A face recognition system based on local feature characterization

Paola Campadelli, Raffaella Lanzarotti

DSI - Università degli Studi di Milano
Via Comelico, 39/41 20135 Milano, Italy
{campadelli,lanzarotti}@dsi.unimi.it

Abstract. A completely automatic face recognition system is presented. The method works on color and gray level images: after having localized the face and the facial features, it determines 16 facial fiducial points, and characterizes them applying a bank of filters which extract the peculiar texture around them (*jets*). Recognition is realized measuring the similarity between the different *jets*. The system is inspired by the elastic bunch graph method, but the fiducial point localization does not require any manual setting.

1 Introduction

Human face recognition has been largely investigated for the last two decades; the most famous approaches adopt eigenfaces and neural networks.

In this paper we present an approach based on another technique: the elastic bunch graphs [3], but with a new and completely automatic method to localize sixteen facial fiducial points (the eyebrow and chin vertices, the nose tip, and the eye and lip corners and upper and lower middle points).

We build three galleries, each one containing an image per person: the frontal, right and left rotated face galleries. Given an image, the system extracts the fiducial points, characterizes them, determines the head pose, and compares the face with the proper gallery images. We observe that, while the face analysis is done on the gray levels only, the fiducial point extractor works on both color and gray level images, even if the one based on color is slightly more precise.

We present encouraging results obtained on databases of up to 200 subjects; 150 of them have been extracted from the FERET database, and 50 from our color image database. The considered sub-set of the FERET database consists in 8 gray level images per person organized according to the angle between the subjects and the camera ($0°$, $\pm 15°$, $\pm 25°$, $\pm 40°$), and where two sets of frontal view images, respectively with neutral and smiling expression, are included. Our database consists in 6 images per person: two frontal, two right rotated and two left rotated, with a miscellaneous of rotation angles and approximately neutral expressions.

2 Face and facial features localization

The first step consists in localizing the face in the image. In [1] we have presented a method which localizes faces in generic color images searching at first all the skin regions, and then validating the ones which contain at least one eye. Regarding gray level images, we have proposed a method [2] that works on images of face foregrounds with a homogeneous and light-colored background. Subsequently, the facial features (eyes, nose, mouth, and chin) are localized [2].

We have tested the methods on 500 color images, and 2000 gray level ones, obtaining correct results in the 95% of the cases.

3 Identification of fiducial points

In this section we describe the steps followed for the determination of the fiducial points. The eyes and mouth are described by two parametric models derived from the deformable templates proposed in [4] with significant variations.

Eyes. In the eye sub-image the iris is first identified with the Hough transform for circumferences and the reflex, often present in it, is eliminated. Without these preliminary steps the deformable template finds very often wrong contours. The template [Fig.1.1], described by 6 parameters $\{x_w, y_w, a, b, c, \theta\}$, is made of two parabolas representing the eye arcs and intersecting at the eye corners.

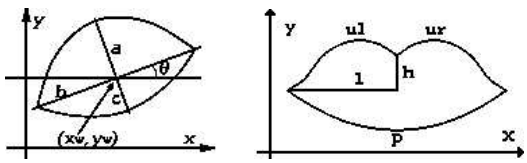

Fig. 1. Eye and Mouth Deformable templates.

As Yuille did, we define an energy function E_t to be minimized. E_t is the sum of three terms which are functions of the template parameters and of the image characteristics (prior information on the eye shape, edges, and 'white' of the eye). For color images, the characteristics are evaluated on the u plane of the CIE-Luv space. More precisely: $E_t = E_{prior} + E_e + E_i$, where:

1. $E_{prior} = \frac{k_1}{2}\left((x_w - x_i)^2 + (y_w - y_i)^2\right) + \frac{k_2}{2}\cdot(b-2r)^2 + \frac{k_3}{2}\left((b-2a)^2 + (a-2c)^2\right)$

   (x_i, y_i) is the iris center and r the iris ray obtained by the Hough transform.

2. $E_e = -\frac{c_1}{|\partial R_w|} \cdot \int_{\partial R_w} \phi_e(\boldsymbol{x}) ds$

   ∂R_w represents the upper and lower parabolas, and ϕ_e is the edge image obtained applying the Sobel edge detector.

