A Face Replacement System Based on Face Pose Estimation

Kuo-Yu Chiu
Department of Electrical Engineering
National Chiao-Tung University,
Hsinchu, Taiwan(R.O.C)
E-mail:Alvin_cgr@hotmail.com*

Shih-Che Chien
Department of Electrical Engineering
National Chiao-Tung University,
Hsinchu, Taiwan(R.O.C)

Sheng-Fuu Lin
Department of Electrical Engineering
National Chiao-Tung University,
Hsinchu, Taiwan(R.O.C)
E-mail:sflin@mail.nctu.edu.tw*

***Abstract*—Face replacement system plays an important role in the entertainment industries. However, most of these systems nowadays are assisted by hand and specific tools. In this paper, a new face replacement system for automatically replacing a face with image processing technique is described. The system is divided into two main parts: facial feature extraction and face pose estimation. In the first part, the face region is determined and the facial features are extracted and located. Eyes, mouth, and chin curve are extracted by their statistical and geometrical properties. These facial features are used as the information for the second part. A neural network is adopted here to classify the face pose according to the feature vectors which are obtained from the different ratio of facial features. From the experiments and some comparisons, they show that this system works better while dealing with different pose, especially for non-frontal face pose.**

Keywords- Facial feature• Face replacement• Neural network• Support vector machine (SVM)

I. Introduction

For entertainment and special effects industries, the ability of automatically replacing a face in a video sequence with that of another person has huge implications. For example, consider a stunt double in full-view of the camera performs a dangerous routine, and the stunt double’s face could be automatically replaced latter with that of the desired actor for each instance by a post-processing. While few of the recent films have achieved good results when performing face replacement on the stunt doubles, there are still some limits, such like the illumination conditions in the environment should be controlled and the stunt double has to wear a special custom-fit mask with reflective markers for tracking [1].

In order to accurately replace a face in a photograph or a frame of video, we separate the system into two main parts. The first part is facial feature extraction and the second part is face pose estimation. Generally, the common approach of face region detection is to detect the face region by using the characteristic of the skin color. After locating the face region, the facial features can be obtained and determined by the geometric relation and statistical information. For example, the most common pre-processing method is to detect skin regions by a built skin tone model. R.L. Hsu et al. [2] proposed a face detection method based on a novel light compensation technique and a nonlinear color trans-formation. Besides, there are still many color models used for the human skin-color [3]-[5]. For example, H. K. Jee et al. [6] used the color, edge, and binary information to detect eye pair from input image with support vector machine. Classifier boost methods are used to detect face region in paper [7]-[8]. However, neural network-based approaches required a large number of face and non-face training examples [9]-[11]. C. Garcia et al. [12] presented a novel face detection based on a convolutional neural architecture, which synthesized simple problem-specific feature extractors. There are also several algorithms for facial feature extraction. C. H. Lin [13] located facial feature points based on deformable templates algorithm. C. Lin [14] used the geometric triangle relation of the eyes and the mouth to locate the face position. Yokoyama [15] synthesized the color and edge information to locate facial feature.

The second part of face replacement system is face pose estimation. It is assumed that the viewpoint is on a fixed location and the face has an unknown pose that needs to be determined by one or more images of the human head. Previous face pose estimation algorithms can be roughly classified into two main categories: window-based approaches [16]-[19] and feature-based approaches [20]-[23]. Window-based approaches extract face block from the input image and analyze the whole block by statistical algorithms. Among window-based approaches, multi-class classification method divides the whole head pose parameter space into several intervals and determine head pose [16]-[17]. For example, Y. M. Li et al. [18] used the technique of support vector regression to estimate the head pose, which could provide crucial information and improve the accuracy of face recognition. L. Zhao et al. [19] trained two neural networks to approximate the functions that map a head from an image to its orientation. Windowed-based approaches have the advantage that they can simplify the face pose estimation problem. However, the face pose is generally coupled with many factors, such as the difference of illumination, skin color, and so on. Therefore, the learning methods listed above require large number of training samples.

On the other hand, the feature-based approaches extract facial features from human face by making use of the 3D structure of human face. These approaches are used to build 3D models for human faces and to match the facial features, such

*Corresponding author

as face contour and the facial components of the 3D face model with their projection on the 2D image. Y. Hu et al. [20] combined facial appearance asymmetry and 3D geometry to estimate face poses. Besides, some sensors are used to improve feature location. For instance, D. Colbry et al. [21] detected key anchor points with 3D face scanner data. These anchor points are used to estimate the pose and then to match the test image to 3D face model. Depth and brightness constraints can be used to locate features and to determine the face pose in some researches [22]-[23].

This paper is organized as follows. The face region detection and facial feature extraction system are introduced in Section 2. Section 3 describes the face pose estimation system. The face replacement system will be exhibited in Section 4. Section 5 shows the experimental results and comparisons. Finally, the conclusions and the future works are drawn in Section 6.

II. Facial Feature Extraction

Facial feature extraction plays an important role in face recognition, facial expression recognition, and face pose estimation. A facial feature extraction system contains two major parts: face region detection and facial feature extraction. According to the skin color model, the candidate face regions can be detected first. Then, the facial features can be extracted by their geometric and statistic properties from the face region. In this section, face region detection and facial feature extraction will be described.

A. Face Region Detection

The first step of the proposed face replacement system is to detect and to track the target face in an image. A skin color model is used here to extract the skin color region which may be a candidate face region. The skin color model is built in *YCbCr* color space [24]. This color space is attractive for skin color modeling because it can separate chrominance from luminance. Hence, an input image is first transformed from RGB color space to *YCbCr* color space. Then the skin-color pixels are obtained by applying threshold values which are obtained from training data. After the skin color region is extracted, the morphological operator and 4-connectivity are then adopted to enhance the possible face region. The larger connected region of skin-color pixels are considered as the face region candidate and the real face region is determined by eye detection. Skin color region with eyes is defined as the face region. SVM classifier [25] is used here to detect eyes. Three sets of eye data are used for training. Eye images with frontal pose (set A) or profile pose (set B) are trained as the positive patterns. For negative patterns, non-eye images (set C) such as nose, lips, and ears are included for eye detection. All the training sets for eye detection are shown in Fig.1.

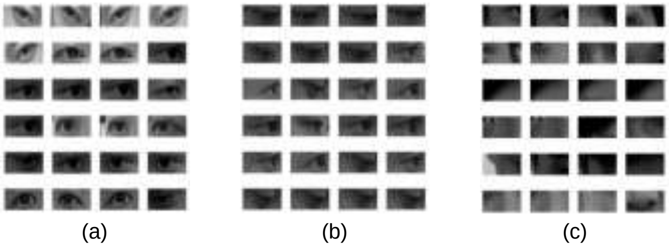


Figure 1. Training data of SVM. (a) Eye images of frontal pose. (b) Eye images of half-profile or profile pose. (c) Non-eye images.

