International Journal of Innovative Computing, Information and Control
Volume **8**, Number **11**, November **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 7729–7746

EMOTION RECOGNITION BY A NOVEL TRIANGULAR FACIAL FEATURE EXTRACTION METHOD

Kuan-Chieh Huang[1], Yau-Hwang Kuo[1,*] and Mong-Fong Horng[2]

[1]Center for Research of E-life Digital Technology
Department of Computer Science and Information Engineering
National Cheng Kung University
No. 1, Daxue Rd., East Dist., Tainan City 701, Taiwan
kchuang@ismp.csie.ncku.edu.tw; *Corresponding author: kuoyh@ismp.csie.ncku.edu.tw

[2]Department of Electronic Engineering
National Kaohsiung University of Applied Sciences
No. 415, Chien Kung Rd., Kaohsiung 807, Taiwan
mfhorng@cc.kuas.edu.tw

Received February 2011; revised February 2012

Abstract. *Recognition of human emotions from facial expressions is highly dependent on the quality of the referred features. However, conventional methods usually were time-consuming and less robust to environmental variations. In this paper, a new triangular facial feature extraction method considering the interactions among facial features is proposed to reduce feature dimensions and avoid the effect of environmental variations as well as noisy facial features. The proposed method first constructs the face shape model of target image using a Modified Active Shape Model (MASM). Unlike traditional ASM, the landmark searching of MASM need not perform gray-level training as a pre-process. A statistical analysis and a genetic algorithm are then respectively employed to extract an optimal set of triangular facial features from the feature points of constructed face shape model. Finally, a neural network classifier is adopted to recognize emotions from the extracted triangular facial features. JAFFE database is used for performance evaluation. According to experimental results, the extracted triangular facial features are really robust to environmental variations. Besides, the triangular facial feature sets extracted by statistical analysis achieved a 65.1% recognition rate on average, which is close to the results conducted by manually selected features. However, the feature dimensions are greatly reduced from several hundreds of facial features to 21 features in the best case by the proposed approach. This reduction much benefits the computational complexity. We can further improve the average recognition rate to 73.9% with the triangular facial feature sets extracted by the proposed genetic algorithm, which is better than legacy approaches.*

Keywords: Human-computer interaction, Emotion recognition, Facial expression, Active shape model, Neural network, Genetic algorithm

1. **Introduction.** The demand on convenient and natural interface for human-computer interaction (HCI) has continued for decades since the birth of the first electronic computer in 1946. Puncher once was the dominated HCI scheme. Keyboard and mouse are the most popular interfaces for contemporary computer users. But, multi-touch is rapidly getting popular in the handheld devices. Obviously, the evolution of HCI technique is still at the rudimentary stage.

Usability, which concerns easy, reliable and efficient interactions as well as intuitive, quick and evident presentation of information, is recognized as the first issue of a successful HCI technique. It is usually evaluated by learnability, familiarity, efficiency, fault-tolerance and satisfaction. Learnability evaluates how easy users can learn to handle an HCI technique. High familiarity denotes that a user is still familiar with an HCI scheme even after a long non-use time. Efficiency concerns with how fast an HCI scheme can be. An interface for HCI will be fault-tolerant if it is robust to users' abnormal operations. A premium HCI should be satisfactory to various kinds of users. However, mouse, keyboard and tracking ball are not so efficient and natural to some kinds of users, especially kids and elders.

Existing HCI schemes are easy to catch the user's intentions for operating a computer. However, most of them cannot touch the sensation of users. User intention is known by understanding key strokes, mouse movements and clicks on touch panel, but user sensation needs more progressive technique to identify because it is tough to measure. Facial emotion recognition is the most popular and promising technique to benefit the acquisition of user sensation and then improve the quality of HCI. However, most existing emotion recognition approaches need to analyze a large number of features and therefore lead to a heavy load of computation. This situation degrades the feasibility of the HCI schemes encompassing ability of emotion recognition.

In this paper, we propose a novel triangular facial feature extraction method which can greatly reduce feature dimensions and improve the performance of emotion recognition whose overall procedure is illustrated in Figure 1. The rest of this paper is organized as follows. Related work is described in Section 2. In Section 3, a feature point detector based on MASM is presented. This detector is invoked to find out the significant feature points for constructing the face shape model of target image. Then, in Section 4, a statistical analysis approach and a genetic algorithm are respectively adopted to implement the triangular facial feature extraction procedure. The experimental results and related

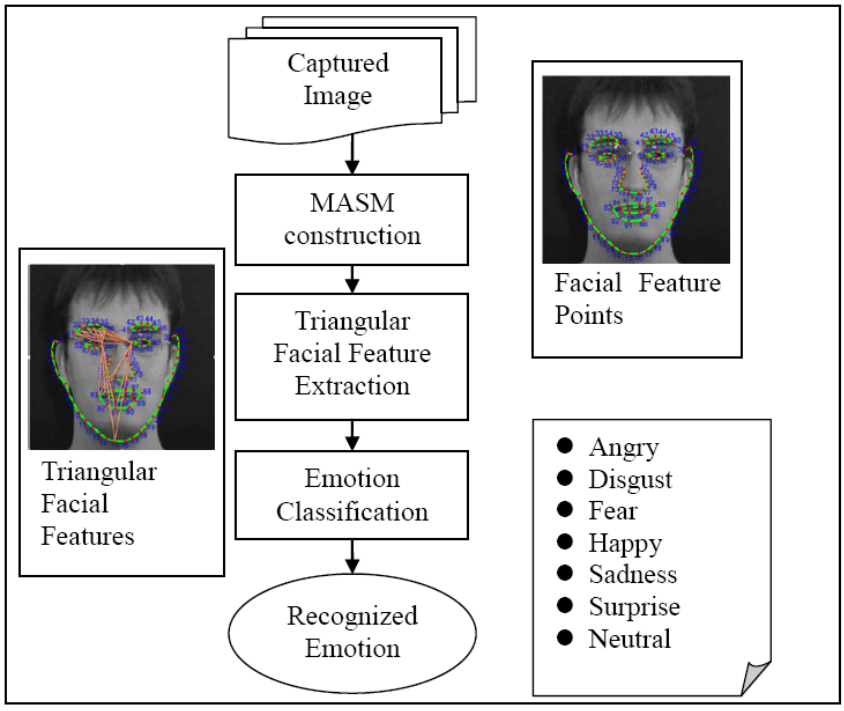


FIGURE 1. Emotion recognition process based on MASM feature detection and the triangular facial feature extraction

performance comparisons are discussed in Section 5. In our experiments, a neural network classifier is employed to recognize seven types of human emotions, including angry, disgust, fear, happy, sadness, surprise, and neutral, from the extracted triangular facial features. Finally, the concluding remarks are presented in Section 6.

2. **Related Work.** Affective communication is one of the abilities of human beings born to have, and it happens between people far earlier than speech and wording. Thus the realization of affective communication in HCI has been pursued for a long time, and the key to its success is how to recognize human emotions accurately and timely. Due to the highly personalized and vague characteristics in expressing emotions by human beings, emotion recognition is very challenging. Moreover, it has attracted lots of research interests from academics and industrials in the fields of distance learning, children education, human-robot interaction, emotion recognition by finger braille, health care, virtual reality, and pervasive computing services [1-3].

Conventionally, facial-expression recognition [4] and vocal signal approaches [5] are mainstreams of emotion recognition. Nowadays, emotion recognition by biometric information [6], such as the breath gas information [7], keyboard input rates [8], or even pressure information from the keyboard [9], was investigated as well. Another research trend is the multi-modal recognition technique which, for instance, combines audio and visual information to achieve better recognition results [13,14]. In the above recognition approaches, some paradigms such as artificial neural networks [10,13,14], Support Vector Machine (SVM) [11,15], rule-based methods [19], and K-nearest neighbors (KNN) [2] methods were widely used for recognizing emotions.

