An Automatic Approach to Facial Feature Extraction for 3-D Face Modeling

Jui-Chen Wu, Yung-Sheng Chen, and I-Cheng Chang

Abstract—Creating a friendly human interface for visual communication has become a popular and significant topic. One key issue is to construct a three-dimensional (3-D) face model as a visual representation. A 3-D face model can present the desired 3-D facial images related to a specified person by using the realistic 3-D structure and texture description. In order to automatically model the human face, we propose a coarse-to-fine method to extract the facial features from a 2-D head-and-shoulder color image and to map these facial features onto a 3-D template model. A skin-color-based scheme is first adopted to extract salient regions for selecting the face region, where the salient region is invariant to different situations such as scaling, rotation, and skewing. To more precisely extract the facial features, face geometry and proportions obtained by training database are used for reference. After locating desired regions such as eyes, lip, nose, and eyebrows, the corresponding facial features in the located regions are extracted based on a feature detection method, where a morphology-based scheme is used to enhance the corner features. Finally the complete facial features are mapped onto the adopted shape model and each facial feature is adjusted automatically to the proper position according to the training data. Our experiments demonstrate the feasibility of the proposed approach.

***Index Terms*—Face segmentation, facial feature location, facial feature extraction, shape model.**

I. Introduction

Development of friendly human interfaces has fast increased [1]-[4] in recent years, and the techniques have been widely applied to many fields such as the videophone, videoconference, and audio and video communications between persons or groups. With the high performance in computing, graphics and networking technologies, real-time face-to-face communications in a virtual world approaches gradually to be realizable. In order to reconstruct an individual head and communicate with user in an efficient way, decreasing the transmission data size and sending few parameters for such a system are necessary. Accordingly, model-based image coding scheme has been proposed by many researchers for future videophone and videoconference services. Instead of transmitting waveform parameters as in conventional coding techniques, model-based image coding uses knowledge about the shape, structure and texture of the object, and to some extent handles the meaning or content of the visual information. Therefore, the face modeling issues has been tackled in various ways such as using a laser scanner or stereo vision techniques, the template-based deformation techniques and so on.

An example of commercial 3-D system based on laser-light scanning is Cyberware Color Digitizer^{TM}. In [5], they developed algorithm that automatically constructs functional models of the heads of human subjects from laser-scanned ranging and reflecting data. However the approach based on 3-D digitization to get range data often requires a high-cost special hardware, and the operation of constructing a 3-D face model is usually complicated. Therefore, an easy and fast way of creating 3-D objects is appreciated. Lee and Thalmann [6] further derived a special person's face model by deforming the template model to fit the appearance of the facial image. The template-based facial modeling technique not only provides a fast and simple method to construct a realistic virtual human face from real images but also reduces the complicated operations for a user in setting the motion parameters. However, the user needs manually select the facial features to derive the face model animation. The objective of this study is therefore addressed to develop an automatic approach to extracting facial features, which can facilitate to further produce the realistic 3-D model by mapping a set of deformable features onto a 3-D template model.

In this paper, a coarse-to-fine method is presented to extract facial features from a 2-D head-and-shoulder color image. The proposed system adopts a skin-color-based scheme to extract salient regions and then finds possible face regions. The salient region feature is invariant to different situations like scaling, rotation, and skewing. In order to extract the facial feature more precisely, the information of face geometry and proportions are used in the training database, which can facilitate to precisely locate the facial regions, such as eyes, lip, nose, and eyebrows. Finally, the facial feature corresponding to each facial region can be found and mapped onto a 3-D generic face model [7]. Our facial model construction method provides the ability of changing the special 3-D facial model for animation

Manuscript received October 12, 2006. This work was supported in part by the Industrial Technology Research Institute and National Science Council, Taiwan, ROC, under the grant number NSC89-2213-E-155-032.

Jui-Chen Wu is with the Department of Electrical Engineering, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li, Taoyuan 320, Taiwan, ROC (e-mail: s919103@mail.yzu.edu.tw).

Yung-Sheng Chen is with the Department of Electrical Engineering, Yuan Ze University, 135 Yuan-Tung Road, Chung-Li, Taoyuan 320, Taiwan, ROC (corresponding author, phone: 886-3-4638800 ext 2409; fax: 886-3-4639355; e-mail: eeyschen@saturn.yzu.edu.tw).

