Human Facial Feature Extraction for Face Interpretation and Recognition

Chung-Lin Huang and Ching-Wen Chen
Institute of Electrical Engineering
National Tsing Hua University
Hsin Chu, Taiwan 30043
Rep. of China

ABSTRACT

This paper presents facial features extraction algorithms which can be used for automated visual interpretation and recognition of human faces. Here, we can capture the contours of eye and mouth by deformable template model because of their analytically describable shapes. However, the shapes of eyebrow, nostril and face are difficult to model using deformable template. We extract them by using active contour model (snake).

I Introduction

Facial feature extraction has become an important issue in automated visual interpretation and recognition of human faces[1-3, 7-9]. Face recognition and interpretation has many applications such as security system, credit-card verification, criminal identifications, and teleconference. Though it is clear that people are good at face identification, it is not at all obvious how faces are encoded or stored, much less how familiarity affects these complex processes. In this paper, we propose a facial feature extraction method for face interpretation and recognition.

In the area of feature selection, the question has been addressed in studies of cue salience in which discrete features such as the eyes, mouth, chin, and nose have been found to be important cues for discrimination and recognition of faces. We develop two different methods to get the contours of eyes, eyebrows, mouth, nose and face. For different facial contours, we use different models to extract them from the original portrait. Because the shapes of eyes and mouth are similar to some geometric figures, we extract them in terms of deformable template model[3]. The other facial features, i.e., eyebrows, nose and face are so variable that we must extract them by active contour model. We illustrate the two models in the following:

(1) *Deformable template model*[3]. The deformable templates are specified by a set of parameters which uses *a priori* knowledge about the expected shape of the features to guide the contour deformation process. The templates are flexible enough to change their sizes and other parameter values, so as to match themselves to the data. The final values of these parameters can be used to describe the features. The deformable templates interact with the image in a dynamic manner. An energy function is defined which contains terms attracting the template to salient features such as peaks and valleys in the image intensity, edges and intensity itself. The minima of the energy function corresponds to the best fit with the image.

(2) *Active contour model (Snake)*[4-6]: Active contour is an energy-minimizing spline guided by the external constraint forces and influenced by image forces that pull it toward features such as lines and edges. Snakes lock onto nearby edges, localizing them accurately. Because snake is an energy-minimizing spline we need to develop energy functions whose local minima comprise the set of alternative solutions available to higher-level processes. Selection of an answer from this set is accomplished by the addition of energy terms that push the model toward the desired solution.

II Feature Extraction using Deformable Template

Before performing feature extraction, we need to know (1) the brightness thresholds that can be used to discriminate the features and other areas in face; and (2) the *rough* contour of each feature as the initial contour for iteration. Here, we use the scale space filter (SSF) to determine the zero-crossing intensity histogram at different scales and rough contour estimation routine (RCER) to roughly find the position of the target contour. Usually, the *rough* (initial) contour is a little larger than the precise contour of each feature. The only exception is the *rough* contour of face which is smaller than its precise contour. The details of the SSF and RCER are discussed in [10].

After the *rough* contour is obtained, the next step of face recognition is to find the physical contour of each feature. Conventional edge detectors can not find facial features (the contours of eye or mouth) accurately from local evidence of edges, because they can not organize local information into a sensible global perception. There is a method to detect the contour of eye using deformable template. This template is specified by a set of parameters which enables *a priori* knowledge about the expected shapes of the features to guide the detection processes. These templates are flexible enough to be able to change their size and shapes by changing analytic parameter values of the template, so as to match themselves to the data. The final values of these parameters can be used to describe the features.

II.1 Eye Template

Here, we discuss the deformable template model that can be applied to our system. The deformable template acts on three representations of the image, as well as on the image itself. The first two representations are peak and valleys in the image intensity and the third is the place where the image intensity changes quickly. Our simplified model are illustrates in Figure 1.