Hence, for an input image as Fig.2a, the skin color region, which may be a candidate face region, can be extracted after applying skin color model, as Fig.2b. Morphological operator and 4-connectivity is used then to eliminate noise and enhance the region shape and boundary, as Fig.2c. The skin color region is defined as a face when the eyes can be found by using SVM-based eye detection, as Fig.2d.

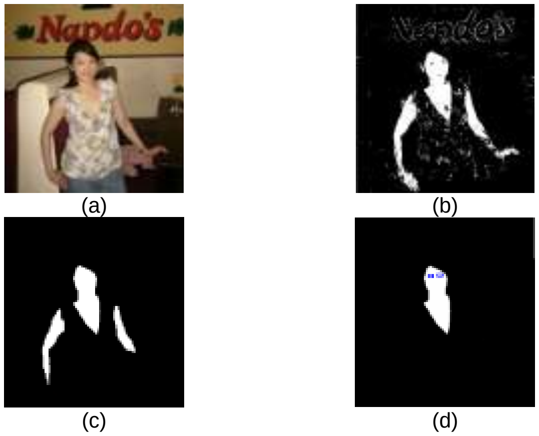


Figure 2. Face region detection. (a) Original image. (b) Binary image after applying the skin color model. (c) Possible face candidate regions after applying morphological operator and 4-connectivity. (d) The remaining face region after applying eye detection.

B. Facial Feature Extraction

After the eyes are located and the face region is determined, the other facial features, such like lips region and the chin curve, can be easily extracted according to their geometrical relationship. In this section, the locating of right tip and left tip of lips region, the construction of chin curve, and the hair region segmentation are described.

To extract the lips region, the property that the lips region is on the lower part of the face and the color of lips region is different from the skin color is considered. Since the lips region is redder than the skin color region, a red to green function *RG*(*x*,*y*) is employed to enhance the difference of lips color and skin color [26]. From the experimental results, the function *RG*(*x*,*y*) is defined as follows:

$$RG(x,y)=\begin{cases}\dfrac{R(x,y)}{G(x,y)+1}, & \text{if} \quad R(x,y)+G(x,y)+B(x,y)>50,\\ 0, & \text{if} \quad R(x,y)+G(x,y)+B(x,y)\le 50.\end{cases} \tag{1}$$

The *RG*(*x*,*y*) has higher value when the value of red channel is larger then the value of green channel, which is probably a pixel of lips region. The possible lips region with higher red value is shown in binary image as Fig.3b. Besides, the edge information is also taken into account here to improve the lips region locating. In the *YCbCr* color space, the Sobel operator is employed to find the horizontal edge in luminance (*Y*) channel. The edge information is shown in Fig.3c. Using the union of redder region and edge information, the left and right tip points of lips region can be determined. The results of left and right tip points of lips region locating is shown in Fig.3d.

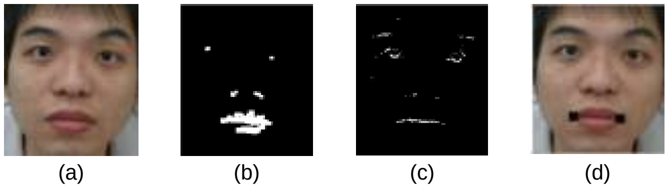


Figure 3. Lips region locating. (a) Original image. (b) Binary image of function *RG*(*x*,*y*). (c) Horizontal edge by using Sobel operator. (d) The left and right tip points of lips region.

The next facial feature which is going to be extracted is the chin curve. There are two advantages to extract the chin curve: one is to separate the head from the neck and the other is to estimate the face pose. Since the chin curve holds strong edge information, a face block image is transformed into gray value image, as Fig.4a, and then the entropy function is applied to measure the edge information. Large entropy value contains more edge information, as shown in Fig.4b. The equation of entropy is defined as follows:

$$E = -\sum_{M,N} P(I_{m,n}) \log P(I_{m,n}) \qquad (2)$$

where $I_{m,n}$ represents the gray level value of point (*m*,*n*) and $P(I_{m,n})$ is the probability density function of $I_{m,n}$. Using the lips position found before and the face block information, the searching region for the chin curve can be pre-defined. Five feature points, x_1, x_2, x_3, x_4, and x_5, are used to represent the chin curve. These five feature points are the intersections of chin curve and horizontal or vertical extended line from lips. The feature points, x_1 and x_5, are the intersections of chin curve and horizontal extended line from left tip of lips and right tip of lips respectively. The feature points, x_2, x_3, and x_4, are the intersections of chin curve and vertical extended line from left tip, middle, and right tip of lips. These feature points are shown in Fig.4c. Since the chin curve may not be symmetric, two quadratic functions, defined as: $y = ax^2 + bx + c$, are adopted here to construct the chin curve. The features x_1, x_2, and x_3 are used to find out the left quadratic function f_{Lc} and the features x_3, x_4, and x_5 are used to find out the right quadratic function f_{Rc} by using lease square method. The result of chin curve fitting is shown in Fig.4d.

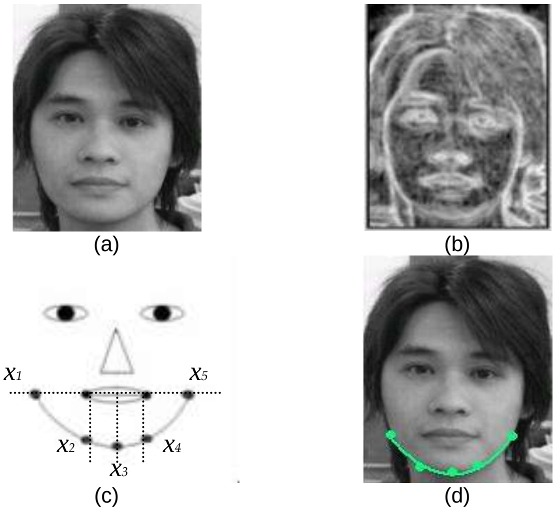


Figure 4. Chin curve construction. (a) Input gray level image. (b) The entropy of input gray level image. (c) Five feature points, x_1, x_2, x_3, x_4, and x_5, represent the most left point to the most right point respectively. (d) The function of chin curve fitting.

Hence, for an input image as Fig.5a, the skin color region can be found first as Fig.5b. Using the information of curve fitting function, the face region can be separated from the neck, as shown in Fig.5c.

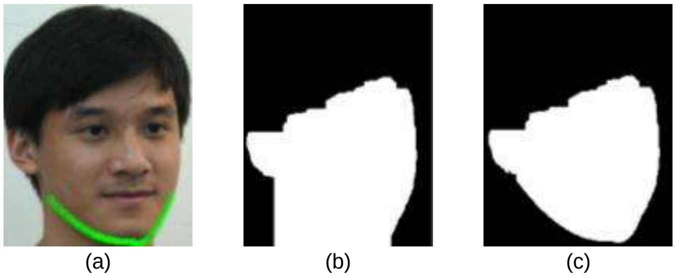


Figure 5. Face region segmentation. (a) Input image with curve fitting function. (b) Skin color region. (c) Face region only by using curve fitting information.

