A feature-based face recognition system

Paola Campadelli, Raffaella Lanzarotti, Chiara Savazzi
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41 20135 Milano, Italy
{campadelli,lanzarotti}@dsi.unimi.it
sc570121@mat.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 or operator intervention.*

1 Introduction

Human face recognition has been largely investigated for the last two decades. It has many practical applications such as access control, security monitoring, and surveillance systems. All these applications aim at nonintrusive system, able to recognize a person with no collaboration, the less possible restrictions, and a match process which is computationally efficient.

Eigenface [4] is one of the most famous face recognition approach. Many papers have presented encouraging results showing that this is a fast, simple, and effective method. The strong drawback is that it is not scale and lighting condition invariant limiting its practical use.

Neural network is another widely studied technique. However, it is suitable for small database only, and, considering it requires many images per person to train the network, its practical use is very limited.

Another approach consists in using graph matching [5]. The idea is to represent a face as a graph, whose nodes, positioned in correspondence to the facial fiducial points, are labelled with a multiresolution description of the surrounding gray level image. This approach is better than the others in terms of rotation invariance, but it is computationally more expensive, and require an initial human intervention.

Again, geometrical feature and template matching techniques have been investigated showing their inadequacy for the management of a reliable face recognition system.

In summary, the face recognition is still an open problem; an effort is required to make the existing techniques suitable for real applications, improving their computational performance and enlarging their field of applicability.

In this paper we present an approach based on the multiresolution technique presented in [5], but with a new and completely automatic method to localize the facial fiducial points.

We build three galleries, each one containing an image per person: the frontal, right and left rotated face galleries. Given a new image, the system extracts the facial fiducial points, characterizes them, determines the head pose and thus the gallery to be used for the comparison, and compares the face with the 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 FERET database consists in 10 gray level images per person organized according to the angle between the subjects and the camera (0°, $\pm 15^\circ$, $\pm 25^\circ$, $\pm 40^\circ$, $\pm 60^\circ$), 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; the angle of acquisition is not known exactly but it varies between 25° and 45°, and the expression is approximately neutral.

The paper is organized as follows: in section 2 the face and facial feature localization techniques are summarized; in sections 3 and 4 the fiducial points extraction and the face normalization procedures are presented; in 5 the face local

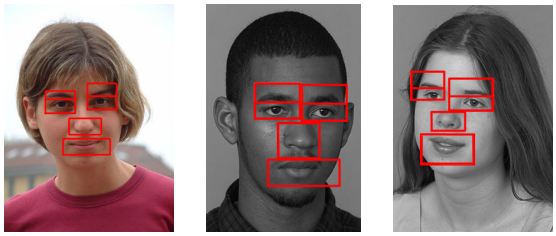

Figure 1. Subdivision results on frontal and rotated faces.

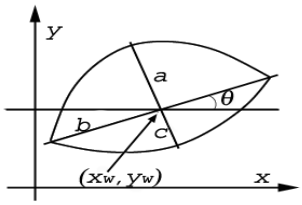

Figure 2. Deformable template of the eye.

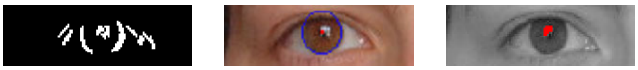

Figure 3. Iris and reflex localization: Input to the Hough transform; Iris description; Detected reflex.

characterization and the similarity measure are described. Finally in sections 6 and 7 results and conclusions are discussed.

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. The method consists in clusterizing the image in three clusters and keeping the one with the intermediate gray-level.

Subsequently, the facial features (eyes, nose, mouth, and chin) are localized [2] [Fig.1]. The shape information is used both for color and gray level images, while colors, when available, allow to tailor the sub-images more precisely.

The face and facial feature localization modules have been tested on 500 color images and 2000 gray level images, showing correct results in the 95% of the cases.

3 Identification of fiducial points

In this section we describe the fundamental steps followed to determine the facial fiducial points. We process each sub-image separately, adopting different instruments for each feature on the basis of its peculiar characteristics. This technique allows to determine 16 fiducial points per image: the eyebrow and chin vertices, the nose tip, and the eye and lip corners and upper and lower middle points.

3.1 Eyes

The eye is described by a parametric model which is derived from the deformable template proposed by Yuille et al. [6] with significant simplifications (6 parameters instead of 11). Once the iris has been identified and a reflex, often present in it, has been eliminated, we use the parametric model for localizing the eye contour. Without this preliminary steps, the deformable template finds very often wrong contours. The model [Fig.2], described by the parameters $\{x_w, y_w, a, b, c, \theta\}$, is made of two parabolas, representing the upper and lower eye arcs, intersecting at the eye corners.

