An Integrated System for Automatic Face Recognition

Maria Paola De Rosa, Alessandro Micarelli, and Giuseppe Sansonetti

Dipartimento di Informatica e Automazione
Laboratorio di Intelligenza Artificiale
Universitá degli Studi "Roma Tre"
Via della Vasca Navale, 79
I-00146 Roma, Italia
{derosa,micarel,gsansone}@dia.uniroma3.it

Abstract. This paper presents an Automated Face Recognition (AFR) system capable of providing satisfactory results even with only one training image per individual. To obtain this result an innovative architecture has been devised with the ability to integrate organically new solutions with well-established, even classic, techniques, i.e., Principal Component Analysis (PCA) and Discrete Cosine Transforms (DCT). The process of identification thereby concludes successfully even under trying circumstances; that is, even in the presence of consistent variations in the orientation, scale and expression of the face under observation. Radial Basis Function (RBF) neural networks are used as classifiers, the output of which converge into a single block that in turn adopts a decisional strategy. Experimental results on the Face Recognition Technology (FERET) database demonstrate the validity of our approach, and invite comparison with other systems of face recognition. . . .

1   Introduction

Automated Face Recognition (AFR) belongs to a wider field of research which involves the design and development of biometric systems, i.e., automatic devices whose objective is to recognise, or to verify, the identity of an individual on the basis of some physiological characteristic. In the case of AFR, the objective is pursued by way of a comparison between a candidate's face and a database of sample faces. Unfortunately, the disruptive factors which may make difficult the task of a AFR system are as diverse in kind as they are plentiful in number: they act in such a way that AFR cannot guarantee results comparable to those obtained by way of physiological characteristics.

The present work is organised as follows: : in section II we present an overview on the state of the art of AFR technology; in section III the overall architecture of the system is presented; in section IV experimental results are reported; whilst in section V a space is reserved for some concluding thoughts and our conjectures on possible future developments.

2   State of the Art

The taxonomies proposed for AFR systems [2] are numerous. Among these there is the work of William A. Barrett [1], who explains the technology for AFR in terms of three principal categories:

1. neural networks;
2. eigenfaces;
3. wavelet/elastic matching.

Our system, for its part, makes use of Principal Component Analysis (PCA), also known as the Karhunen-Loève Transform (KLT). This mehod is widespread in the field of signal representation. Sirovich and Kirby were the first to apply PCA to face recognition [3]. Turk and Pentland were afterwards to utilise PCA in the development of the *eigenface* method [4].

As we shall see further on, use is also made here of the Discrete Cosine Transform (DCT) relative to the pre-processing phase. It is by no means unusual that DCT should appear frequently in AFR literature, even if in a form substantially different from that being proposed here .

As far as classificatory operations are concerned, Radial Basis Function neural networks (RBF) have been utilised. The reason for such a choice stems from the fact that input patterns configurations - if they are to be classifiable at all - demand the formation of complex, decision-making regions, unobtainable by way of simpler, less computation intensive linear classifiers.

3   Description of the System

Figure 1 illustrates the typical configuration of a Single-Feature Neural Network (SFNN) system, i.e., a system that extracts its features from a single domain and that employs a single neural network as classifier.   The proposed architec-

Fig. 1. Typical architecture.

ture, on the other hand, is given in Figure 2. Unlike systems of a traditional type, ours is developed in five stages: localisation of a face within the image, extraction of salient features (involving two "extractors" running parallel on two difernt domains, one based on Combined Projection, the other on Discrete Cosine Transform), dimensionality reduction by means of PCA, classification and decisional strategy. The localisation of the face pertains to the pre-processing stage and, as such, is beyond the scope of the present work.
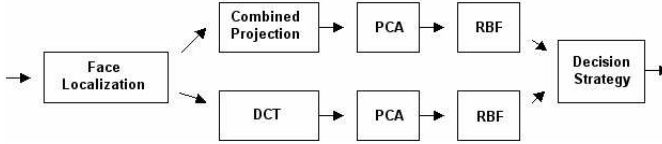
Fig. 2. Our architecture.