After the face region is found, the hair region can be defined easily. It is known that the hair region is above the face region. Hence, if an appropriate block above the face region is chosen and the skin color region is neglected, the remaining pixels, as Fig.6a, can be used as the seeds for seed region growing (SRG) algorithm. The hair region then can be extracted. The hair region extraction result is shown in Fig.6b.

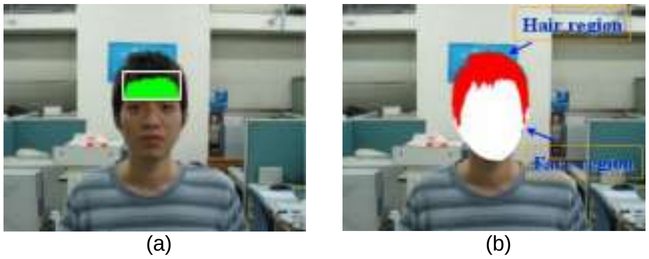


Figure 6. Hair region extraction. (a) The remaining pixels are used as the seeds for SRG after the skin color region is neglected. (b) The result of hair region extraction.

III. POSE ESTIMATION

In this section, how to estimate the face pose is detailed. All the different face poses are described with three angle parameters, namely the yaw angle *α*, the tilt angle *β*, and the roll angle *γ*, as shown in Fig.7.

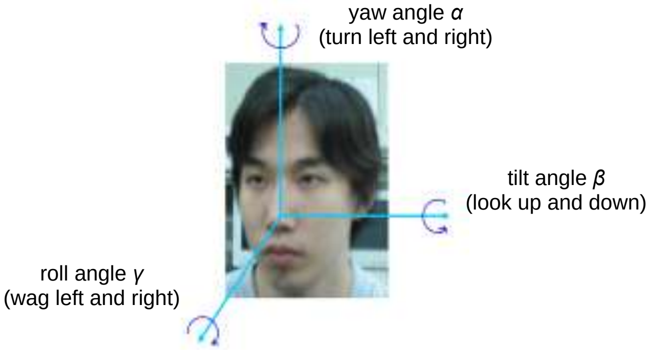
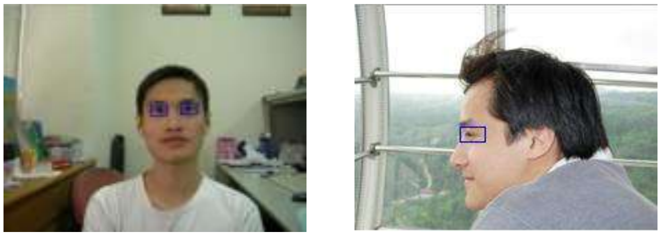


Figure 7. Parameters of face pose estimation.

Since the profile pose is much different from the frontal pose, two different methods are proposed here. All the different poses are roughly divided into two classes according to the number of eyes extracted in SVM-based eye detection system. When there are two eyes extracted, the face pose belongs to class A which is more frontal. Otherwise, if only one eye is extracted, then the face pose belongs to class B which is more profile. The examples of class A and class B are shown in Fig.8a and Fig.8b respectively.



(a) (b)

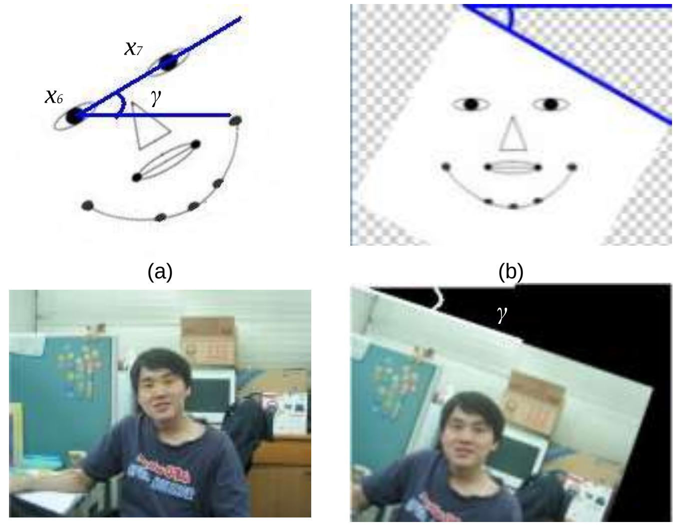
Figure 8. Two kinds of face pose. (a) Frontal face pose with two eyes extracted. (b) Profile face pose with only one eye extracted.

A. Pose Angle Estimation of Class A

For an input image of class A, it will be normalized and rotated first so that the line crossing two eyes is horizontal. In other words, the roll angle *γ* of the input face should be found out first. The roll angle *γ* is defined as the elevation or depression angle from left eye. Using the relative vertical and horizontal distance of the two eyes, the roll angle *γ* can be obtained. Set x_6 and x_7 as the center of left eye and right eye respectively as shown in Fig. 9a, the roll angle *γ* is defined by:

$$\gamma = \tan^{-1}\left(\frac{y_{x_7} - y_{x_6}}{x_{x_7} - x_{x_6}}\right) \tag{3}$$

where *x* and *y* represent the x-coordinate and y-coordinate respectively. Using the information of roll angle *γ*, the image can be rotated to horizontal as Fig. 9b. For an input image Fig. 9c, the normalization result is shown in Fig. 9d.



(a) (b)

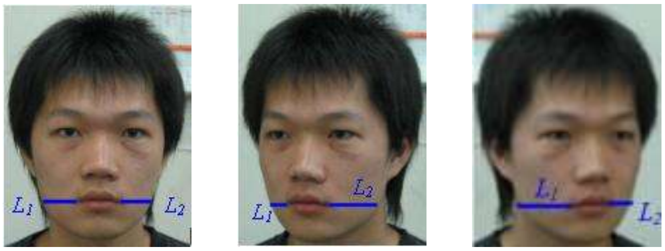
(c) (d)

Figure 9. The roll angle *γ*. (a) The definition of x_6 and x_7. (b) The rotated image with horizontal eyes. (c) Input image. (d) Normalization result of input image (c).

After retrieving the roll angle information, the face can be normalized to horizontal. Five scalars, v_1, v_2, v_3, v_4, and v_5, are used as the input of neural network to estimate the face pose in class A. The first scalar v_1 is defined as:

$$v_1 = \frac{L_1}{L_2} \tag{4}$$

where L_1 is the horizontal distance between the left tip of lips and the constructed chin curve f_c and L_2 is the distance between the right tip of lips and f_c. The scalar v_1 is relative to the yaw angle *α*. It is close to 1 when the yaw angle $\alpha \approx 90^\circ$, as Fig. 10a. When the face turns to right as Fig. 10b, L_1 is smaller than L_2 and the scalar v_1 is smaller than 1. Contrarily, when the face turns to left as Fig. 10c, L_1 is larger than L_2 and the scalar v_1 is larger than 1.



(a) (b) (c)

Figure 10. The relationship between scalar v_1 and the yaw angle *α*. (a) Scalar v_1 is close to 1 when the face is frontal. (b) Scalar v_1 is smaller than 1 when the face turns to right. (c) Scalar v_1 is larger than 1 when the face turns to left.

