AN AUTOMATIC FEATURE-BASED FACE RECOGNITION SYSTEM

Stefano Arca, Paola Campadelli, Raffaella Lanzarotti

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

ABSTRACT

In this paper a completely automatic face recognition system is presented. The method, working on color images, determines 19 facial fiducial points, and characterizes them applying a bank of Gabor filters. The system is inspired by the Elastic Bunch Graph method [13], but it is completely automatic and computationally more efficient.

Keywords: Face recognition, Fiducial points.

1. INTRODUCTION

Human face recognition has been largely investigated for the last two decades; in the recent literature there has been a shift from global [8, 11, 12], to local [6, 9, 10] approaches which have proven to perform better.

In this paper we present a local approach, based on the idea presented in [13], but with a new and completely automatic method to localize the facial fiducial points. We consider a set of 19 points: the eyebrow and chin vertices, the nose tip and lateral extremes, the eye and lip corners and upper and lower middle points, and the mid-point between the two eyes.

To cope with images of people in frontal, right and left rotated pose, the system builds different galleries, each one containing one image per person. This choice has been driven by the observation that the recognition is more robust when the angular disparity between the gallery images and the test ones is at most of 15° [5].

Both the gallery and the test images are characterized applying a bank of Gabor filters in correspondence to each fiducial point. Thus, given a test image, the system computes its face characterization (the *jets vector* J), selects the reference gallery, G, on the basis of the pose estimated for the test image, and measures the similarity between J and the *jets vectors* of the images in G. The face is recognized to be the one in the gallery which maximizes the similarity measure.

We have experimented the whole system on two databases: the XM2VTS [1] and ours (UniMiDb).

The XM2VTS database consists of images of frontal people with neutral expression acquired over a homogeneous dark background; the illumination is uniform, and the image scale is fixed (about (230×300) pixels). For our experiments, we have considered 750 images, that is all those of people without glasses.

The UniMiDb database consists of 400 color images of very different scales (from (25×30) to (500×650) pixels), with homogeneous and light-colored background, under frontal illumination. Faces can be either in frontal position, rotated around the head vertical axis of 30° at most, or tilted laterally of about 10°.

2. FACIAL FEATURES AND FIDUCIAL POINTS LOCALIZATION

The first step consists in detecting the face in the image and localizing the corresponding facial features (eyes, nose, mouth, and chin). In [3, 4, 7] we have proposed a scale-independent method that deals with images acquired with uniform, frontal and diffuse illumination, with head rotation around the vertical axis up to 45°, and lateral tilt of about 10°; moreover it is assumed that the mouth is closed and the eyes are opened and without glasses. At first the image is clustered into three or two clusters in case of clear or dark background respectively, and the *skin region* is extracted keeping the region with the intermediate gray-level representant in case of three clusters, or the region with the clearest representant in case of two clusters. Subsequently, limiting the search to the found *skin region*, the eyes are looked for adopting a technique based on a neural network classifier, the mouth is localized exploiting its peculiar color and shape, and the nose, eyebrows and chin are detected on the basis of the eyes and mouth positions [Fig.1].

We have experimented this method on 1150 color images, reporting correct localization of all the features in the 95% of the cases. Only in the 1.1% of the images both eyes are incorrectly localized making all the other feature localization fail.

To determine the facial fiducial points we process each sub-image separately, adopting different techniques according to the feature peculiarities.

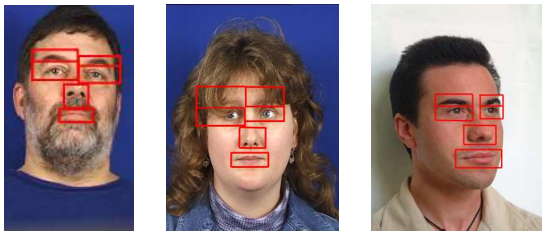


Fig. 1. Facial feature localization.

Fig. 2. Some results on the two databases.

The **eye** is described by a parametric model which is inspired by the deformable template proposed by Yuille and others [14] but with significant simplifications (6 parameters instead of 11) and variations: we have made a great deal to improve the template initialization, and to exploit the color information.