3.1   Combined Projection

First proposed by Jainxin Whu and Zhi-Hua Zhou as an extension of PCA [5], Combined Projection is here utilised so that the system might better withstand changes in projection and depth rotation. It is based on the combination of the original image with its horizontal and vertical projections. It is expressed as

$$P_\alpha(x,y) = \frac{P(x,y) + \alpha Mp(x,y)}{1 + \alpha} \tag{1}$$

$P(x, y)$ is the original image (such a function represents the intensity value of every single pixel in the image), with $x \in [1, n]$, $y \in [1, m]$ e $P(x, y) \in [0, 255]$, while $Mp(x, y)$ is the *map of characteristics*.

$$Mp(x,y) = \frac{H_p(y)V_p(x)}{nm\bar{P}}, \tag{2}$$

Parameter α is fixed at 0,38 on the basis of tests carried out on 400 images relative to 40 people.

3.2   Discrete Cosine Trasform

One of the original characteristics of our system is the manner in which it combines the Discrete Cosine Transform (DCT) with Principal Component Analysis (PCA). Such a combination allows us, on the one hand, to pick up the salient features of a generic face, on the other, the discrepancies which emerge between faces that belong to different individuals.

Consider the image of a face $f(x, y)$ as a bi-dimensional array $m \times n$ of grey values. As such, the image may be represented by a point in a N-dimensional space, with $N = m \cdot n$. It follows, then, that a group of images can be represented by a set of points in the same space. These, as one may well expect, need not be distributed throughout the entire space of images. They may be concentrated, rather, in a limited zone that we shall call the *face space*. The points which fall within this space may represent images relative to different individuals, as well as images relative to just one individual taken under varying conditions and poses.

3.3   Principal Component Analysis

As one can clearly see from figure 2, the images elaborated either by means of Combined Projection (indicated P_α) or by DCT (indicated F) are subject to PCA. PCA models represent the faces to be recognised as the weighted sum of eigenvalues as extracted from the vectorial matrix of a set of facial images. In the case of Combined Projection, the covariance matrix turns out to be

$$C = \frac{1}{M} \sum_{i=1}^{M} \iota_i \iota_i^{\mathbf{T}} = P_\alpha c P_\alpha c^{\mathbf{T}},$$

where ι_i is the vector constituted by the columns of the P_α matrix, whilst $P_\alpha c$ is the matrix of diverse samples of a single subject. The covariance matrix, therefore, represents the range of variability of a single class, i.e., of a single person.

Initially, when only trial photos are contained in the database, the covariance matrix is constituted by a single-vector column. A single image, in short, is all the system described here requires in order to recognise an individual.

If, during the recognition phase, the system identifies the input-subject as one already present in the database, the covariance matrix is enriched straightaway since the representation of the photo in question is added to it and becomes another sample of the same type. In other words, the variability of one class is increased. The first d eigenvectors of the covariance matrix constitute the basis of the "face space" (indicated by $\sum_{i=1}^{d} \mathbf{u}_i = U$). The projection of a sample into this space is, therefore, indicated by the following expression

$$\hat{\iota} = U_d^{\mathbf{T}} \iota = \sum_{i=1}^{d} \mathbf{u}_i^{\mathbf{T}} \mathbf{x} \tag{3}$$

The system as described here does not devise a single "face space" for all the subject present in the gallery; rather, each and every subject-class is assigned its own space. The basis of every space, therefore, is given by the eigenvectors of the covariance matrix of the subject-class itself. This permits us to highlight the peculiar features of the single individual as opposed to those it shares with all the others. This facilitates its own identification within the database.

The same treatment may be applied to those images created by DCT: in this case, F (the DCT of the input image) replace P_α.

3.4   Classification and Decisional Strategy

Classification is a parallel action on two different domains: on the one combined projection plus PCA, on the other, DCT plus PCA. Both of these domains, in turn, must be previously subdivided into clusters such that each image can be assigned its proper classification upon input. The clusters, once determined, are represented by the gaussian functions of the RBF neural networks which actually

do the classification [8]. K-means [9] is the algorithm used for clustering, where k depends on the number of the eigenfaces used.

The gaussian outputs of every classifier constitute the two partial scales. Each scale provides the similarity value between the probe and every sample of the gallery. The similarity value is just the gaussian output. The two scales are then combined on the basis of the weights (similarity values). The total similarity value is the mean of the weights given by the two scales. The probe is considered to belong to the class with highest total similarity value.