The iris is localized adopting the Hough Transform for circumferences. The transform is applied to the binarized horizontal derivative of the eye gray-level image [Fig.3.1] since it highlights the two iris vertical sides, which are the best references to find the iris border. Among the found circumferences, we choose the one which obtained the most of the votes [Fig.3.2]. We then proceed searching for the reflex inside the found circumference; if present, it is characterized by high values in the gray level image [Fig.3.3].

Once the iris is detected, we start looking for the shape of the eye. Thus, 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, since in this color plane the information we are looking for (edges and 'white' of the eye) are strengthened and clearer.

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 previously by the Hough transform.

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

   ∂R_w represents the upper and lower parabolas, and ϕ_e is the edge image.

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

   R_w is the region enclosed between the two parabolas, and ϕ_i is the weighted image obtained following these steps:

   (a) **For color images:** threshold the u plane with a global threshold $th = (max(u) - 0.1 \cdot max(u))$, obtaining a binary image. If the number of white pixels on a fixed neighborhood of the iris is less than *(area iris/5)*, we lower the threshold until the condition is verified. We count the white pixels both in a left and right neighborhood of the iris and we compare them. If one of them is more than double of the other, we decrease the threshold on the side with less white pixels until the condition is verified. **For gray level images:** threshold the left and right neighborhood of the iris, putting to 1 the 10% of the clearest pixels.

   (b) For every pixel p, $\phi_i(p)$ is set to 255 if p is white, to -100 if p is black.

A remark has to be done about the ϕ_i construction. Regarding color images, in general the white of the eye corresponds to the highest u values; however it is necessary to consider two exceptions: the case where a lighter region is present in the image, and the case where a shadow darkens one part of the white of the eye. In the first case using a global threshold we would discard the white of the eye, and in the second case only one half of it would be found. We thus detect these particular situations checking the iris neighborhood, and applying, if necessary, an adaptive thresholding. In case of gray level images, it is always necessary to limit the search area to the iris neighborhood, since the skin is often very clear in the luminance plane.

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

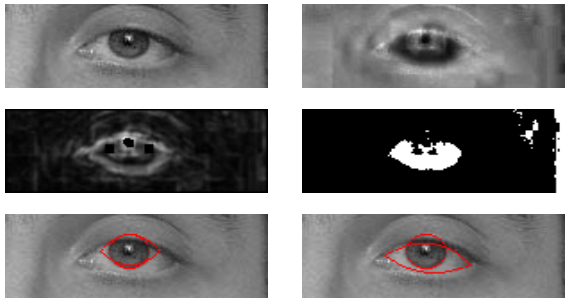

Figure 4. Eye image processing (from left to right, top to bottom): Original image; u plane; Gradient of the u plane; Binarization of the u plane; Model initialization; Final result.

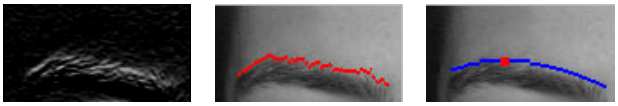

Figure 5. Eyebrow image processing: Vertical derivative; Upper border; Parabola approximation.

3.2 Eyebrows

A good description of the eyebrow is given by the best parabola which approximates it. In order to find it, we calculate the vertical derivative [Fig.5.1] of the eyebrow gray-level image, we binarize it keeping the 10% of the highest derivative values, and we extract the upper border [Fig.5.2] of the obtained regions. We then proceed fitting these points to a parabola and extracting the upper vertex [Fig.5.3].

3.3 Nose

The nose is characterized by very simple and generic properties which are true independently of the pose: 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. 7].

3.4 Mouth

Our goal for the mouth is the determination of its corners and of its upper and lower middle points. To get a more robust estimate, the entire border is determined adopting

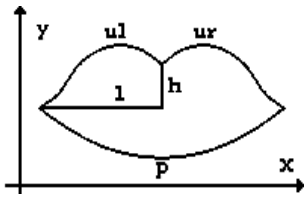

Figure 6. Deformable template of the mouth.

a parametric model derived from the deformable template proposed by Yuille et al. [6] with significant variations.

