

A face recognition system dealing with expression variant faces

Stefano Arca*, Paola Campadelli, Raffaella Lanzarotti, Giuseppe Lipori
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico, 39/41 20135 Milano, Italy
{arca,campadelli,lanzarotti,lipori}@dsi.unimi.it

Abstract

In this paper we present the generalization of the automatic face recognition system, presented in [2], making it able to deal with different expressions and the presence of spectacles. The system behavior is measured on the FRGC 1.0 and the XM2VTS databases, addressing some of the current challenges highlighted by Phillips and others [7]. We set up experiments having in the gallery, for each subject, both one still image and multi-still images with different expressions. Besides, we investigate how the image resolution affects the performances and test the scalability of the system doubling the gallery size.

1. Introduction

The problem of face recognition has been largely investigated for the last two decades. According to [6] and [8] the existing systems work fairly only under strict assumptions: most of them deal with frontal faces, with no occlusion, neutral expression, uniform background, homogeneous illumination, and known scale. Their performance decreases drastically when even one of these aspects is uncontrolled.

To evaluate the progress achieved in the last five years, Phillips and others [7] launched the Face Recognition Grand Challenge (FRGC) consisting of several experiments which cover a wide range of aspects (3D face recognition systems, single and multi-still images, variations in expression, lighting conditions, etc).

Referring to some of these challenging problems, in this paper we examine the behavior of a generalization of the face identification system we presented in [2]: the original method, developed for color images,

showed to be efficient, and robust to head rotations and scale variations, while it made the assumptions of neutral expression, and no spectacles which are now removed. To test the robustness of the generalized system to the spectacle presence we refer to the XM2VTS database, while to test its behavior dealing with non-neutral expressions we adopt the controlled images of the FRGC 1.0 public database. The images represent frontal faces, taken under two lighting conditions, homogeneous background, with either neutral or smiling expression. We set up two classes of experiments: the first, with one image per subject in the gallery, aims to measure the system behavior when the subjects in the gallery and in the test have either the same or different expressions. The second class aims to evaluate the increase of the recognition rate when the gallery consists of multi-still images with different expressions for each subject.

Furthermore, according to Phillips and others [7], there are two other points which deserve some considerations. The first concerns the image resolution: nowadays the technology allows both to acquire and to process high resolution images; the question is whether this further information improves the system performance or not. The second crucial point regards the scalability of a system, that is how the gallery size affects the system performance. Thus, we set up two further experiments to investigate these topics.

In the following, we briefly describe the new modules which generalize our face recognition system (section 2), and then we describe the experiments carried out and the results obtained (section 3).

2. Method description

In [2] we presented a component-based face recognition system which does not need any training session. It works on color images and consists of three modules: the first localizes the face and the facial features (eyes,

*Work partially supported by the PASCAL Network of Excellence under EC grant no. 506778.

eyebrows, nose, mouth, and chin); the second processes the feature sub-images and extracts precisely 24 facial fiducial points¹; the third characterizes, compares and recognizes the faces applying a bank of Gabor filters in correspondence to the found fiducial points.

In order to generalize the system, the steps which need to be modified are the ones which extract the fiducial points of eyes and mouth: the method we proposed determines them robustly adopting the deformable templates, which give very precise results in controlled situations (neutral expression and no partial occlusions), but they are unreliable in dealing with spectacles or open mouths. Thus, we introduce a module to determine if spectacles are present or not, and to classify a mouth as open or closed. This information allow us to adopt techniques more flexible than deformable templates when required.

2.1. Detection and process of eyes with spectacles

The eventual presence of spectacles is recognized on the basis of the analysis of the edges in the eye sub-images; the method is based on the consideration that when spectacles are present, the edges are stronger and involve more pixels (Figure 1).



Figure 1. Edges with and without spectacles.

To determine the eye fiducial points in images of people wearing glasses, we exploit the information given by the eye localizer, and some statistical considerations which relate the intra-ocular distance and the mean width and height of the eyes. Regarding the eye localizer [3], we observe that it localizes with high precision the eye centers by means of an SVM: testing the method on both the XM2VTS [1] and Controlled FRGC databases, we obtained that in the 94.5% of the images the distance between the correct eye centers (ground truth) and the points automatically determined is less or equal than 0.1 times the intra-ocular distance.

To infer the eye fiducial points having the eye centers, we examined 1500 frontal images from different databases obtaining that the mean ratio between the

¹Such points are: the eyebrow and chin vertices, the nose tip, its lateral extremes and its vertical mid point, the eye and lip corners and their upper and lower middle points, the mid-point between the two eyes, and four points on the cheeks

intra-ocular distance and the eye width is 2.5 with a standard deviation of 0.2, while the mean ratio with the eye height is 5.6 and its standard deviation is 0.9.

