FACIAL FEATURE EXTRACTION USING GENETIC ALGORITHMS

Bilgin Eşme, Bülent Sankur, Emin Anarım
Boğaziçi University, Electrical Engineering Dept., Bebek, 80815, İstanbul
{ esme, sankur } @boun.edu.tr

ABSTRACT

Face models are used in such applications as videotelephone, graphic animation and automatic answering devices. Extraction and localization of facial features is the first step in constructing and adapting face models. Typical facial features are the eyes, the lips, the chin contour, and the nostrils. In this work, novel deformable templates in combination with genetic algorithms are used to capture eyes and lips contours.

1 INTRODUCTION

In this work, we present an automatic eye and mouth contour extraction system. This system is considered mainly for very low bitrate videotelephone compression, where the face is assumed to be the only object in the scene and it occupies the majority of the frame. In the first step, the locations of the eyes and the mouth are estimated by using the symmetry property of the human face. In the second step, deformable templates, that are simplified derivatives of those of Yuille *et.al*[3] are thrown to these locations which adapt themselves as guided by *genetic matings.*

2 SYMMETRY AXIS

Under the assumption that the human face is symmetric, an *axis of symmetry* is found by a pixelwise scoring method with tilt adjustments limited to a range of 20^o(Fig[1]). The method is similar to the one introduced in[1] but is improved in two aspects: most effects of the non-facial background are removed and thus is more robust, and the calculations are done not on the image itself but on the lower resolution image by a factor of 4. To find the axis of symmetry, the ordinate of the center of symmetry is fixed at $y = \frac{Y}{2}$. This axis is then allowed to shift laterally and to rotate by small increments through the ranges of χ and θ. The image $\mathcal{I}(x, y)$ with lower resolution and dimensions $X \times Y$ is scanned between $[-20^o, 20^o]$ for integer values of θ. For every (χ, θ) parameter pair, a symmetry score is obtained using Eq[1]. In this algorithm, for every (x_o, y_o) point on the left hand side of the axis, the mirror image point, namely (x_s, y_s) is found with respect to the parameters (χ_k, θ_k). N is the total number of points that are used for calculating the symmetry measure which in general less than $X \times Y$, since after the rotation, some points may be left out of the frame. In other words, mirror symmetry comparison is done with all those pixels for which a counterpart can be found within the borders of the image. The highest symmetry is denoted by the smallest value of $S(\chi, \theta)$, so the symmetry axis is determined as the (χ_k, θ_k) parameters that make $S(\chi, \theta)$ function minimum.

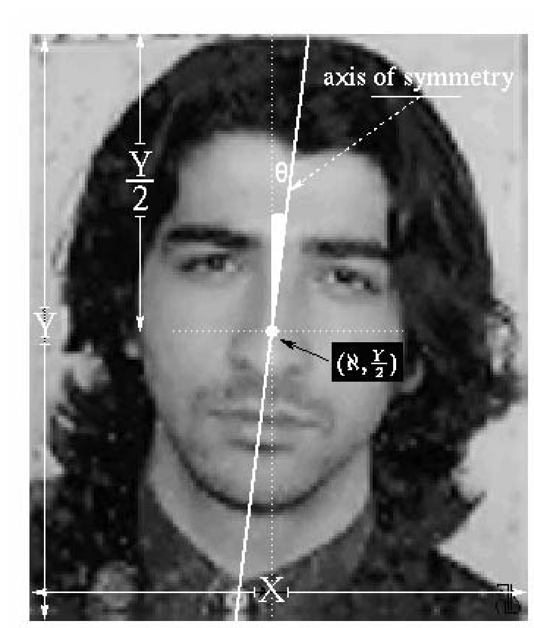


Figure 1: Axis of symmetry

$$x_i = \frac{\chi + \frac{X \tan\theta}{2} + \tan^2\theta x_o - \tan\theta y_o}{1 + \tan^2\theta}$$

$$y_i = y_o - \tan\theta x_o + \tan\theta x_i$$

$$x_s = (x_i - x_o) + x_i$$

$$y_s = (y_i - y_o) + y_i$$

$$S(\chi, \theta) = \frac{1}{N} \sum_{y=\frac{Y}{4}}^{\frac{3Y}{4}} \sum_{x=\chi-\frac{X}{6}}^{\chi+\frac{X}{6}} \mid \mathcal{I}(x_s, y_s) - \mathcal{I}(x_o, y_o) \mid \quad (1)$$