Given an eye sub-image, at first the iris is localized, allowing to initialize the eye deformable template in a plausible position and scale, which are two required conditions to obtain correct results; to this end we adopt the Hough transform for circumferences; afterwards the eventual presence of a reflex is determined and removed, since otherwise the strong reflex edges would make the template converge to them, bringing to a wrong eye description.

The eye model is made of two parabolas, representing the upper and lower eye arcs, and intersecting at the eye corners. To adapt the generic template to a specific eye, we minimize an energy function E_t which is a function of the template parameters and of the image characteristics (prior information on the eye shape, edges, and 'white' of the eye). The characteristics are evaluated on the u plane of the CIE-Luv space, where the information we are looking for is strengthened.

Regarding the **nose**, we extract its profile, that is the set of points with the highest symmetry and high luminance values, and we take as nose tip the clearest point in this set.

Our goal for the **mouth** is the determination of its corners and of its upper and lower middle points. Since we assume the mouth closed, the most robust way to estimate its corners is to determine the 'lip cut', and to take its extremes. The lip cut is characterized by high vertical derivative values, and low gray level values; combining this information we obtain a robust localization. Once determined the mouth corners, we derive the upper and lower middle points as a function of the corner positions and the lips length.

This simple method gives results comparable to those obtained adopting deformable templates (see [4, 7]) for determining the whole mouth border.

For both the **eyebrow** and the **chin** we extract the vertices of the parabolas which best approximate their shapes. The parabolas are found applying the Hough transform to the edge pixels obtained with a vertical derivative operator for the eyebrows and a non-linear edge detector for the chin.

3. POSE ESTIMATION AND RELIABLE FIDUCIAL POINT SELECTION

Given a test image, we estimate the head pose in order to compare it with the proper gallery constituted by either frontal, left or right rotated faces. To this end, we consider the triangle T defined by the nose tip, N, and the two external eye corners, E_{sx} and E_{dx}, and we determine the pose on the base of the ratio $r = \overline{NE_{dx}}/\overline{NE_{sx}}$ [Fig. 3].

Fig. 3. Segments used to determine the head pose.

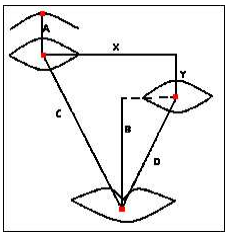
Moreover we observe that the area of the triangle T can be used to infer robustly the image scale; since the face characterization module (section 4) needs to deal with images of a fixed scale, we have to reduce the images to almost a common size. We thus scale all the images so that the triangle area is of 2000 pixels, making them comparable.

3.1. Selection of the reliable fiducial points and error correction

Both the feature localization and the fiducial point estimation introduce some errors. However it happens very seldom that all the features are wrongly localized or described; this observation can be very useful since, if we manage to recognize automatically which fiducial points have been wrongly determined, we can discard them and base the face recognition on the remaining ones.

To this end, we have considered 200 normalized images whose features had been correctly localized and

the fiducial points well determined, and we have statistically estimated both the typical feature dimensions and their relative positions (figure 4,*right*), considering in particular the distances drawn in figure 4. On the basis of this information, we have derived the rules that follow to discard the unreliable fiducial points.



Measure	μ	σ
Mouth Area	940	91
Eye Area	296	28
X	63	5
Y	0	1
A	18	3
C	81	8
D	80	8

Fig. 4. Means (μ) and variances (σ) of the distances considered for the fiducial point selection.

We organize the rule description according to the examined feature:

- **Eye:**

  Eliminate it if its area is not in the range $(\mu(Eye\ Area) \pm 2 \cdot \sigma(Eye\ Area))$ or if the ratio between its height and width is greater than 0.7.

  If the two eyes are unaligned ($Y > 7$) then eliminate the one whose distance from the mouth (C or D) is further from the corresponding mean value.

  If the two eyes are too close to each other ($X < \mu(X) - 2 \cdot \sigma(X)$) maintain the one further from the vertical axis passing through the mouth centroid.

