FACIAL FEATURE EXTRACTION AND IMAGE WARPING USING PCA BASED STATISTIC MODEL

Zhong Xue [1], *Stan Z. Li* [2], *Eam Khwang Teoh* [1]

[1] School of Electrical & Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore, 639798

[2] Microsoft Research China
Zhi Chun Road, Haidian District
Beijing, 100080, P.R. China

Emails: zhong_xue@yahoo.com, eekteoh@ntu.edu.sg, szli@microsoft.com

ABSTRACT

A new algorithm is proposed to extract the facial features and estimate the control points for facial image warping using the Principle Component Analysis (PCA) based statistic face model. In this algorithm, first a full-face model consisting the contour points and the control points is built. Based on a number of manually marked training samples, the prior distribution of the full-face model can be obtained by using the PCA. Given an input face image, first the contour points are obtained by using the recently developed Bayesian Shape Model (BSM), and then the control points are estimated from the contour points. Finally, the extracted face path is normalized using the piece-wise affine triangle warping algorithm. Experimental results illustrate the effectiveness of the proposed algorithm.

1. INTRODUCTION

In face recognition, one of the main procedures is first matching and extracting the face patch, and then warping it to the standard view and the normal expression. In this process, the accuracy in extraction and modeling the face is crucial, *e.g.* in [1], the normalized frontal view face is not visually satisfied since there is no face extraction performed.

The Active Shape/Appearance Models (ASM & AAM) have been demonstrated to be successful in facial feature matching[2, 3]. However, in the ASM, the reconstructed prototype or its geometrical transformed version is regarded as the final matching result, and hence the accuracy of matching is largely depended on the training samples selected. Provided there were enough samples, ASM can globally match the object shapes very well, but with lower local accuracy, *i.e.* the performance deteriorates under local shape variations. To overcome this disadvantage, the Bayesian Shape Model (BSM) is developed [4], where both the global and local shape deformations are considered: the global shape variations are modeled by the prior distribution of the marked samples and the local deformations are constrained by the measures between the deformable model and the prototype. In this way, object shapes can be extracted with improved accuracy.

Generally, the feature points used by the ASM or BSM are the outline contours of the facial features, *i.e.* the outlines of the face, eyebrows, eyes, nose and mouth (see Fig.1). Although these points are enough to describe the shape of a face, they are insufficient in dealing with face image warping/normalization. To solve the problem, in [3], the Delauney triangulation is utilized to partition the area that is surrounded by the contour points into a set of triangles. However, this algorithm does not consider the characteristics of the expression/action of the face, and the triangles obtained differ greatly when the position of the contour points changes.

In this paper, a 2-D full-face model is built based on the *CANDIDE* model, which consists of the contour points and the control points. The contour points include the outline of face, eyebrows, eyes, nose and mouse; the control points are the feature points used for facial image warping. Based on a number of manually marked training samples, the prior distribution of the full-face model is obtained by using PCA. Given an input face image, first the contour points of the face can be matched accurately by using the recently developed Bayesian Shape Model (BSM), and the control points are determined by the matching results according to the prior distribution of the full-face model. As a result, the facial image warping can be performed very easily using the piece-wise affine algorithm based on the full face model.

Experimental results for normalizing the BSM matching results are presented to demonstrate the effectiveness of the proposed algorithm.

2. MODELING THE FULL FACE USING PCA

The full-face model consists of two point sets: 1. Contour points: the outline contours of the face, the eyebrows, eyes,

0-7803-6725-1/01/$10.00 ©2001 IEEE

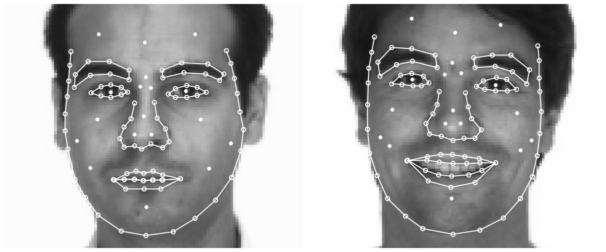

Fig. 1. Two examples of the manually marked face. 'o': contour points, '*': control points.

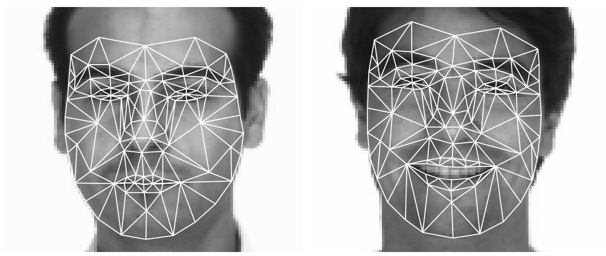

Fig. 2. Link the model points to triangles.

nose and mouse; 2. Control points: the points that are useful for face image warping, but lack image features.