Although emotion recognition schemes become diversified, emotion recognition from facial expressions is still the most promising one. One of its major challenges is feature extraction. The well-known feature extraction methods include optical flow [18], Principal Component Analysis (PCA) [13], Gabor Wavelet Transform (GWT) [15], adaptive thresholds [16], and Active Shape Model (ASM) [21]. Appearance-based approaches using PCA or GWT usually require much computation time. On the other hand, feature-based approaches often suffer from image-processing issues like color distortion [16,19] and adaptive thresholds due to environmental variations. Furthermore, most approaches assume facial features to be non-interactive or independent, making it difficult to ensure recognition accuracy. In contrast, ASM is a statistical appearance model which considers the interactions of feature points. ASM was originally proposed to derive a model that describes the typical shapes and variability of resistors [17]. Many studies applied and extended ASM to facial expression recognition [19], medical image analysis [21] and video tracking [22]. ASM needs to be trained by related object examples. The main construction algorithm of ASM comprises four stages: shape representation, shape training, gray-level training, and landmark searching. The first stage defines the representation scheme of the target object shape as a set of feature points. According to the representation scheme, the next two stages are performed to construct a Point Distribution Model (PDM) of the object shape with its mean geometry and statistical variation obtained from a training set. The last stage is based on the PDM and the gray-level profiles constructed by gray-level training to search landmarks for identifying the accurate object shape in a new image. Among the four main stages, gray-level training is crucial to landmark searching which is the most critical. However, the gray levels of an image are easily affected by conditions of the environment, and their variations may lead to unsatisfactory feature extraction [21]. To overcome this problem, we develop the Modified Active Shape Model (ASM)-based method to detect feature points and to offer a faster landmark searching. The proposed MASM presents the same shape representation derived by the shape training schemes

with no prior training. Based on the feature points detected by MASM, we further develop a triangular facial feature extraction approach to greatly reduce feature dimensions and avoid the influence of noisy facial features.

3. **Detection of Facial Feature Points Using MASM.** In the proposed emotion recognition procedure as shown in Figure 1, the first step is to construct a MASM-based face shape model for the target image. This model will be represented with a suitable number of feature points as depicted in Figure 2. According to our experiments, 100 feature points are sufficient to construct a face shape model. However, they should be significant enough, for example, the edges and shapes of eyes, nose, lip, ear and face. In order to acquire significant MSAM construction results, we perform a MASM training procedure to derive a face shape model as illustrated in Figure 3 and explained Section 3.1. Referring to the learned shape model, the landmark searching procedure of MASM for a target face image is depicted in Figure 4 and explained in Section 3.2. This procedure features some novel ideas such as statistical searching and the tuning of search range for facial expression features extraction.