- **Mouth:**

  Eliminate it if its area is not in the range $(\mu(Mouth\ Area) \pm 2 \cdot \sigma(Mouth\ Area))$ or if its mid-point abscissa is not within the abscissae of the two eye mid-points.

- **Eyebrow, Nose and Chin:**

  In case both eyes have been eliminated, eliminate the two eyebrows too.

  In case the corresponding eye has already been eliminated, compare the eyebrow with the other one, and maintain it only if they are aligned.

  Eliminate the eyebrow fiducial point if it is either too distant from the centroid of the corresponding eye ($A > \mu(A) + 2 \cdot \sigma(A)$) or too unaligned (its abscissa is not within the eye corners).

  Eliminate the nose or the chin fiducial points if their abscissae are not within the mouth corners.

- **Whole images:**

  Discard the whole image if both the mouth and at least one eye have been eliminated.

The thresholds used in the described rules have been chosen so that no correct fiducial point is rejected.

This module, applied to all the outputs obtained on the XM2VTS and UniMiDb databases (1150 images), has allowed to discard completely the 1.1% of the images, that is the ones where the feature localization module failed in determining all the features.

The remaining images are all considered reliable for the recognition. In the 76.6% of the cases, all the 16 fiducial points have been maintained, while in the other 22.3% either isolated points (9.8%) or a complete feature have been discarded (12.5%).

In order to recover some of the discarded fiducial points, we search for them once more, exploiting both the *a priori* knowledge on the feature relative positions and dimensions, and the gathered information on the reliable fiducial points. This process has allowed to increase the number of images on which all the 16 fiducial points are maintained (from the 76.6 to the 92%), and it has reduced systematically the number of images on which some fiducial points are discarded (4.2%). We eliminate one eye or the mouth only in the 1.3% and the 0.6% of the cases respectively.

3.2. Inference of additional fiducial points

Once determined the reliable fiducial points, we infer from them additional ones, in order to gather more information for the recognition. In particular we are interested in exploring the regions in correspondence to the mean point between the eyes and the nose lateral extremes [Fig. 5].

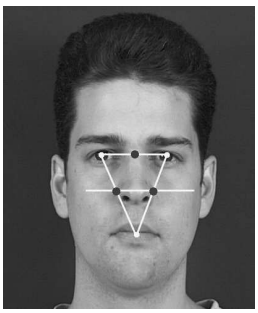


Fig. 5. Inferred points.

Thus, referring to the calculated fiducial points, the *eyes mean point* can be easily derived taking the mid point of the segment which extremes are the internal corners of the left and right eyes.

The *nose lateral extremes*, are determined as the intersection between the straight line passing through the mouth mid-point and the corresponding eye mid-points, and the horizontal straight line passing through the nose tip. In case of rotated faces, only one nose extreme is visible. Thus, on the basis on the estimated

pose, we decide whether to consider both the nose lateral extremes or only one.

Experimentally we have observed that, for neutral expression faces, the three additional fiducial points allow to increase the recognition rate systematically; on the contrary, in case of non-neutral expression, the nose lateral extremes provoke a loss of performances.

4. FACE CHARACTERIZATION

Once the fiducial points have been extracted, the pose determined, and the face rescaled, we proceed characterizing each fiducial point in terms of the surrounding gray level portion of image.

Following the idea of Wiskott [13], to characterize a fiducial point, we convolve the portion of gray image around it with the following bank of 40 *Gabor kernels* (we consider 5 frequencies and 8 orientations):

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

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 maintained 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. We define the similarity between two *jets vectors* as the average over the similarities between pairs of corresponding *jets*:

$$S_v(V^1, V^2) = \frac{1}{N}\sum_{n=1}^{N} S(J_n^1, J_n^2),$$

where N is the number of jets defined in both the two images, and

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

5. EXPERIMENTAL RESULTS

In the following we report the experiments carried out on the XM2VTS and the UniMiDb databases; in these experiments there is no human intervention: the algorithm starts selecting among the found fiducial points the reliable ones (section 3.1), and, on the basis of this selection, it constructs the galleries and the test sets.