I-Cheng Chang is with the Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien 97401, Taiwan, ROC (e-mail: icchang@mail.ndhu.edu.tw).

(Advance online publication: 24 May 2007)

applications. Fig. 1 outlines our automatic face modeling procedure.

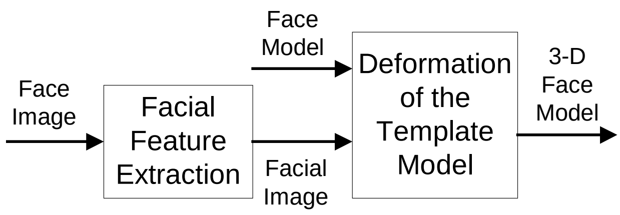


Fig. 1 Our face modeling procedure.

II. Proposed Approach

The presented modeling procedure is an integrated approach that can automatically extract the facial features from a 2-D image. The 3-D face is then modeled and generated via a deformable 3-D template. The main parts of this approach are described in this section. The face region is first segmented according to YCrCb color space, the facial features in the segmented face region are then located and extracted by means of a morphology technique. The details of the algorithms are described in the following subsections.

A. Face Segmentation

Many face segmentation algorithms have been developed and discussed in the past ten years. The method proposed by Chai and Nagan [8] is quite suitable for our current study. We modify their method to locate the desired facial region for further extracting facial features. The face segmentation algorithm includes analysis of skin color, skin density regularization, and geometric filtering.

Our face segmentation algorithm uses the color information to segment the facial regions. The pixels of the input image can be classified into skin color and non-skin color. The facial region has a special color distribution, which differs significantly from those of the background objects. Hence a skin-color reference filtering in YCrCb color space is adopted. In the CIE chrominance, a skin-color region can be identified by the presence of a certain set of chrominance (i.e., Cr and Cb) values, which belongs to a narrow and consistent distribution. Let $Skin_{Rcr}$ and $Skin_{Rcb}$ denote respectively the range of Cr and Cb values that correspond to skin color which subsequently define our skin-color reference filter. According to our experiments, the suitable ranges are found and given by $Skin_{Rcr} = [133,\ 173]$ and $Skin_{Rcb} = [77,\ 127]$ as skin-color reference filter. We have experimentally found that these filters are robust against different types of skin color. Consider an input image of $W \times H$ pixels, the facial region may be detected by using these ranges of chrominance values. In other word, we use it to filter out non-facial region of the source image. The skin segmentation can be defined as follows.

$$Skin_{color}(x,y) = \begin{cases} 1, & \text{if } Cr(x,y) \in Skin_{Rcr} \text{ and } Cb(x,y) \in Skin_{Rcb} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The point (x,y) is classified as a skin pixel and set to 1 if both the Cr and Cb values for the pixel fall inside their respective ranges $Skin_{Rcr}$ and $Skin_{Rcb}$. Otherwise, the point is classified as a non-skin pixel and set to 0.

However after performing skin segmentation, the segmented regions which may include other areas where the chrominance values coincide with those of skin color and noises. Those skin color pixels are like some objects with skin-color appearance in the background; whereas noises are like some small holes on the segmented facial region such as eyes and mouth. In order to distinguish between these two areas, we need to identify the segmented regions. The skin density distribution of skin-color pixels is used for this purpose. Let $(W/4) \times (H/4)$ array containing skin density values, called skin density map $Skin_{Density}(x,y)$, be computed as

$$Skin_{Density}(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} Skin_{color}(4x+i, 4y+j), \quad (2)$$

where $x = 0, \ldots, W/4-1$ and $y = 0, \ldots, H/4-1$. It divides the bitmap representing the skin-color pixels into non-overlapping groups of 4×4 pixels. The number of skin-color pixels within each group is counted and assigned to the corresponding point of skin density map.

In accordance with the skin density value, we classify each point into three types, namely, zero ($Skin_{Density} = 0$), intermediate ($0 < Skin_{Density} < 16$), and full ($Skin_{Density} = 16$). The case of zero density is displayed with white, intermediate density with gray, and full density with black, respectively. Once the density map is derived, the further process of skin density regularization algorithm containing the following three steps will be performed.