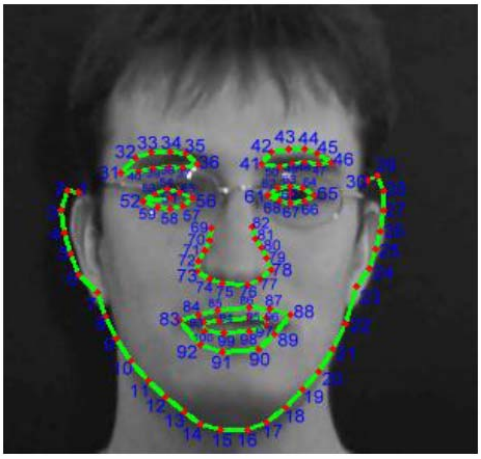


FIGURE 2. A facial image example with 100 feature points

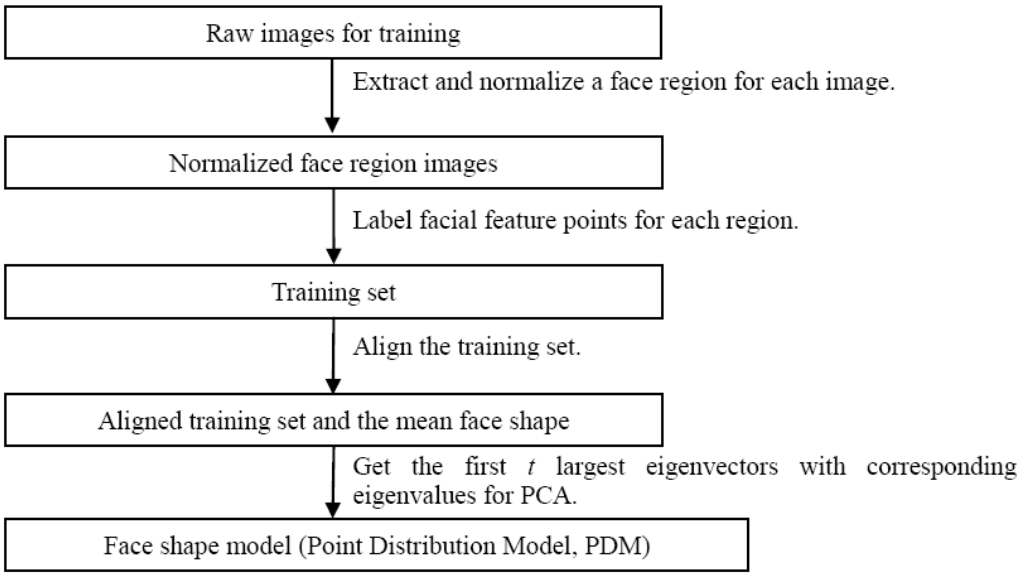


FIGURE 3. Flowchart of MASM face shape training

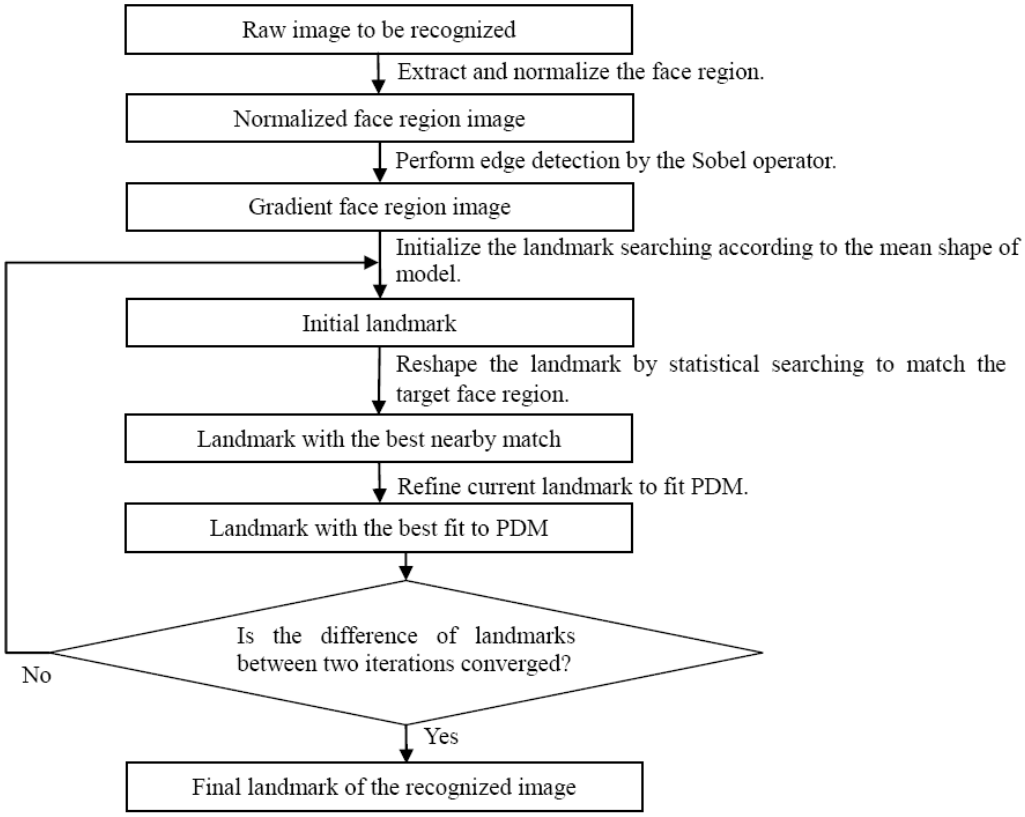


FIGURE 4. Flowchart of MASM landmark searching

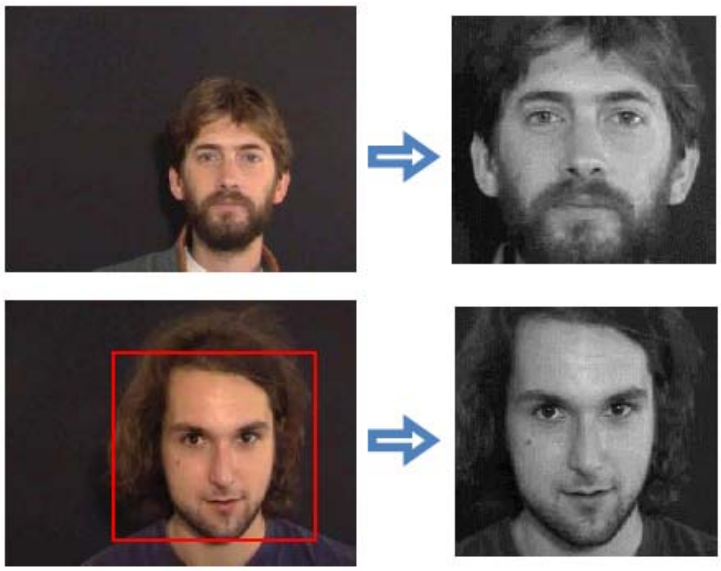


FIGURE 5. Extracting and normalizing frontal faces

3.1. **Building face shape model for emotion recognition.** In the MASM training procedure, we adopt a set of training images with different faces (100 facial expression images in our experiment) to construct the face shape model. The face region of each training image is extracted and then normalized into an image with uniform size of 300×300 pixels. Figure 5 shows the face detection and normalization results of two example images. After normalization, 100 feature points are manually selected from the face region

as a training pattern. The extraction of face region can reduce the influence of background image in face shape training, and it is implemented by Viola and Jones' method [20] to reduce the search space. The normalization can simplify later triangular facial feature extraction and avoid the effect of environmental variations, such as different distances between the camera and the target person, on feature extraction.

With the 100 face shape training patterns, we adopt the iterative approach suggested by Bookstein [21] to compute the PDM of the face shape model:

Step 1: Select a face shape training pattern as the initial mean shape.
Step 2: Align all training patterns with respect to current mean shape using the rotation and scaling operations.
Step 3: Calculate the new mean shape from those aligned patterns using the arithmetic mean operation.
Step 4: If the new mean shape is significantly different from the original mean shape, return to Step 2.
Step 5: Get the PDM of face shape using PCA.

The PDM f comprises a mean shape $\bar{f}$ with 100 feature points and its allowed distortion $P \cdot b$ where P is the first t eigenvectors of the Principal Component Analysis (PCA) for the aligned training patterns with respect to $\bar{f}$, and b is the weighting vector, also called shape parameters, for the eigenvectors. Figures 6(a)-6(d) depict an example for the MASM training process.

3.2. **Statistical landmark searching.** The landmark searching of MASM, as illustrated in Figures 6(e)-6(h), is performed to construct the face model for an incoming target image. Figure 6(e) shows the result after executing face region detection, extraction and

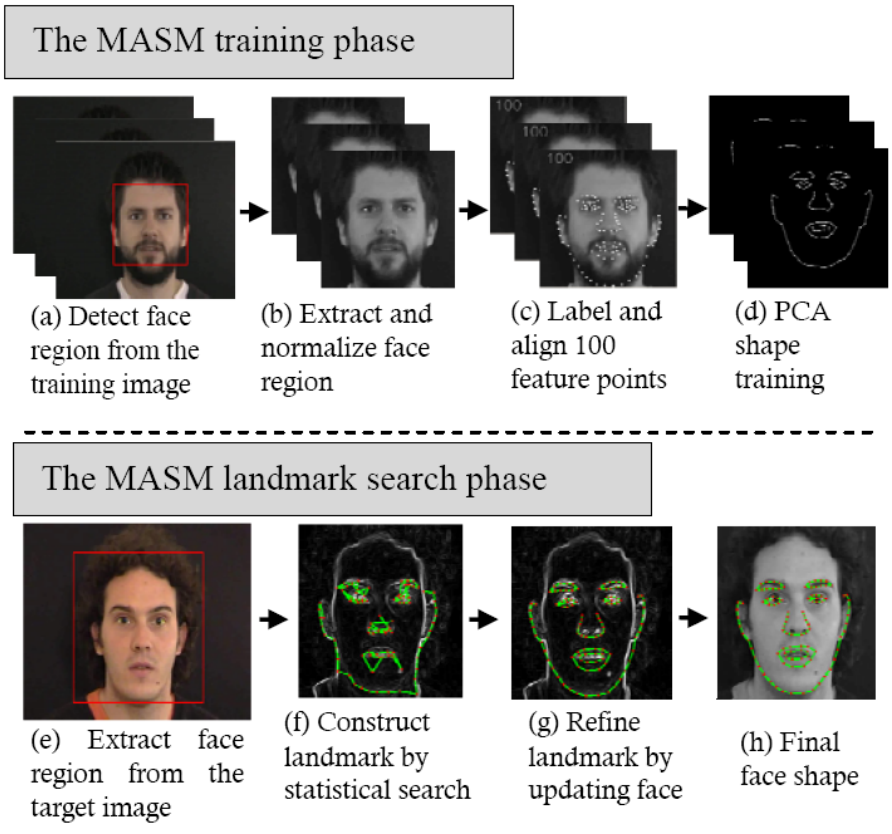
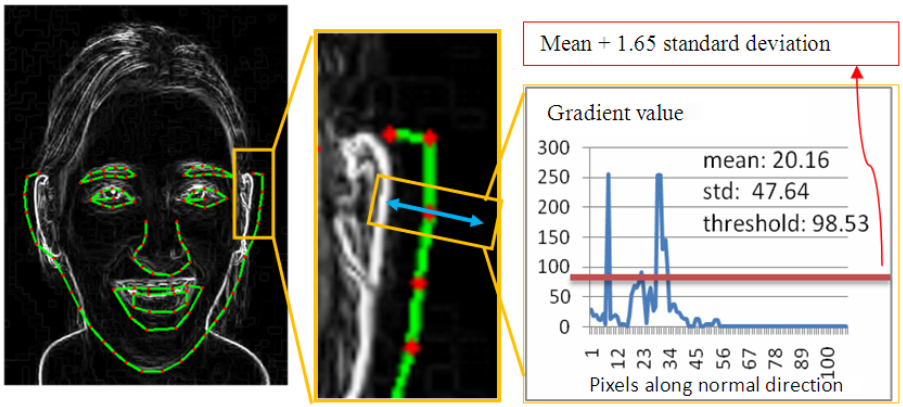


FIGURE 6. Examples of MASM face shape training and landmark searching

normalization like the shape training phase. The mean face shape $\bar{f}$ obtained in the training phase is then labeled on the normalized face region as initial landmark. Since emotions are subtle in facial features, a tiny facial feature variation may convey a different emotion. In this paper, the accuracy of labeling initial landmark is achieved by referring to four critical feature points, namely the two mouth corners and the inner endpoints of both eyebrows. To detect these four feature points, the mouth and eyebrow areas are transformed using the adaptive binary and erosion approach and the target feature points are searched according to some empirical rules. Then, we align $\bar{f}$ such that the total Euclidean distance between the detected feature points (u) and their corresponding ones ($u_{\overline{f}}$) in the aligned $\bar{f}(T(\bar{f}))$ is minimized; that is, the values of alignment parameters (X_t, Y_t, s, θ) are determined such that $|u - u_{\overline{f}}|$ is minimized. T denotes the overall shape alignment transformation with respect to the X_t and Y_t for shape translation, s for scaling, and θ for rotation.

