Similarity Searches in Face Databases

Annalisa Franco and Dario Maio

DEIS – Università di Bologna, Viale Risorgimento, 2 – 40136 Bologna - Italy
{annalisa.franco,dario.maio}@unibo.it

Abstract. In this paper the problem of similarity searches in face databases is addressed. An approach based on relevance feedback is proposed to iteratively improve the query result. The approach is suitable both to supervised and unsupervised contexts. The efficacy of the learning procedures are confirmed by the results obtained on publicly available databases of faces.

1 Introduction

Face is one of the most studied biometric characteristics and a huge literature exists about face recognition approaches [12]. Face recognition relies on the availability of a labeled database of faces, i.e. the identity of each face image is known, thus allowing the creation of a template representative of the user; recognition is performed by comparing the unknown input face image with the stored templates. In this work a different scenario is addressed where a similarity search is more appropriate than a direct identification. Several real applications fall in this category. For example in video-surveillance applications usually large quantities of images are gathered and stored to be examined in case of need (e.g. image sequences acquired at a bank entrance could be analyzed after a robbery: an image of the robber could be compared with other images in the attempt of finding frames where the face is more clearly visible). Another important application in the law enforcement field is the typical mug-shot search problem where a witness has in mind the image of a subject and a support tool could help to leaf through a database of suspects. In this case a query image could not be available, so that typical face recognition approaches cannot be applied.

Our framework was designed with the objective of providing a valid and flexible tool to perform similarity searches in face databases. The proposed system exploits relevance feedback techniques to gradually improve the search result. The search process is iterative and the information obtained at each iteration is used to improve the result in the subsequent search steps. The proposed system is very flexible being able to work in different modalities: supervised by the user (with or without a query image) and unsupervised. In the former, like in traditional feedback techniques [13], the system gathers information from the user judgment to improve the results, in the latter, which represents the main contribution of this work, an unsupervised iterative mechanism is exploited to improve the initial result without requiring the user intervention. The two modalities can either work independently or be combined to reduce the user effort needed to obtain a satisfactory result.

P. Foggia, C. Sansone, and M. Vento (Eds.): ICIAP 2009, LNCS 5716, pp. 443–450, 2009.

Many relevance feedback techniques have been proposed in the literature for similarity searches in generic image databases [13] where the user supervision is exploited to gradually improve the performance. Quite a few works have been proposed to address the specific problem of similarity searches in face databases. Supervised feedback techniques for mental face retrieval have been proposed in [4] [2]. We believe that in this application the user supervision cannot be completely avoided, but on the other hand some automatic mechanisms are needed to speed up the search process in very large databases.

We previously proposed a supervised relevance feedback approach in [5] [6]. This work starts from the same basic representation, but the method is here significantly extended with the possibility of performing unsupervised feedback iterations that does not require the user intervention. This paper is organized as follows: in section 2 the proposed framework is described, in section 3 the experiments carried out and the results obtained are presented, and finally in section 4 some concluding remarks are given.

2 System Overview

The processing flow is analogous to that of a traditional system for similarity searches based on relevance feedback. A query image is presented to the system that returns a first set of results, based usually on a "nearest neighbor" criterion. The result is then analyzed (either by the user in the *supervised* modality or automatically in the *unsupervised* one), thus determining an updating of the search parameters, subsequently used to produce a new set of results.

In the supervised search the query image is optional: if it is not available the system randomly selects an initial set of images from the database and the user can start directly indicating the most similar to the results he has in mind; this kind of search allows to perform the so called *mental face retrieval.*

2.1 Query Representation

Similarly to [5], a vector model is adopted here, that is each image is represented as a point in a multidimensional space, and a subspace-based representation is used for the query. The information about the relevant images obtained at each iteration is organized in clusters and represented by a mixture of linear subspaces [7]. The use of a mixture of linear subspaces allows to obtain a more robust representation and to deal with the non-linearity that characterizes this problem [5]. A single KL subspace [7] can be defined as follows. Let $P = \{\mathbf{x}_i \in \Re^n | i = 1,..,m\}$ be a set of images represented by n-dimensional feature vectors. The k-dimensional eigenspace S_P related to P is obtained by selecting the first k eigenvectors from the KL transformed space of P, $S_P = [\bar{\mathbf{x}}, \Phi, \Lambda, m]$, where $\bar{\mathbf{x}}$ is the mean vector, Φ is the matrix of the first k eigenvectors of the data covariance matrix and Λ the matrix of related eigenvalues. The value of k is a parameter bounded by the dimension n of the feature space and the number m of available samples. In our experiments no optimizations of the parameter k have been conducted and it has been fixed to $min(3,m\text{-}1)$ for all the subspaces, which is a reasonable value in consideration of the limited number of samples available. Given a subspace S_P, the distance of a pattern from the subspace, called *distance from space*, is defined as:

$$d_{FS}(\mathbf{x}, S_P) = \sqrt{\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 - \|\Phi^{\mathrm{T}}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2}.$$