Skin Density Regularization Algorithm

Step 1: Discard all points at the boundary of the skin density map.

Step 2: Erode any full density point if it is surrounded with less than five full-density points in its local 3×3 neighborhood.

Step 3: Dilate any point of either zero or intermediate density if there are more than two full-density points in its local 3×3 neighborhood.

After performing this process, the skin density map will be converted into a bitmap image representing the density regularization result according to following equation. For the result of skin density regularization algorithm, one is shown with black and the other is shown with white.

$$Skin_{Density\_Reqularization}(x,y) = \begin{cases} 1, & \text{if } Skin_{Density} = 16 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Since the source image used in our study contains a head-and-shoulder person with a simple background, which will be effectively used in modeling 3-D human face, the human face size may be further considered as fixed size. Let the connected component analysis be first performed on the result of skin density regularization, the face candidate can then be selected by the geometric filtering defined as below.

$$Face_{candidate} = \bigcup_{i=1}^{n} \left\{ O_{(i)} \mid W_{O_{(i)}} > 150 \cap H_{O_{(i)}} > 250 \right\} \qquad (4)$$

Where $\left\{ O_{(1)}, O_{(2)}, \ldots, O_{(n)} \right\}$ is the obtained set of connected objects by applying the connected component procedure to the processed image. Their width and height of the connected objects are $\left\{ W_{O_{(1)}}, W_{O_{(2)}}, \ldots, W_{O_{(n)}} \right\}$ and $\left\{ H_{O_{(1)}}, H_{O_{(2)}}, \ldots, H_{O_{(n)}} \right\}$. After performing our face segmentation algorithm, the segmented face region and background can be obtained and illustrated in Fig. 2 (a) and 2(b), respectively.

(a) (b)

Fig. 2 Result of face segmentation.

B. *Facial Feature Locating*

After face segmentation, facial features are extracted next from the segmented image. In our approach, a coarse-to-fine method, which can increase the system efficiency to extract facial features, is presented. The coarse-to-fine method is first to locate facial features on the segmented image, and then to extract the facial feature from each located facial feature region. We locate facial features based on geometrical relationships between facial components such as nose, eyes, mouth and so on. For these facial features, the degree of darkness of eyebrow depends strongly on the density or color, whereas that of the nostril relies on how much they are occluded. As to the shape of mouth, it is frequently displayed in various appearances and size variances due to facial expressions. In addition, the eyes can be considered as a salient and quite stable feature of a face. Based on these observations, we locate eyes, lip, and eyebrows as well as nose sequentially. The method of locating facial features is described as follows.

We detect eyes based on its color different from the segmented image, and its symmetry property. The eye-analogue image is first found by filtering out the non-skin color pixel from the segmented face image. We then apply the connected component analysis to eye-analogue image $EAI_{i|i=0,\ldots,m}$ and compute the center point EAI_i^{center} and area EAI_i^{area} for each segment. Finally we utilize the relationships of the sizes and positions of eyes to design matching rules as described below.

Rule 1: The area of two eye-analogue segments must be nearest.

Rule 2: The distance between the center points of two eye-analogue segments must be larger than a fixed number.

Rule 3: One eye must be located by extension of a small range from the other eye.

These simple rules can well extract the eyes from eye-analogue image. After locating the eyes, we then locate lip region from the image. We observe that the lip intensity differs from the background. Therefore, we present a scheme for locating lips based on the use of chromatic information. Previous work [9] is applied to obtain lip area by detecting pixels being satisfactory to the following condition.

$$1.2 \le R/G \le 1.5 \qquad (5)$$

Where R and G are the red and green color components respectively. Through the above method, the candidates of extracted lip region may also include some noises where the other regions may have the same intensity value with that of the lip's region. Therefore in order to remove these noises, some steps including isolating, opening and closing operations are used for locating the lip shape.

Since facial features have relative geometry relations onto a face, to make more precisely, a face geometry relation for different environments, lighting, and orientation is designed with a set of faces. Combining this face geometry and the presented procedure of locating facial features, the final result can be obtained and illustrated in Fig. 3.