In order to evaluate the reliability of the eye fiducial points determined with this method, we measured the impact they have on our face recognition system: we set up two experiments referring to the whole XM2VTS database. Regarding the images of people wearing glasses, in the first we consider the ground truth of the eye fiducial points [2], while in the second we use the eye fiducial points automatically determined. The results show a decrease of performance of less than the 1% (that is from the 95.6% to the 94.8%).

2.2. Detection and process of open mouths

At first we want to determine whether the mouth is open or closed; to highlight the lips, we apply in the YC_bC_r color space the following transformation to the mouth sub-image:

$$MouthMap = (255 - (C_r - C_b)) \cdot C_r^2.$$

Then, we look for the mouth corners: we threshold the *MouthMap* keeping the 20% of its highest values, and we take the most lateral extremes (Figure 2).



Figure 2. Thresholded *MouthMap* of an open and a closed mouth.

To classify automatically a mouth as ‘open’ or ‘closed’ we examine the rectangular portion R of the thresholded *MouthMap*, included between the two corners and high as $1/8$ of the distance between them. In case of close mouth most of the pixels in R are set to 1, that is they belong to the lips, while in case of open mouth there is a large number of pixels set to 0, that is corresponding to the mouth inside.

In case of open mouth, we adopt the snakes [5] since we verified that, when suitably initialized, they can robustly describe very different shapes. We anchor the snake to the mouth corners and let it adapting to the *MouthMap* (Figure 3).

3. Experiments

The experiments are carried out on the FRGC public database, taking into account the controlled images,

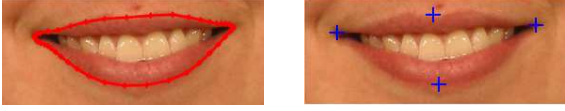


Figure 3. The mouth contour and its fiducial points.

that is the ones acquired in a studio setting, with uniform background, taken under two lighting conditions and with either neutral or smiling expressions. The images have high resolution, represent 275 subjects in frontal position, and are organized in *subject sessions*, consisting of four images, two with neutral and two with smiling expression. The number of sessions varies from subject to subject, between one to seven.

The first experiment we present (*Exp1 - Neutral Gallery*) is the very classic face recognition problem: recognize any test image, either with neutral or smiling expression, given a gallery of one image per subject with neutral expression. In this framework the gallery consists of 275 images (for each subject we extracted from the first session the first image with neutral expression), and the test set consists of the 792 images of the second session divided according to the expressions.

Table 1 shows the recognition results where the performances are evaluated according to the *Cumulative Match Characteristic* CMC metric presented in [4] and defined as:

$$P_R(r) = \frac{|C(r)|}{|T|} \cdot 100$$

where $C(r)$ is the set of images in the test set T that are recognized at *rank* r or better.

<i>Exp1 - Neutral Gallery</i>					
Test Set	$P_R(1)$	$P_R(2)$	$P_R(3)$	$P_R(4)$	$P_R(5)$
$T_{Neutral}$	96.0%	98.0%	98.5%	99.0%	99.0%
$T_{Smiling}$	80.8%	88.9%	89.9%	91.9%	93.9%

Table 1. Recognition rate referring to a gallery with one image per subject with neutral expression, and test images representing subjects both neutral and smiling.

The second experiment we report (*Exp1 - Smiling Gallery*) is very similar to the first, except for the expression of the subjects in the gallery. As before the gallery consists of 275 images, but in this case for each

subject we extracted the first image with smiling expression from the first session. Results are reported in Table 2.

<i>Exp1 - Smiling Gallery</i>					
Test Set	$P_R(1)$	$P_R(2)$	$P_R(3)$	$P_R(4)$	$P_R(5)$
$T_{Neutral}$	71.6%	80.3%	83.1%	84.1%	86.9%
$T_{Smiling}$	94.4%	96.2%	96.5%	97.2%	98.3%

Table 2. Recognition rate referring to a gallery with one image per subject with smiling expression, and test images representing subjects both neutral and smiling.

Comparing Tables 1 and 2, one can observe that, as expected, the recognition rate is significantly higher when faces with the same expression are compared, being better in case of neutral expression. Regarding the comparison of faces with different expression, what is unexpected is the significant decrease in the recognition rate when the images in the gallery represent smiling faces and the test images are neutral faces (compare second row of Table 1 with the first of Table 2).

A second type of experiments aims to test the system behavior referring to multi-still images. We built the gallery taking into account all the 1100 images acquired in the first session (four per subject); regarding the test set, we set up two cases: in the first (*Exp2 - 1 vs 4*) we use the test set defined in *Exp1*; in the second (*Exp2 - 4 vs 4*) we refer to all the four images per subject of the second session as a unique probe.