Adopting $T(\bar{f})$ as the initial landmark, we then carry out a statistical searching procedure for creating a refined landmark more matched with the contour of the target face region, as illustrated in Figure 6(f). The search is executed on a gradient image obtained by implementing Sobel operator to the target face region, and Figure 7 illustrates the detailed procedure with an example. As can be seen, each feature point of the best nearby landmark in the gradient face region is searched along the normal direction through the corresponding feature point (vertex) in the contour of the initial landmark. This approach is different from ASM which searches for new landmark by referring to the gray-level profile obtained from the training phase. Within a specific search range, which will be discussed in next subsection, any pixel along the normal direction whose gradient exceeds the mean value with at least z times the standard deviation is selected as a candidate feature point (in this study, z is set as 1.65). Finally, the candidate nearest to the vertex is selected. The face shape expressed by the best nearby landmark may be not as smooth as the mean shape of PDM. Thus, further shape optimization is necessary to get a landmark best fitting to both the target face region and PDM, as shown in Figure 6(g). Since adjusting the shape parameters can result in various face shape expressions compliant with PDM, we choose the one (named as $B(\bar{f})$) most similar to the best nearby landmark as the best fitting landmark. In other words, the shape optimization is carried out by finding a set of shape parameters b such that $B(\bar{f}) = \overline{f} + P \cdot b$ and the total Euclidean distance between $B(\bar{f})$ and $T(\bar{f})$, $\left|B(\overline{f}) - T(\overline{f})\right|$, is minimized.



(a) Gradient face region by edge detection (b) Normal direction (c) Distribution of pixel gradients

FIGURE 7. Statistical landmark searching of MASM

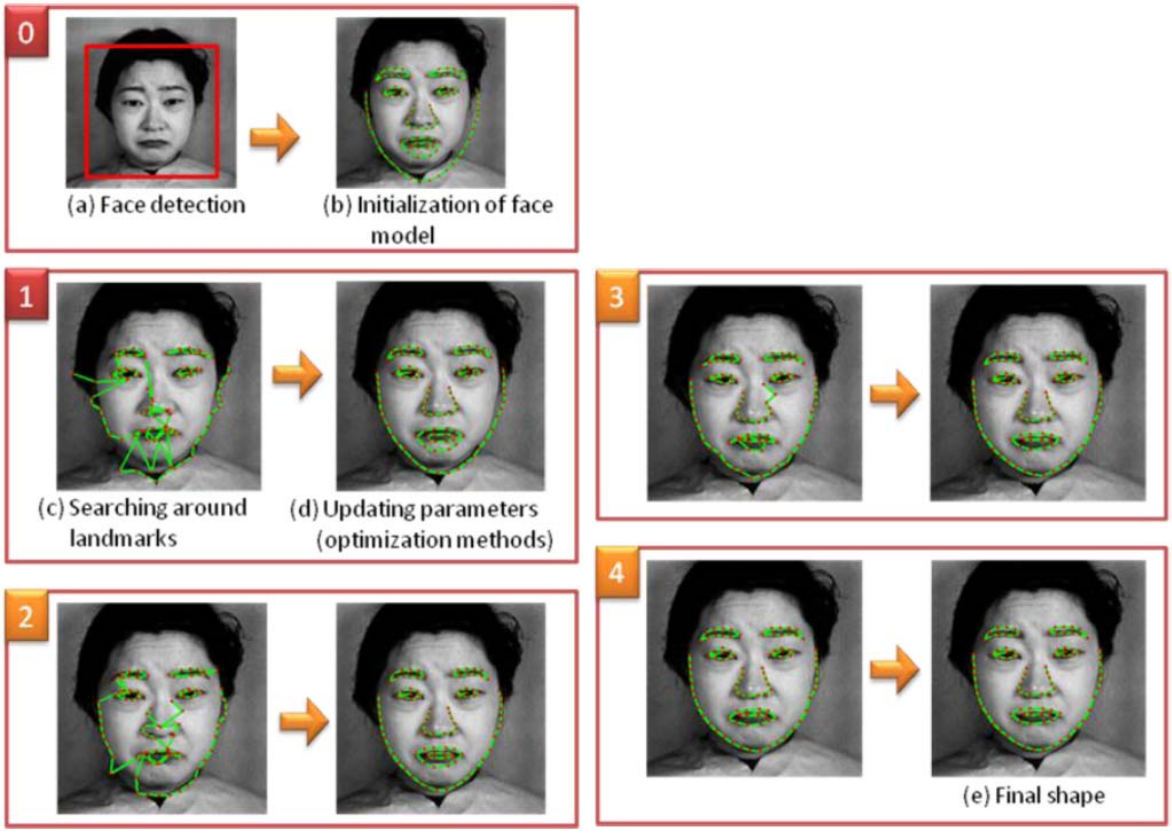


FIGURE 8. Example illustrating iterative optimization process for landmark searching

As illustrated in Figure 4, the landmark searching procedure will be repeatedly executed until the landmarks of two consecutive iterations converge. According to our experimental results, four iterations are enough for the search of the best fitting landmark. Figure 8 illustrates an example of the whole iterative optimization process.

3.3. **Adaptive search range-tuning.** In searching the best nearby landmark, we adopt an adaptive search range scheme to reduce the search time. The search range is defined as a straight-line search area along the normal direction, and the current feature point is the midpoint of the search range. To have good search performance, the initial search range is set to be greater than that of conventional ASM. Then, with increase in search cycles, the search range is decreased gradually in a linear or non-linear manner. This procedure is called the linear or non-linear range-tuning scheme as illustrated in Equations (1) and (2) respectively. In these two equations, *Range(epoch)* denotes the search range in each epoch. $Range_{\max}$ is a predefined value for the maximum search range (0.25 and 0.2 times the normalized face region for the non-linear and linear schemes respectively, in our experiment), and $Range_{\min}$ is a predefined value for the minimum search range (5 pixels for both schemes here). $Epoch_{const}$ is a predefined upper-bound of the number of search cycles.

$$Range_{linear}(epoch) = Range_{\max} - (epoch \times UnitRange_{linear}) \tag{1}$$

where $UnitRange_{linear} = (Range_{\max} - Range_{\min})/epoch_{const}$.

$$Range_{nonlinear}(epoch) = Range_{\max} \times (UnitRange_{nonlinear})^{epoch} \tag{2}$$

where $UnitRange_{nonlinear} = (Range_{\min}/Range_{\max})^{1/epoch_{const}}$.

Figure 9 depicts the linear and the non-linear search range curves in our experiment, where the search ranges of both schemes converge to $Range_{\min}$ in the 4th epoch. Thereafter, the search range is fixed to be 1 pixel for the fine-tuning of feature points.

In this paper, experiments are performed with the open database, eNTERFACE'05 [23]. First, we investigate the effect of cycle (epoch) counts and the number of eigenweights. As shown in Figures 10 and 11, the search lasting four epochs usually results in the smallest root mean square errors (RMSE) for both methods and the linear method performs better than the non-linear method. When determining the number of eigenweights, the criterion of 90% cumulative eigenvector energy is adopted. Table 1 lists the values of parameters used in our experiments.

3.4. **Comparisons between ASM and MASM.** Table 2 illustrates the algorithmic differences between ASM and MASM, and Table 3 compares their performance. As seen in the comparison, MASM performs better in landmark searching than ASM.