The second scalar v_2 is defined as the ratio of L_3 and L_4:

$$v_2 = \frac{L_3}{L_4} \tag{5}$$

where L_3 is the vertical distance between the middle point of two eyes, defined as x_8, and the constructed chin curve, and L_4

is the vertical distance between the center of the lips and x_8, as Fig. 11a. The scalar v_2 is relative to the tilt angle β as Fig. 11b. The third scalar v_3 is defined as:

$$v_3 = \frac{L_3}{L_5} \tag{6}$$

where L_5 represents the distance of x_6 and x_7. The scalar v_3 is relative to the tilt angle β, as Fig. 11c, and the yaw angle α, as Fig. 11d, simultaneously.

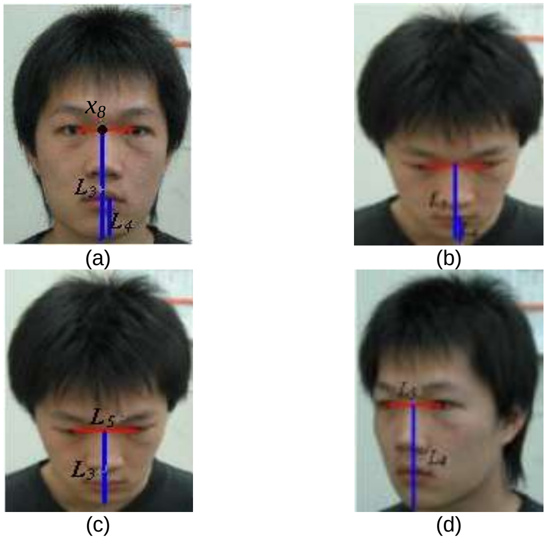


Figure 11. The relationship between scalars, v_2 and v_3, and pose parameter, tilt angle β and yaw angle α. (a) The definitions of x_8, L_3 and L_4. (b) The relationship between scalar v_2 and tilt angle β. (c) The relationship between scalar v_3 and tilt angle β. (d) The relationship between scalar v_3 and yaw angle α.

Before defining the last two scalars, another two parameters, L_6 and L_7, are defined first. Connecting the feature point x_3 of the chin curve and two tip points of the lips, the extended lines will intersect the extended line crossing x_6 and x_7 with two intersections. These two intersections are defined as x_9 and x_{10} from left to right respectively as Fig. 12a. Parameter L_6 is then defined as the distance between x_6 and x_9, and L_7 is the distance between x_7 and x_{10}. The definitions of parameters L_6 and L_7 are shown in Fig. 12b. Then, the forth scalars v_4 is defined as:

$$v_4 = \frac{L_6 \cdot L_7}{L_5} \tag{7}$$

and the last scalars v_5 is defined as:

$$v_5 = \frac{L_6}{L_7}. \tag{8}$$

Scalar v_4 is relative to tilt angle β, as shown in Fig. 12c, and the scalar v_5 is relative to yaw angle α, as shown in Fig. 12d.

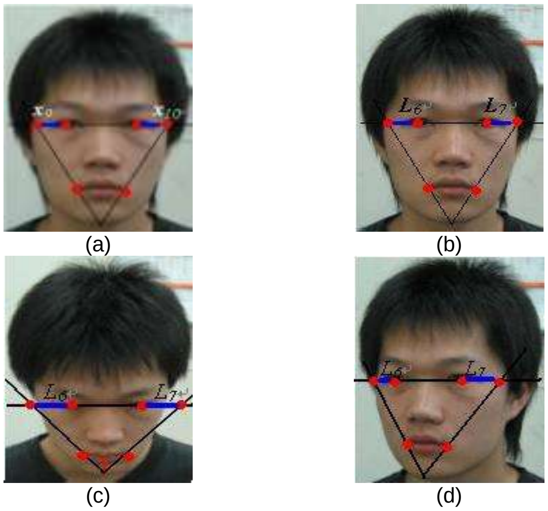


Figure 12. The relationship between scalars, v_4 and v_5, and pose parameter, tilt angle β and yaw angle α. (a) The definitions of x_9 and x_{10}. (b) The definitions of L_6 and L_7. (c) The relationship between scalar v_4 and tilt angle β. (d) The relationship between scalar v_5 and yaw angle α.

B. Pose Angle Estimation of Class B

The face is classified to class B if only one eye can be found when applying eye detection. For the face in class B, there are also five scalars used as the input of neural network to estimate the face pose. Feature points x_{11} and x_{12} represent the intersection points of face edge and the horizontal extended line crossing the eye and the lips respectively. Feature point x_{13} the tip point of chin curve which is found with the largest curvature and feature point x_{14} is the only extracted eye center. With these four feature points which are shown in Fig. 13a, the first scalar v_1' is defined as:

$$v_1' = \frac{L_9}{L_8} \tag{9}$$

where L_8 is the distance between x_{14} and face edge x_{11} and L_9 is the distance between x_{14} and the middle point of the lips. These two parameters are shown in Fig. 13b. The first scalar v_1' is relative to the yaw angle α as Fig. 13c.

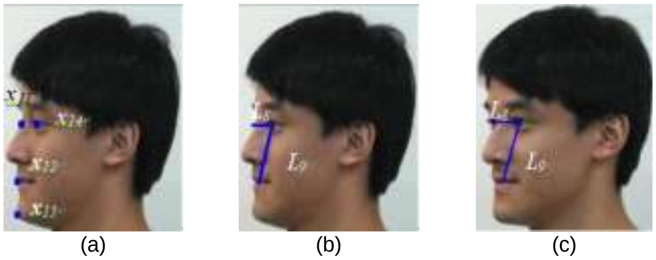


Figure 13. The relationship between scalar v'_1 and yaw angle α. (a) The definition of feature points x_{11}, x_{12}, x_{13}, and x_{14}. (b) The definition of parameters L_8 and L_9. (c) The scalar v'_1 is relative to the yaw angle α.

The scalar v_2' is the slope of the line crossing x_{12} and x_{13} as shown in Fig. 14a and it is defined by:

$$v_2' = m_{\overline{x_{12}x_{13}}} \tag{10}$$

where m represents slop. The scalar v'_2 is relative to the tilt angle β as Fig. 14b. The scalar v'_3 is the angle θ which is defined by:

$$\theta = \angle x_{11}x_{14}x_{12},\ 0^\circ < \theta < 90^\circ \tag{11}$$

and it is shown in Fig. 14c. The scalar v'_3 is relative to the tilt angle β as Fig. 14d and the yaw angle α as Fig. 14e, simultaneously.