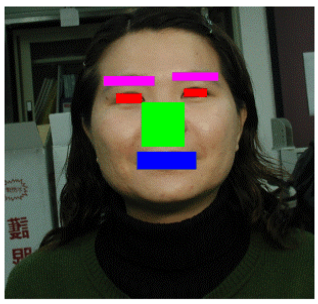


Fig. 3 Final result of locating facial features.

C. *Facial Feature Extraction*

To detect facial features from each facial region, we further use a morphology-based scheme to extract high-contrast regions as guide to locate corner features such as lip corner, eyes corner, and so on. According to each corner feature, the complete facial features can be obtained based on the adopted shape model. In the following, we first present eye, lip, and nose features extraction. Then, the extraction of eyebrow features is described. Finally, face contour is extracted based on some key feature points.

1) Eye and Lip Feature Extraction: In the past, most approaches adopt edge-based method to locate features. However, such an approach is easily affected by noise. In addition, the shape of a mouth or eye may be displayed in various appearances and size variances due to facial expressions. Fortunately, the corner feature is quite stable in various facial expressions. Hence the high-contrast feature is adopted for detecting corner feature. After getting corner features, the corresponding shape mode is utilized for detecting complete facial features. In what follows, details of the morphology-based scheme to locate high constant features are described.

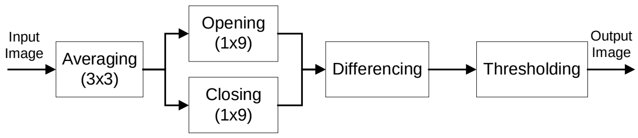


Fig. 4 Morphology operations.

The whole procedure of our morphology-based feature extraction is depicted in Fig. 4. In order to eliminate noise, an averaging operation with a structure element $S_{3,3}$ is applied first. Then the closing and opening operations with a structure element $S_{1,9}$ are performed on the averaged image such that the images I_c and I_o can be obtained. To detect horizontally edges, a differencing operation is further applied to the images I_c and I_o. All possible horizontally edges can be extracted with a thresholding operation. Finally a labeling process is performed to extract the horizontally segments. Through the above operations, all facial corner features can be well extracted as shown in Fig. 5.

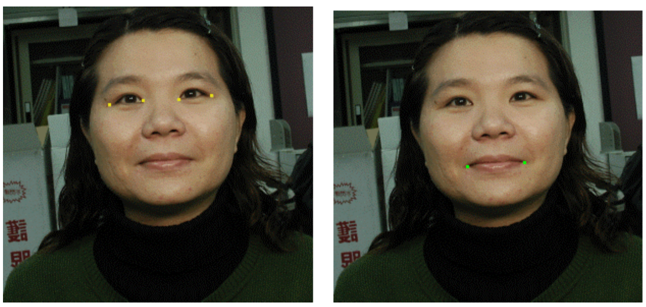


Fig. 5 Extraction results of corner features.

After obtaining corner features, the complete facial features are detected by a shape model. Since some key points are useful and must be extracted for face modeling, we use the color characteristics to detect the key points. Due to the high contrast of colors near to the edge parts of eyeball, it is suitable to apply skin color segmentation for estimating lower and upper eyelid positions. For the extraction of lip features, the horizontal edge and color information are used in our scheme, where Prewitt (horizontal) operation is used to extract upper lip features. However since the lower lip feature is sensitive by edge detection, it is alternatively estimated by color classification and symmetry property. Since the lip intensity in red is stronger than the surrounding skin, the lip area is first detected according to following equation. The pixel classification may be given as follows.

$$R_{Cr} = [145,170] \cap R_{Cb} = [104,120] \qquad (6)$$

Where R_{Cr} and R_{Cb} denote the corresponding ranges of Cr and Cb in YCrCb color space. Through this color filtering stage, the lip shape can be obtained completely. Although the mouth may also be displayed in various appearances, it inherently possesses a symmetry property. Hence the lower lip features can be detected easily. Fig. 6 shows the final eyes and lip feature detection results, which are regarded as facial key points.