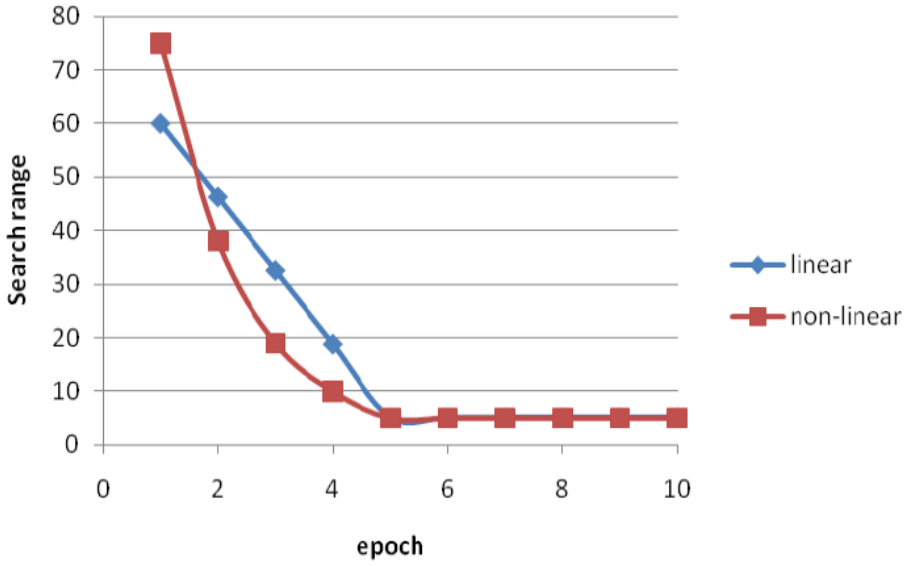


FIGURE 9. Search range curves of adaptive range-tuning methods

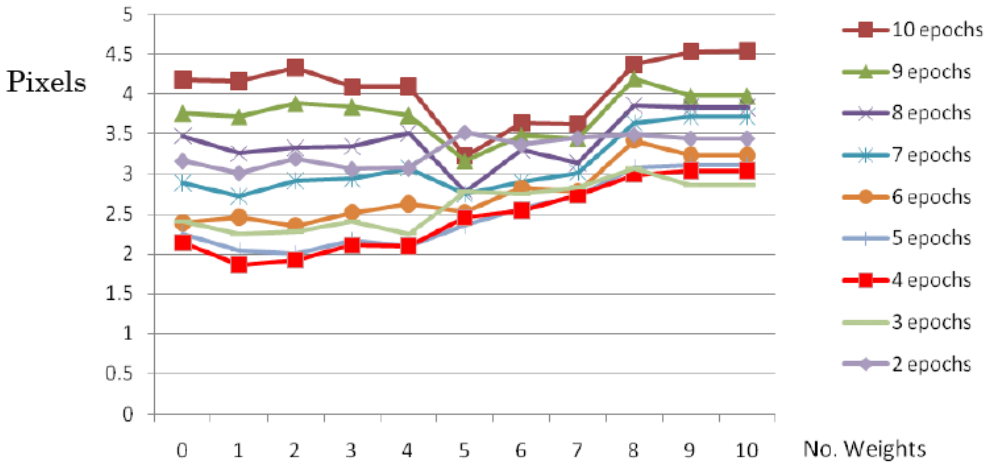


FIGURE 10. Root mean square errors of linear adaptive range-tuning

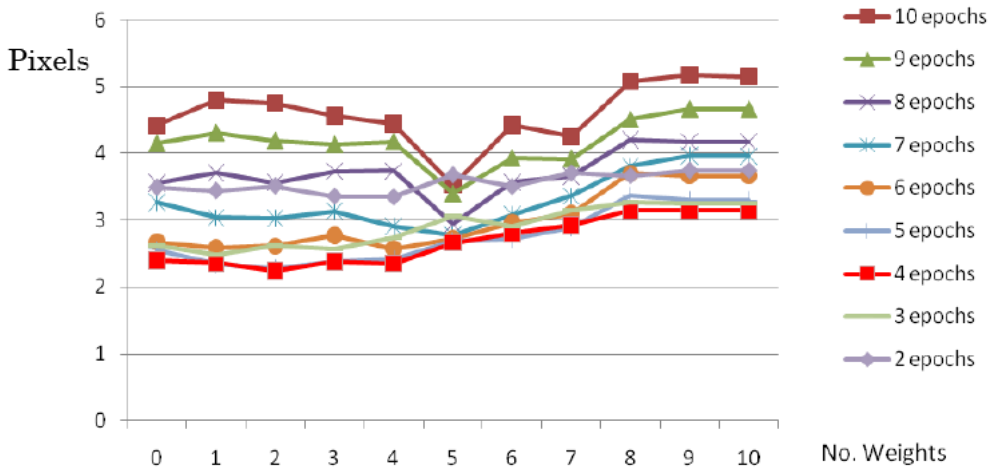


FIGURE 11. Root mean square errors of non-linear adaptive range-tuning

TABLE 1. Values of parameters for evaluating MASM performance

Parameters	**Values**
Normalized image size	300×300 pixels
Eigen weights	90% cumulative energy
Number of epochs	4
Search range	0.2 times face region
Search range-tuning scheme	Linear

TABLE 2. Algorithmic differences between ASM and MASM

	ASM	**MASM**
Gray-level training	Necessary	Not necessary
Edge detection	Not necessary	Necessary
Face detection preprocessing	No	Yes
Determination of search range	Fixed	Adaptive range-tuning
Determination of search cycles	Empirical	Empirical

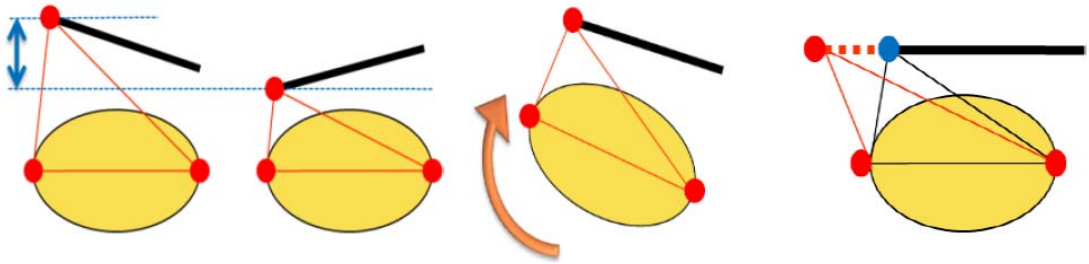
TABLE 3. Comparison of performance between ASM and MASM

	N. features	**Execution time (ms)**	**RMSE (pixels)**
ASM	68	30.27	8.104
MASM	100	43.29	7.119

4. **Emotion Recognition Using Triangular Facial Features.** Using the feature point-based face model to recognize emotions may suffer from a large number of features and the influence of noisy facial features. Therefore, in this paper, we further propose the concept of triangular facial feature for emotion recognition, and develop two triangular facial feature extraction methods which can reduce the feature dimensions and benefit the efficiency of emotion classification. The triangular features are extracted according to the result of MASM landmark searching.

4.1. **Extracting triangular facial features by statistical analysis.** Since expressions of eyebrows are crucial to emotion recognition, we propose the concept of using triangular features to distinguish between different emotions according to the features around the eyebrows. Figure 12 depicts some cases to explain the usefulness of triangular facial features. As illustrated in Figure 12(a), the height of a triangular feature is a good reference for classifying different emotions. The eye may sometimes be rotated slightly, as shown in Figure 12(b). Fortunately, the triangular feature reveals the invariant property for rotation when its shape is only slightly changed. Finally, as illustrated in Figure 12(c), even though we use MASM to extract facial features, some noises are still unavoidable. However, the property of triangular feature can overcome this problem since its area is reserved by the same height and base length of the triangle. These cases reveal that maintaining the stability of the triangle base is advantageous to the triangular feature scheme.

To select triangular facial features with a stable triangle base, we observe the characteristics of facial muscles from the images in eNTERFACE'05 and label their feature points. We then analyze statistically the positions of 102 labeled feature points for all images and their standard deviations as shown in Figure 13. In this figure, indices 1-102 denote the x-axis values of the 102 labeled feature points and indices 103-204 denote their y-axis values, where different colors represent different facial regions. The difference between 100 feature points in MASM construction and 102 feature points in triangular feature extraction will be explained later.