After having determined the mouth corners, we use the parametric model for localizing the mouth contour. The model [Fig.6], described by the parameters $\{l, h, a_p, a_{ul}, b_{ul}, a_{ur}, b_{ur}\}$, is made of one parabola, p, for the lower lip, and two cubics, ul and ur, for the upper lip.

To determine the mouth corners, we extract the mouth cut: in correspondence to it, there are high values of the vertical derivative and low values in the gray level image. Combining this information, we obtain the mouth cut and, taking its extremes, the correct corner positions.

Thus, we define two energy functions to be minimized. 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 image edges. The model is modified in two epochs considering respectively the E_i and the E_e functions.

More precisely:

1. $E_i = c_2 \int_R \phi_i(\vec{x}) dA$
   where R is the region enclosed among the 3 curves and ϕ_i is the binary image obtained in the following way:
   - **For gray level images**, the Mouth Map is the negative of the mouth sub-image itself, while **for color images**, the Mouth Map is determined as follows:

     $$MouthMap = (255 - (C_r - C_b)) \cdot C_r^2$$
   - Mouth Map binarization, using a clustering algorithm.
   - For every pixel p, $\phi_i(p)$ is set to 255 if p is white, to -80 if p is black.
2. $E_e = c_1(-\frac{100}{|ul|} \int_{ul} \phi_e(\vec{x}) ds - \frac{100}{|ur|} \int_{ur} \phi_e(\vec{x}) ds - \frac{10}{|p|} \int_p \phi_e(\vec{x}) ds)$
   where ϕ_e is the edge image calculated on the gray level image.

3.5 Chin localization and description

To localize and describe the chin, we determine the best parabola which approximates it and its vertex. The chin shape information is extracted applying to the gray level image a non-linear edge detector. At this stage, knowing the positions and the dimensions of the facial features, we can restrict significantly the parabola search area, and in particular the position of its vertex, obtaining a good chin approximation [Fig.7].

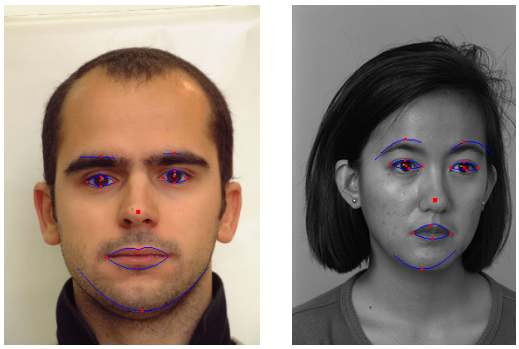

Figure 7. Examples of facial feature and fiducial point description.

3.6 Experimental results

The fiducial point detection has been tested on the subset of images on which the face and facial feature localization succeeded. 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 around the vertical axis is at most of $45°$; the illumination is not too low, and does not create particular shadows on the faces. We observe that, when processing images of faces rotated of $60°$, all the features are localized correctly, but the fiducial points of the most hidden eye are not well defined. Since we are able to recognize this degree of rotation (see next paragraph), in the recognition phase we do not consider the characteristics associated to such fiducial points.

A consideration about the face expression has to be done: the algorithm does not work on people with either the mouth opened or the eyes closed; to deal with these situations it would be necessary to introduce different characterizations.

4 Pose determination and face dimension normalization

In order to discriminate the images on the basis of the head rotation, we compare the length of the segments which connect the nose tip and the two external eye corners. 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. In particular we conclude that the head is left rotated if the length of the segment linking the left eye to the nose is less than 0.8 times the length of the other segment, and vice versa. This information will be used to choose the suitable gallery (frontal, right or left rotated head) for the search. Moreover, if the ratio between the two segments is less than 0.3, we conclude that the head is rotated around the vertical axis of approximately $60°$, we thus discard the fiducial points of the most hidden eye in the recognition module.

Finally, in order to compare faces acquired at different scales, it is necessary to normalize them. The most robust information regarding the face scale is given by the area of the triangle defined by the nose tip and the two external eye corners; we thus scale all the images so that the triangle area is of 2000 pixels.

5 Face characterization

Once the fiducial points have been extracted, the face rescaled, and the pose determined, we proceed characterizing each fiducial point in terms of the surrounding gray level image. Two techniques have been experimented: the Gabor wavelet transform and the steerable Gaussian first derivative basis filters. Both of them are biologically motivated and in both cases a bank of filters is adopted, varying the orientations and the frequencies/scales. The first technique has shown greater robustness with respect to rotation and little error in the fiducial point localization. We thus describe it only.

