An efficient method to detect facial fiducial points for face recognition

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. It consists of two main modules: in the first, the facial fiducial points are localized, and in the second the face is characterized applying a bank of Gabor filters in correspondence to the found fiducial points. This method is an evolution of the one we have presented in [2]: the fiducial point estimation is more efficient and self-correcting, and the face characterization modified.

1. Introduction

Human face recognition has been largely investigated for the last two decades [9]; in this period many face recognition techniques (FRTs) have been proposed [7, 4, 8] and most of them cannot leave apart from localizing the facial features and/or the corresponding fiducial points, determining the feature localization manually.

In [2] we have already presented a FRT, which localizes automatically 16 fiducial points: the eyebrow and chin vertices, the nose tip, the eye and lip corners and upper and lower middle points. Here we improve the method making the fiducial point estimation more efficient, introducing a module for the self-evaluation of eventual errors in the feature description, and modifying the face characterization. These changes have led to a considerable improvement of both the feature description and of the recognition results.

Each image is characterized applying a bank of Gabor filters in correspondence to each fiducial point.

Our system builds three galleries, each one containing one image per person in frontal, right and left rotated pose respectively, observing that the recognition is more robust when the angular disparity between the gallery images and the test ones is at most of 15° [3].

Given a test image, the system computes its face characterization, selects the reference gallery, G, on the basis of the pose estimated for the test image, and maximizes a suitably defined similarity function (section 5.2). We have experimented the whole system on 750 images (all those of people without glasses) taken from the XM2VTS [1], and ours, the UniMiDb database, containing 400 images.

2. Feature localization

The first step consists in detecting the face in the image and localizing the corresponding facial features (eyes, nose, mouth, and chin). In [2, 5] 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.

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.

We observe this module allows us to estimate the face scale, and to normalize the corresponding image to about (230×300) pixels, which is a size compatible with the dimensions of the filters used in the following steps.

3. Feature description

Given the feature sub-images, we process each of them separately, with the aim of extracting the most characteristic fiducial points. In [2] we have presented a method to determine robustly and efficiently the fiducial points associated to the eyebrows, the nose and the chin; regarding the eyes and the mouth we adopted the deformable template technique which estimates the whole feature contour, but which is computationally very expensive. Here we propose an efficient alternative for the eyes, while for the mouth we consider the mouth corners used for the template initialization and we derive the upper and lower middle points as a function of them.

Given the eye sub-image, we apply the Gaussian first derivatives basis filters with standard deviation $\sigma = 1$ and considering 8 directions ($\theta = j\pi/8, j = 0, ..., 7$). The resulting 8 images I_j are then thresholded putting to 1 the $n\%$ of the pixels with the highest values; n is set to 15 for the images I_j with $j = 3, 4, 5$ (that is the ones corresponding to the horizontal borders), and to 30 in the other cases; this is done since the horizontal borders are stronger than the others, and applying a more selective threshold to them allows to balance the contribution in all the directions. The obtained binary images are then combined determining the *border image* B which always represents the eye, although also other components might be included (such as the eyebrow or the hair).

Subsequently, we calculate the vertical and horizontal projections of B and we extract the portion S of image corresponding to the largest intervals different from 0 in both projections. S represents almost exclusively the eye, which can be identified selecting the largest connected region r; finally we dilate r with a structuring element (3×3), so that eventual closed but disconnected regions are joint to r. The dilated region d corresponds with high reliability to the eye, whose fiducial points can be easily determined taking the upper, lower, left and right extremes of d.

This method has been tested on 2500 eye sub-images automatically extracted, detecting in all the cases the eye region, and giving the correct fiducial points in the 98% of the cases, thus improving the results obtained with deformable template; in the remaining 2% of the images at least one fiducial point has been wrongly determined.

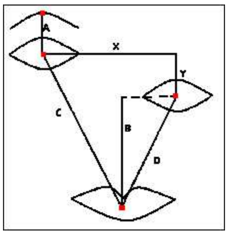
4. Scale estimation and error correction

The fiducial point estimation gives good results, although it can be further improved: it happens very seldom that, given a face image, all its features are wrongly localized or described, thus if we manage to recognize automatically which fiducial points have been wrongly determined, we can try to recover them on the basis of the positions and dimensions of the reliable features. In case of a second failure, the wrong fiducial points will be discarded for the recognition.

In this section we present a module which detects the eventual errors and tries to deal with them. To this aim a first fundamental step is a more precise scale estimation of the face image which can be done calculating the area of the triangle T defined by the nose tip, N, and the two external eye corners, E_{sx} and E_{dx}. In order to reduce all the face images to almost a common size, we scale them so that the triangle area is of 2000 pixels. We observe that the triangle T is used also to estimate the head pose in order to compare the face image with the proper gallery constituted by either frontal, left or right rotated faces. The pose is determined on the base of the ratio $r = \overline{NE_{dx}}/\overline{NE_{sx}}$.

4.1. Error detection

In order to estimate both the typical feature dimensions and their relative positions (figure 1), we have considered 200 normalized images whose features had been correctly localized and the fiducial points well determined. 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

(a) (b)

Figure 1. (a) Distances considered for the fiducial point selection; (b) Means (μ) and variances (σ) of the considered measures.