3. $E_i = -c_2 \int_{R_w} \phi_i(\boldsymbol{x}) ds$

   where R_w is the region enclosed between the two parabolas, and ϕ_i is the binary image obtained applying an adaptive thresholding able to balance the number of white pixels on both sides of the iris. $\phi_i(p)$ is set to 255 if p is white, to -100 if p is black.

After this, we follow the Yuille's work obtaining a good eye description, and we extract from it the two eye corners and the upper and lower middle points.

Fig. 2. Eye image processing: ϕ_e obtained from the u plane; ϕ_i obtained binarizing the u plane; Final result.

Mouth. In the mouth sub-image we calculate the mouth corners, and the entire border adopting a parametric model.

The mouth corners correspond to the extreme of the mouth cut, obtained combining the image vertical derivative, and the image low values [2].

The mouth model [Fig.1.2] is parameterized by $\{l, h, a_p, a_{ul}, b_{ul},\ a_{ur}, b_{ur}\}$, and is made of one parabola, p, for the lower lip, and two cubics, ul and ur, for the upper lip. Two energy functions to be minimized are defined; both of them are functions of the template parameters but the first, E_i, depends on the image colors/gray levels, while the second, E_e, depends on the edge image (I_{Edges}). The model is modified in two epochs considering respectively the E_i and the E_e functions. More precisely:

1. $E_i = c_2 \int_R \phi_i(\boldsymbol{x}) dA$

   where R is the region enclosed among the 3 curves and ϕ_i is the binary image obtained clustering in 2 clusters the MouthMap (for gray level images, the MouthMap is the negative of the mouth sub-images itself, while for color images it is: $MouthMap = (255 - (C_r - C_b)) \cdot C_r^2$), and setting a pixel to 255 if it is white, to -80 if it is black.

2. $E_e = c_1(-\frac{100}{|ul|} \int_{ul} \phi_e(\boldsymbol{x}) ds - \frac{100}{|ur|} \int_{ur} \phi_e(\boldsymbol{x}) ds - \frac{10}{|p|} \int_p \phi_e(\boldsymbol{x}) ds), \quad \phi_e = I_{Edges}.$

Eyebrow and Chin. The eyebrow and chin description consists in the best parabola which approximates their vertical derivative [Fig. 3].

Nose. The nose is characterized by very simple and generic properties: the nose has a 'base' which gray levels contrast significantly with the neighbor regions; moreover, the nose profile can be characterized as the set of points with the highest symmetry and high luminance values; finally we can say that the nose tip lies on the nose profile, above the nose base line, and is bright [Fig. 3].

Fig. 3. Examples of facial feature and fiducial point description.

3.1 Experimental results

The method has shown very good performances (error of 1 or 2 pixels) under some commonly accepted assumptions: the head image dimensions are not lower than (100×100) pixels; the head rotation is at most of $45°$; the mouth are closed and the eyes opened; the illumination is not too low, and does not create particular shadows on the faces.

We observe that errors of one or two pixels do not constitute critical problems for the subsequent steps, since both the face characterization and recognition are not based on the fiducial points punctual values, but on a local analysis of the regions around them, making the system more robust.

4 Face dimension normalization and pose determination

The previous steps have dealt with faces of any scale and different orientation; however the face characterization and recognition are very sensitive to these kind of variations. We thus proceed rescaling the images to a common size and determining the head rotation. To these purposes, we consider the triangle whose vertices are the nose tip (N_t) and the eye external corners (C_1, C_2). We first normalize the image so that the triangle area is of 2000 pixels; subsequently we compare the length of the two segments which connect N_t to C_1 and C_2 respectively. In case of frontal image, the two segments are approximately of the same length, while, when the head is rotated, the two lengths vary greatly and according to the rotation side. We thus recognize three different poses: frontal, right and left rotated.

5 Face characterization

In order to characterize the fiducial points, we have experimented two techniques: the Gabor wavelet transform and the steerable Gaussian first derivative basis

filters. The first technique has shown greater robustness with respect to rotation and little error in the fiducial point localization. We thus describe it only.

To characterize a fiducial point, we convolve the portion of gray image around it with a bank of *Gabor kernels*; following the idea of Wiskott [3], 5 different frequencies and 8 orientations are exployed. The obtained 40 coefficients are complex numbers. A *jet* J is obtained considering the magnitude parts only.