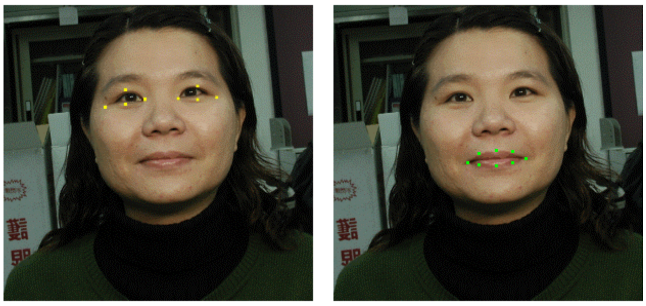


Fig. 6 Results of facial key points.

We utilize these facial key points to further derive the other facial features. The shape models shown in Table 1 are used to fit the facial features for each set of facial key points. The best position candidate is selected according to a high energy criterion. Finally, the complete facial features are extracted and shown in Fig. 7.

Table 1 Facial shape models.

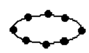
Components	Eye	Lip
Descriptions		

Fig. 7 Results of facial features.

2) Nose Feature Extraction: Inside the nose region, we first extract the line under the nostril features whose color is darker than the surrounding skin. Since the morphology-based method can extract high contrast to the background in the image, we adopt it to extract the portion having high contrast with respective to background and regard it as nostril feature. After extracting the high contrast features, the line under the nostril is estimated and extracted as shown in Fig. 8.

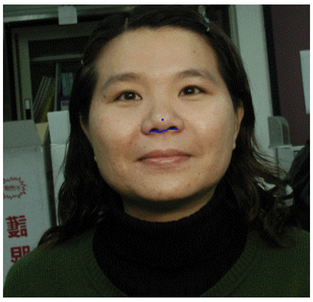


Fig. 8 Extraction of nose feature.

3) Eyebrow Feature Extraction: Inside these areas, the eyebrow is estimated assuming that its color is darker than the surrounding skin. Since the degree of darkness of eyebrow depends on the density or color, we first adopt erosion operation to enhance this property and then segment it by binarizing [10] the luminance image inside the eyebrow area. The extracted result is shown in Fig. 9.

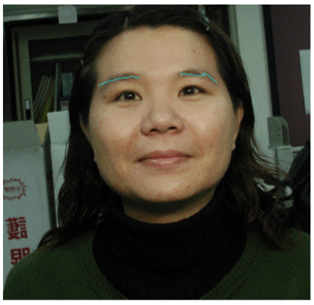


Fig. 9 Extraction of eyebrow features.

4) Face Contour Extraction: The face localization stage may contain the neck, it is necessary to find the face contour for further modeling. The face image brightness is involved in our method to extract face contour. First, we select eight control points inside the face image based on the positions of facial features as shown in Fig. 10. Fig. 11(a) shows the extracted initial face contour by the defined positions. Here we observe that the face image brightness gradually decreases from central portion of face toward its boundary zone, and the changing rate of brightness increases toward the face boundary zone. On account of these properties, we introduce the following image convergence function for each control point *i*.

$$C_{image}(i) = \alpha C_{bri}(i) + \beta C_{dif}(i) \tag{7}$$

Where α and β denote weighting factors. $C_{bri}(i)$ is the mean value representing the brightness per pixel on the edge connecting to control points. $C_{dif}(i)$ is the sum of all the brightness differences between two adjacent pixels on the edge *i*. Fig. 11(b) shows the final face contour obtained by this procedure.

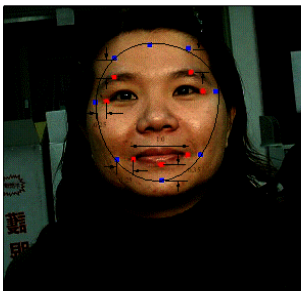


Fig. 10 Definition of initial face contour.

(a) (b)

Fig. 11 Results of face contouring.