3 PREPROCESSING

The image is subjected to a local maximum operation defined as follows:

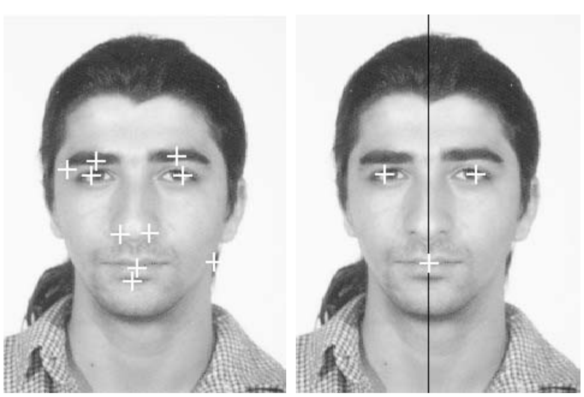
$$\mathcal{I}_e(x, y) = u(\mathcal{I}(x, y) - M_t) \times 255 \quad (2)$$

where $\mathcal{I}_e(x, y)$ is the output of the local maximum operator, M_t is the average of the window centered in the pixel, and u is the unit step function. Window with dimensions $m \times m$ is scanned over the image and each pixel is binarized to the value 255 if its gray value is above that of the window mean, and to 0 otherwise(Fig[2.b]).

a b

Figure 2: (a) The original image with axis of symmetry Found, (b) The binarized map

Window sizes in the range between $10 - 18$ are found to be most satisfactory.



a b

Figure 3: (a) The center of masses, (b) The eyes located

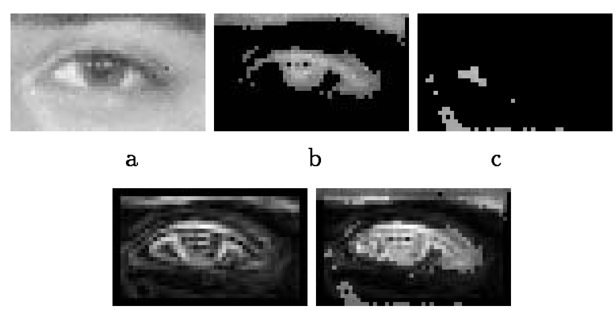
These segments are labeled by the connected components algorithm and the center of mass of each labeled segment is calculated(Fig[3.a])[4]. The set of these center of masses turns out to be useful information to identify those segments comprising the sought facial objects. The centroids are selectively eliminated based on their likelihood of corresponding to facial object locations. This combinational search is aided by various anthropometric measures[7].

4 ENERGY FIELD

To implement the template fitting around the eyes and lip contours, various energy fields have to be generated. In this work, we used only one energy field resulting from the fusion of peak, valley and edge fields. We have found out that the fused energy field is a more parsimonious representation of information and it results in a more stable training.

Since the locations of the eyes and mouth were already estimated, there's no need to extract the energy field from the whole image. Thus, rectangular frames, large enough to include these facial objects are drawn and processing is done only within these areas.

The energy field is composed of the valley map (the dark regions), the peak map (light regions) and edge field. Valley map comprises information about the shades around the eye and mouth regions. The white regions of the eye are captured by the peak map data, and the edge information is used to detect the line structures around the mouth and the eyes. These three sources of information are fused to give a combined energy field instead of treating them seperately with different cost functions as is done by X.Xie *et.al*[2]. In Fig[4], the results of preprocessing operations are illustrated.



a b c

d e

Figure 4: (a) Original eye rectangle, (b)Valley map, (c) Peak map, (d)Edge field, (e) Fused energy field

The valley map $\mathcal{D}(x, y)$ is a variation of local minimum operator, and is obtained by windowing the framed image:

$$\mathcal{D}(x, y) = \left(\frac{u(I_B(x, y) - \mathcal{I}(x, y)) \times |I_B(x, y) - \mathcal{I}(x, y)|}{I_B(x, y)} \right) \cdot 255 \quad (3)$$

