International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 22, No. 3 (2008) 515–534

World Scientific
www.worldscientific.com

EXTRACTING FACES AND FACIAL FEATURES FROM COLOR IMAGES

FRANK Y. SHIH*, SHOUXIAN CHENG and CHAO-FA CHUANG

College of Computing Sciences
New Jersey Institute of Technology
Newark, NJ 07102, USA
**shih@njit.edu*

PATRICK S. P. WANG

College of Computer and Information Science
Northeastern University
Boston, MA 02115, USA

In this paper, we present image processing and pattern recognition techniques to extract human faces and facial features from color images. First, we segment a color image into skin and non-skin regions by a Gaussian skin-color model. Then, we apply mathematical morphology and region filling techniques for noise removal and hole filling. We determine whether a skin region is a face candidate by its size and shape. Principle component analysis (PCA) is used to verify face candidates. We create an ellipse model to locate eyes and mouths areas roughly, and apply the support vector machine (SVM) to classify them. Finally, we develop knowledge rules to verify eyes. Experimental results show that our algorithm achieves the accuracy rate of 96.7% in face detection and 90.0% in facial feature extraction.

Keywords: Face extraction; facial feature; mathematical morphology; principle component analysis; support vector machine.

1. Introduction

Automatic human face and facial feature extraction plays an important role in person identification in the areas of video surveillance, human computer interaction and access control.[1,2,5,18–20,24] To extract facial features we need first to detect human faces in images. Yang *et al.*[27] classified face detection methods into four categories: knowledge-based, feature invariant, template matching and appearance-based. Sung and Poggio[21] used Gaussian clusters to model the distribution of face and non-face patterns. Rowley *et al.*[14] designed a neural network-based algorithm to detect frontal-view faces in grayscale images. Schneiderman and Kanade[15,16]

applied Bayesian classifier to estimate the joint probability of local appearance and the position of face patterns. Li and Zhang[11] proposed multiview face detection algorithms using FloatBoost learning method. Wu *et al.*[26] built two fuzzy models to extract skin and hair colors using Farnsworth's perceptually uniform color system. Hsieh *et al.*[8] applied clustering-based splitting to extract face regions and performed statistics-based face verification.

Researchers have also developed techniques for locating facial features, such as eyes and mouths. Lam and Yan[10] used snake model, corner detection and cost functions to detect eyes. However, they needed to initialize the snake manually and could only handle the images with a single person or a plain background. Feng and Yuen[6] proposed eye detection using three cues: relatively low intensity of eye regions, orientation of the line connecting two eye centers, and response of convolving the eye variance filter with the face image. Huang and Wechsler[9] developed optimal wavelet packets for eye representation and Radial Basis Functions Neural Networks (RBFNN) for eye classification. Cai *et al.*[3] used template matching to locate eyes and mouths. Wang and Yuan[25] applied wavelet decomposition to detect facial features and three-layer back-propagation neural networks to classify eyes. You *et al.*[28] presented a method of face representation using non-tensor product wavelets.

In this paper, we present image processing and pattern recognition techniques to extract faces and facial features extraction in color images, which may contain many people and complex background. First, we use a color skin model to delete background and non-skin regions. Then, we use mathematical morphology, size, shape and principle component analysis (PCA) to verify face candidates. We use an ellipse model to select facial features, and apply the support vector machine (SVM) to classify eyes and mouths. Finally, we develop knowledge rules to verify them.

The rest of this paper is organized as follows. The overall design scheme is presented in Sec. 2. We describe the face extraction algorithm in Sec. 3 and the facial feature extraction algorithm in Sec. 4. We present the experimental results in Sec. 5. Finally, we draw conclusions in Sec. 6.

2. Overall Design Scheme

Figure 1 shows the outline of the overall design scheme. Firstly, we segment a color image into skin and non-skin regions by a 2-D Gaussian skin-color model. Secondly, we apply the mathematical morphology technique to remove noises and the region-filling technique to fill holes. Thirdly, we use the shape, size and PCA information to verify face candidates. Fourthly, we establish an ellipse model to locate the area of interest (AOI) of eyes and mouths. Fifthly, we apply automatic thresholding and morphological opening to discard non-facial feature pixels. Sixthly, we perform the SVM classification on the AOI of eyes and mouths. Finally, we further verify the eye and mouth candidates based on knowledge rules.

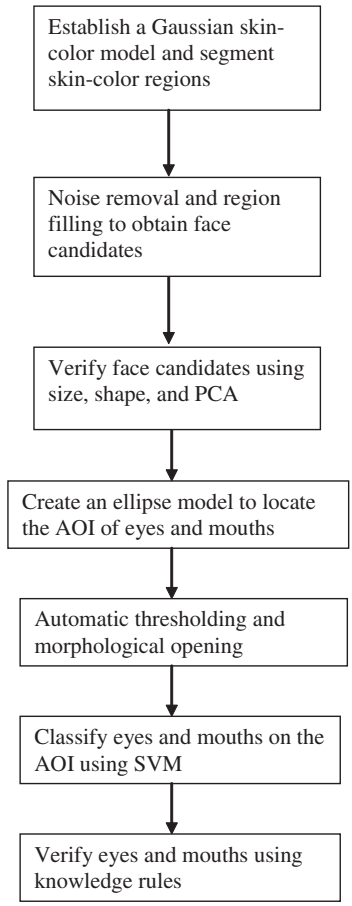


Fig. 1. Outline of the overall design scheme.