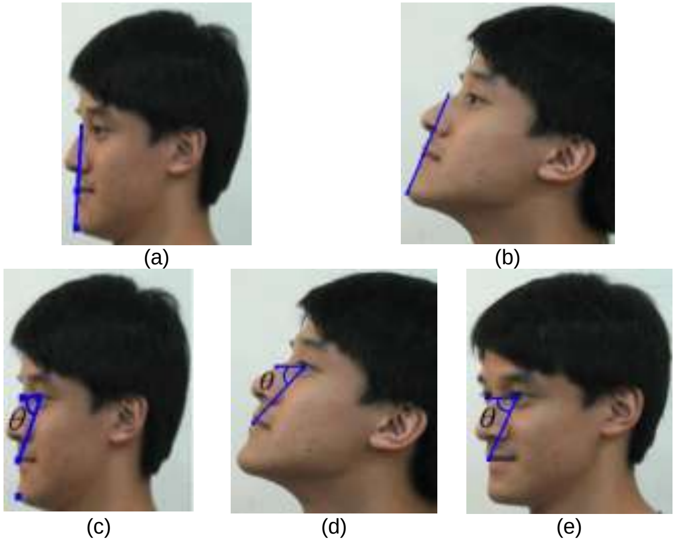


Figure 14. The relationship between scalars, v'_2 and v'_3, and pose parameter, tilt angle β and yaw angle α. (a) The line crossing x_{12} and x_{13}. (b) The scalar v'_2 is relative to the tilt angle β. (c) The definition of angle θ. (d) The scalar v'_3 is relative to the tilt angle β. (e) The scalar v'_3 is relative to the yaw angle α.

Connecting x_{14} with middle point and right tip point of lips, the extended line will intersect the horizontal line passing x_8 with two intersections. L_{10} is defined as the distance between these two intersections as Fig. 15a. Then the scalar v'_4 is defined as:

$$v'_4 = L_{10} \cdot L_8 \tag{12}$$

and the scalar v'_5 is defined as:

$$v'_5 = \frac{L_{10}}{L_9}. \tag{13}$$

The scalar v'_4 and v'_5 are relative to the tilt angle β, Fig. 15b, and the yaw angle α, Fig. 15c, simultaneously.

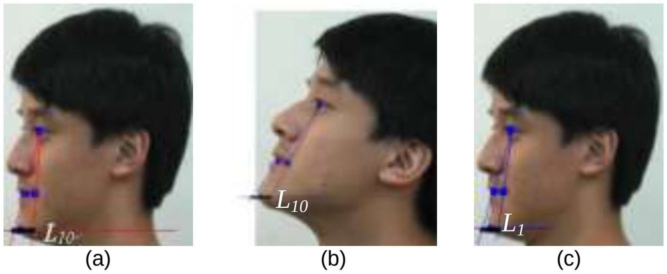


Figure 15. The relationship between scalars and pose parameter (a) The definition of L_{10}. (b) The scalar v'_4 is relative to the tilt angle β. (c) The scalar v'_5 is relative to the yaw angle α.

IV. Face Replacement

In this section, the procedure of face replacement is detailed. For an input target face, the face pose is estimated first and then the face with similar face pose is chosen from the database as the source face to replace the target face. However, there are some problems when replacing the face, such like the mismatch of face size, face position, face pose angle, and skin color. Hence, image warping and shifting are adopted first to adjust the source face so that it is much similar as the target face. Color consistency and image blending are used later to reduce the discontinuousness due to the replacement. All the details are described below.

A. Image Warping and Shifting

After the face pose angle of target face is determined and the face region of source face is segmented, the target face is going to be replaced by the source face. However, the resolution, face size, and face pose angle may not be exactly the same. Hence, image warping is adopted here to deal with this problem.

Image warping is applied according to features matching. It is a spatial transformation that includes shifting, scaling, and rotating. In this paper, an affine matrix with bilinear interpolation is used to achieve image warping. The affine transformation matrix is defined by:

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{14}$$

where (X',Y') is the feature point coordinate of target face, (X,Y) is the feature point coordinate of source face, and $m_1,\ldots,m_6$ are parameters. For faces in class A, six feature points, two eyes (x_6 and x_7), the center of lips, and feature points x_1, x_3, and x_5 of chin curve, are used to solve the matrix by the least square method as Fig. 16a, while four feature points, x_{11}, x_{12}, x_{13}, and x_{14}, are used for faces in class B as Fig. 16b.

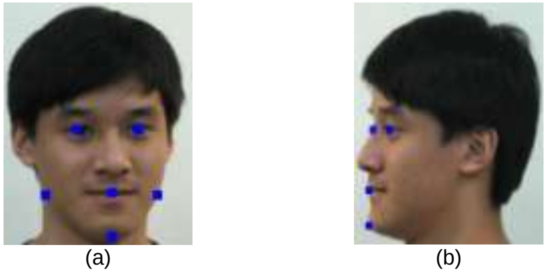


Figure 16. Feature points for image warping. (a) Six feature points are used for class A. (b) Four feature points are used for class B.

After the source face is warped, a suitable position is going to be found to replace the target face. A better face replacement is achieved when more face and hair regions are matched in both source face and target face. The source face is first pasted on so that the coordinates of pasted feature point are the same for source face and target face. The pasted feature point is chosen as the middle point of chin curve, x_3, for class A and tip points of chin curve, x_{13}, for class B. Later, the pasted source face is shifted around the pasted feature point to a best position with most matching points. A matching degree function $M(x,y)$ for a pasting point (x,y) is used to evaluate the degree of matching, which is defined as:

$$M(x,y)=\sum_{(i,j)\in I}[h(F_s(i,j),F_t(i,j))+h(H_s(i,j),H_t(i,j))] \qquad (15)$$

where $F_s(i,j)$ and $F_t(i,j)$ are binary face images which have value 1 only for face region pixel in source and target images respectively, $H_s(i,j)$ and $H_t(i,j)$ are binary hair images which have value 1 only for hair region pixel in source and target images, and I is the region of interest which is larger than the pasted region. The function $h(a,b)$ in equation (15) is defined by:

$$h(a,b)=\begin{cases}+1, & \text{if } a=b=1,\\ 0, & \text{if } a=b=0,\\ -1, & \text{if } a\neq b.\end{cases} \qquad (16)$$

For each point near the pasted feature point, the matching degree can be calculated. The point with highest matching degree will be chosen as the best position to paste the source face on. For example, the face region (white) and hair region (red) for source face image and target face image are shown in Fig. 17a and 17b respectively. When the target face is randomly pasted by the source face as Fig. 17c, there are more “-1”, denoted as the red region, and less “+1”, denoted as the white region. This means that the matching degree is low. After calculating all the matching degree of nearby points, the best pasting point with most “+1” and least “-1” can be found, as Fig. 17d.