In this equation, u is the unit step function, $\mathcal{I}(x,y)$ is the original image frame, $I_B(x,y)$ is the window average of the incident pixel and obtained as:

$$I_B(x,y) = \frac{1}{K^2}\sum_{i=0}^{K}\sum_{j=0}^{K}\mathcal{I}(x+i,y+j) \qquad (4)$$

where K is a window size to be chosen around 10. In this method, for every pixel, a score is given proportional to the amount by which a pixel's gray value is less than the window average. This score is normalized to the range of 0 to 255. If the gray value of the pixel $\mathcal{I}(x,y)$ is greater than the window average, $\mathcal{D}(x,y)$ is assigned to 0. The peak map is obtained the same manner to be used for the eye regions only as in Eq[5].

$$\mathcal{L}(x,y) = \left(\frac{u(\mathcal{I}(x,y) - I_B(x,y)) \times |I_B(x,y) - \mathcal{I}(x,y)|}{I_B(x,y)}\right)\cdot 255 \qquad (5)$$

The edge map is obtained by Sobel operator without using any threshold. The magnitudes of the edges are normalized and quantized between $[0,255]$ range. The fused image is composed by addding these three fields and normalizing the output to $0-255$ range.

5 TEMPLATES AND GENETIC DEFORMATION

The *deformable templates* used in this work were first proposed by Yuille *et.al.*[3], and have since been modified by X.Xie *et.al*[2]. However, the template used in our work is a simplified version since it is restricted to finding only the eye contours, hence it contains fewer parameters. The template is composed of two second degree polynomials.

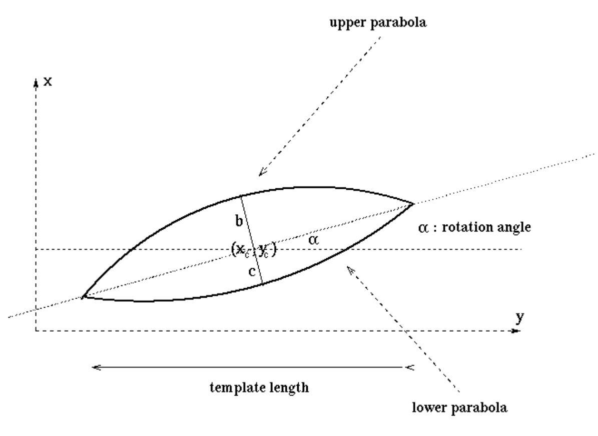


Figure 5: The template

There are six parameters describing the eye template: the spatial coordinates of the centroid, (x_c, y_c); b and c, the focal distances of upper and lower curvatures, respectively, to the center point; L, the length of the template and the rotation angle, α. The feature vectors formed from these six parameters are the *chromosomes* of one molecule[5][6]. Each of the six chromosomes consists of two bits; 00 corresponds to a parameter decrement, 01 and 10 are no action cases and 11 is a parameter increment. For example a chromosome string as 101000110010, indicates a decrement in the 3^{rd} and 5^{th} parameter, and an increment in 4^{th} oneFig[6].

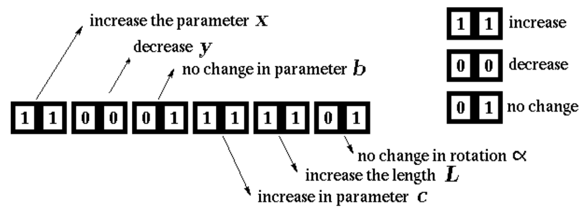


Figure 6: The coding of the genes

Initially, 100 genes were created randomly. These genes deform the template and the effect of each deformation vector is scored by the fitness function C(Eq[6]). After each generation, the 4 most fit individuals are chosen as offsprings, and the rest of the population is divided into four to mate with them. Mating is done by offspring's transfer of one randomly chosen chromosome to the ordinary individual(Fig[7]).

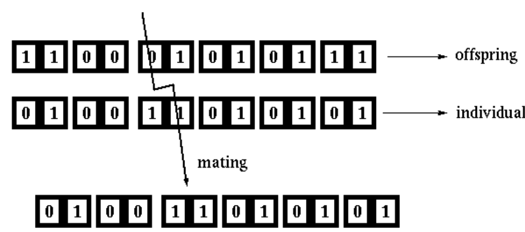


Figure 7: The mating rule

After one generation, the %10 least successful genes of the population are killed and replaced with other randomly created genes. The template is considered converged whenever negligible change occur between generations. The cost function C can be expressed as:

$$C = (1-W)\frac{M_t - E_i}{E_i} + W\frac{E_t - E_i}{E_i} \qquad (6)$$

where E_i is the energy confined within the template, M_t is the maximum energy that can be confined by the template which is equal to *the area of the template* $\times$ 255; E_t is the total energy present in the rectangular region and W is a weight set in the range $[0,1]$ to achieve desired balance between the two additive terms. The first term encourages expansion and cover all the energy, and the second term encourages the template to be more compact.