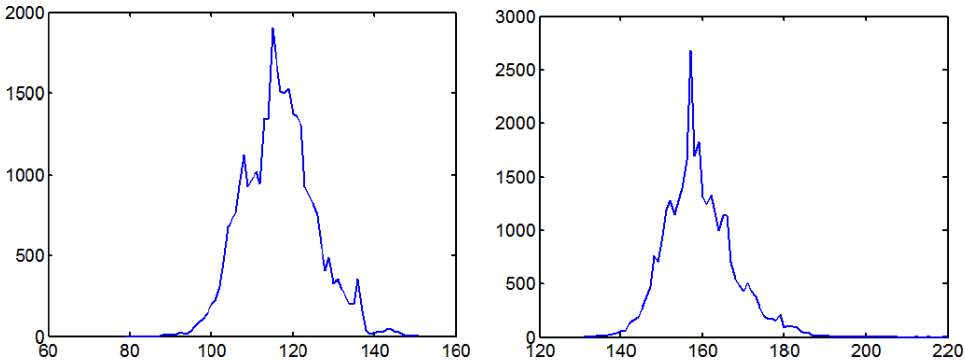
3. Face Extraction

We adopt the YC_bC_r color space for quick skin-color segmentation because it is perceptually uniform.[4] The input color image is converted into a luminance part Y and two chrominance parts C_b and C_r, where C_b represents the difference between the blue component and a reference value, and C_r represents the difference between the red component and a reference value. The Y, C_b and C_r can be obtained from RGB values in the range of $[0, 1]$ as

$$Y = 16 + 65.481 * R + 128.553^{*}G + 24.966 * B \tag{1}$$

$$C_b = 128 - 39.797 * R - 74.203 * G + 112 * B \tag{2}$$

$$C_r = 128 + 112 * R - 93.786 * G - 18.214 * B. \tag{3}$$



(a) (b)

Fig. 2. The histograms of (a) C_b and (b) C_r.

We randomly selected 80 color face images from the Internet and extracted the face skin patch of size 20×20 each to establish the 2-D Gaussian skin-color model in C_b and C_r color spaces. Figure 2 shows the histogram of overall C_b and C_r components. The parameters are calculated using the maximum likelihood method as

$$\boldsymbol{\mu} = \begin{bmatrix} \bar{C}_b \\ \bar{C}_r \end{bmatrix} = \begin{bmatrix} 116.88 \\ 158.71 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 74.19 & -43.73 \\ -43.73 & 82.76 \end{bmatrix}, \quad \rho = -0.5581, \tag{4}$$

where $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is the covariance matrix, and ρ is the correlation coefficient.

Then, for each input pixel a probability value is calculated to indicate its likelihood belonging to the skin color. The skin-likelihood probabilities for the whole image are normalized in the range of [1, 100]. In Fig. 3, (a) shows an input color image, and (b) shows the skin likelihood image. The next step is to segment the likelihood image into skin and non-skin regions. We adopt Otsu's algorithm[12] for automatic threshold selection. The threshold is chosen in order to maximize the discrimination ratio of σ_B^2/σ_w^2, where σ_B^2 is the between-class variance and σ_w^2 is the within-class variance. Figure 3(c) shows the segmented skin image.

Since the segmented skin image contains noises and holes, we must perform image cleaning. Mathematical morphology provides an effective tool in image processing and analysis.[7,17] Mathematical Morphology, based on set theory, can explore object shapes by designing variant structuring elements. The basic morphological operators include dilation, erosion, opening and closing. With an input image A, the result after region filling process, by a structuring element B, which is set to be the flat "disk" of size 5×5, we perform the morphological opening. The morphological dilation of A by B, denoted $A \oplus B$, is defined as

$$A \oplus B = \{a + b | \exists a \in A \text{ and } b \in B\}. \tag{5}$$

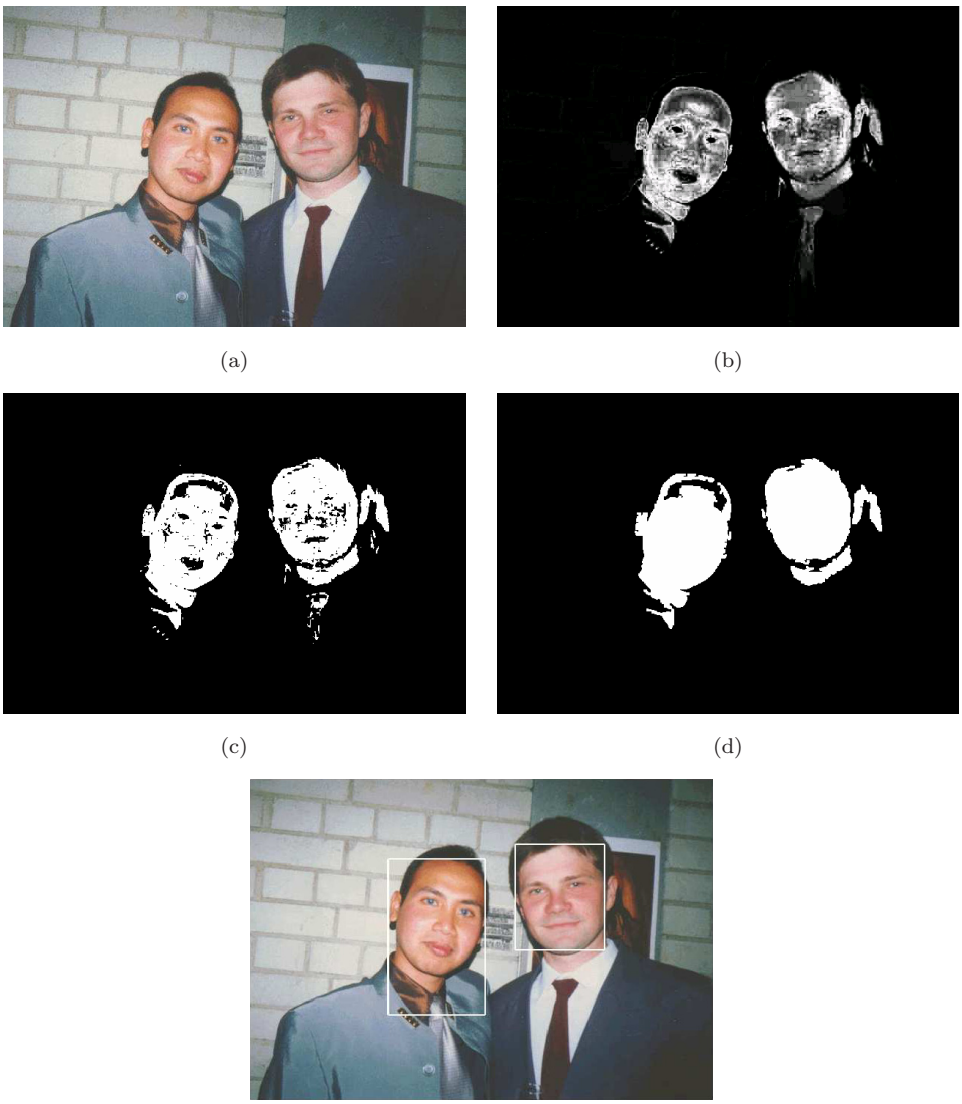


Fig. 3. (a) An input color image, (b) the skin likelihood image, (c) the segmented skin image, and (d) the image after morphological opening and region filling, and (e) the result after PCA classification.