4   Experimental Results

In this section the reader will find the results obtained by means of the experimental tests devised by FERET evaluation methodology [6]. These are then further compared with the results obtained by other systems that were subject to the same type of testing. The other systems in question were developed at MIT Media Laboratory, Rutgers University, and the University of Southern California (USC).

In considering the data we wish once again to draw the reader's attention to one of the peculiarities of our system, i.e., that its training is carried out by means of a single image per person. For a more exacting comparison, therefore, our system ought to have been pitted against others with analogous characteristics.

Two different models were used to perform the aforementioned experiments. The first, the so-called *closed universe*, ensures that every facial input is contained within a gallery. The second, the so-called *open universe*, allows for the usage of candidates that are not present in the training set.

For the first test, the so-called *large gallery test*, a closed universe model was used. More precisely, we used a gallery of 317 frontal images pertaining to as many individuals and a test set comprising 316 faces. Two different images for each individual were taken into consideration, one for the training stage, the other for the test itself. Now, to evaluate the capacity of any generic algorithm to identify an image in input, the FERET group considers it opportune to evaluate the percentage by which the correct answer appears among the first k provided by the systems's output. Thus the diagram in Figure 3 puts the generic k position along the x-axis relative to the *Cumulative Match Score* (*CMS*). The latter is obtained from the relation between the R_k number of tests in which the correct answer appears among the first k provided by the system, and the total number P of probes. It may thus be observed, for example, how our system provided the correct answer relative to the first five positions in 95% of the attempts undertaken.

In the next test, the *rotation test*, the closed universe was used yet again. More precisely, a gallery was comprised of 40 images with a total of 225 tests whereby the individual could appear either in a frontal or in a rotated 90 degree position. In Figure 4 appears the pertinent diagram on whose axes one can find the same quantities of the large gallery test. It is noteworthy how our system proved most robust with respect to deep rotations, and that only the system originating from

USC obtained better results. The positive outcome is essentially attributable to Combined Projection, which allows the system to recognise the individual on the basis of a single and, what is more, frontal image.

The final test required by FERET methodology is the *false alarm test*. In this test an open universe was used, which allows for the possibility that the individual need not belong to the gallery. A false alarm occurs every time the algorithm recognises a "member" of the gallery who, quite the contrary, does not belong. The execution of such a test requires a gallery of 25 faces and a test set comprising 305 individuals. In the related diagram (see fig. 5) the values assumed by a quantity known as *Receiver Operating Characteristic* (ROC) are provided. ROC expresses the relation that exists between the probability of a false alarm P_f, and the probability of a correct identification P_i. Figure 5 provides us with the ROC curve pegged to a variable threshold. The latter simply expresses the value of rate of recognition below which the individual is deemed to be extraneous to the gallery. The curve is obtained by simply calculating P_f and P_i for every variable value. It may be observed that our system works better in these cases than that developed at Rutgers University, but less well in comparison with the remaining two (especially for low threshold values). This is clearly one of the aspects that must be improved upon in the new year.

5   Conclusions

In this paper, we have proposed a framework capable of integrating innovative solutions with well established methods the likes of which includes DCT, PCA and the RBF networks. It is a system that is quite capable of guaranteeing satisfactory results even with a training of one image per individual. It has, moreover, proved itself to be rather robust when confronted with variations in scale, expression and lighting. As far as future developments are concerned, the false alarm test has made it most evident to us that our system clearly needs refining relative to the capacity of recognizing de facto extraneous individuals. Other areas that could use improving, notwithstanding the encouraging experimental results, is the management of ample depth rotations and consistent oscillations in lighting. The experimental results obtained from the subjection of our system to the most taxing of tests (those devised by FERET) are evidence enough of the validity of our approach.