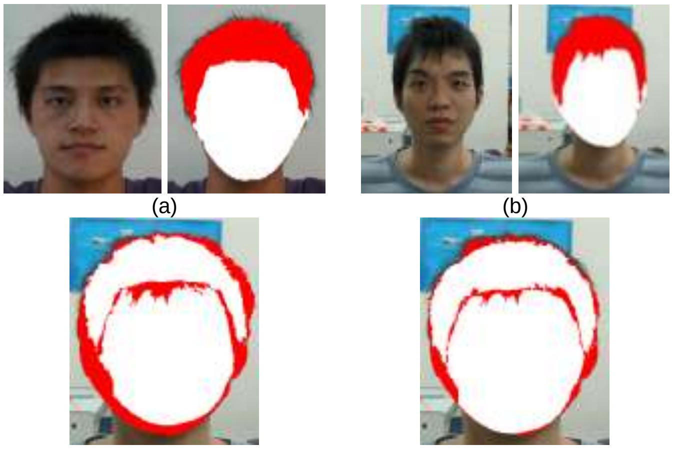

(a) (b) (c) (d)

Figure 17. Image shifting according to matching degree. (a) Source face image (b) Target face image (c) Face replacement with lower matching degree. (d) Face replacement with highest matching degree.

B. Color Consistency and Image Blending

Because of the difference of luminance and human races, the skin color of target face may not be similar to the source face. To solve the problem, skin color consistency is adopted here. The histogram of both source face and target face are analyzed first and the mean of skin color of target face is shifted to the same value as the mean of source face. For example, the source face as Fig. 18a is darker than the target face as Fig. 18b. If the face replacement is applied without adopting skin color consistency, the skin color of face region and necks region of the result is different, as shown in Fig. 18c. To avoid this situation, the mean of histogram of the target face is shifted to the same value as the source face, as Fig. 18d. Then, the skin color of the face region and necks region will be similar after replacement. The result of face replacement with skin color consistency is shown in Fig. 18e.

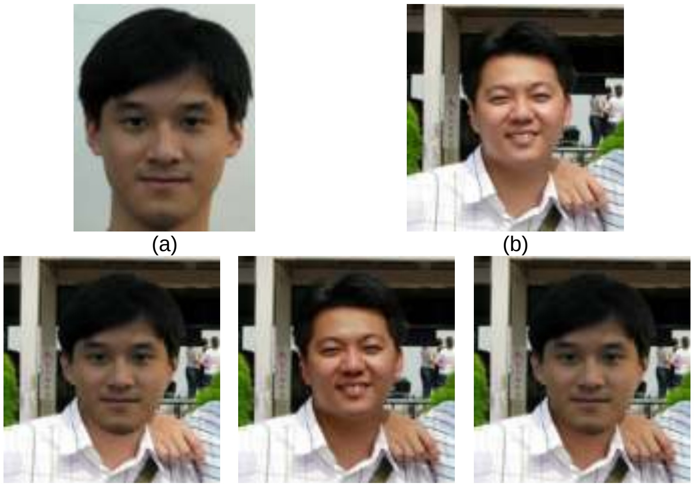

(a) (b) (c) (d) (e)

Figure 18. Skin color consistency. (a) Source Face. (b) Target Face. (c) Face replacement without skin color consistency. (d) The mean of histogram of target face is shifted to the same value as the source face. (e) Face replacement with skin color consistency.

Finally, an image blending method is applied to deal with the boundary problem. Though the source skin color is changed so that it is consistent with target face, there is still boundary problem when the source face replaces the target face because of discontinuousness. The objective of image blending is to smooth boundary by using interpolation. The hyperbolic tangent is used as the weight function:

$$\tanh(x)=\frac{\sinh(x)}{\cosh(x)}=\frac{e^x-e^{-x}}{e^x+e^{-x}}. \qquad (17)$$

The horizontal interpolation is described as:

$$I(x,Y)=w_{h(x)}L(x,Y)+(1-w_h(x))R(x,Y) \qquad (18)$$

and the vertical interpolation is described as:

$$I(X,y)=w_{v(y)}D(X,y)+(1-w_h(y))U(X,y) \qquad (19)$$

where $I(x,y)$ is the boundary point; $L(x,Y)$, $R(x,Y)$, $U(X,y)$, and $D(X,y)$ represent the left, right, up, and down image respectively. The result of image blending is exhibited in Fig. 19. These images are not applied with color consistency, so the boundary is sharper because of face replacement. However, it can be seen that the image in Fig. 19b with image blending has smoother boundary than the one in Fig. 19a without image blending.

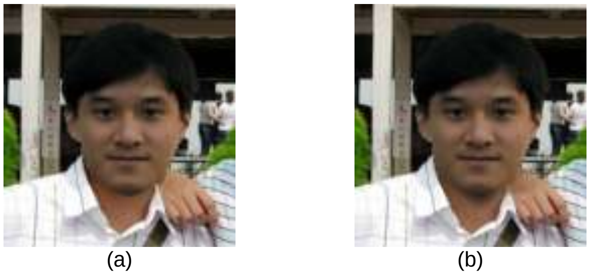


Figure 19. Image blending. (a) Without Image blending. (b) With image blending.

V. Experimental Results

In this section, the results of face pose estimation and face replacement will be shown. Some analyses and comparisons will also be made in this section.

A. *Face Pose Estimation*

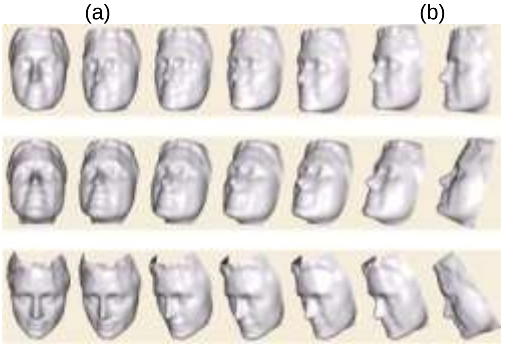
To verify the accuracy of the face pose estimation system, face images under various poses are collected and tested. In the database, the face poses are divided into 21 classes according to different yaw angle *α* and tilt angle *β*. The face pose is estimated by multi-class classification based on neural network. The yaw angle is divided into 7 intervals and the tilt angle is divided into 3 intervals, as shown in Fig. 20a and Fig. 20b respectively. Because the face poses of turning right and turning left are the same by applying a reflection matrix, only left profile face poses are considered.

α_1	-10° ~10°
α_2	10° ~25°
α_3	25° ~40°
α_4	40° ~55°
α_5	55° ~70°
α_6	70° ~85°
α_7	85° ~90°

(a)

β_1	-7° ~7°
β_2	7° ~20°
β_3	-7° ~-20°

(b)



(c)

Figure 20. Intervals of different face angles and different face poses. (a) 7 intervals of different *α*. Positive value represents turning left and the magnitude represents the turning degree. (b) 3 intervals of different *β*. Positive value represents looking up and the magnitude represents the turning degree. (c) 21 Different face poses.

There are 1680 images from 80 people in the database for training and another 1680 images from other 80 people are used for testing. The accuracy rate of pose estimation, R_{PE}, is defined by:

$$R_{PE} = \frac{\text{Total Photos - Failure estimation}}{\text{Total Photos}}. \qquad (20)$$

Therefore, the accuracy rate of face pose estimation can be calculated. There are some other face pose estimation methods, such as pose estimation method based on Support Vector Regression (SVR) using Principal Component Analysis (PCA) [18] and Neural Network (NN) based approach [19]. The comparisons of face pose estimation methods are shown in Fig. 21.