(a) (b) (c)

FIGURE 12. Different cases of triangular feature

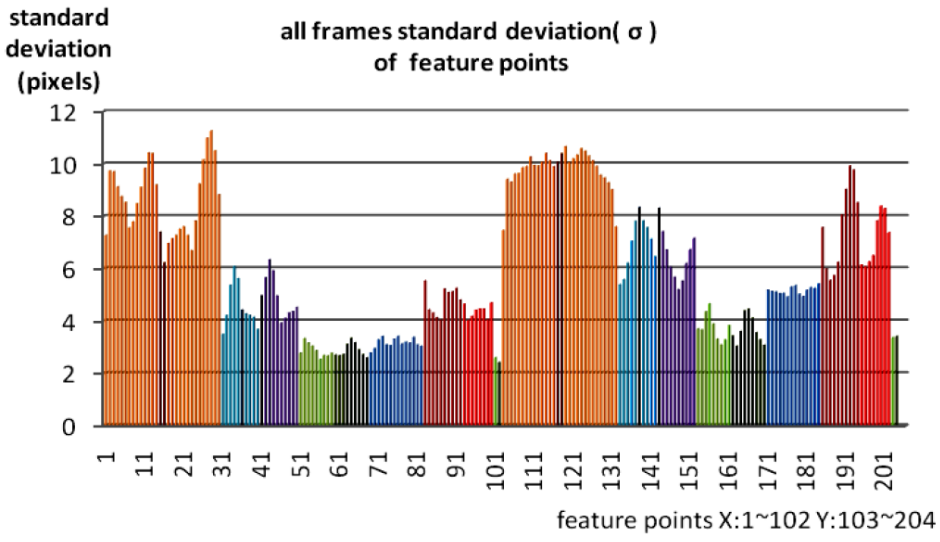


FIGURE 13. Standard deviations of positions of 102 feature points

Referring to Figure 14, we find that the inside corners of both eyes have relatively lower standard deviations. In other words, these two points, marked as 56 and 61, are relatively stable, so they are suitable for being selected as the two endpoints of the triangle base. For the features with a stable triangle base, their height in different facial images will be expressive in recognizing emotions. For example, the endpoints of both eyebrows marked as 36 and 41 in Figure 14 are distinguishable in the vertical direction. We also notice that these two points have lower x-axis standard deviation than their y-axis standard deviation. Therefore, we intend to find the feature points which are not only expressive, meaningful and differentiable in various emotions, but also stable in the same emotion. Finally, 48 triangles are selected as listed in Table 4 and some of them are shown in Figure 14. Among the triangles, some are on the horizontal triangle base $(56, 61)$ as illustrated in Figure 14(a), while others are on the vertical triangle base $(101, 102)$ as illustrated in Figure 14(b). Feature points 101 and 102 are derived features from points 56 and 61, respectively. The four points form a square with triangle bases $(56, 61)$ and $(101, 102)$ equal in length and perpendicular to each other. The reason of adding two additional feature points is to consider the expressiveness of horizontal movements in triangle features for emotion recognition. Therefore, we need a stable vertical triangle base. It is a good manner that we derive it from a stable horizontal base in an area invariant to emotions. Obviously, points 101 and 102 are the most suitable. We also consider the phenomenon that different persons may have different face size ratios. Therefore, all the 48 features

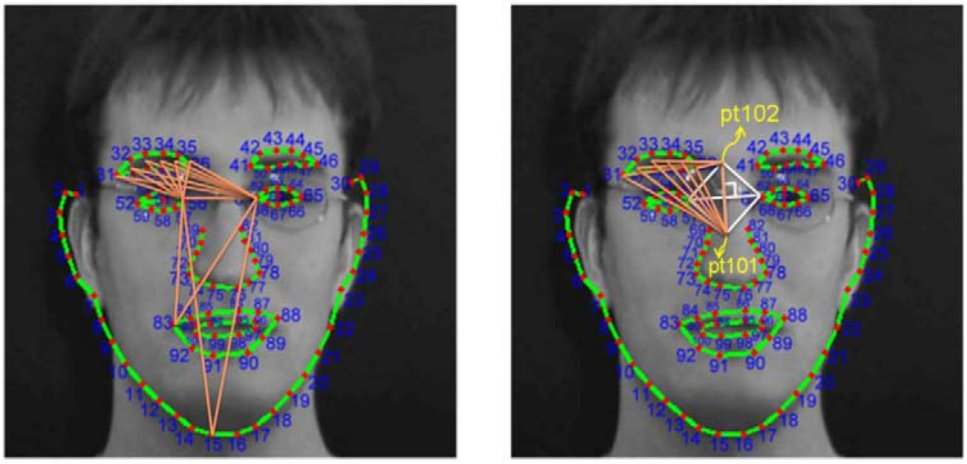


FIGURE 14. Some selected triangles by statistical analysis

TABLE 4. All 48 triangular features in the example image

1	$(56, 61, 15)$	**13**	$(56, 61, 98)$	**25**	$(101, 102, 15)$	**37**	$(101, 102, 98)$
2	$(56, 61, 16)$	**14**	$(56, 61, 99)$	**26**	$(101, 102, 16)$	**38**	$(101, 102, 99)$
3	$(56, 61, 36)$	**15**	$(56, 61, (35 + 37)/2)$	**27**	$(101, 102, 36)$	**39**	$(101, 102, (35 + 37)/2)$
4	$(56, 61, 41)$	**16**	$(56, 61, (34 + 38)/2)$	**28**	$(101, 102, 41)$	**40**	$(101, 102, (34 + 38)/2)$
5	$(56, 61, 83)$	**17**	$(56, 61, (33 + 39)/2)$	**29**	$(101, 102, 83)$	**41**	$(101, 102, (33 + 39)/2)$
6	$(56, 61, 84)$	**18**	$(56, 61, (32 + 40)/2)$	**30**	$(101, 102, 84)$	**42**	$(101, 102, (32 + 40)/2)$
7	$(56, 61, 87)$	**19**	$(56, 61, 31)$	**31**	$(101, 102, 87)$	**43**	$(101, 102, 31)$
8	$(56, 61, 88)$	**20**	$(56, 61, (42 + 50)/2)$	**32**	$(101, 102, 88)$	**44**	$(101, 102, (42 + 50)/2)$
9	$(56, 61, 90)$	**21**	$(56, 61, (43 + 49)/2)$	**33**	$(101, 102, 90)$	**45**	$(101, 102, (43 + 49)/2)$
10	$(56, 61, 91)$	**22**	$(56, 61, (44 + 48)/2)$	**34**	$(101, 102, 91)$	**46**	$(101, 102, (44 + 48)/2)$
11	$(56, 61, 94)$	**23**	$(56, 61, (45 + 47)/2)$	**35**	$(101, 102, 94)$	**47**	$(101, 102, (45 + 47)/2)$
12	$(56, 61, 95)$	**24**	$(56, 61, 46)$	**36**	$(101, 102, 95)$	**48**	$(101, 102, 46)$

TABLE 5. Baseline triangles

1	(101, 102, 56)	**3**	(56, 61, 69)	**5**	(56, 61, 81)
2	(101, 102, 61)	**4**	(56, 61, 70)	**6**	(56, 61, 82)

are divided by the average area of six baseline triangles which are listed in Table 5. These triangles are selected because they are relatively invariant for different emotions.

4.2. **Extraction of triangular facial features by genetic optimization approach.** Feature extraction by statistical analysis requires much manual effort. Here, we transform the selection of triangles into an optimization problem and solve it using a genetic optimization procedure [24,25]. For solving this discrete-type optimization problem and speeding up the optimization process, a binary-coded genetic algorithm with elite mechanism is adopted. Each chromosome consists of a set of facial feature points which are the elements for constructing triangular feature. The reason of applying elite mechanism to genetic algorithm is to keep best solution and speed up convergence in the training phase. The objective of optimization is to obtain a set of triangular features with properties that not only maximize the variances of facial expressions in different emotions of the same subject, but also minimize the variances of facial expressions for the same emotion of different subjects. We thus define the fitness function as follows.

$$\text{The Fitness function} = \frac{\sigma_{Inter}}{1 + \sigma_{Intra}} \tag{3}$$

where

$$\sigma_{Inter} = \frac{\sum_{s=1}^{M} \sum_{t=1}^{N} \sqrt{\frac{1}{E} \sum_{p=1}^{E} \left(T_t^p - \overline{T_t}\right)^2}}{M \times N} \tag{4}$$

$$\sigma_{Intra} = \frac{\sum_{p=1}^{E} \sum_{t=1}^{N} \sqrt{\frac{1}{M} \sum_{s=1}^{M} \left(T_t^s - \overline{T_t}\right)^2}}{E \times N} \tag{5}$$

In (4), σ_{Inter} denotes the variances of facial expressions in different emotions of the same subject. In (5), σ_{Intra} denotes the variances of facial expressions for the same emotion in different subjects, M is the number of subjects, E is the number of emotions, N is the number of triangular features, T_t^p denotes the triangular feature t of the emotion p, $\overline{T_t}$ denotes the average of the triangular feature t, and T_t^s denotes the triangular feature t of the subject s. In (3), 1 is added to the denominator to avoid the problem of dividing by zero.