Denoting the full shape model of the face as $\bar{s}$, N refers to the number of the points,

$$\bar{s} = [\begin{array}{c} \bar{f} \\ \bar{c} \end{array}]. \tag{1}$$

$\bar{f}$ represents the contour points and $\bar{c}$ indicates the control points. Fig.1 shows two examples of the marked face, where 'o' is the contour point representing the outlines of the face, eyebrows, eyes, nose and mouth respectively and '*' stands for the control point.

The PCA is utilized to build the shape model of the face. First, all the sample face images are manually marked and normalized/aligned to a standard view by using least square errors method. Then, we use the following steps to model the shape variations.

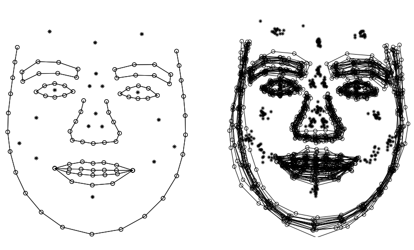

Fig. 3. The mean and normalized face models.

i). Compute the mean of the data, $\bar{s} = \frac{1}{D}\sum_{i=1}^{D}\bar{s}_i$, where D is the number of samples.

ii). Computer the covariance of the data,

$$M = \frac{1}{D-1}\sum_{i=1}^{D}(\bar{s}_i - \bar{s}_0)\cdot(\bar{s}_i - \bar{s}_0)^T. \tag{2}$$

iii). Compute the eigenvectors, ϕ_i, and corresponding eigenvalues e_i of M (sorted so that $e_i \geq e_{i+1}$).

Therefore, the shape of a face can be reconstructed from the mean $\bar{s}_0$ and the shape parameter, w_s:

$$\bar{s} = \bar{s}_0 + \Phi w_s \tag{3}$$

where Φ is the matrix containing the t eigenvectors corresponding to the largest eigenvalues, $\Phi = [\phi_1|\phi_2|...|\phi_t]$. When $\bar{s}$ is known, the shape parameter w is given by

$$w_s = \Phi^T(\bar{s} - \bar{s}_0) \tag{4}$$

Normalizing input face image removes the effect of position/expressions and gives more reliable recognition results. It can be performed by extracting the facial features and determining the control points of the face model, and then applying image warping.

3. MATCHING FACIAL FEATURES USING BSM

The BSM formulates the matching of a deformable model to the object in a given image as maximizing *a posteriori* (MAP) estimation problem [5]. According to the Bayesian estimation, the joint posterior distribution of f and $\bar{f}$, $p(f,\bar{f}|d)$, is

$$p(f,\bar{f}|d) = \frac{p(d|f)p(f,\bar{f})}{p(d)} \tag{5}$$

where $p(d|f) = p(d|f,\bar{f})$ is the likelihood distribution of input image data d.

$$p(f,\bar{f}) = p(f|\bar{f})p(\bar{f}) \tag{6}$$

is the joint prior distribution of f and $\bar{f}$.

For a given image d, the MAP estimates, f_{MAP} and $\bar{f}_{MAP}$, are defined as,

$$\begin{aligned}\{f_{MAP},\bar{f}_{MAP}\} &= \arg\max_{f,\bar{f}}\{p(f,\bar{f}|d)\} \\ &= \arg\max_{f,\bar{f}}\{\frac{p(d|f)p(f|\bar{f})p(\bar{f})}{p(d)}\}\end{aligned} \tag{7}$$

where $p(\bar{f})$ is the distribution of the prototype $\bar{f}$, $p(f|\bar{f})$ describes the variations between f and $\bar{f}$, and $p(d|f)$ indicates the matching between f and the salient features of the object in image d. Provided the densities in Eq.(7) can be modeled as Gibb's distribution, maximizing the posterior distribution is equivalent to minimizing the corresponding energy function of the contour:

$$\{f_{MAP}, \bar{f}_{MAP}\} = \arg\min_{f,\bar{f}} \{E_{contour}\} \tag{8}$$

where $E_{contour} = E_{con} + E_{int} + E_{ext}$.

The constraint energy term E_{con} of the prototype contour is caused by the prior distribution, $p(\bar{f})$, which can be approximated by $p(\bar{s})$ and modeled by a single Gaussian distribution [6].

$$p(\bar{f}) \doteq p(\bar{s}) = \frac{\exp(-\frac{1}{2}\sum_{i=1}^{t}\frac{w_i^2}{e_i})}{(2\pi)^{t/2}\prod_{k=1}^{t} e_k^{1/2}} \tag{9}$$

where w is given in Eq.(4). Since $\bar{c}$ is unknown, it is set to $\bar{c}_0$. Experiments show that this approximation is acceptable because generally the variations of $\bar{f}$ can represent the variations of the full-face model. Therefore, the constraint energy is denoted as

$$E_{con} = \frac{1}{2}\sum_{i=1}^{t}\frac{w_i^2}{e_i} \tag{10}$$

The variations of the prototypes contour is limited by the plausible area of the corresponding shape parameters w,

$$\sum_{i=1}^{t}\frac{w_i^2}{e_i} \leq M_t \tag{11}$$

The threshold, M_t, is chosen using the χ^2 distribution [2].