The morphological erosion of A by B, denoted $A \ominus B$, is defined as

$$A \ominus B = \{Z | (B)_Z \subseteq A\} \tag{6}$$

where the subscript "Z" denotes a set is translated by Z. The morphological opening is defined as

$$A \circ B = (A \ominus B) \oplus B. \tag{7}$$

It is used to remove small objects from the image while preserving the shape and size of larger objects. The initial erosion removes the small details, but it also darkens the image. The subsequent dilation again increases the brightness of the image without reintroducing the details removed by erosion. We use the morphological opening to remove noises followed by a region-filling algorithm to fill holes. The result is shown in Fig. 3(d).

Since face regions occupy significant areas, we remove the connected components of small sizes. Based on experimental results, we adopt 100 pixels as the threshold value. Furthermore, we calculate the width-to-length ratio of the smallest rectangle that encloses the component. If the ratio is in the range of [0.6, 2.2], the component is taken as a face candidate. Sometimes, the face candidates are connected to other skin regions; for example, in Fig. 4, the woman's face is connected to hands. In order to extract face regions correctly, we verify multiple squared windows of different sizes determined by the size and shape of the skin region. The size of the first window is chosen as the small value of the maximum width and the maximum height. The maximum width is defined as the maximum number of connected pixels in the horizontal direction. The maximum height is defined as the maximum number of connected pixels in the vertical direction. The interval of the window size is set to be 10%; i.e. the size of the next window is 90% of the previous window. The searching processes continue until the size reaches the small value of the minimum width and the minimum height or reaches 10 if the value is less than 10. In addition, for each face candidate the ratio of the area of the skin component to the area of the rectangle is calculated. We remove the face candidate if the ratio is smaller than a reasonable percentage (we use 55% according to experiments).

The color components C_b and C_r are used only once in the 2-D Gaussian skin model for skin classification. After that, we simply process grayscale images. We use eigenfaces to represent face features because the face images do not change radically in projection onto face space, whereas the projection of non-face images appears quite different.[22]

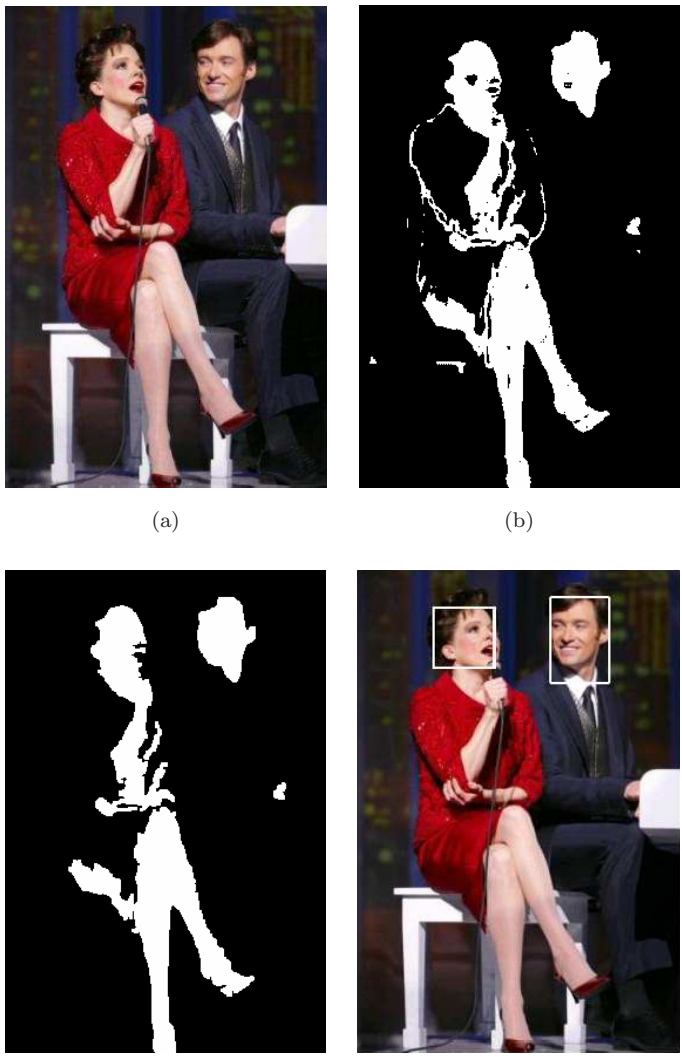
For training, a total of 450 faces of size 32×32 representing 50 people with nine poses ($+60°$, $+40°$, $+25°$, $+15°$, $0°$, $-15°$, $-25°$, $-40°$, $-60°$) per person are used. Figure 5 shows an example of a face image with nine poses. Let each of 450 face vectors be $\mathbf{x}_i$ ($i = 1, 2, \ldots, 450$) of dimension 1024×1. We first calculate the mean and the standard deviation, and then normalize each vector to be mean 0 and variance 1 as

$$\frac{\mathbf{x}_i - \text{mean}(\mathbf{x}_i)}{\text{std}(\mathbf{x}_i)}, \tag{8}$$

where $\text{mean}(\mathbf{x}_i)$ indicates the mean value of all components of vector $\mathbf{x}_i$, and $\text{std}(\mathbf{x}_i)$ indicates the standard deviation of all components of vector $\mathbf{x}_i$.

We calculate the mean of all normalized vectors as

$$\bar{\mathbf{x}} = \frac{1}{450} \sum_{i=1}^{450} \mathbf{x}_i. \tag{9}$$



(a) (b)

(c) (d)

Fig. 4. (a) An input color image, (b) the skin likelihood image, (c) the image after morphological opening and region filling, and (d) the result after PCA classification.

Fig. 5. An example of a face image with nine poses.