D. *Mapping Generic Face Model*

In this section, we introduce an approach of deforming the template model to fit the human face image. The control points on the template model are manually selected, and correspond to the points of extracted facial features in a facial image. After obtaining all facial feature points, control points on the template model will be adjusted to coincide with the features. A 3-D human face model is then constructed by deforming the template model to fit the facial images. At the first, the features on the facial images are extracted and their corresponding positions on the projection of the template model are selected. Once the control points are selected, their corresponding nearest 3-D vertices on the template model will be estimated, and then the control points will be reshaped by projecting the estimated vertices onto the projection of the template model. The selecting order needs not to be the same as the one in the facial image. Topological constraints of the facial features are used to establish the correspondence between the features and control points automatically. The silhouettes of the facial images are matched first with the projection of the template model. Then the correspondences of mouth, eyes are established. The lip separation, eyebrows and nose are matched finally based on their spatial relationship. After adjusting the template model to fit the appearance of the special person's face, the texture mapping for the 3-D model is performed. However, the images captured under different perspectives may cause the visual defects at the boundary where the images stitch due to different lighting conditions, varying pose, and changing camera setting. To eliminate the shading inconsistency problem, a multilevel texture blending process [7] is applied to the images for extracting textures and texture boundary pixels. The main idea is to iteratively subdivide the triangles on the 3-D face model and corresponding triangles on the 2-D texture map, until each triangle reaches some designated size. The texture value at each end point of the triangle in 2-D face image is assigned to the corresponding vertex on the 3-D face model.

III. Experimental Results

The proposed system is implemented on a MS-Window based PC with Pentium VI 2.8G CPU inside and the programming environment is under Visual C++ 6.0. The image acquisition device is a commercial digital camera. The flowchart of the system is shown in Fig. 12.

A. Face Facial Feature Extraction

Several images have been used to verify our feature extraction algorithm, where some results depicting different facial expressions are shown in Fig. 13. From these results, our algorithm can well detect the facial features even the face has slight rotation such as Fig. 13(c) shows. They demonstrate that our algorithms can extract the facial features effectively.

B. 3-D Model Generation

A 3-D template model and facial image are first fed into the system. The adopted template model [7] is adopted and given in Fig. 14. The template model is then deformed to fit the extracted facial features in the images. Finally textures from the images are extracted, blended and mapped onto the deformed model. In our approach, the facial features are automatically extracted including face segmentation, facial features location and extraction. Fig. 15 shows an example of texture mapped face model. Fig. 16 shows different views of the reconstructed 3-D human face model.

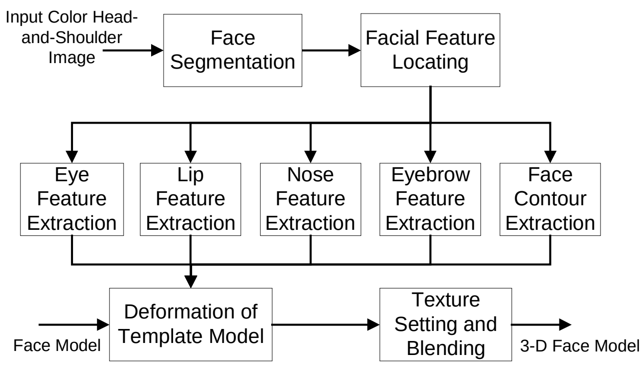


Fig. 12 Flowchart of our system.

Fig. 13 Some results of feature extraction.

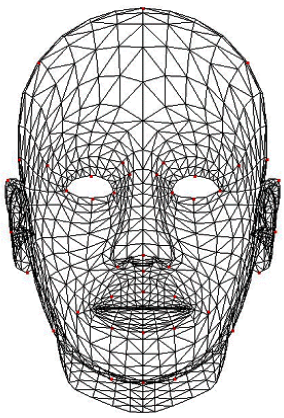


Fig. 14 The 3-D facial template model [7].

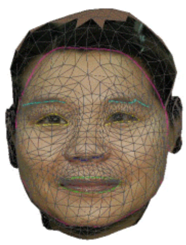


Fig. 15 A texture mapped face model.

Fig. 16 Different views of the reconstructed 3-D face model.

IV. CONCLUSIONS

In this paper, we have presented an automatic approach to the extraction of facial features, which can facilitate to produce the realistic 3-D face model by means of mapping a set of deformable features onto a 3-D template model. Based on a 2-D head-and-shoulder color image, the main stages of our approach include face segmentation, facial feature locating, facial feature extraction, and face contour extraction. Our experiments have confirmed the feasibility of the proposed system and the ability of changing the special 3-D facial model for further animation applications.