Regarding the XM2VTS database, we have constructed one gallery (in this database there are only frontal images) and three test sets, each one containing one image per subject and organized according to their quality: in the gallery we put the image with the highest number of reliable fiducial points, in the first test set (T1) we put the second best image according to this criterion, and so on.

Given this setting, we have obtained a 188-subject gallery, an equivalent test set T1, and the other two test sets T2 and T3 with 179 and 168 images respectively. These values are obtained considering that in the original 750-image set, there are 202 subjects, but among them we have only one image for 12 subjects, which cannot be used for the recognition; furthermore we have to exclude 8 images which are completely discarded by the module for the selection of the reliable fiducial points, and this excludes other 2 subjects. The test sets T2 and T3 have even a lower number of images because not always we have three and, even more difficult, four available images.

The obtained results are reported in table 1.

Test Set	% First rank	% First 5 ranks
T1	95	98
T2	93	97
T3	90	96

Table 1. Recognition results obtaining referring to the 188-subject gallery, and exploiting all the maintained fiducial points.

Such results show the system works well even with a subset of fiducial points (T2 and T3); moreover we notice that the wrong matches depend on the face expression: most of them take place when the test image represents a face whose eyes look either up, down or laterally, and when the mouth shape is not neutral.

Regarding the UniMiDb, we recall it consists of both frontal and rotated faces; thus, according to the head pose, we construct three galleries of 50 people each. This is automatically done: given all the images of a subject, we cluster them according to the head pose, and we select for each pose the image with the highest number of reliable fiducial points (which will be used as gallery image) and the second best image (which will be used as test image). We obtain three 50-subject galleries and 150 test images; the recognition experiment has given the 96% of hits and the 98% of correct matches among the first five candidates.

These high performances, with respect to the ones obtained with the XM2VTS database, are due to both the lower number of images in the galleries and the neutrality of the face expression which characterizes all the UniMiDb database.

6. DISCUSSION

We have presented a completely automatic system able to recognize a face image against a closed gallery.

Experimenting the method on 1150 images, it discards completely only the 1.1% of the images, while on the others it achieves high recognition performances. The method has shown to be robust to head rotations, while it is quite sensitive to face expression variations which make the feature appearance change greatly.

A direct comparison of our system with others cannot be done, because of the databases we have adopted: most of the face recognition techniques presented in the literature work on gray level images, showing their experimental results on gray level image databases such as the FERET. However, comparing the percentage of recognition, what we can conclude is that our system performances are similar to the ones reported by well known approaches such as the Elastic Bunch Graph Matching [13], the PCA [10], or LDA [15], although our method is completely automatic, robust to head rotations and to scale variations, and, being local-based, it can be extended to deal with partial occlusions.

7. REFERENCES

[1] The xm2vts database. *Web address: http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/*, 2001.

[2] A.Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *CVPR1994*, pages 84–91.

[3] S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on local feature analysis. *AVBPA2003, Lecture Notes in Computer Science*, 2688:182–189.

[4] P. Campadelli and R. Lanzarotti. Fiducial point localization in color images of face foregrounds. *Image and Vision Computing (to appear)*.

[5] P. Campadelli, R. Lanzarotti, and C. Savazzi. A feature-based face recognition system. *ICIAP2003*, pages 68–73.

[6] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global vesus component-based approach. *ICCV2001*, pages 688–694.

[7] R. Lanzarotti. *Facial feature detection and description*. PhD thesis, Università degli Studi di Milano, Web address: http://homes.dsi.unimi.it/ lanzarot/, 2003.

[8] A.M. Martinez and A.C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.

[9] J. Matas, K. Jonsson, and J. Kittler. Fast face localisation and verification. *Image and Vision Computing*, 17(8):575–581, 1999.

[10] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[11] E. Osuna, R. Freund, and F. Girosi. Training support vector machines:an application to face detection. *CVPR1997*.

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

[13] 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.

[14] 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.

[15] W. Zhao, R. Chellappa, and P.J. Phillips. Subspace linear discriminant analysis for face recognition. Technical Report CAR-TR-914, University of Maryland, 1999.