2.2 Feature Extraction

To represent the face images, texture features are used. No color information is exploited since the role of color in the analysis of faces is not as important as in typical image retrieval problems; indeed most of the face recognition approaches known in the literature use features derived from gray level images. The Local Binary Pattern operator (LBP) originally introduced in [9] is used in this work. The LBP operator was designed to operate on gray level images, so that if the database contains color images they have first to be converted to gray scale. The LBP operator analyzes the eight neighbors of a pixel, using the value of the central pixel as a threshold. If a neighbor pixel has a higher or equal gray value than the center pixel, then a 1 is assigned to that pixel, otherwise it gets a 0. The LBP code for the central pixel is then obtained by concatenating the eight 1 or 0 to a binary code. To obtain the final feature vector the approach proposed in [1] is used. The face image is divided into regions, and for each region a histogram with all possible labels is constructed. This means that every bin in a histogram represents a pattern and contains its frequency in the related region. The feature vector is then constructed by concatenating the local histograms to one global histogram.

2.3 Supervised Feedback

In the supervised modality, if a query image is available, the initial result is obtained by a simple nearest neighbor search in the feature space: the Euclidean distance between the images in the database and the query is calculated and the *r* nearest images are selected. If the query image is not available, a set of *r* random images are extracted from the database. As a future work more effective techniques to extract the initial image set will be studied to select heterogeneous images representative of the whole database content (e.g. men and women, young and old persons, etc.).

At each iteration the user can express a judgment about the result obtained, selecting the images considered “relevant” to the query. The images selected are organized in disjoint subsets according to a K-means clustering algorithm [8]; starting from each subset a KL subspace is calculated thus obtaining the mixture of linear subspaces constituting the new query representation. To obtain a new result the images in the database are compared to the mixture of subspaces according to the *distance from set of spaces* metric, and the *r* nearest images are selected.

2.4 Unsupervised Feedback

The procedure used in the unsupervised modality is more complicated and represents the main novel contribution of this work. In this case no interaction with the user is required, and the system has to automatically evaluate the result obtained at each iteration. A scoring mechanism has been designed for this purpose. At the beginning of a query, the score of each image is initialized to 0 and is successively updated according to different factors that will be detailed later. At the end of each iteration the *r* images with the highest score are shown in the result.

The scoring process is based on the assumption that the *r* images retrieved are very similar to the query. The images are thus organized into disjoint clusters which are sorted according to their distance from the query image. The distance of each image is then calculated taking into account its distance from the query and the position of the related cluster, thus rewarding the images in the clusters nearest to the query. Finally the *r* nearest images to the query, according to this distance are executed as second level queries: the images that belong to the result of both the original query and the second level query receive an additional score. This last step is outlined in Fig. 1.

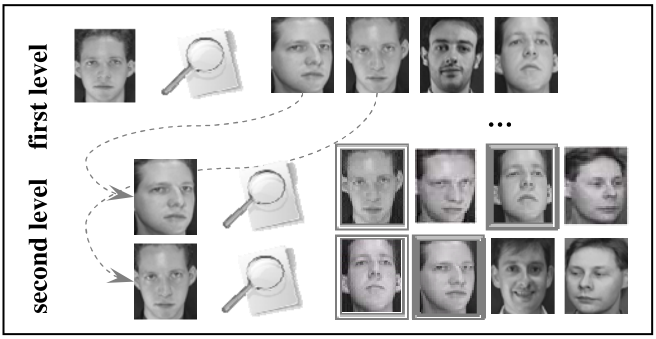

Fig. 1. Visual representation second level query step. The images with frame will receive an additional score

The detailed procedure is now described. Let's indicate with:

- $Y = \{\mathbf{y}_1, \mathbf{y}_2, .., \boldsymbol{y}_\mathrm{n}\}$ the set of images in the database, where each image is represented by the LBP features extracted (section 2.2);
- $X^i = \{\mathbf{x}_1^i, .., \mathbf{x}_r^i\}, X^i \subseteq Y$ the set of *r* images retrieved at iteration *i*;
- $S^i = \{s^i(\mathbf{y}_1), s^i(\mathbf{y}_2), \ldots, s^i(\mathbf{y}_n)\}$ the set of scores accumulated by each image, respectively, during the *i* iterations;
- $\mathbf{q}$ the feature vector representative of the query image.

