A face recognition system based on local feature analysis

Stefano Arca, Paola Campadelli, Raffaella Lanzarotti

Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41 20135 Milano, Italy
as486518@mat.unimi.it
{campadelli,lanzarotti}@dsi.unimi.it

Abstract. In this paper a completely automatic face recognition system is presented. The system is inspired by the elastic bunch graph method, but the fiducial point localization is completely different and does not require any operator intervention. Each fiducial point is characterized applying a bank of filters which extracts the peculiar texture around it (*jet*). The performances of the steerable Gaussian first derivatives basis filters are compared to the ones of the Gabor wavelet transform, showing similar results when images of faces in approximately the same pose are compared.

1 Introduction

Human face recognition has been largely investigated for the last two decades. It has a wide application field ranging from the commercial to the security one.

Eigenface [6] is one of the most famous face recognition approach; it is a fast, simple, and effective method, but not scale and light condition invariant. 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 often unable. Another approach [7], [5] represents 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 much better than the others in terms of rotation, light and scale invariance, but a manual definition of appropriate grid structures and a semi-manual process of bunch graph acquisition are required.

In this paper we present an approach based on the multiresolution technique presented in [7], but with a new and completely automatic approach to localize the facial fiducial points; besides, the Gabor wavelet analysis [7] is compared to a technique based on steerable Gaussian first derivative basis filters [4]. Such representations have shown similar results, above all when images of faces in approximately the same pose are compared.

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 localizes the face and the facial features, extracts the facial fiducial points, determines the

head pose and thus the gallery to be used for the comparison, computes the face local characterization, and compares the face with the gallery images. We observe that while the fiducial point extractor uses color information, the face analysis is done on the gray levels only. Encouraging preliminary results obtained on a small database consisting in 300 images of 50 subjects are presented.

2 Face and facial features localization

On color images of face foregrounds, we localize the face with a very precise skin map [2]. Briefly, we characterize the skin colors, represented in the YPbPr color space, with a bidimensional Gaussian mixture model, then we identify the pixels in the image which have a high probability of being skin and iteratively grow the corresponding regions, making the expansion stop at the face borders. In figure 1 the algorithm outputs on three images are presented.

Fig. 1. Some skin map localization results

To localize the eyes we combine two different types of information: the eye chromatic characterization, and the binary image obtained clustering into two clusters the gray level image corresponding to the skin map [2].

When the two eyes are found we estimate the eventual head tilt determining the tilt of the line passing through the center of the two eye regions. This allows us to 'straight' the head and deal in the next steps with a vertical head.

In the straightened head the eyebrow regions are localized upon the eye regions, while the search of mouth and nose is circumscribed to the square under the eyes: at first we extract the shape information, which allows to localize the features roughly, and then we use the color information to tailor the feature sub-images more precisely [3].

We conclude this module with the chin description, determining the best parabola which approximates it and its vertex. The chin shape information is extracted applying a non-linear edge detector. 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. 4].

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 [8] 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.2.1], 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.

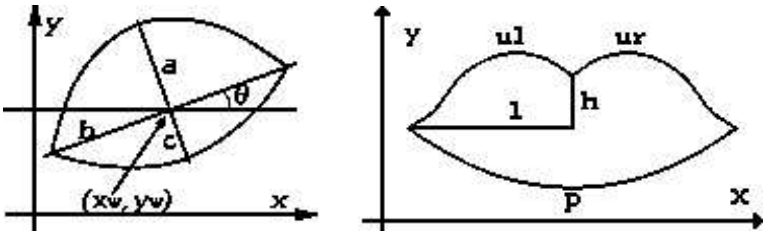

Fig. 2. Deformable templates: **1 Eye.** Parameterized by a, the upper parabola height, c, the lower parabola height, b, the eye half width, θ, the orientation angle, and (x_w, y_w) the eye center. **2 Mouth.** Parameterized by h, the upper lip height, l, the lip half width, a_p, a_{ul}, b_{ul}, a_{ur}, b_{ur}, the polynomials coefficients.

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). The image 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 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.

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. 3. Eye image processing: u plane gradient; u plane binarization; Final result.

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

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 CIE-Luv color space. Combining this information, we obtain the mouth cut and, taking its extremes, the correct corner position [3].

Thus, we define the mouth model [Fig.2.2], parameterized by $\{l, h, a_p, a_{ul}, b_{ul}, a_{ur}, b_{ur}\}$, and 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, 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. $E_i = c_2 \int_R \phi_i(\boldsymbol{x}) dA$

   where R is the region enclosed among the 3 curves and ϕ_i is the binary image obtained clustering in 2 clusters the MouthMap ($MouthMap = (255 - (C_r - C_b)) \cdot C_r^2$) and setting a pixel to 255 if it is white, to -80 if it is black.

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

   where ϕ_e is the edge image.

Eyebrow. The eyebrow description consists in the best parabola which approximates it: we calculate the vertical derivative of the eyebrow gray-level image, we binarize it keeping the 10% of the highest derivative values, and we extract the upper border of the obtained regions. We then proceed fitting these points to a parabola and extracting the upper vertex [Fig. 4].

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. 4].