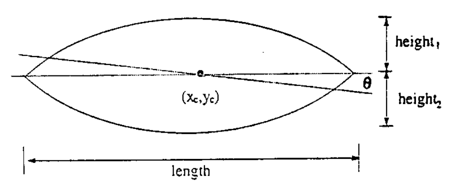

Figure 1. Simplified Eye Template.

Because the rough contour of mouth is pre-determined by RCER, we assume that the possible ranges of *length*, $height_1$, $height_2$ and *orientation* are: (a) *length* ϵ $[l-\alpha_1, l+\alpha_2]$, (b) $height_1$ ϵ $[h_1-\beta_1, h_1+\beta_2]$, (c) $height_2$ ϵ $[h_2-\gamma_1, h_2+\gamma_2]$, (d) *orientation* ϵ $[\theta-\theta_1, \theta+\theta_2]$. The tolerance α_i, β_i, γ_i and θ_i, $i=1,2$, are determined by RCER. (P_x, P_y) represents the centroid point. (x', y') is the translated and rotated version of (x, y) and their relationships are given by:

$$x-P_x=x'\cos\theta+y'\sin\theta, \quad y-P_y=-x'\sin\theta+y'\cos\theta \qquad (1)$$

The total energy function is defined as:

0-8186-2915-0/92 $3.00 © 1992 IEEE

$$E_{total} = E_{edge} + E_{white} + E_{black} \tag{2}$$

The E_{edge}, E_{white} and E_{black} are defined in the following: (i) The edge potentials are given by the integral over the curves of upper and lower parabola divided by the length:

$$E_{edge} = -\frac{w_1}{Upper\text{-}length}\int_{Upper\text{-}bound}\phi_{edge}(x,y)\ dS - \frac{w_2}{Lower\text{-}length}\int_{Lower\text{-}bound}\phi_{edge}(x,y)\ dS \tag{3}$$

where *Upper-bound and Lower-bound* represent upper part and lower part of eye, $\phi_{edge}(x,y)$ represents the edge response of point (x,y). (ii) The potentials of white and black points are defined as the integral over the area bounded by the upper and lower parabola divided by the area:

$$E_{w\text{-}b}=\frac{1}{Area}\iint_{Para\text{-}area}(-w_b N_{black}(x,y)+w_w N_{white}(x,y))dA \tag{4}$$

where $N_{black}(x,y)$ *and* $N_{white}(x,y)$ represent the numbers of black and white points, w_b *and* w_w are weights related with black and white points. In order not to be affected by improper threshold (θ), we define the tolerance (ϵ) for determining the black and white points (for Eq. 4) as:

$$\begin{aligned}&P(x,y)\ \text{is a Black point if}\ I(x,y)\le\theta-\epsilon\\&P(x,y)\ \text{is a White point if}\ I(x,y)\ge\theta+\epsilon\\&P(x,y)\ \text{is an umbignous point if}\ \theta-\epsilon\le I(x,y)\le\theta+\epsilon\end{aligned} \tag{5}$$

Having known the *rough* contour via RCER, we can extract the *precise* contour of eye. Using the energy functions defined above, we calculate the energy in the range of little modulations of *length, height*$_1$, *height*$_2$, *orientation*. When the minimum energy value takes place, the *precise* contour is extracted.

II.2 Mouth Template

In the whole features of front view face, the role of the mouth is relatively important. The mouth template is shown in Fig.2. We assume the possible ranges of *length, length*$_u$, *height*$_s$, *height*$_d$ and *height*$_u$ as: (a) $length \in [l-\alpha_1, l+\alpha_2]$, (b) $length_u \in [l_u-\beta_1, l_u+\beta_2]$, (c) $height_s \in [h_s-\gamma_1, h_s+\gamma_2]$, (d) $height_d \in [h_d-\delta_1, h_d+\delta_2]$ and (e) $height_u \in [h_u-\sigma_1, h_u+\sigma_2]$. The tolerance, α_i, β_i, γ_i, δ_i, σ_i, $i=1, 2$, are determined by RCER.