On the other hand, BSM uses a transformational invariant internal energy term, which describes degree of matching between the deformable model f in the image domain and the prototype $\bar{f}$ in the shape domain. Mathematically, $\bar{f}$ and f are related by $f_i = A\bar{f}_i + T + \epsilon, \ \ (1 \leq i \leq N)$, where A is the transformational matrix, T is a translation vector. The invariant internal energy term is defined as (see [7] about the detail of Eq.(14):

$$\begin{aligned} E_{int}(f|\bar{f}) &= E_{gint}(f|\bar{f}) + \frac{1}{N}\sum_{i=1}^{N} E_{lint}(f_i|\bar{f}) && (12) \\ E_{gint}(f|\bar{f}) &= \frac{1}{N}\sum_{i=1}^{N}[(\bar{f}_i - \hat{A}^{-1}(f_i - \hat{\mathbf{t}}))^T \\ &\quad \cdot(\bar{f}_i - \hat{A}^{-1}(f_i - \hat{\mathbf{t}}))] && (13) \\ E_{lint}(f_i|\bar{f}) &= \frac{(S_1 + S_2)AREA_{proto}}{AREA_{aligned}} && (14) \end{aligned}$$

The external energy term can be calculated as follows [8].

First, the image $d = \{d(x,y)\}$ is smoothed using Gaussian function, $G_\sigma(x,y)$, $d_\sigma(x,y) = G_\sigma(x,y) * d(x,y)$. Second, the normalized gradient of the smoothed image d_σ at each pixel location (x,y), is computed:

$$d_\sigma^g(x,y) = (d_{\sigma x}^g(x,y), d_{\sigma y}^g(x,y)), (||d_\sigma^g(x,y)|| \in [0,1]).$$

, and finally, the external energy is defined as:

$$E_{ext}(d \mid f) = \sum_{i=1}^{N}(1 - ||d_\sigma^g(x_i,y_i)||)|\mathbf{n}\cdot\mathbf{h}| \tag{15}$$

$\mathbf{h}(x_i,y_i)$ is the direction (unit vector) of the gradient $d_\sigma^g(x_i,y_i)$, and $\mathbf{n}(x_i,y_i)$ indicates the normal vector of the contour f at point $f_i = (x_i,y_i)$, with $||\mathbf{n}(x_i,y_i)|| = 1$ and $\mathbf{n}(x_i,y_i) = [\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}]\mathbf{v}_i/||\mathbf{v}_i||$, $\mathbf{v}_i = \frac{f_{i+1}-f_i}{||f_{i+1}-f_i||} + \frac{f_i - f_{i-1}}{||f_i - f_{i-1}||}$ is the tangent vector of contour f at point f_i.

4. DETERMINING THE CONTROL POINTS FOR FACIAL IMAGE WARPING

Using the BSM facial feature extraction algorithm, the matching results of the contour points f and $\bar{f}$, and the pose relationship A and T (between f and $\bar{f}$) can be obtained. Then, the contour points in the shape domain is calculated by

$$\bar{f}' = A^{-1}(f - T). \tag{16}$$

The algorithm to determine the control points corresponding to $\bar{f}'$ is to estimate the shape parameter $\hat{w}_s$, so that the contour points $\hat{\bar{f}}$ of the reconstructed model $\hat{\bar{s}} = \begin{bmatrix} \hat{\bar{f}} \\ \hat{\bar{c}} \end{bmatrix}$ match $\bar{f}'$ closely, and hence the corresponding reconstructed control points $\hat{\bar{c}}$ can be regarded as the estimate of the control points. The detail algorithm is follows,

i). The estimation problem is to find out $\hat{w}_s$, so that the result $\hat{\bar{f}}$ matches $\bar{f}'$ in the sense of least square errors. From Eq.(3), $\hat{\bar{f}}$ can be calculated by

$$\begin{aligned} \hat{\bar{f}} &= [I\ \mathbf{0}]\begin{bmatrix} \hat{\bar{f}} \\ \hat{\bar{c}} \end{bmatrix} = [I\ \mathbf{0}](\bar{s}_0 + \Phi\hat{w}_s) \\ &= [I\ \mathbf{0}]\bar{s}_0 + [I\ \mathbf{0}]\Phi\hat{w}_s && (17) \end{aligned}$$