Fig. 4. Examples of facial feature and fiducial point description

3.1 Experimental results

The algorithm described in section 2 has been tested on two databases: the 'Champion database'(940 images) [1], and ours (500 images). The face and feature localization works well in the 90% of the 'Champion database' images, and in the 95% of ours. The failures are caused by the skin color model inadequacy, the presence of strong shadows, and a wrong eye detection.

The fiducial point detection has been tested on our database only, showing 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 $45°$.

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; in these situations it would be necessary to introduce different characterizations.

4 Face characterization

4.1 Pose determination and face dimension normalization

Once the fiducial points have been extracted, we normalize the face images scaling them so that the area of the triangle defined by the nose tip and the two external eye corners is of 2000 pixels. Moreover, 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: if they are approximately of the same length, we conclude that the face is frontal; on the contrary 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.

We finally proceed characterizing each fiducial point in terms of the surrounding gray level image adopting the two techniques described below.

4.2 Gabor wavelet transform

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 [7], 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:

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

4.3 Steerable Gaussian first derivative basis filters

In this case the filter bank is derived from the Gaussian first derivative basis filters, varying the scale and the orientation: as in the previous case we have considered 5 scales and 8 orientations. The *jets* consist in the real coefficients obtained applying the filters to the gray image around each fiducial point. Using these filters, the similarity measure described above has bad performance; we thus resorted to the Euclidean distance between vectors, which of course, has to be minimized.

5 Experimental results

We have experimented the whole face recognition system on a database of 50 subjects. For each of them three images are catalogued in the galleries according to the pose. In such images all the fiducial points are visible.

We try to recognize frontal, right and left rotated face images referring to the three different galleries. For each experiment, we set three possible situations: all the fiducial points are visible, the eyes and eyebrows are hidden, and finally the case where the mouth and the chin are hidden. The experiments are carried out

using both the Gabor wavelet filters and the steerable Gaussian first derivative basis filters. The results are reported in the tables 1 and 2.

Gallery	**Test**	**First rank**		**First 5 ranks**	
		Gabor	Gauss	Gabor	Gauss
Frontal	Frontal	96%	98%	96%	98%
Frontal	Right	82%	74%	92%	88%
Frontal	Left	88%	84%	92%	96%
Right	Right	98%	92%	100%	98%
Right	Left	52%	34%	78%	56%
Right	Flipped Left	74%	48%	90%	72%
Left	Left	94%	86%	98%	92%
Left	Right	57%	36%	70%	68%
Left	Flipped Right	68%	30%	84%	54%

Table 1. Recognition results obtained when all the fiducial points are visible.

		Eyes, eyebrows occluded				***Mouth, chin occluded***			
Gallery	**Test**	**First rank**		**First 5 ranks**		**First rank**		**First 5 ranks**	
		Gabor	Gauss	Gabor	Gauss	Gabor	Gauss	Gabor	Gauss
Frontal	Frontal	92%	78%	98%	98%	96%	96%	96%	98%
Frontal	Right	58%	42%	78%	74%	72%	60%	88%	82%
Frontal	Left	46%	42%	82%	70%	78%	70%	86%	98%
Right	Right	88%	84%	94%	98%	96%	92%	100%	96%
Left	Left	88%	74%	96%	84%	92%	78%	100%	90%

Table 2. Recognition results obtained having only part of the fiducial points.

We observe that, when images in the same pose are compared, the two filter banks behave approximately in the same way. Moreover we observe that, independently of the adopted filter banks, the system works better when images in the same pose class are compared. This result highlights the importance of having an automatic face pose estimator which allows us to refer each time to the correct gallery.

Finally we observe that the eyes and eyebrows occlusion causes a greater loss of performances than the occlusion of the mouth and the chin, meaning that most of the discriminant face characteristics are in the upper part of the faces.

6 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 algorithm has been tested on about 1500 images performing correctly in more than the 90% of the cases. The algorithm for the fiducial point extraction, tested on 500 images, detects the points with high accuracy (errors of 1 or 2 pixels are negligible). Regarding these algorithms, we are working to enrich the skin color model, to make the eye detection more robust, and to generalize the fiducial point detection to deal with opened mouth and closed eyes; finally we are extending the method to deal with partial occluded faces and gray level images too.

The whole face recognition system has been tested on a database of 300 images of 50 subjects. Even if the database is small, we can affirm that our fiducial point extractor allows to obtain the same recognition performances as the elastic bunch graph used in [7], while being completely automatic.

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.

We are now interested in enlarging the image database, in order to analyze the system behavior in more challenging situations.

References

1. The champion database. *Web address: http://www.libfind.unl.edu/alumni/events/breackfast-for-champions.htm*, 2001.
2. P. Campadelli, F. Cusmai, and R. Lanzarotti. A color based method for face detection. *Submitted*, 2003.
3. 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.
4. J.H. Elder and S.W. Zucker. Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Learning*, 20(7):699–716, July 1998.
5. B.S. Manjunath and R. Chellappa. A feature based approach to face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 373–378, June 1992.
6. M. Turk and A. Pentland. Face recognition using eigenfaces. *Journal of cognitive neuroscience*, 3(1), 1991.
7. 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.
8. 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.