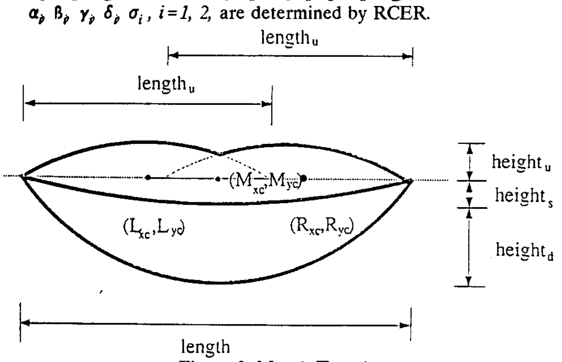

Figure 2. Mouth Template.

The mouth template is represented as three parts:
(i) The equation of middle lips, (i,e, crack between upper and lower lips) is given as:

$$y'=height_s-\frac{height_s}{length^2}\cdot x'^2\ ,\ x'\in[M_{xc}-\frac{length}{2},\ M_{xc}+\frac{length}{2}] \tag{6}$$

(ii) The equation of the lower lip is defined as:

$$y'=height_d-\frac{height_d}{length^2}\cdot x'^2,\ x'\in[M_{xc}-\frac{length}{2},\ M_{xc}+\frac{length}{2}] \tag{7}$$

(iii) The equation of the upper lip consists of two parts:

$$\begin{aligned}&y'=height_u-\frac{height_u}{length_u^2}\cdot x'^2,\ x'\in[L_{xc}-\frac{length_u}{2},\ M_{xc}]\\&y'=height_u-\frac{height_u}{length_u^2}\cdot x'^2,\ x'\in[M_{xc},\ R_{xc}+length_u]\end{aligned} \tag{8}$$

Because of the effect of the brightness in the picture-taking period, the middle of the lower lip may not be apparent. RCER cannot find the approximate height of lower lip. Fortunately, the length of the mouth still can be found by RCER. With the *a priori* knowledge of mouth shape (i.e. the height of lower lip is between one-fourth and one-sixth of the mouth's length) we can define the mouth contour energy function for the *precise* contour extraction algorithm. Mouth contour energy function consists of edge terms E_{edge} and black terms E_{black}. The edge term dominates at the edge area, whereas the black term encloses as many black points belonging to mouth as possible. They are described as:

$$E_{total} = E_{edge} + E_{black} \tag{9}$$

(i) The *edge energy function* consists of three parts: middle lip (gap between lips), lower lip and upper lip separated at philtrum. The equation for middle lip part is

$$E_{edge}=-\frac{w_{lower}}{Lower\text{-}length}\int_{Lower}\phi_{edge}(x,y)\ dS -\frac{w_{lef}}{Lef\text{-}length}\int_{Lef}\phi_{edge}(x,y)\ dS-\frac{w_{righ}}{Righ\text{-}length}\int_{Righ}\phi_{edge}(x,y)\ dS \tag{10}$$

where *Lower* represents the lower boundary of mouth, *Lef* represents the left part of upper lip, *Righ* represents the right part of upper lip, and $\phi_{edge}(x,y)$ represents edge response of point *(x,y)*.

(ii) The *black energy function* is designed to enclose as many black points belonged to mouth as possible, it is defined as:

$$E_{black} = \frac{1}{Area}\iint_{Lbound}^{Ubound}-w_{black}N_{black}(x,y)\ dA + \frac{1}{Mid\text{-}length}\int_{Mid}-w_{mid}N_{black}(x,y)\ dS \tag{11}$$

where *Lbound* represents lower lips, *Ubound* represents upper lip, *Mid* represents gap between lips. The number of black points, N_{black}(x,y), is defined by Eq. 5. The weights w_{black}, w_{mid}, w_{lower}, w_{lef} and w_{righ} are experimentally determined.