6 RESULTS

We have presented an algorithm for lip and eyes contour extraction which has been tested on a large repertoire of facial images. This algorithm is applicable to scenes where human face is the dominant object. Furthermore, it must not be rotated or tilted to the degree of its symmetry axis being out of sight. Estimation of the symmetry axis is followed by local minima operator which yields seeds of candidate regions. These are then reduced to only three, that are the eyes and mouth segments by a search method based on antropometric measures. Deformable templates, initialized around these segments are driven by a genetic algorithm and a combined energy field. Such a template fitting is demonstrated in Fig[8].

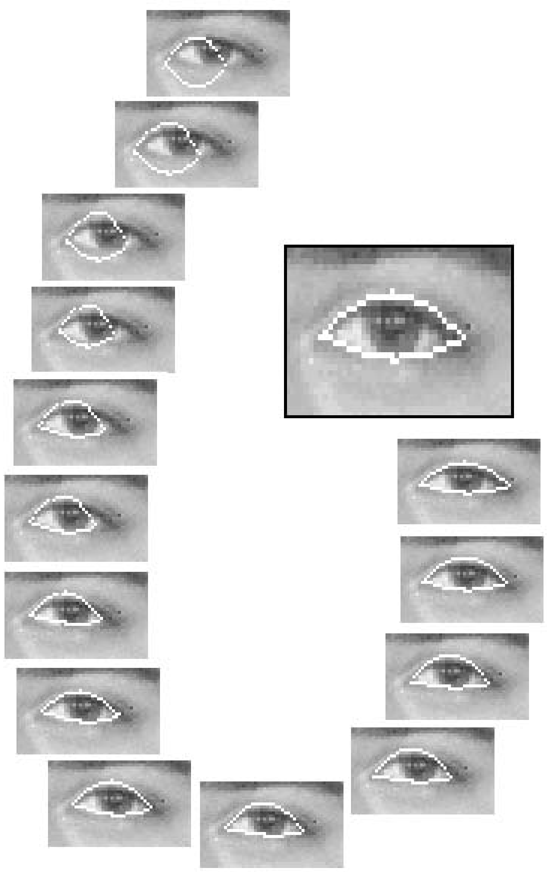


Figure 8: The epochs of the genetic search engine

Genetic algorithms are found out to be useful in contour extraction, especially because they avoid being trapped by local minima.

The symmetry axis method is also instrumental as the first step in the localization process. The algorithm is not sensitive to noise or any kind of cluttering background. The images with spectacles also yielded successful results, but occlusion problems cannot yet be dealt with.

References

[1] Antonio J.Colmenarez, Thomas S.Huang, "Frontal View Face Detection," *SPIE*, Vol.2501 pp. 90-98 1990

[2] X.Xie, R.Sudhakar and H.Zhuang, "On Improving Eye Extraction Using Deformable Templates," *Pattern Recognition* , Vol.27 No.6 1994

[3] A.L.Yuille, P.W.Hallinan and D.S.Cohen, "Feature Extraction from Faces Using Deformable Templates," *Int. J. Comput. Vision,*, 1989

[4] M.J.T.Reinders, P.J.L. van Beek, B.Sankur, J.C.A. van der Lubbe, "Facial Feature Localization and Adaptation of a Generic Face Model for Model Based Coding," *TUDELFT Tech. Report,*, Holland

[5] John Holland, "Genetic Algorithms," *Scientific American,* July, 1992

[6] David E.Goldberg, "The Existential Pleasures of Genetic Algorithms," *IlliGAL Report No.94008 University of Illinois at Urbana,* November, 1994

[7] Rama Chellapala, Charles L.Wilson, Saad Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. of IEEE.,*, vol.83 no.5 pp. 705-740, 1995

Figure 9: Extracted features