	The proposed Method	PCA+SVR	Neural Networks
$[\alpha_1, \alpha_4]$ (pure background)	**88.75 %**	**87.5 %**	86.45 %
$[\alpha_5, \alpha_7]$ (pure background)	**89.16 %**	**86.17 %**	85.67 %
$[\alpha_1, \alpha_4]$ (complex background)	**86.94 %**	**85.69 %**	85.56 %
$[\alpha_5, \alpha_7]$ (complex background)	87.5 %	82 %	81.67 %

Figure 21. Accuracy comparisons of face pose estimation.

B. *Face Replacement*

In this section, the results of face replacement are shown. Various conditions are considered to test the robustness of this automatic face replacement system, such like wearing glasses, different resolution, different luminance, different skin color, different yaw angle, different roll angle, and different tilt angle. It can be seen from the results that this system performs well while dealing with these conditions.

In Fig. 22, wearing glasses or not is discussed. The target face with glasses, as Fig. 22b, is replaced by the source face without glasses, as Fig. 22a. Since the target face region is replaced by the entire source face, wearing glasses or not will not affect the results.

When the face size and luminance of target face and source face are different, the face size and the skin color will be adjusted. The source face in Fig. 23a will be resized to fit the target face by the affine matrix according to the facial feature matching. Color consistency method is also applied in this case. It can adjust the skin color of the target face in Fig. 23b so that the skin color of target face is similar to the source face and the result would be better after replacement. From the result in Fig. 23c, it can be seen that the skin color of target face are adjusted and shifted to the similar value as source face, especially for the neck region. Since the skin color and the face size are adjusted, the replacement result is more nature.

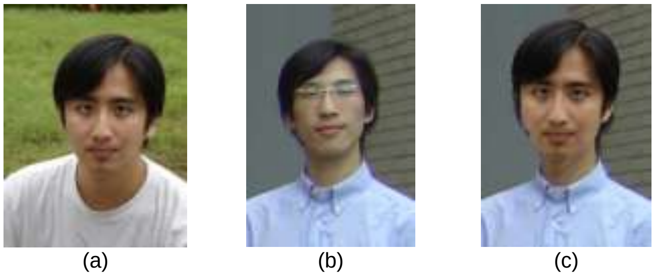


Figure 22. Face replacement result when considering the glasses. (a) Source face image without glasses. (b) Target face image with glasses. (c) The replacement result.

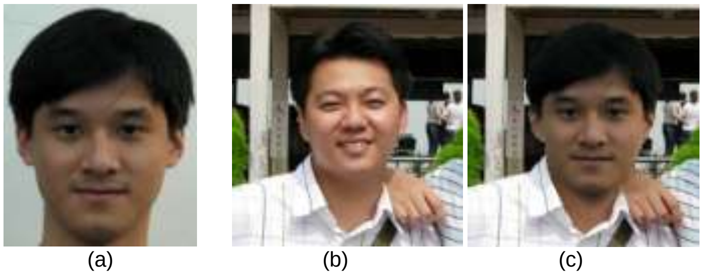


Figure 23. Face replacement result when considering different face size and illumance. (a) Source face image with thiner face region and darker skin color. (b) Target face image with wider face region and brighter skin color. (c) The replacement result.

While dealing with the profile pose, such as the face with 90 degrees of yaw angle as Fig. 24a, a face image with similar face pose is chosen from the database to replace the target face. The result would be poor if the replacement is done by only adopting an affine matrix without a proper face pose. From the result shown in Fig. 24b, it can be seen that this system performs well while dealing with profile pose, even if the face has 90 degrees of yaw angle.

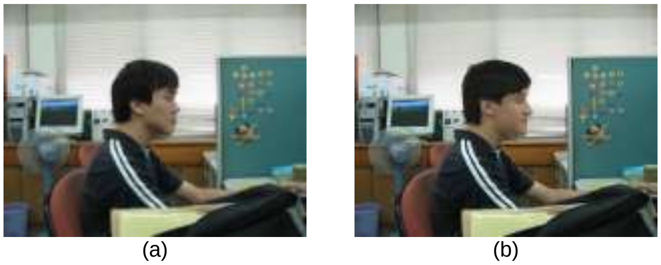


Figure 24. Face replacement result when considering the profile face. (a) Target face image. (b) The result of face replacement.

Like the profile pose, when dealing with the face with tilt angle such as Fig. 25b, a proper source face will be found from the database first. According to the face pose estimation system, a face with most similar face pose is chosen, as Fig. 25a. After applying a reflection matrix, the face pose of source face is almost the same as the target face. With color consistency method, the replacement can be done even though there are tilt angles and yaw angles at the same time for a target face. The face replacement result is shown in Fig. 25c.

When considering the target face with a roll angle, such as Fig. 26b, the roll angle is calculated first according to two eyes. After the roll angle is found and a similar pose is chosen from the database for the source face as Fig. 26a, a rigid transformation is adopted to rotate the source image such that the roll angle of source face is the same as the roll angle of target face. In Fig. 26c, it can be seen that the replacement is done with a rigid transformation.

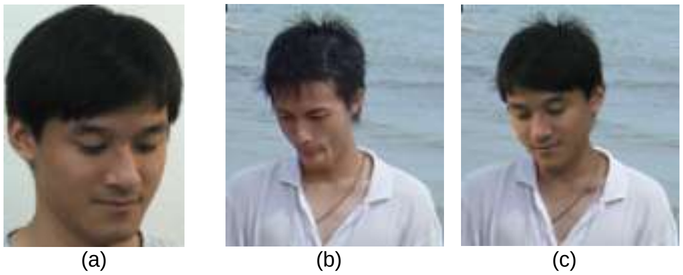


Figure 25. Face replacement result when considering the face with tilt angle. (a) Source face image (b) Target face image with a tilt angle. (c) The replacement result.

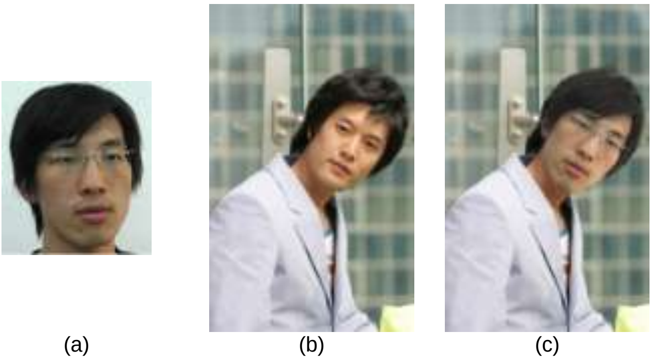


Figure 26. Face replacement result when considering the face with roll angle. (a) Source face image. (b) Target face image with a roll angle. (c) The replacement result.

VI. Conclusions and Future Work

Face replacement system plays an important role in the entertainment industries. In this paper, a face replacement system based on image processing and face pose estimation is described. Various conditions are considered when replacing the face, such as different yaw angle, roll angle, tilt angle, skin color, luminance, and face size. The experiment results show that this face replacement system has good performance while dealing the conditions listed before. In the future, facial expression would be a further challenge task to be considered. Along with the facial expression, the results of face replacement will be realer and the system will be much more powerful and useful in entertainment industries.