III Feature Extraction using Active Contour

The shapes of eyebrow, nostril and face, unlike eye and mouth, are much more different for different people and their contours cannot be captured by using deformable template. Here, we apply a model called "snake" or "active contour models" to extract the contours of eyebrow, nostril and face. A snake is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges. Snakes are active contour models. They lock onto the nearby edges and localize them accurately. The energy function of active contour model is defined[5] as:

$$\begin{aligned} E_{snake} &= \int_0^1 E_{snake}(v(s))dS \\ &= \int_0^1 [E_{internal}(v(s)) + E_{images}(v(s)) + E_{constraint}(v(s))]dS \end{aligned} \tag{12}$$

where $v(s)$ represents the position of snake, $E_{internal}$ represents the internal energy of the contour due to bending, E_{images} gives rise to the image forces, and $E_{constraint}$ represents the external energy.

III.1 Energy of Active Contour

The original active contour model is user-interactive. The advantage of user-interactive is that the final form of the snake can be influenced by feedback from a higher level process. As the algorithm iterates, the energy terms can be adjusted by higher level processes to obtain a local minima that seems most useful to that process. However, there are some problems with minimization procedure. Here, we use a faster iteration method (Greedy Algorithm)[11] for the minimal energy iteration process of the active contour model. Although this model still has disadvantage which it cannot guarantee global minima, The iteration speed is much faster. Our definition of active contour energy is illustrated as:

$$\begin{aligned} E_{total} = \int_0^1 (\alpha(s)E_{continuity}(v(s)) &+ \beta(s)E_{curvature}(v(s)) \\ &+ \gamma(s)E_{images}(v(s)))\, dS \end{aligned} \tag{13}$$

However, it has to consider the problem of bunching up on a strong portion happened in active contour. This problem occurs during the iterative process, contour points will accumulate at certain strong portion of the active contour. The definition of $v(s)$ is similar to Eq. 12 and the following approximations are used:

$$\left|\frac{dv_i}{ds}\right| \approx |v_i - v_{i-1}|^2, \quad \left|\frac{d^2 v_i}{ds^2}\right| \approx |v_{i-1} - 2v_i + v_{i+1}|^2 \tag{14}$$

where these formulas assume that the points on the contour are spaced at unit interval and the parameter is arc length. One effect of this is that unevenly spaced points will have higher curvature.

(A) Continuity Force. The first derivative, $|v_i - v_{i-1}|^2$, causes the curve to shrink. It is actually minimizing the distance between points. It also contributes to the problem of points bunching up on strong portions of the contour. Therefore, we derive a term which encourages even spacing of the points that satisfy the original goal of first-order continuity without the effect of shrinking. Here, this term uses difference between the average distance between points, d, and this distance between the two points under consideration: $|d - |v_i - v_{i-1}||$. Thus, points having distance near the average will have the minimum value. At the end of each iteration, a new value of d is computed.

(B) Curvature Force. Since the formulation of the continuity term causes the points to be relatively evenly spaced, $|v_{i-1} - 2v_i + v_{i+1}|^2$ gives a reasonable and quick estimate of curvature. This term, like the continuity term, is normalized by dividing the largest value in the neighborhood, giving a number from 0 to 1.

(C) Image Force. E_{images} is the image force which is defined by the following operations.

i *For eight-neighbors, we have eight image energy measurements (Mag).*

ii *To normalize the image energy measurements, we select the minimum (Min) and Maximum (Max) terms from those eight measurements, and then do the calculation i.e., (Min-Mag)/(Max-Min) to obtain image force.*

At the end of each iteration, the curvature is determined at each point on the new contour. If the value is larger than threshold, β_i is set to 0 for the next iteration. The greedy algorithm is applied for fast convergence. The energy function is computed for the current location of v_i and each of its neighbors. The neighbor having the smallest value is chosen as the new position of v_i. Here, we apply the greedy algorithm for the feature extractions of eyebrow, nostril and face.