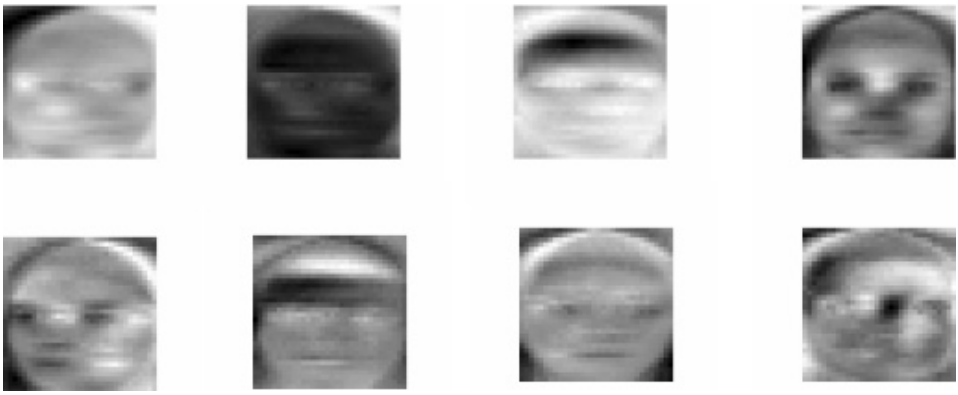


Fig. 6. Examples of eigenfaces.

We calculate the covariance matrix of all normalized vectors as

$$\mathbf{Y} = \sum_{i=1}^{450} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}}, \tag{10}$$

where "T" denotes the transpose of a matrix. Then, we calculate the eigenvectors and eigenvalues of $\mathbf{Y}$.

We construct a matrix $\mathbf{U}$ of size 1024×20 which contains the first 20 most important eigenvectors. These eigenvectors represent the most face-like images. Some eigenfaces are shown in Fig. 6. Each face, as a vector of dimension 20×1, is calculated by

$$\text{face}\{i\} = \mathbf{U}^{\mathbf{T}}(\mathbf{x}_i - \overline{\mathbf{x}})\,. \tag{11}$$

To test a face candidate image, we obtain its PCA representation by Eq. (11). For face candidates of different sizes, we normalize them into 32×32. We calculate the distances between the face candidate and each of the 450 training faces. If the minimum distance is larger than a threshold (we use 90 based on experimental results), the image is removed. Therefore, some false positives can be eliminated. Figures 3(e) and 4(d) show the face detection results after the PCA classification.

4. Facial Feature Extraction

After face images are extracted, we perform the detection of facial features, such as eyes and mouths. To locate facial features and eliminate false positives quickly, we develop an ellipse model to fit in the face. Let w denote the width of the face image extracted from the face detection algorithm. An ellipse is inserted in the face image by using w as the minor axis and $h = 1.55\,w$ as the major axis. We use the eye rectangle whose horizontal lines are located in between $\frac{2}{10}\,h$ and $\frac{9}{20}\,h$ away from the top of the ellipse, as illustrated in Fig. 7. Similarly, we use the mouth

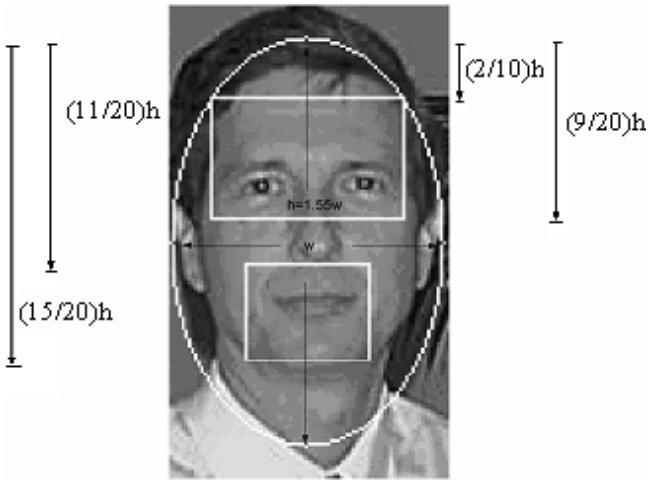


Fig. 7. The ellipse and the rectangles of facial features.

rectangle whose horizontal lines are located in between $\frac{11}{20}\,h$ and $\frac{15}{20}\,h$. By connecting two eyes and the mouth, we are able to form a triangle to enclose the nose. Note that by comparing to eyes and mouth, the nose plays a less important role in face identification.

After extracting the rectangles of facial features, we apply automatic thresholding and morphological opening to discard non-facial feature pixels. We observe that facial features are darker than non-facial features. Therefore, we remove non-facial feature pixels to speed up the succeeding SVM[23] verification and decrease the existence of false positives. Figure 8 shows the resulting facial feature pixels.

We adopt the FERET face database[13] to train eyes and mouths in the SVM. It contains different pose variations from $-90°$ (profile left) to $90°$ (profile right). In experiments, we select face images from $-25°$ to $25°$ as our training set. Figure 9 shows some images of different pose variations in our training set.

The SVM performs pattern classification by applying a variety of kernel functions as the possible sets of approximating functions, optimizing the dual quadratic programming problem, and applying structural risk minimization as the inductive principle.[23] Different types of SVM classifiers are used depending upon input patterns. A linear maximal margin classifier is used for linearly separable data, a linear soft margin classifier is used for overlapping classes, and a nonlinear classifier is used for classes that are overlapped as well as separable by nonlinear hyperplanes. We adopt the linearly nonseparable SVM as our classification tool for facial feature classification since it can produce the best result.

Suppose that we have a set of training data, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. Each $\mathbf{x}_i$ belongs to one of two classes $y_i \in \{-1, 1\}$. Our objective is to separate the training data into two classes with a minimal error. To accomplish this, we use non-negative slack variables ξ_i, $i = 1, 2, \ldots, 4000$. Thus, the two-class input data can be constructed

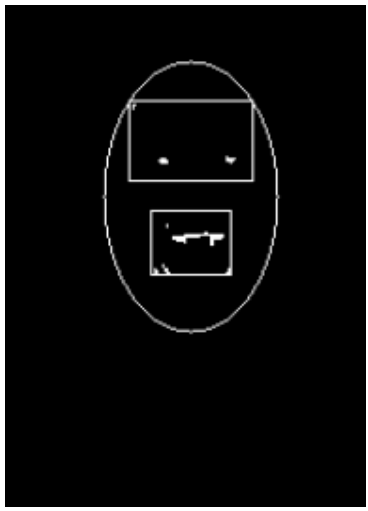


Fig. 8. The resulting facial feature pixels.

