlett MS, Movellan, JR (2010) Facial expression recognition using Gabor motion energy filters. IEEE Comput Soc Conf Comput Vis Pattern Recognit Workshops, 42–47

49. Wu T, Fu S, Yang G (2012) Survey of the facial expression recognition research. Adv Brain Inspired Cogn Syst Lect Notes Comput Sci 7366:392–402
50. Zhao G, Pietikainen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans Pattern Anal Mach Intell 29(6):915–928
51. Zhao G, Pietikäinen M (2009) Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern Recogn Lett 30(12):1117–1127
52. Zhu X, Ramanan D (2012) Face detection, pose estimation, and landmark localization in the wild, Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island
53. Zilu Y, Jingwen L, Youwei Z (2008) Facial expression recognition based on two dimensional feature extraction. Int. Conf. Software Process, pp 1440–1444

Chen-Chiung Hsieh received his B.S., M.S., and Ph.D. degrees in the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1986, 1988, and 1992, respectively. During Dec. 1992 to Jan. 2004, he was with the Institute for Information Industry (III) as a vice director. From Dec. 2004 to Jan. 2006, he joined Acer Inc. as a senior director. He is presently an Associate Professor in the Department of Computer Science and Engineering at Tatung University, Taipei, Taiwan. His research area is mainly focused in image and multimedia processing.

Mei-Hua Hsih received her M.S. degree in the Department of Business Administration, SooChow University, Taipei, Tawan, in 2010. Currently, she is a Ph.D. student in the Department of Industrial Design, Tatung University, Taipei, Taiwan. Her research interests are in the fields of customer relationship management, leisure recreation, leisure food and beverage management.

Meng-Kai Jiang received his B.S. and M. S. degrees in the Department of Computer Science and Engineering, Tatung University, in 2008 and 2010, respectively. His research interests include image processing and video surveillance.

Yun-Maw Cheng received his M.S. and Ph.D. degrees in the Department of Computer Science, University of Glasgow, UK, respectively in 1999 and 2001. Currently, he is an assistant professor in the Department of Computer science and Engineering, Tatung University, Taipei, Taiwan. My research interests lie in the general area of Human-Computer Interaction, specifically in the fields of Ubiquitous Computing, Context-Aware Computing, Emotion-Aware/Affective Computing, and Social Aspects of Computing.

En-Hui Liang received his Ph.D. degrees in the Department of Computer Science, Syracus University, New York, US, in 1991. Currently, he is an Associate Professor in the Department of Information Management, TamKang University, New Taipei City, Taiwan. His research area is mainly focused in image and multimedia processing.