Applying the Gabor wavelet transform to all the facial fiducial points, we obtain the face characterization, consisting in a *jets vector* of $40 \times N$ real coefficients where N is the number of visible fiducial points.

To recognize a face image I we compute a similarity measure between its *jets vector* and the ones of all the images G_i in the corresponding gallery, and we associate I to the G_i which maximizes the measure of similarity. Being J_n^i the *n-th* jet of the *jets vector* i, we define the similarity between two *jets vector* as:

$$S_v(V^1, V^2) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{J_n^1 \cdot J_n^2}{\|J_n^1\| \|J_n^2\|}$$

6 Experimental results

We have experimented the whole face recognition system on databases of 50, 100, 150, and 200 subjects. For each of them, three images are catalogued in the galleries according to the pose. Regarding the FERET database, the frontal and neutral expression image set, and the $\pm 40°$ image sets are used as gallery images, while the other are used to test the system. In the following, we report the most significant results.

At first, in order to highlight the system behavior according to the different rotation angles, we report the experiments carried out referring to the FERET images only. In particular, we give all the details for the most challenging experiment, and summary results for the other cases.

Analyzing the results exhibited in table 1, we notice that the system is more robust to little rotation disparity (e.g. second line) than to expression variations (first line). However, incrementing the rotation angle disparity, the performances decrease (e.g. sixth line); it thus arises the importance of having an automatic face pose estimator which allows to compare each test image to the gallery with the less angle disparity. The bold lines in the table reflect this choice.

In table 2 we report the recognition performance obtained using subsets of fiducial points. We remark that most of the discriminating face characteristics are in the upper part of the faces, above all if the face expression varies significantly (see first line of the table).

Finally we report the results obtained, referring to a gallery of 200 subjects (150 from the FERET database and 50 from our color image database), and using all the fiducial points to test the system. In the 92.5% of the cases the best match corresponded to the right person, while in 95.6% of the cases the correct person's face was in the top five candidate matches.

Gallery	Test	Best rank	In top 5
0°	**0°**	**70**	**82**
0°	**+15°**	**94**	**96**
0°	**−15°**	**95**	**97**
0°	+25°	90	96
0°	−25°	93	96
+40°	+15°	90	96
+40°	**+25°**	**96**	**98**
−40°	−15°	78	93
−40°	**−25°**	**95**	**96**

Table 1. Recognition results obtained referring to the 150-subject galleries, and exploiting all the 16 fiducial points.

Gallery	Test	Best rank		In top 5	
		EEH	MCH	EEH	MCH
0°	0°	26	68	46	79
0°	+15°	93	93	95	96
0°	−15°	89	92	95	95
+40°	+25°	81	91	91	95
−40°	−25°	72	93	82	95

Table 2. Recognition results obtained referring to the 150-subject galleries, in case of partial occlusions: **EEH**: eyes and eyebrows hidden; **MCH**: mouth and chin hidden.

7 Conclusions

We have presented a system that, given a face image, extracts the facial fiducial points, determines the head pose, normalizes the image, characterizes it with its *jets vector*, and compares it with the ones in the corresponding gallery. The image is recognized to be the most similar one in the gallery.

The facial feature detection and description methods have been tested on 2500 face foregrounds images detecting the fiducial points with high accuracy (errors of 1 or 2 pixels are negligible) in 93% of the images.

The whole face recognition system has been tested on a database of 1500 images of 200 subjects. We can affirm that our fiducial point extractor allows to obtain the same recognition performances as the elastic bunch graph used in [3], while being completely automatic.

References

1. P. Campadelli, F. Cusmai, and R. Lanzarotti. A color based method for face detection. *Submitted*, 2003.
2. P. Campadelli and R. Lanzarotti. Localization of facial features and fiducial points. *Processings of the International Conference Visualisation, Imaging and image Processing (VIIP2002), Malaga (Spagna)*, pages 491–495, 2002.
3. L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. In L.C. Jain et al., editor, *Intelligent biometric techniques in fingerprints and face recognition*, pages 355–396. CRC Press, 1999.
4. A.L. Yuille, P.W. Hallinan, and D.S. Cohen. Feature extraction from faces using deformable templates. *International journal of computer vision*, 8(2):99–111, 1992.