5.1 Gabor wavelet transform

Following the idea of Wiskott [5], to characterize a fiducial point, we convolve the portion of gray image around it with the following bank of *Gabor kernels*:

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[exp(i\vec{k}_j\vec{x}) - exp\left(-\frac{\sigma^2}{2}\right)\right]$$

in the shape of plane waves with wave vector $\vec{k}_j$, restricted by a Gaussian envelope function. We employ a discrete set of 5 different frequencies, index $\nu = 0, ..., 4$, and 8 orientations, index $\mu = 0, ..., 7$,

$$\vec{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_\nu \cos\varphi_\mu \\ k_\nu \sin\varphi_\mu \end{pmatrix}, k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \varphi_\mu = \mu\frac{\pi}{8}$$

with index $j = \mu + 8\nu$. The width σ/k of the Gaussian is controlled by the parameter $\sigma = 2\pi$. We observe that the kernels are DC-*free*, that is $\int \psi_j(\vec{x})d^2\vec{x} = 0$ allowing to deal with different illumination conditions.

The obtained 40 coefficients are complex numbers. A *jet* [3] 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 $40 \times M$ real coefficients where M is the number of considered fiducial points.

To recognize a new 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. We define the similarity between two *jets vector* as the average over the similarities between pairs of corresponding *jets*:

$$S_v(V^1, V^2) = \frac{1}{M}\sum_n S(J_n^1, J_n^2)$$

where

$$S(J_n^1, J_n^2) = \frac{\sum_i J_i^1 J_i^2}{\sqrt{\sum_i (J_i^1)^2 \sum_i (J_i^2)^2}} \qquad (i = 0, ..., 39)$$

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 experimental 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.

In table 1 we report the recognition results obtained exploiting all the fiducial points (with the exception of the ones corresponding to the most hidden eye on faces rotated of $60°$), and referring to the 150-subject galleries.

We observe 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.

As it can be expected, the performances obtained in the same conditions, but with smaller galleries (50 and 100 subjects) are higher. For example, for the frontal to frontal comparisons, they increase of 12% and 7% respectively.

Gallery	Test	% First rank	% First 5 ranks
0°	**0° Smiling**	**70**	**82**
0°	**+15°**	**94**	**96**
0°	**−15°**	**95**	**97**
0°	+25°	90	96
0°	−25°	93	96
0°	+60°	59	78
0°	−60°	62	84
+40°	+15°	90	96
+40°	**+25°**	**96**	**98**
+40°	**+60°**	**92**	**96**
−40°	−15°	78	93
−40°	**−25°**	**95**	**96**
−40°	**−60°**	**93**	**95**

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

In table 2 we report the results obtained referring again to the 150-subject galleries, but considering subsets of the fiducial points; two cases have been experimented: the first considers nose, mouth and chin only (NMC), and the second takes into account eyebrows, eyes, and nose (EEN).

Gallery	Test	% First rank		% First 5 ranks	
		NMC	EEN	NMC	EEN
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°	+60°	58	87	75	94
−40°	−25°	72	93	82	95
−40°	−60°	61	80	88	94

Table 2. Recognition results obtained referring to the 150-subject galleries, having only subsets of characteristics.

We observe that the eye and eyebrow occlusion causes a greater loss of performances than the occlusion of mouth and chin, meaning 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).

Regarding the 50 and 100 subject galleries, while for the EEN experiments we obtained only slight increments, for the NMC one we obtained significant success increments (12% and 6% respectively).

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 available images and 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.

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 1800 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 [5], while being completely automatic. Moreover the use of a smaller fiducial point set allows to reduce the computational costs.

A last consideration regards the choice of the factors to consider for the recognition. In this paper we have used the information linked with the fiducial points only, but we believe that an improvement can be obtained taking also into account the information given by the feature geometry and by the position of intermediate points.

References

[1] P. Campadelli, F. Cusmai, and R. Lanzarotti. A color based method for face detection. *To appear in International Symposium on Telecommunications (IST2003), Isfahan, (Iran)*, 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] J. Koenderink and A. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[4] M. Turk and A. Pentland. Face recognition using eigenfaces. *Journal of cognitive neuroscience*, 3(1), 1991.

[5] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. In L. J. et al., editor, *Intelligent biometric techniques in fingerprints and face recognition*, pages 355–396. CRC Press, 1999.

[6] A. Yuille, P. Hallinan, and D. Cohen. Feature extraction from faces using deformable templates. *International journal of computer vision*, 8(2):99–111, 1992.