REFERENCES

[1] N. M. Thalman, P. Kalra, and M. Escher, "Face to virtual face," *Proc. of the IEEE*, vol. 86, no.5, May 1998, pp. 870-883.

[2] W. H. Leung and T. Chen, "Creating a multiuser 3-D virtual environment," *IEEE Signal Processing Magazine*, May 2001, pp. 9-16.

[3] Y. Lu and B. Yuan, "A deforming method for model-based 3-D individualized face from photographs," *Proceedings of ICSP*, vol.2, 2000, pp.1255-1258.

[4] S. Morishima, "Face analysis and synthesis," *IEEE Signal Processing Magazine*, vol. 18, no. 3, May 2001, pp. 26-34.

[5] W. S. Lee, P. Kalra and N. M. Thalmann P., "Model based face reconstruction for animation," *Proc. Multimedia Modeling*, 1997, pp. 323-338.

[6] W. S. Lee and N. M. Thalmann, "From real faces to virtual faces: problems and solutions," *Proc. 3IA'98, Limoges (FRANCE)*, 1998, pp. 5-19.

[7] H. L. Chou and C. C. Chen, "A photo-realistic facial model generation method using the template-based deformation technique," *Proc. of 16th IPPR Conference on Computer Vision, Graphics, and Image Processing*, Kinmen, Taiwan, Aug. 2003, pp. 584-589.

[8] D. Chai and K. N. Nagan, "Face segmentation using skin-color map in videophone applications," *IEEE Trans. on Circuits and System for Video Technology*, vol. 9, no. 4, June 1999, pp.551-564.

[9] T. Wark and S. Sridharn, "A syntactic approach to automatic lip feature extraction for speaker identification," *IEEE Proc. of International Conference on Acoustics, Speech and Signal Processing*, vol. 6, 1998, pp. 3693-3696.

[10] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 2, Feb. 1988, pp. 233-260.

Jui-Chen Wu was born in Taiwan, R.O.C., on November 22, 1976. She received the B.S. degree from Chung Yuan Christian University and the M.S. degree from Yuan Ze University, Chung-Li, Taiwan, in 2000 and 2002, respectively. She is currently working toward the Ph.D. degree in the Department of Electrical Engineering, Yuan Ze University, Chung-Li, Taiwan, R.O.C.

Her research interests include image processing and pattern recognition.

Yung-Sheng Chen was born in Taiwan, R.O.C., on June 30, 1961. He received the B.S. degree from Chung Yuan Christian University, Chung-Li, Taiwan, in 1983 and the M.S. and Ph.D. degrees from National Tsing Hua University, Hsinchu, Taiwan, in 1985, and 1989, respectively, all in electrical engineering.

In 1991, he joined the Electrical Engineering Department, Yuan Ze Institute of Technology, Chung-Li, where he is now a Professor. His research interests include human visual perception, computer vision and graphics, circuit design, and website design.

Dr. Chen received the Best Paper Award from the Chinese Institute of Engineers in 1989 and an Outstanding Teaching Award from Yuan Ze University in 2005. He has been listed in the Who's Who of the World since 1998 and awarded with The Millennium Medal from The Who's Who Institute in 2001. He is a member of the IEEE, and the IPPR of Taiwan, R.O.C.

I-Cheng Chang was born in Taiwan, R.O.C., in 1965. He received the B.S. degree in nuclear engineering in 1987, and the M.S. and Ph.D. degrees in electrical engineering in 1991 and 1999, respectively, all from National Tsing Hua University, Hsinchu, Taiwan.

In 1999, he joined Opto-Electronics & Systems Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, as an engineer and project leader. In the autumn of 2003, he joined the Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien, Taiwan, and he is currently an assistant professor. His research interests include image/video processing, computer vision and graphics, and multimedia system design.

Dr. Chang received the Annual Best Paper Award from Journal of Information Science and Engineering in 2002, and the Research Awards from Industrial Technology Research Institute in 2002 and 2003. He is a member of the IEEE, and the IPPR of Taiwan, R.O.C.