Fig. 9. The training set of different pose variations.

using Eqs. (12) and (13), where $\mathbf{w}$ denotes a vector and b denotes a scalar.

$$\begin{cases} \mathbf{w} \cdot \mathbf{x_i} + b \geq 1 - \xi_i & \text{if } y_i = 1, \\ \mathbf{w} \cdot \mathbf{x_i} + b \leq -1 + \xi_i & \text{if } y_i = -1. \end{cases} \tag{12}$$

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad i = 1, 2, \ldots, 4000. \tag{13}$$

Obtaining the soft margin hyperplane in the linear nonseparable case requires that the following minimization problem be solved:

$$\text{Minimize:} \quad \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C\left(\sum_{i=1}^{4000} \xi_i\right),$$

$$\text{Subject to:} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, 4000$$

$$\xi_i \geq 0, \quad i = 1, 2, \ldots, 4000$$

where C denotes a penalty or regularization parameter.

Minimizing the Lagrangian function with respect to $\mathbf{w}, b$, and ξ_i can be transformed to its dual problem as follows:

$$\text{Maximize:} \quad \sum_{i=1}^{4000} \alpha_i - \frac{1}{2} \sum_{i=1}^{4000} \sum_{j=1}^{4000} \alpha_i \alpha_j y_i y_j \mathrm{x}_i \mathrm{x}_j$$

$$\text{Subject to:} \quad \sum_{i=1}^{4000} y_i \alpha_i = 0, \quad 0 \le \alpha_i \le C.$$

By solving the dual problem, we obtain the soft margin hyperplane by

$$\mathbf{w}_0 = \sum_{i=1}^{4000} \alpha_i y_i \mathbf{x_i} \tag{14}$$

$$b_0 = y_i - \mathbf{w}_0 \cdot \mathbf{x}_i. \tag{15}$$

From Eq. (14), we obtain the hyperplane by all the training data $\mathbf{x}_i$ that have the corresponding attributes of $\alpha_i > 0$. These support vectors can be divided into two categories. The first category has the attribute of $\alpha_i < \mathrm{C}$ and $\xi_i = 0$. These support vectors lie at the distance $\frac{1}{\|\mathrm{W}\|}$ from the optimal separating hyperplane. The second category has the attribute of $\alpha_i = \mathrm{C}$. The support vectors are either correctly classified with a distance smaller than $\frac{1}{\|\mathrm{W}\|}$ from the optimal separating hyperplane if $0 < \xi_i \le 1$, or misclassified if $\xi_i > 1$.

In the SVM training of eye classification, we select 800 eye and 10,000 noneye images from the FERET database.[13] These images are of size 19×39 which is approximately the ratio $1:2$ in terms of height-to-width. To determine the number of training samples which is most suitable for eye classification, we test five cases: 3000, 4000, 6000, 8000 and 10,000 samples. From experiments, we obtain the best eye classification rate of 99.85% by using 4000 samples.

The training and testing images are of size 19×39, which is equivalent to 741 dimensions in classification. We take advantage of PCA to reduce the dimensionality. To determine the number of eigenvalues, we test seven cases: 20, 30, 40, 50, 60, 70 and 80 eigenvalues. From experimental results, we conclude that 60 eigenvalues are the best trade-off in between computational time and accuracy rate.

Furthermore, we test different degrees of polynomial kernel functions: 2, 3 and 4 using 60 eigenvalues on 4000 samples, and obtain accuracy rates respectively as 98.42%, 98.48% and 98.51%. From results, we observe that the linear SVM (degree of 2) performs almost the same as nonlinear SVMs. Therefore, we choose the linear SVM to save computational time.

For the SVM training in mouth classification, we select 400 mouth and 10,000 nonmouth images from the FERET database. These images are of size 31×62, which is approximately the ratio $1:2$ in terms of height-to-width. We also test five cases: 3000, 4000, 6000, 8000 and 10,000 samples. From experiments, we obtain the best eye classification rate of 96.50% using 4000 samples.

We use PCA to reduce the dimensionality and test seven cases: 20, 30, 40, 50, 60, 70 and 80 eigenvalues. From results, we conclude that using 70 eigenvalues can

reduce computational time and maintain the classification rate of 95.50%. Furthermore, we test different degrees of polynomial kernel functions and conclude that the linear SVM classifier performs the best accuracy rate.

For mouth classification, we choose four different sizes due to the size variation of facial features in a face image. Let the width of the face rectangle be w. We use the four widths of the facial feature rectangle as $\frac{1}{3}w$, $\frac{1}{4}w$, $\frac{1}{5}w$ and $\frac{1}{6}w$. The height is approximately half of its width. To allow tolerance, we include vertical shift of 2-pixel up and 2-pixel down of the rectangle.

Furthermore, we apply the following six rules to eliminate false positives:

(1) From the four sizes used in each eye and mouth, we choose the one having the largest occurrence. For example, if two of $\frac{1}{3}w$, one of $\frac{1}{5}w$, and one of $\frac{1}{6}w$ are detected as eyes, we choose $\frac{1}{3}w$ as the eye size.
(2) Both eyes have similar gray levels. We calculate the average intensity from a set of eye samples. If the intensity of an eye candidate is within $\pm 30\%$ of the average intensity, it is considered to be a real eye.
(3) The distance between both eyes is within a certain range of head width. It should be within the range of 30% to 80% of the width of the face rectangle.
(4) The line connecting both eyes is nearly horizontal (i.e. the line slope is within $\pm 15°$).
(5) The mouth is located vertically in between two eyes.
(6) The three lines connecting two eyes and a mouth form an approximate isosceles triangle.

Figure 10 shows some results of facial feature extraction.