In *Exp2 - 1 vs 4* the system compares the test image with the gallery, evaluating the similarity score as described in [2]. Let us denote it as $s_{j,i}$ where $j = 1, \dots, 275$, and $i = 1, \dots, 4$, then we compute the global score S_j for each subject j of the gallery in two different ways:

$$S_j = \max_i(s_{j,i}) \quad (1)$$

$$S_j = \sum_{i=1, \dots, 4} s_{j,i} \quad (2)$$

A person is recognized as the one which corresponds to the maximum global score. On this database, the two methods used to evaluate the global scores gave very similar results. In Table 3 we report those obtained with the global score (2), since the recognition rate is 0.5% higher than with score (1).

In the second situation (*Exp2 - 4 vs 4*), for each subject j and each test image k ($k = 1, \dots, 4$) we com-

<i>Exp2 - 1 vs 4</i>					
Test Set	$P_R(1)$	$P_R(2)$	$P_R(3)$	$P_R(4)$	$P_R(5)$
$T_{Neutral}$	96.8%	98.5%	98.8%	99.0%	100%
$T_{Smiling}$	98.3%	98.8%	99.8%	99.8%	100%

Table 3. Recognition rate referring to a gallery with four image per subject, and test images are both neutral and smiling faces.

pute the global scores S_j^k , as described above. The final score is obtained as $F_j = \sum_k S_j^k$, and the subject is recognized as the one with the highest final score.

In Table 4 we report the results obtained using the global score (2) as before.

<i>Exp2 - 4 vs 4</i>					
Test Set	$P_R(1)$	$P_R(2)$	$P_R(3)$	$P_R(4)$	$P_R(5)$
Test	99.5%	99.5%	99.5%	99.5%	100%

Table 4. Recognition rate referring to a gallery and a test set with four images per subject.

Comparing the results in Tables 3 and 4 with those in Tables 1 and 2 we can observe that having a multi-still gallery with different expressions of the same subject, increases the overall recognition rate significantly, since each test image has a higher probability to find in the gallery a closer one. Having multi-still images also in the test we obtained an even higher performance.

Furthermore, we are interested in measuring how the image resolution and the gallery size affects the system performances. Regarding the first question, we have varied the image scale, so that the intra-ocular distance varies from 50 to 250 pixels with a step of 25, varying the Gabor filter supports accordingly. We observe that on average the behavior does not change significantly in the range 75 - 250, while if the intra-ocular distance is lower than 75 pixels the performance decreases. For our experiment we scaled the images to an intra-ocular distance of 150 pixels.

To test the scalability of the system, we double the gallery size, adding to the 275 subjects of the FRGC the 295 subjects of the XM2VTS. We test the system with the $T_{Neutral}$ of the *Exp1*, and with the 885 images of the XM2VTS not inserted in the gallery.

We can say that the system performance are not affected doubling the gallery size, since on both the databases the recognition rate does not decrease.

4. Conclusions

In this paper we presented a generalization of a face recognition system, which can now deal with head rotations, scale variations, presence of spectacles, and both neutral and smiling expressions.

The system has been tested on the FRGC and XM2VTS databases, allowing to conclude that: first, having multi-still images at least in the gallery helps significantly in improving the performances, above all if dealing with different expressions; second, the image resolution affects the system performance only if it is low (intra-ocular distance lower than 75 pixels); finally, we observe that doubling the gallery size, that is from 275 to 570 subjects, the system performance does not decrease, showing a certain scalability.

We are now interested in testing the system on the FRGC uncontrolled sub-set, in order to evaluate also the robustness to blurred images, with strong shadows and different illuminations.

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 automatically determined facial fiducial points. *Pattern Recognition*, 39(3):432–443, 2006.
- [3] P. Campadelli, R. Lanzarotti, and G. Lipori. Eye localization and face recognition. *RAIRO, to be published*, 2006.
- [4] P. Grother, R. Micheals, and P. Phillips. Face recognition vendor test 2002 performance metrics. *Proceedings AVBPA 2003*, pages 937–945, 2003.
- [5] G. Hamarneh. Image segmentation with constrained snakes. *Swedish Image Analysis Society Newsletter SS-ABlaskan*, 8:5–6, 2000.
- [6] J. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone. Face recognition vendor test 2002: overview and summary. *[Online]*, Available: <http://www.frvt.org>, 2003.
- [7] P. Phillips, P. Flynn, T. Scruggs, and K. Bowyer. Overview of the face recognition grand challenge. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [8] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM, Computing Surveys*, 35(4):399–458, 2003.