The average fitness curve with 50000 epochs and various initial settings (e.g., random seed, crossover and mutation rate) is shown in Figure 15. The results show that the searching process can converge in early epochs by the genetic algorithm with elite mechanism. Figure 16 depicts the 21 optimal triangular facial features extracted by the genetic algorithm with the defined fitness function. Besides, to evaluate the computational cost, we implement and evaluate the proposed method under an Intel i7 2.93 GHz with 8GB RAM. The implementation needs 2353.85 seconds for training and 0.0003 seconds for recognition on average. Because the training phase needs not be run on the fly in emotion recognition, the proposed method is real-time for sure. In this work, we find that the computation for facial emotion recognition using triangular features is relatively simple but effective. Therefore, the proposed method is feasible in real-world applications.

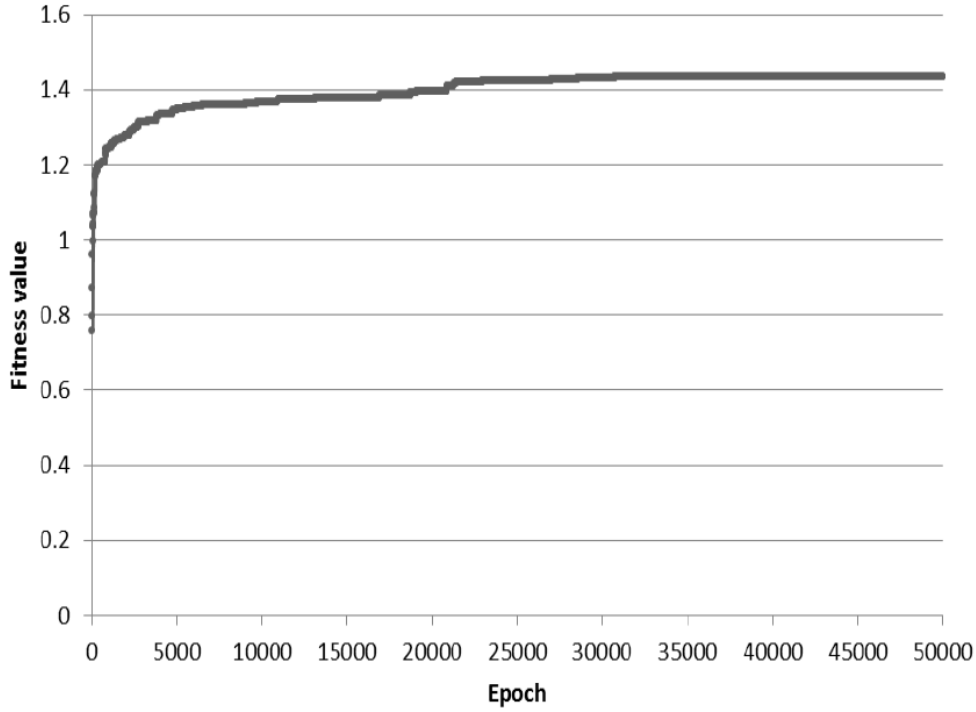


FIGURE 15. The convergence curve of searching optimal triagular feature

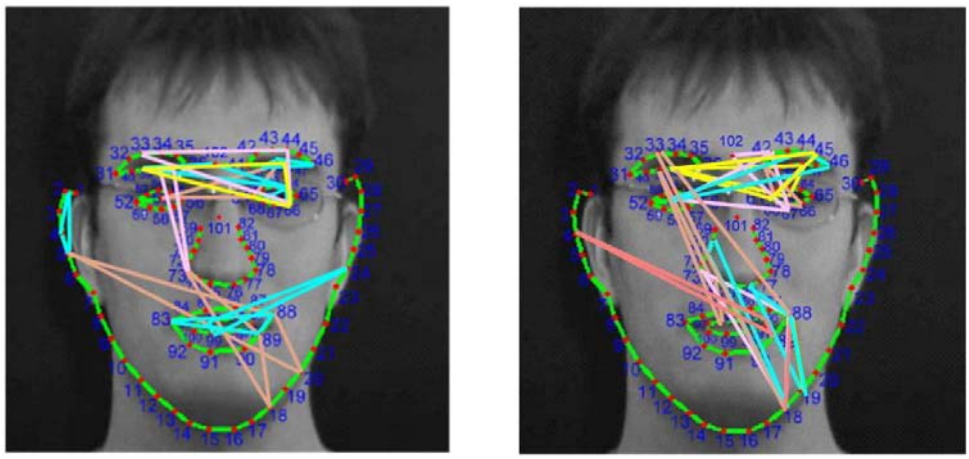


FIGURE 16. 21 triangular features extracted by the proposed genetic algorithm

5. **Experimental Results.** In this paper, both person-independent and person-dependent experiments are performed for the Japanese Female Facial Expression (JAFFE) database [26]. In the person-independent experiment, we do the cross-validation experiment; the facial expressions of nine subjects are used as training patterns while those of another subject are adopted as test patterns. A neural network with back-propagation algorithm is employed to recognize emotions using the extracted triangular facial features.

In Table 6, a comparison of the average recognition rates with some existing methods is shown. As seen, the triangular features selected by genetic optimization yield better recognition accuracy, 78.24%, than the other methods with features extracted manually. Moreover, the number of features required by the proposed method is remarkably reduced. When the MASM-based feature point detection procedure is invoked, the genetic optimization method achieves a recognition rate greater than 70%. On average, the genetic

optimization method yields a better recognition rate about 5% higher than that achieved by statistical analysis.

Tables 7 and 8 list the individual recognition rates of different subjects through statistical and genetic triangular feature selection respectively. These two tables show that different subjects express emotions diversely so it is hard to achieve similar recognition rates for them using the same feature set. However, because the proposed objective function with the genetic feature selection method considers more information about different emotions of the same subject and the variances of facial expressions for the same emotion

TABLE 6. Comparison of recognition rates for person-independent experiments

	Method		**Dim.**	**Constraint**	**Accuracy**
	Feature Extraction	**Classifier**			
Lyons [26]	Gabor wavelet + PCA	LDA	1020	34 manually selected fiducial points	75.00%
Shinohara [28]	HLAC + Fisher weight maps	LDA	1280	Face extracted manually	69.40%
Feng [29]	LBP	Coarse-to-Fine classification	19200	Pupil positions obtained manually	77.00%
Horikawa [30]	kCCA	KNN	400	Face extracted manually	67.00%
The proposed method	MASM + statistical feature selection	Neural Network	48	Face extracted by [20]	65.11%
	MASM + genetic feature selection	Neural Network	21		73.90%
	Features extracted manually + statistical feature selection	Neural Network	48	Features extracted manually	72.63%
	Features extracted manually + genetic feature selection	Neural Network	21		78.24%

TABLE 7. Individual recognition rates of subjects in JAFFE by statistical feature selection

Subject	**Features extracted manually**	**Features extracted by MASM**
1	78.26%	69.57%
2	81.82%	72.73%
3	77.27%	72.73%
4	70.00%	60.00%
5	61.90%	57.14%
6	66.67%	57.14%
7	75.00%	65.00%
8	71.43%	61.90%
9	66.67%	66.67%
10	77.27%	68.18%
Average	72.63%	65.11%

TABLE 8. Individual recognition rates of subjects in JAFFE by genetic feature selection

Subject	Features extracted manually	Features extracted by MASM
1	78.26%	75.91%
2	86.36%	81.82%
3	81.82%	75.00%
4	75.00%	70.00%
5	71.43%	69.21%
6	71.43%	69.21%
7	75.00%	73.91%
8	76.19%	73.91%
9	71.43%	68.18%
10	95.45%	81.82%
Average	78.24%	73.90%

TABLE 9. Comparison of recognition rates for person-dependent experiments

	Method		Dim.	Constraint	Accuracy
	Feature Extraction	Classifier			
Lyons [26]	Gabor wavelet + PCA	LDA	1020	34 manually selected fiducial points	92.00%
Deng [31]	Gabor wavelet + PCA + LDA	Minimum distance	7680	Facial feature points detected manually	97.30%
Liao [32]	LBP	SVM(RBF)	4096	Face extracted manually	94.59%
Horikawa [30]	kCCA	KNN	400	Face extracted manually	97.30%
Shan [33]	Boosted-LBP	SVM(RBF)	2478	Face extracted manually	81.00%
The proposed method	Features extracted by MASM + optimal feature selection	Neural Network	21	Face extracted by [20]	86.96%
	Features extracted manually + optimal feature selection	Neural Network	21	Features extracted manually	95.65%

in different subjects, it can extract more useful and general features to improve the accuracy of emotion recognition. In Table 9, a comparison for person-dependent experiment is presented. Obviously, the proposed method significantly reduces feature dimensions with respect to the other methods but still demonstrates good recognition results.