Acknowledgment

This work was supported in part by the National Science Council, Taiwan, R.O.C., under Grant NSC 95-2221-E-009-214.

References

[1] J. Kleiser: Kleiser-Walczak on the one. DOI: http://www.kwcc.com/works/ff/the_one.html

[2] Hsu, R.L., Abdel-Mottaleb, M., Jain, A.K.: Face detection in color images. IEEE Trans. Pattern Analysis and Machine Intelligence. 24(5), 696-706 (2002)

[3] Zhu, X., Yang, J., Waibel, A.: Segmenting hands of arbitrary color. In: Proceedings of 4th IEEE Conf. Automatic Face and Gesture Recognition, pp. 446-453, 2000

[4] Jee, H.K., Lee, K., Pan, S.B.: Eye and face detection using SVM. In: Proceedings of IEEE Conf. Intelligent Sensors, Sensor Networks and Information Processing, pp. 577-580, 2004

[5] Lin, Y.Y., Liu, T.L.: Robust face detection with multi-class boosting. In: Proceedings of IEEE Computer Society Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 680-687, 2005

[6] Huang, C., Al, H.Z., Wu, B., Lao, S.H.: Boosting nested cascade detector for multi-view face detection. In: Proceedings of 17th IEEE International Conf. Pattern Recognition, vol. 2, pp. 415-418, 2004

[7] Bingulac S.P.: On the Compatibility of Adaptive Controllers. In: Proceedings of Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994

[8] MacQueen J.: Some Methods for Classification Analysis of Multivariate Observations. In: Proceedings of Fifth Berkeley Symp. Math. Statistics and Probability, pp. 281-297, 1967

[9] Fu, H.C., Lai, P.S., Lou, R.S., Pao, H.T.: Face detection and eye localization by neural network based color segmentation. In: Proceedings of IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X, vol. 2, pp. 507-516, 2000

[10] Garcia, C., Delakis, M.: Convolutional face finder: a neural architecture for fast and robust face detection. IEEE Trans. Pattern Analysis and Machine Intelligence. **26**(11), 1408-1423 (2004)

[11] Lin, C.H., Wu, J.L.: Automatic facial feature extraction by genetic algorithms. IEEE Trans. Image Processing. , **8**(6), 834-845 (1999)

[12] Lin, C., Fan, K.C.: Human face detection using geometric triangle relationship. In: Proceedings of 15th IEEE Int. Conf. Pattern Recognition, vol. 2, pp. 941-944, Barcelona, Spain 2000

[13] Yokoyama, T., Wu, H., Yachida, M.: Automatic detection of facial feature points and contours. In: Proceedings of 5th IEEE Int. Workshop on Robot and Human Communication. pp. 335-340, Tsukuba, Japan 1996

[14] Yang, Z.G., Ai, H.Z., Okamoto, T., Lao, S.H.: Multi-view face pose classification by tree-structured classifier. In: Proceedings of IEEE International Conf. Image Processing, vol. 2, pp. 358-361, 2005

[15] Li, S.Z., Fu, Q.D., Scholkopf, B., Cheng, Y.M., Zhang, H.J.: Kernel machine based learning for multi-view face detection and pose estimation. In: Proceedings of 8th IEEE International Conf. Computer Vision, vol. 2, pp. 674-679, Vancouver, BC, Canada 2001

[16] Li, Y.M., Gong, S.G., Liddell, H.: Support vector regression and classification based multi-view face detection and recognition. In: Proceedings of 4th IEEE International Conf. Automatic Face and Gesture Recognition, pp. 300-305, Grenble, France 2000

[17] Zhao, L., Pingali, G., Carlbom, I.: Real-time head orientation estimation using neural networks. In: Proceedings of IEEE International Conf. Image Processing, vol.1, pp. 297-300, 2002

[18] Hu, Y., Chen, L.B., Zhou, Y., Zhang, H.J.: Estimating face pose by facial asymmetry and geometry. In: Proceedings of 6th IEEE Conf. Automatic Face and Gesture Recognition, pp. 651-656, 2004

[19] Colbry, D., Stockman, G., Jain, A.: Detection of anchor points for 3D face verification. IEEE Conf. Computer Vision and Pattern Recognition, 3, 118-125 (2005)

[20] Covell, M., Rahini, A., Harville, M., Darrell, J.: Articulated pose estimation using brightness- and depth-constancy constraints. In: Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 438-445, 2000

[21] Harville, M., Rahimi, A., Darrell, T., Gordon, G., Woodfill, J.: 3D pose tracking with linear depth and brightness constraints. In: Proceedings of 7th IEEE International Conf. Computer Vision, vol. 1, pp. 206-213, Kerkyra, Greece 1999

[22] Chai, D., Ngan, K.N.: Face segmentation using skin-color map in videophone applications. IEEE Trans. on circuits and systems for video technology, 9(4), 551-564 (1999)

[23] Vapnik, V.: The Nature of Statistical Learning Theory. New York: Springer, 1995

[24] Eveno, N., Caplier, A. Coulon, P.Y.: A new color transformation for lips segmentation. In: Proceedings of 4th IEEE Workshop on Multimedia Signal Processing, pp. 3-8, Cannes, France 2001

[25] Nugroho, H., Takahashi, S., Ooi, Y., Ozawa, S.: Detecting human face from monocular image sequences by genetic algorithms. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 4, 2533-2536 (1997)

AUTHORS PROFILE

Sheng-Fuu Lin (S’84–M’88) was born in Tainan, R.O.C., in 1954. He received the B.S. and M.S. degrees in mathematics from National Taiwan Normal University in 1976 and 1979, respectively, the M.S. degree in computer science from the University of Maryland, College Park, in 1985, and the Ph.D. degree in electrical engineering from the University of Illinois, Champaign, in 1988. Since 1988, he has been on the faculty of the Department of Electrical and Control Engineering at National Chiao Tung University, Hsinchu, Taiwan, where he is currently a Professor. His research interests include image processing, image recognition, fuzzy theory, automatic target recognition, and scheduling.

Shih-Che Chien was born in Chiayi, R.O.C., in 1978. He received the B.E. degree in electronic engineering from the Nation Chung Cheng University, in 2002. He is currently pursuing the M.E. and Ph.D. degree in the Department of Electrical and Control Engineering, the National Chiao Tung University, Hsinchu, Taiwan. His current research interests include image processing, image recognition, fuzzy theory, 3D image processing, intelligent transportation system, and animation.

Kuo-Yu Chiu was born in Hsinchu, R.O.C., in 1981. He received the B.E. degree in electrical and control engineering from National Chiao Tung University, Hsinchu, Taiwan, R.O.C, in 2003. He is currently pursuing the Ph. D. degree in the Department of Electrical and Control Engineering, the National Chiao Tung University, Hsinchu, Taiwan. His current research interests include image processing, face recognition, face replacement, intelligent transportation system, and machine learning.