5. Experimental Results

We randomly selected 213 color images from the Internet. They consist of one hundred and ten 1-person, forty-six 2-person, twenty-six 3-person, and thirty-one 4-or-more-person images. These images include outdoor and indoor scenery, shadow and lighting variations. The background varies from uniform wall to complex natural scene. The faces appear at different positions with different sizes in images. Some faces are with poses up to 45° and orientations up to 15°. The facial expressions include smile, angry, sad and surprise. Some images showing these variations are given in Fig. 11.

Table 1 summarizes the face extraction results. The number of *Correct Detections* (CDs) is based on 213 color images, which contain 486 faces with variations in size, pose, expression and illumination. The *Detection Rate* (DR) is defined as the ratio of the number of correctly detected faces to the total number of faces in all the images. The *False Positive Rate* (FPR) is defined as the ratio of the number of detected false positives to the total number of faces. The *False Negative Rate* (FNR) is defined as the ratio of the number of false negatives to the total number of faces. We obtain the face detection rate 96.7% over a total of 486 faces

Fig. 10. Some results of facial feature extraction.

with FPR 7.4% and FNR 3.3%. Figure 11 shows the detected results in different cases: (a) many faces in an image, (b) faces of different sizes, (c) and (d) faces with pose and orientation variations, and (e) a shadowed face. Note that these faces also present different expressions.

Table 2 lists the comparison of our method with five other methods. The CMU and MIT methods used grayscale images. The method in Ref. 8 used color images, but their testing data were unavailable. Therefore, we created the testing database by randomly downloading color images from the Internet. Note that our method need not search for face candidates at all locations with different sizes as the methods in Refs. 14–16 and 21. As compared to Ref. 8, our method could detect faces connected with other skin-like regions as demonstrated in Fig. 4. The CMU face detection system is available at http://www.vasc.ri.cmu.edu/cgi-bin/demos/findface.cgi. We sent grayscale images to the CMU system for processing and obtained their results as shown in Fig. 12, where some false positives in nonskin regions were present.

Table 3 summarizes the results of eye and mouth detection. The accuracy rate is obtained by using the ratio of the total number of correctly detected eyes and mouths to the total number of eyes and mouths, which is equal to 1269 : 1410.

We implemented our algorithms using the Matlab language on Sun Blade1000 workstation under Solaris 5.8. The CPU used is Sparcv9+vis2. The running time

Fig. 11. Some face extraction examples show variations in face images.

Table 1. Face extraction results.

	No. of Images	No. of Faces	No. of CD	No. of FP	No. of FN	DR (%)	FPR (%)	FNR (%)
1-person	110	110	107	7	3	97.3	6.4	2.7
2-person	46	92	90	10	2	97.8	10.9	2.2
3-person	26	78	73	9	5	93.6	11.5	6.4
Others	31	206	200	10	6	97.1	4.8	2.9
Total	213	486	470	36	16	96.7	7.4	3.3

CD: Correct Detection, FP: False Positive, FN: False Negative, DR: Detection Rate, FPR: False Positive Rate, FNR: False Negative Rate.

Table 2. Performance comparisons of our method and five other methods.

System	Detection Rate	False Positives	Test set
MIT Distribution-based[21]	81.9%	8.4%	23 grayscale images containing 155 frontal faces.
CMU Neural Network[14]	84.5%	5.2%	23 grayscale images containing 155 frontal faces.
CMU Bayes[15]	91.2%	8.8%	20 grayscale images containing 136 faces.
CMU Bayes[16]	85.5%	20.63%	208 grayscale images containing 441 faces.
Color Clustering[8]	70.3%	6.6%	60 color images containing 118 faces.
Our System	96.7%	7.4%	213 color images containing 486 faces.

depends on the image size and the number of people in an image. For instance, the running time for a 4-person image of size 210×172 is 5.0 s for face extraction and 9.3 s for eye and mouth extraction.

There could be some situations that our face detection algorithm may fail. First, some face region might not be detected as a skin region due to its color distortion caused by lighting or shadowing. Second, some part of background might be detected as a skin region. Third, if two or more faces are too close to be merged into a connected component, the error may occur. In Fig. 13, (a) shows that a face is missed due to its face region being detected as a nonskin region, and (b) shows that the hand region is detected as a false positive.

For facial feature extraction, our eye and mouth extraction algorithm works well on 1-person images (i.e. the detection rate of 95.3%). Our algorithm can detect eyes and mouths on tilted heads. The detection rate decreases as the number of persons increases in an image. Other factors causing misclassification may be that the eyes are closed, too small or too much tilted.