III.2 Feature Extraction of Eyebrow

After the *rough* contour of eyebrow has been obtained by RCER, we go for the *precise* shape extraction. The energy function for extracting the shape of eyebrow is:

$$\begin{aligned} &\sum_{i=1}^{n} (\alpha_i E_{continuity} + \beta_i E_{curvature} + \gamma_i E_{images}) \\ &= \sum_{i=1}^{n} \left[\alpha_i \frac{|d_{mean} - |v_i' - v_{i-1}||}{Largest_{icon}} + \beta_i \frac{(|v_{i-1} - 2v_i' + v_{i+1}|)}{Largest_{icur}} \right. \\ &\qquad \left. + \gamma_i \frac{(Edge_{iMin} - Edge_{iValue}')}{(Edge_{iMax} - Edge_{iMin})} \right] \end{aligned} \tag{15}$$

where v'_i is the next location of v_i for the next iteration,
$neigh(v_i)$ denotes one of the eight neighbors of v_i.
d_{mean} represents the average length between points,
$Largest_{icon}$ represents largest of 8 measurements $\{|d_{mean} - |neigh(v_i) - v_{i-1}||\}$,
$Largest_{icur}$ represents largest of 8 measurements $\{|v_{i-1} - 2neigh(v_i) + v_{i+1}|\}$,
$Edge_{iValue}$ represents edge response of v_i,
$Edge_{iMin}/Edge_{iMax}$ represents the largest/smallest edge response in 8 neighbors of v_i.

We use the above energy function and Greedy Algorithm for contour extraction. Because RCER estimates the *rough* contour as the initial contour for the snake, it should converge more quickly and more precisely. Because the complexity of the optimization problem for the multi-dimensional energy function, mathematically, we cannot identify the global minima . However, we can excite the snake and observe its second convergent place. If second convergent place is the same as first position, we assume that it converges at an accurate place.

III.3 Boundary Extraction of a Face

Unlike eyebrow extraction, the boundary extraction of a face is more time-consuming because the *rough* contour of a face is estimated too coarsely by RCER. With the positions of the precise contours of eyebrow, mouth, RCER may estimate the *rough* contour as accurate as possible. The energy function associated with boundary extraction of a face is defined as

$$E_{face} = \sum_{i=1}^{n} (\alpha_i E_{continuity} + \beta_i E_{curvature} + \gamma_i E_{images}) \quad (16)$$

where $E_{continuity}$ and $E_{curvature}$ are defined in Eq. 15, E_{images} is defined as:

$$E_{images} = w_{lines} E_{lines} + w_{edge} E_{edge} \quad (17)$$

The goal of E_{lines} is to attract more white points and less dark ones inside the active contour, whereas, the goal of E_{edge} is to attract more edge points on the active contour boundary.

$$E_{line} = \begin{cases} -Mag_{white}, & \text{if } B(x,y)=0 \text{ and } I(x,y) \geq \theta+\epsilon \\ Mag_{dark}, & \text{if } B(x,y)=0 \text{ and } I(x,y) < \theta-\epsilon \\ -Mag_{white}, & \text{if } B(x,y)=1 \text{ and } I(x,y) \geq \theta+\epsilon \\ Mag_{dark}, & \text{if } B(x,y)=1 \text{ and } I(x,y) < \theta-\epsilon \\ -Mag_{contour}, & \text{if } B(x,y)=255 \text{ and } I(x,y) \geq \theta+\epsilon \\ Mag_{contour}, & \text{if } B(x,y)=255 \text{ and } I(x,y) < \theta-\epsilon \\ 0 \quad , & \text{otherwise} \end{cases} \quad (18)$$

$$E_{edge} = |I(x,y) - \frac{1}{8}[I(x+1,y+1)+I(x+1,y)+I(x+1,y-1)+ I(x,y+1)+I(x,y-1)+I(x-1,y+1)+I(x-1,y)+I(x-1,y-1)]| \quad (19)$$

where $I(x,y)$ represents the intensity value of point (x,y), *threshold* (θ) *determination* for the face is mentioned section II, *tolerance* (ϵ) is experimentally determined.