6   Acknowledgments

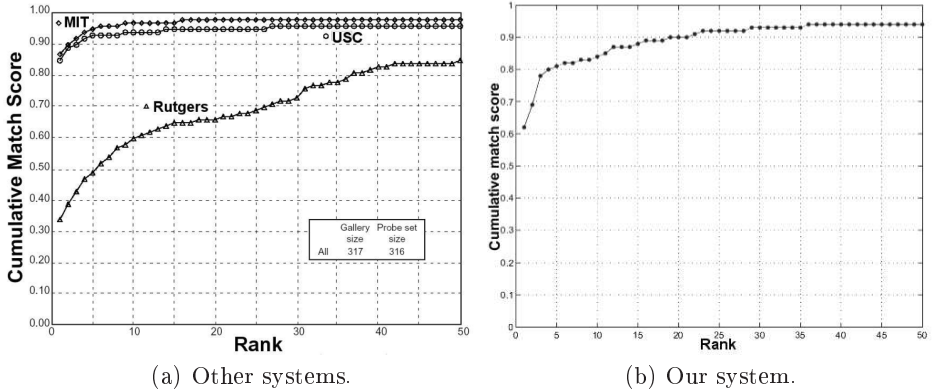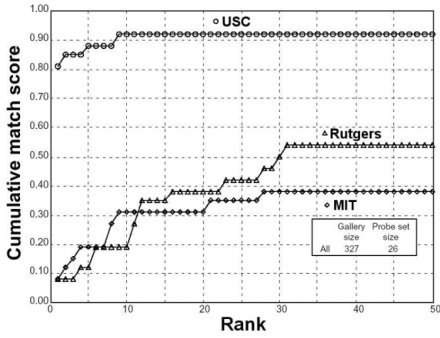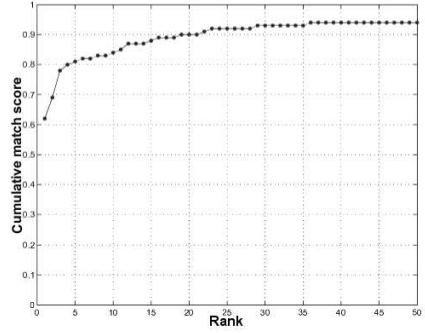(a) Other systems.   (b) Our system.

Fig. 3. Large gallery test.

References

1. Barrett,W.A.: A survey of Face Recognition Algorithms and Testing Results. Conference Record of the Thirty-First Asilomar **1** (1998) 301–305
2. Brunelli, R. and Poggio,T.: Face recognition: Features versus Template. IEEE, Transaction on Pattern Analysis and Machine intelligence **15** (1993) 1042–1052
3. Sirovich, L. and Kirby, M.: Low dimensional procedure for the characterization of human face. J.Opt.Soc.Am.A. **4** (1987) 519–524
4. Turk, M.A. and Pentland, A.P.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86
5. Wu, J. and Zhou,Z.H.: Face recognition with one training image per person. Pattern Recognition Letters (2002)
6. Rizvi, S.; Phillips, P. and Moon, H.: The feret verification testing protocol for face recognition algorithms. Journal of Cognitive Neuroscience (1999)
7. Martinez, A.M. and Kak, A.C.: PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence. **23** (2001) 228–233
8. Looney, C.G.: Radial Basis Functional Link Nets as Learning Fuzzy Systems. url = "citeseer.nj.nec.com/322716.html"
9. Fraley,C. and Raftery,A.: Model-Based Clustering, Discriminant Analysis, and Density Estimation. C. Fraley and A. Raftery, Model-Based Clustering, Discriminant Analysis, and Density Estimation, University of Washington, Dept. of Stats., Tech. Rep. No. 380, Oct. 2000.
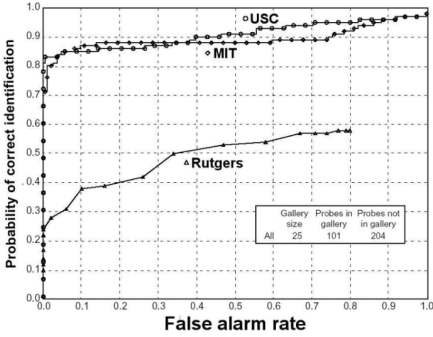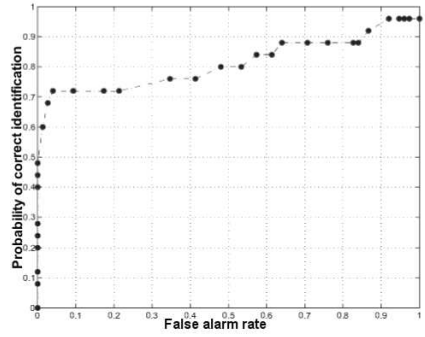
(a) Other systems.   (b) Our system.

Fig. 4. Rotation test.

(a) Other systems.   (b) Our system.

Fig. 5. False alarm test.