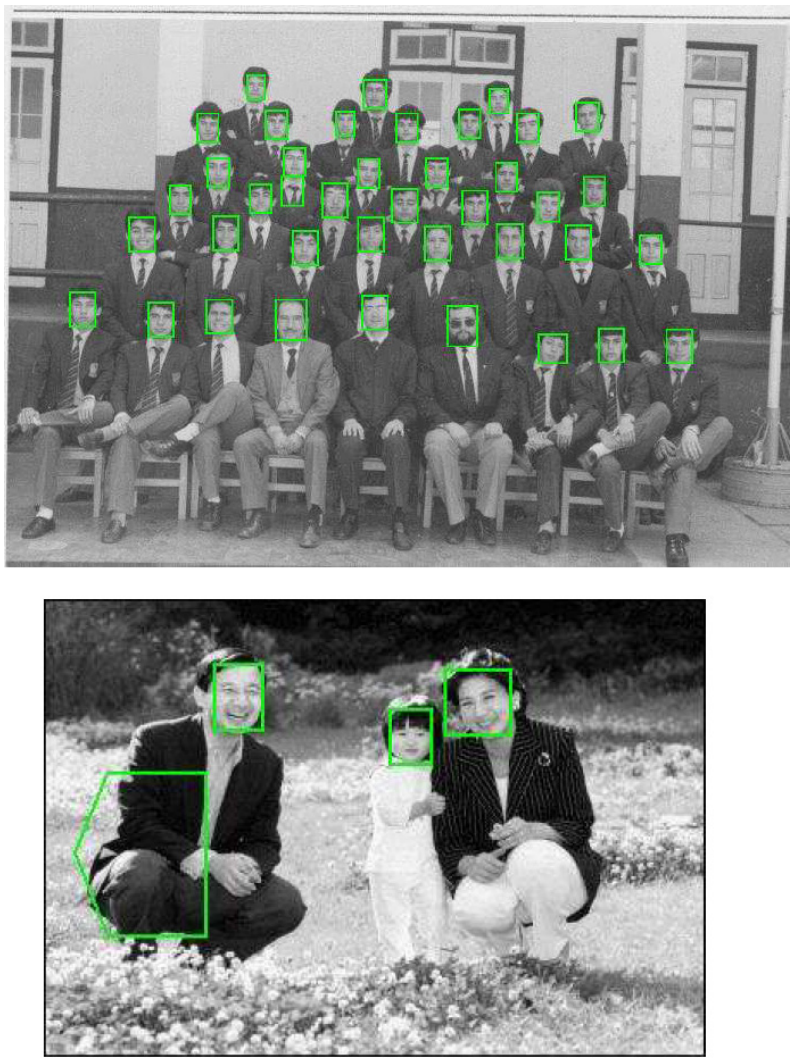


Fig. 12. Two results of using CMU face detection system show that false positives occur in non-skin color regions.

Table 3. Eye and mouth detection results.

Images	Detection Rate (%)
1-person	95.3
2-person	92.8
3-person	90.4
Others	85.8
Total	90.0

(a) (b)

Fig. 13. (a) A face is missed because the face region is not detected as a skin region, and (b) the hand is detected as a false positive.

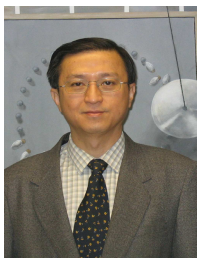
6. Conclusions

We have presented the image processing and pattern recognition techniques to extract human faces and facial features from color images, which may contain many people or complex background. Experimental results show that our face extraction method achieved the accuracy rate of 96.7%, and for our eye and mouth extraction it was 90.0%. Compared to others, our method need not search for the face candidates at all the locations. We will continue to improve the accuracy rates by using more sophisticated techniques in the future.

References

1. S. Brahnam, C. Chuang, F. Y. Shih and M. Slack, Machine recognition and representation of neonatal facial displays of acute pain, *Int. J. Artif. Intell. Med.* **36** (2006) 211–222.
2. S. Brahnam, C. Chuang, F. Y. Shih and M. Slack, SVM classification of neonatal facial images of pain, *Fuzzy Logic and Applications*, eds. I. Bloch, A. Petrosino and A. G. B. Tettamanzi, Lecture Notes in Computer Science, Vol. 3849 (2006), pp. 121–128.
3. J. Cai, A. Goshtasby and C. Yu, Detecting human faces in color images, *Proc. Int. Workshop on Multi-Media Database Management Systems* (1998), pp. 124–131.
4. D. Chai and A. Bouzerdoum, A Bayesian approach to skin color classification in YCbCr color space, *Proc. TENCON 2000* (2000), pp. 421–424.
5. C. Chuang and F. Y. Shih, Recognizing facial action units using independent component analysis and support vector machine, *Patt. Recog.* **39** (2006) 1795–1798.

6. G. C. Feng and P. C. Yuen, Multi cues eye detection on gray intensity image, *Patt. Recog.* **34** (2001) 1033–1046.
7. R. M. Haralick, S. K. Sternberg and X. Zhuang, Image analysis using mathematical morphology, *IEEE Trans. Patt. Anal. Mach. Intell.* **9** (1987) 532–550.
8. I.-S. Hsieh, K.-C. Fan and C. Lin, A statistic approach to the detection of human faces in color nature scene, *Patt. Recog.* **35** (2002) 1583–1596.
9. J. Huang and H. Wechsler, Eye detection using optimal wavelet packets and radial basis functions (RBFs), *Int. J. Patt. Recog. Artif. Intell.* **13** (1999) 1009–1026.
10. K.-M. Lam and H. Yan, Locating and extracting the eye in human face images, *Patt. Recog.* **29** (1996) 771–779.
11. S. Z. Li and Z. Q. Zhang, FloatBoost learning and statistical face detection, *IEEE Trans. Patt. Anal. Mach. Intell.* **26**(9) (2004) 1112–1123.
12. N. Otsu, A threshold selection method from gray-level histogram, *IEEE Trans. Syst. Man Cybern.* **9**(1) (1979) 62–66.
13. P. J. Phillips, H. Moon, P. J. Rauss and S. Rizvi, The FERET evaluation methodology for face recognition, *IEEE Trans. Patt. Anal. Mach. Intell.* **22** (2000) 1090–1104.
14. H. Rowley, S. Baluja and T. Kanade, Neural network-based face detection, *IEEE Trans. Patt. Anal. Mach. Intell.* **20** (1998) 23–38.
15. H. Schneiderman and T. Kanade, Probabilistic modeling of local appearance and spatial relationships for object recognition, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (1998), pp. 45–51.
16. H. Schneiderman and T. Kanade, A statistical method for 3D object detection applied to faces and cars, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2000), pp. 746–751.
17. F. Y. Shih and O. R. Mitchell, Threshold decomposition of grayscale morphology into binary morphology, *IEEE Trans. Patt. Anal. Mach. Intell.* **11** (1989) 31–42.
18. F. Y. Shih and C. Chuang, Automatic extraction of head and face boundaries and facial features, *Inform. Sci.* **158** (2004) 117–130.
19. F. Y. Shih, Y. Fu and K. Zhang, Multi-view face identification and pose estimation using B-spline interpolation, *Inform. Sci.* **169** (2005) 189–204.
20. F. Y. Shih, K. Zhang and Y. Fu, A hybrid two-phase algorithm for face recognition, *Patt. Recog. Artif. Intell.* **18** (2004) 1423–1435.
21. K.-K. Sung and T. Poggio, Example-based learning for view-based human face detection, *IEEE Trans. Patt. Anal. Mach. Intell.* **20** (1998) 39–51.
22. M. A. Turk and A. P. Pentland, Face recognition using eigenfaces, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* Maui, Hawaii (1991), pp. 586–591.
23. V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer-Verlag, New York, 1995).
24. P. S. Wang and S. N. Yanushkevich, Biometric technologies and applications, *Proc. IASTED Int. Conf. Artificial Intelligence and Applications*, Innsbruck, Austria (2007), pp. 226–231.
25. Y. Wang and B. Yuan, A novel approach for human face detection from color images under complex background, *Patt. Recogn.* **34** (2001) 1983–1992.
26. H. Wu, Q. Chen and M. Yachida, Face detection from color images using a fuzzy pattern matching method, *IEEE Trans. Patt. Anal. Mach. Intell.* **21** (1999) 557–563.
27. M.-H. Yang, D. J. Kriegman and N. Ahuja, Detecting faces in images: a survey, *IEEE Trans. Patt. Anal. Mach. Intell.* **24** (2002) 34–58.
28. D. Zhang, X. G. You, Q. H. Chen, P. S. Wang and Y. Y. Tang, Face representation by using non-tensor product wavelets, *Proc. Int. Conf. Pattern Recognition*, Hong Kong (2006), pp. 503–506.