where I is a unity matrix and $\mathbf{0}$ is a zero matrix. Denoting $\bar{f}_0 = [I\ \mathbf{0}]\bar{s}_0$, $\Phi_f = [I\ \mathbf{0}]\Phi$, and noting that $\bar{f}_0$, Φ_f, and $\bar{f}'$ are known, the shape parameter $\hat{w}_s$ can be estimated using

$$\hat{w}_s = {\Phi_f}^T(\bar{f}' - \bar{f}_0) \tag{18}$$

ii). The control points $\hat{\bar{c}}$ can be calculated through $\hat{w}_s$,

$$\hat{\bar{c}} = [\mathbf{0}\ I](\bar{s}_0 + \Phi\hat{w}_s) \tag{19}$$

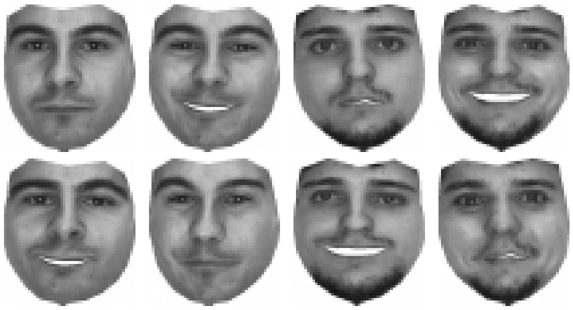

Fig. 4. The original (upper row) and the warped (lower row) face images.

iii). The final full-face contour in the shape domain is represented by $\hat{\hat{s}} = \begin{bmatrix} \bar{f}' \\ \hat{\hat{c}} \end{bmatrix}$ and hence its transformed counterpart in the image domain, $s = \begin{bmatrix} f \\ c \end{bmatrix}$ is $s \approx \hat{s} = A\hat{\hat{s}} + T$.

Fig.4 shows several examples of the warping results using the proposed face model. The upper row is the original faces, while the bottom row is the warped images of the upper row correspondingly. For example, the first column of the images illustrates changing the expression to smile, and the second column stops the smile. It can be seen from the figure that the face model describes the face well and by image warping, the expressions of face can be removed/changed, which is very useful for face recognition.

5. EXPERIMENTAL RESULTS

The experiments are carried out using the BSM toolkit developed in the Matlab 5.3 Environment. Given an input face image, the BSM is utilized to match the contour points of the face model. The matching results are then utilized to estimate the control points of the face model to obtain the extracted full face in the image domain. Finally, an image warping algorithm is performed to transfer the extracted face into the mean of the face model in the shape domain. Fig.5 presents examples of the experiment results using the proposed algorithm. The left column is the input face image, the middle column shows the extracted contour points, and the normalized faces are plotted in the right column.

6. CONCLUSION

In this paper, a full-face model consisting the contour points and the control points is built based on PCA, and an algorithm is proposed to estimate the control points from the contour points that are matched by the BSM facial feature extraction algorithm. Then the warping and normalization of the extracted face patch can be performed by using piecewise affine algorithm. Experimental results demonstrate good performance of the proposed algorithm.

Fig. 5. Facial feature extraction and face normalization/warping results.

7. REFERENCES

[1] W. Zhao and R. Chellappa, “Sfs based view synthesis for robust face recognition,” in *4th IEEE Conference on Automatic Face and Gesture Recognition, Grenoble, France*, 2000, pp. 285 –292.

[2] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, “Active shape models - their training and application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[3] T. Cootes, G.J. Edwards, and C.J. Taylor, “Active appearance models,” in *Proceeding of 5th European Conference on Computer Vision*, 1998, vol. 2, pp. 484–498.

[4] Z. Xue, Stan Z. Li, J.W. Lu, and E.K. Teoh, “Bayesian model for extracting facial features,” in *Sixth International Conference on Control, Automation, Robotics & Vision, ICARCV 2000, Dec., Singapore*, 2000.

[5] Stan. Z. Li and J. Lu, “Modeling Bayesian estimation for deformable contours,” in *Proceedings of 7th IEEE International Conference on Computer Vision, Kerkyra, Greece.*, 1999, pp. 991–996.

[6] B. Moghaddam and A. Pentland, “Probabilistic visual learninig for object representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[7] H.H.S. Ip and D. Shen, “An affine-invariant active contour model (AI-Snake) for model-based segmentation,” *Image and Vision Computing*, vol. 16, no. 2, pp. 125–146, 1998.

[8] A.L. Yuille, P.W. Hallinan, and D.S. Cohen, “Feature extraction from faces using deformable templates,” *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, 1992.