To extract the boundary of a face without beard, snake iterates and moves toward chin because the E_{face} decreases constantly. The convergent process of snake is based on greedy algorithm.

III.4 Feature Extraction of Nostrils

In our picture-taking environment, the shape of nose is very difficult to extract. Nose is a convex feature, which may generate highlight and cause ambiguous intensity variations. We give up the nose contour extraction and try to find the nostrils. However, not everyone's nostrils are noticeable. Therefore, the first step of extracting nostril is to justify whether the appearance of nostril is noticeable. If it is, we go on extracting them. If it is not, we notify recognition process that nostrils of this person are not noticeable. Therefore, the feature vector, used for face identification, will not consist of the components associated with nostrils. RCER obtains the *rough* contour of nostril. To prevent from extracting false *rough* contour, we add a constraint which is length-to-width ratio. If the ratio is not beyond threshold, we assume the *rough* contour is accurate.

IV Experimental Results and Conclusion

Figure 3 shows the features extracted for one of the many portraits done in our experiment. There are some white curves on each image, these curves represent the extracted contours of front-view facial features. From these experiments, we know our method is fast(about 1-2 minutes on Solbourne Workstation) and robust. Our model can automatically identify the facial features of these portraits without any user involvements.

The facial feature extractions using deformable template and active contour are novel and the classification accuracy is relatively high. Our method has three contributions: Firstly, our method use the active contour to find the eyebrows, nostrils, and face accurately. Secondly, our algorithms overcome the problems of tilting and scaling. Thirdly, our model requires no user involvement because RCER can find the initial contour for each feature automatically.

REFERENCES

1. V. Govindaraju, S. N. Srihari, and P. B. Sher, Proc. of 3rd ICCV, pp.718-721, 1990.
2. D. Tezopoulos and K. Waters, Proc. 3rd ICCV, pp.727-732, 1990.
3. A. L. Yuille, D. S. Cohen and P. W. Hallinan, Proc. CVPR, pp.104-109, 1989.
4. M. Berger and R. Mohr, Proc. 10th ICPR, pp.847-, 1990.
5. M. Kass, A. Witkin and D. Terzopoulos, Inter. J. Computer Vision, pp.321-331, 1988.
6. A. A. Amini, T. E. Weymouth and R. C. Jain, IEEE Trans. PAMI,vol.12,no.9, pp.855-867, 1990.
7. L. D. Harmon, M. K. Khan, Richard Lasch and P. F. Ramig, Pattern Recognition, vol.13, no.2, pp.97-110, 1981.
8. G. J. Kaufman, Jr and K. J. Breeding, IEEE Trans. SMC, Vol.SMC-6, pp.113-120, 1976.
9. O. Nakamura, S. Mathur and T. Minami, Pattern Recognition, vol.24, no.3, pp.263-272, 1991.
10. C.W. Chen, M.S. Thesis, National Tsing-Hua University, Hsin-CHu, Taiwan, ROC, June, 1991.
11. D. J. Williams and M. Shah, Proc. of 3rd ICCV, pp.592-595, 1990.
12. T. Y. Cheng, and C. L. Huang, 1991 Workshop on CVGIP, Tainan, Taiwan, ROC, Aug. 25-27, 1991.

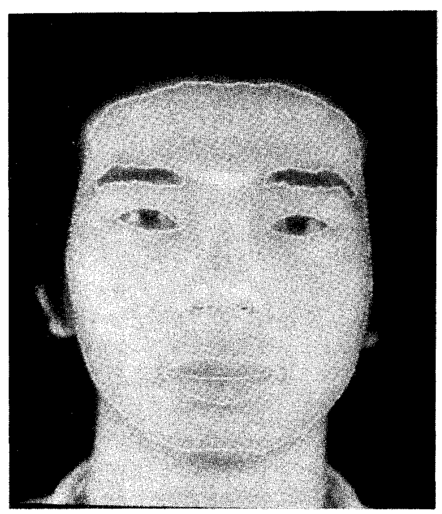

Figure 3. Extracted Features from the Portrait.