Frank Y. Shih received the B.S. degree from National Cheng-Kung University, Taiwan, in 1980, the M.S. degree from the State University of New York at Stony Brook, in 1984, and the Ph.D. degree from Purdue University, West Lafayette, Indiana, in 1987, all in electrical engineering. He is presently a professor jointly appointed in the Department of Computer Science, the Department of Electrical and Computer Engineering, and the Department of Biomedical Engineering at New Jersey Institute of Technology, Newark, NJ. He currently serves as the Director of Computer Vision Laboratory.

Dr. Shih is currently on the Editorial Board of the *Int. J. Pattern Recognition*, the *Int. J. Pattern Recognition Letters*, the *Int. J. Pattern Recognition and Artificial Intelligence*, the *Int. J. Recent Patents on Engineering*, the *Int. J. Recent Patents on Computer Science*, the *Int. J. Internet Protocol Technology*, and the *J. Internet Technology*. Dr. Shih has contributed as a steering member, committee member, and session chair for numerous professional conferences and workshops. He was the recipient of the Research Initiation Award from the National Science Foundation in 1991. He won the Honorable Mention Award from the International Pattern Recognition Society for Outstanding Paper and also won the Best Paper Award in the International Symposium on Multimedia Information Processing. He has received several awards for distinguished research at New Jersey Institute of Technology. He has served several times on the Proposal Review Panel of the National Science Foundation.

Dr. Shih holds the research fellow for the American Biographical Institute and the IEEE senior membership. He has published one book (Digital Watermarking and Steganography, CRC Press, 2007), seven book chapters and over 190 technical papers, including 92 in well-known prestigious journals.

His current research interests include image processing, computer vision, sensor networks, pattern recognition, bioinformatics, information security, robotics, fuzzy logic, and neural networks.

Shouxian Cheng received the B.S. degree in Material Science from Beijing University of Chemical Technology, Beijing, China, in 1995, and the M.S. and Ph.D. degrees in computer science from New Jersey Institute of Technology, respectively in 2001 and 2005.

His research interests include image processing, pattern recognition and machine learning.

Chao-Fa Chuang received M.B.A degree from National Chung Hsing University, Taiwan, and Ph.D. degree in computer science from New Jersey Institute of Technology in January 2006. He is now working as a system analyst at the enfoTech & Consulting, Inc.

His research interests include image processing, pattern recognition, artificial intelligence, and machine learning.

Patrick S. P. Wang is tenured full professor of computer science at College of Computer and Information Science of Northeastern University, Boston, MA, USA. He is *Fellow* of IAPR (International Association of Pattern Recognition) and co-founding chief editor of IJPRAI (*International Journal of Pattern Recognition and Artificial Intelligence*) and WSP MPAI Book Series. Dr. Wang has been elected as Otto-Von-Guericke Distinguished Guest Professor at the Graphics and Simulations Lab of the University of Magdeburg (near Berlin) of Deutschland (Germany). Dr. Wang has been visiting at University of Calgary's Biometrics Technology Laboratory, Calgary, Canada as *iCORE* (Informatics Circle of Research Excellence) visiting Professor, where he published his 23rd book in July 2007, entitled: *Image Pattern Recognition — Synthesis and Analysis in Biometrics* by World Scientific Publishing Co. He is now *Zi-Jiang Visiting Chair Professor* at the Department of Computer Science and Technology, East China Normal University, Shanghai, China. Professor Wang has also served as Honorary Advisor Professor of Xiamen University and Sichuan University of China.

Dr. Wang got his B.S. in electronic engineering from National Chiao Tung University (Hsin Chu), M.S. in electrical engineering from National Taiwan University, M.S. in information and computer science from Georgia Institute of Technology, Ph.D. in computer science from Oregon State University. He has been Visiting Scientist at MIT AI Lab, and Research Consultant at MIT Sloan School, as well as adjunct faculty of Computer Science at Harvard University Extension School. For the past decades, he has been invited to give lectures in many countries in America, Asia and Europe. He has served as the Chair of IAPR-SSPR TC-2 (Structural and Syntactic Pattern Recognition) for 4 years (1992–1996). He has also organized many international conferences, including Data Mining (MLDM2003 and 2005) at Leipzig and BT2004, Int. Workshop on Biometric Technology, 2004, Calgary, Canada. Dr. Wang was general conference co-chair of the IAPR's 18th ICPR International Conference on Pattern Recognition, August, 2006 at Hong Kong, China. Prof. Wang serves as Program Vice-Chair and Distinguished International Advisor of the *IEEE-BIBE2007* at Harvard Medical. He gave a keynote plenary lecture on *Biometrics Intelligence Information Systems and Applications* and was honored and received a Plaque of *Outstanding Achievement Award* from Conference General Chair Prof. Jack Yang and IEEE-SMC President Prof. Larry Hall. He also gave a tutorial *on IPR and Biometrics* at the *IEEE-SMC 2007*, Montreal, Canada, October 2007. Dr. Wang was also invited to give a keynote speech and tutorial at IAS 2007, Manchester, UK, and e-Forensics 2008 at Adelaide, Australia.

His other literature works on music and poems have received many awards including the prestigious World Journal Prose Competition Award, and his book *Harvard Meditation Melody* (*哈佛冥想曲*) has won the prestigious Taiwan Provincial Better Publication Award.