We organize the rule description according to the examined feature:

1. Eye: eliminate it if its area is not in the range *(μ(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* < *μ(X)*− 2 · *σ(X)*) maintain the one further from the vertical axis passing through the mouth centroid.

2. Mouth: eliminate it if its area is not in the range *(μ(Mouth Area) $\pm$ 2 $\cdot$ σ(Mouth Area))* or if its mid-point abscissa is not within the ones of the two eye mid-points.

3. 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.

4. Whole images: discard the whole image if 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.

In order to recover the discarded fiducial points, we search for them once more, exploiting both the gathered information on the reliable fiducial points, and the *a priori* knowledge on the feature relative positions and dimensions. For example, in case one eye has been discarded, we determine its rough position on the base of the mouth and the other eye positions.

This module, applied to all the outputs obtained on the XM2VTS and UniMiDb databases (1150 images), has allowed to reduce the discarded fiducial points, see table 1.

Kind of Error	% Detected	% Unsolved
Whole Image	1.1%	0.7%
Two Eyes	0.8%	0.1%
One Eye	5.1%	0.5%
Mouth	4.3%	1.2%
Eyebrows, Nose, Chin	7.7%	5.2%
No Error	81.0%	92.3%

Table 1. Percentage of detected errors before and after the corrections.

5. Face recognition

5.1. Inference of additional fiducial points

Given 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. 2].

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.

5.2. 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 [8], a Jet of 40 coefficients is assigned to each fiducial point, convolving the portion of gray image around the point with the following bank of 40 *Gabor kernels* (we consider 5 frequencies and 8 orientations):

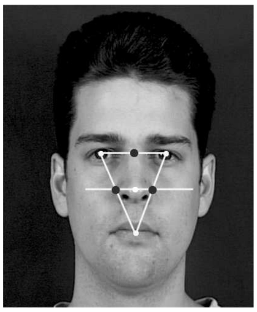


Figure 2. The dark points are inferred by intersection of straight lines through the extracted fiducial points.

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

Given a test image T, we look for "the closest" image in the gallery G proceeding as follows:

- for each fiducial point i, and for each image $k \in G$, compute the similarity between corresponding Jets:

$$S^{k,i} = S(J^{T,i}, J^{k,i}) = \frac{\sum_z J_z^{T,i} J_z^{k,i}}{\sqrt{\sum_z (J_z^{T,i})^2 \sum_z (J_z^{k,i})^2}}$$

  where $z = 0, ..., 39$.

- for each i, order the values $\{S^{k,i}\}$, and assign to each a weight $w^{k,i}$ as a function of its ordered position p.

  The weight $w^{k,i} = f(p)$ is determined as:

$$f(p) = c \cdot [\ln(x + y) - \ln(x + p)],$$

  where $y = \frac{|G|}{4}$, $x = e^{-\frac{1}{2}}$, and c is a normalization factor.

- for each gallery image k, consider the set, *Best10*, of the 10 fiducial points which have got the highest weights, and determine the score:

$$\text{Score}(k) = \sum_{i \in \textit{Best10}} w^{k,i} S^{k,i}.$$

  The face represented in T is recognized as the one in the image k with the highest score.

This technique gives better results than the computation of the average over all the similarities [5], since it allows to discard wrong matches on single points.

5.3. Experimental results

We report here the experiments carried out on the UniMiDb database and on the subset of the XM2VTS database consisting of all the 750 images of people without glasses.

In the first experiment we consider frontal images only, taking into account all the XM2VTS images and the frontal ones of the UniMiDb. We refer to a 236-subject gallery (50 of the UniMiDb and 186 of the XM2VTS respectively), constructed choosing among the images of a subject the one with the most neutral expression. Moreover, we have constructed different test sets referring to the images not used for the gallery construction: T1 contains all the remaining UniMiDb images, while T2, T3 and T4 cluster the XM2VTS images so that each test set contains at most one image per subject and grouping the images on the basis of the face expression (T2 contains the face with the most neutral expression, while T4 contains the images representing the less neutral faces). The obtained results are reported in the following table.

Test Set	% First rank	% First 5 ranks
T1	96	96
T2	98	100
T3	97	99
T4	93	97

This experiment shows the system is quite sensitive to feature appearance variations, such as smiling mouth or turned eyes (test set T4) which is the main limit of local filter approaches. In order to test the system robustness to head rotations, we have carried out another experiment, referring to the UniMiDb which consists of both frontal and rotated faces; 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.

6. Discussion

We have presented a completely automatic system able to recognize a face image against a closed gallery. The method has shown to be efficient, robust to head rotations, while it is quite sensitive to face expression variations which make the feature appearance change greatly. It is based on a module for the feature extraction and description, which is self-correcting, and determines with high reliability (table 1) the correct fiducial points. This module can be very useful for appearance-based method too, to automate the face description and morphing.

A direct comparison of our system with others cannot be done: 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, we can conclude that our system performances are similar to the ones reported by well known approaches such as the Elastic Bunch Graph Matching [8], the PCA [6], or LDA [9], but 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.

References

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

[2] 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.

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

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

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

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

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

[8] 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.

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