6. **Conclusion.** Emotion recognition has become increasingly important for friendly human-computer interaction. Among various studies, the recognition of human emotions from facial expressions is the new but challenging approach. However, the recognition performance is highly dependent on the quality of the referred facial expression features. In this paper, we propose a novel triangular facial feature extraction method with the advantages of (1) robustness, even in cases of feature rotation and noisy facial features and (2) significant dimension reduction in features. The proposed feature extraction method

can adopt either the statistical analysis method or the genetic optimization method to determine an optimal set of triangular facial features. Experimental results reveal its excellent accuracy in emotion recognition. An efficient approach, MASM, to automatic detection of facial feature points for the subsequent triangular feature extraction is also developed in this paper to avoid the effects of environmental variations. MASM adopts an edge-oriented feature point extraction scheme instead of the gray-level learning scheme used by ASM. The experimental results also confirm its efficiency.

Acknowledgement. The authors would like to thank the National Science Council, Taiwan, for supporting this research under contract no. NSC100-2221-E-006-251-MY3. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Q. Luo and H. Tan, Facial and speech recognition emotion in distance education system, *International Conference on Intelligent Pervasive Computing*, pp.483-486, 2007.

[2] Z. Zeng, J. Tu, M. Liu, T. S. Huang, B. Pianfetti, D. Roth and S. Levinson, Audio-visual affect recognition, *IEEE Transactions on Multimedia*, vol.9, no.2, pp.424-428, 2007.

[3] R. W. Picard, *Affective Computing*, MIT Press, 1997.

[4] S. Sonoh, K. Horio, S. Aou and T. Yamakawa, An emotional expression model inspired by the amygdala, *International Journal of Innovative Computing, Information and Control*, vol.5, no.5, pp.1147-1160, 2009.

[5] T. Pao, Y. Chen and J. Yeh, Emotion recognition and evaluation from mandarin speech signals, *International Journal of Innovative Computing, Information and Control*, vol.4, no.7, pp.1695-1709, 2008.

[6] P. S. Aleksic and A. K. Katsaggelos, Audio-visual biometrics, *Proc. of the IEEE*, vol.94, no.11, pp.2025-2044, 2006.

[7] K. Takahashi and I. Sugimoto, Feasibility of emotion recognition from breath gas information, *IEEE International Conference on Advanced Intelligent Mechatronics*, pp.625-630, 2008.

[8] G. A. Tsihrintzis, M. Virvou, E. Alepis and I. O. Stathopoulou, Towards improving visual-facial emotion recognition through use of complementary keyboard-stroke pattern information, *International Conference on Information Technology*, pp.32-37, 2008.

[9] H. R. Lv, Z. L. Lin, W. J. Yin and J. Dong, Emotion recognition on pressure sensor keyboard, *International Conference on Digital Object Identifier*, pp.1089-1092, 2008.

[10] Y. Wang and L. Guan, Recognizing human emotion from audiovisual information, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.1125-1128, 2005.

[11] M. J. Han, J. H. Hsu, K. T. Song and F. Y. Chang, A new information fusion method for SVM-based robotic audio-visual emotion recognition, *IEEE International Conference on Systems, Man and Cybernetics*, pp.2656-2661, 2007.

[12] M. Fujita, On activating human communications with pet-type robot AIBO, *Proc. of the IEEE*, vol.92, no.11, pp.1804-1813, 2004.

[13] M. H. Bindu, P. Gupta and U. S. Tiwary, Cognitive model-based emotion recognition from facial expressions for live human computer interaction, *IEEE Symposium on Computational Intelligence in Image and Signal Processing*, pp.351-356, 2007.

[14] Y. Yoshitomi, S. Kim, T. Kawano and T. Kitazoe, Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face, *IEEE International Workshop on Robot and Human Interactive Communication*, pp.178-183, 2000.

[15] I. Buciu, C. Kotrropoulus and I. Pitas, ICA and Gabor representation for facial expression recognition, *International Conference on Image Processing*, vol.3, pp.855-858, 2003.

[16] B. Scassellati, Eye finding via face detection for a foveated active vision system, *National Conference on Artificial Intelligence*, pp.969-976, 1998.

[17] T. F. Cootes, G. J. Taylor, D. Cooper and J. Graham, Active shape models – Their training and application, *Computer Vision and Image Understanding*, vol.61, no.1, pp.38-59, 1995.

[18] K. Anderson and P. W. McOwan, A real-time automated system for the recognition of human facial expressions, *IEEE Transactions on Systems, Man, and Cybernetics Part B*, vol.36, pp.96-105, 2006.

[19] N. Esau, E. Wetzel, L. Kleinjohann and B. Kleinjohann, Real-time facial expression recognition using a fuzzy emotion model, *IEEE International Conference on Fuzzy Systems*, pp.351-356, 2007.
[20] P. Viola and M. J. Jones, Rapid object detection using a boosted cascade of simple features, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.511-518, 2001.
[21] F. L. Bookstein, Landmark methods for forms without landmarks: Localizing group differences in outline shape, *Medical Image Analysis*, vol.1, no.3, pp.225-243, 1997.
[22] T. F. Cootes and C. J. Taylor, Statistical models of appearance for computer vision, *Tech. Report*, University of Manchester, 2001.
[23] O. Martin, I. Kotsia, B. Macq and I. Pitas, The eNTERFACE'05 audio-visual emotion database, *IEEE International Conference on Data Engineering Workshops*, 2006.
[24] J. Holland, *Adaption in Natural and Artificial Systems*, MIT Press, 1975.
[25] Z. Michalewicz, *Genetic Algorithm + Data Structures = Evolution Programs*, Springer-Verlag, 1996.
[26] M. J. Lyons, S. Akamatsu, M. Kamachi and J. Gyoba, Coding facial expressions with Gabor wavelets, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.200-205, 1998.
[27] M. Lyons, J. Budynek and S. Akamastu, Automatic classification of single facial images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.21, pp.1357-1362, 1999.
[28] Y. Shinohara and N. Otsu, Facial expression recognition using fisher weight maps, *IEEE Conference on Automatic Face and Gesture Recognition*, pp.499-504, 2004.
[29] X. Feng, Facial expression recognition based on local binary patterns and coarse-to-fine classification, *International Conference on Computer and Information Technology*, pp.178-183, 2004.
[30] Y. Horikawa, Facial expression recognition using KCCA with combining correlation kernels and Kansei information, *International Conference on Computation Science and Its Applications*, pp.489-498, 2007.
[31] H. B. Deng, L. W. Jin, L. X. Zhen and J. C. Huang, A new facial expression recognition method based on local Gabor filter bank and PCA plus LDA, *International Journal of Information Technology*, pp.86-96, 2005.
[32] S. Liao, W. Fan, C. S. Chung and D. Y. Yeung, Facial expression recognition using advanced local binary patterns, Tsallis entropies and global appearance features, *IEEE International Conference on Image Processing*, pp.665-668, 2006.
[33] C. Shan, S. Gong and P. W. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, *Image and Vision Computing*, pp.803-816, 2009.