The result X^{i+1} at iteration *i*+1 is calculated as follows:

- Partition the image set X^i into p disjoint subsets $P_1, P_2, \ldots, P_p$ using the K-means clustering algorithm [8];
- Calculate a KL subspace S_{P_i} for each subset $P_i, i = 1, .., p$;
- Sort the p subspaces according to their distance from the query $d_{FS}(\mathbf{q}, S_{P_i})$;
- For each image in the database $\mathbf{y}_i \in Y$ calculate the distance d_{ij} from each subspace S_{P_j} in the sorted list of subspaces ($d_{ij} = d_{FS}(\boldsymbol{y}_i, S_{P_j})$).

  Let $S_{P_{j^*}}$ be the nearest subspace to $\mathbf{y}_i$: $j^* = \underset{j}{argmin}\, d_{FS}\left(\mathbf{y}_i, S_{P_j}\right)$
- Calculate for each image $\mathbf{y}_i \in Y$ the composite weighted distance: $wd_i = \left(\frac{d_{ij^*} + \|\mathbf{q} - \mathbf{y}_i\|_2}{2}\right) \times \frac{j^*}{p}$

- Sort the images $\mathbf{y}_i \in Y$ in increasing order of distance wd_i, thus obtaining the sorted set $Y' = \{\mathbf{y}'_1, \mathbf{y}'_2, .., \mathbf{y}'_n\}$.
- Update $s^{i+1}(\mathbf{y}'_i) = s^i(\mathbf{y}'_i) + scoreFunc(i) \;\; i = 1, .., r$
- Take the first r images of $\mathbf{y}'_i \in Y', i = 1, .., r$ and find for each of them the *r1* nearest images $Y'' \subseteq Y$, where *r1* represents the number of images to score in the second level queries.
- Increase the score of each image $\mathbf{y}'_i \in (Y' \cap Y'')$ according to the following formula: $s^{i+1}(\mathbf{y}'_i) = s^i(\mathbf{y}'_i) + scoreFunc(i) \times \frac{1}{j^*+1}$

where j^* indicates the subspace $S_{P_{j^*}}$ nearest to $\mathbf{y}'_i$, and the subscript *i* refers to the position of the image $\mathbf{y}'_i$ in the sorted set Y'.

The new result X^{i+1} at iteration *i*+1will include the *r* images with the highest score. The score function assigns a score according to the position *i* of an image in the result set. In this work the following formula has been used: $scoreFunc(i) = \frac{1}{c^i}$ where *c* is a constant value (5 in our experiments). This score function rapidly decreases thus rewarding mainly the images in the first positions.

3 Experimental Results

Some experiments have been carried out on two face databases:

- the ORL database of faces [10] containing 400 different images related to 40 individuals (10 images for each subject);
- the Faces96 database from the Essex facial images database [3] containing 3040 images of 152 individuals (20 per subject); the images are characterized by a complex background and large variations of head scale and position, image lighting and expression. The face has been automatically detected using the Viola Jones face detector [11].

In order to perform the experiments, the images in each database are partitioned into disjoint subsets containing respectively the first half and the second half of the images of each individual: the first one is used as database to search on, the second one is used as query set. The performance is measured in terms of *precision* defined as the percentage of relevant images returned by the system after a fixed number of iterations. An image is considered relevant to a query if it belongs to the same class. It is worth noting that the information about the class of the images is not used during the search process; it is only exploited to finally calculate the precision. The results reported refer to the average precision obtained over the query set.

The unsupervised search procedure is first analyzed. In particular in Fig. 2 the average precision is reported for different parameters setup as a function of the number of iterations for the two databases.

The two most critical parameters are considered in this experiment: the number *r* of images to retrieve, and the number *r1* of images to score in the second level queries. Of course the precision is higher when a higher number of images is retrieved, but the results in the graph clearly show that, independently of the parameters setup, the unsupervised search procedure allows to gradually increase the precision. A term

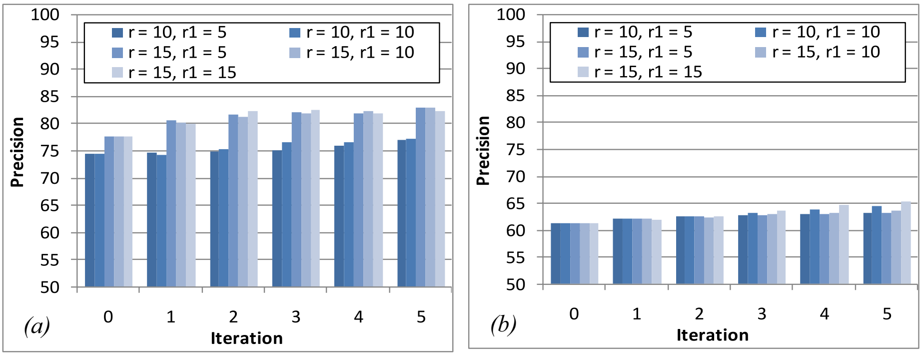

Fig. 2. Precision of the *unsupervised* search approach as a function of the number of iterations: *(a)* ORL database, *(b)* Faces96 database

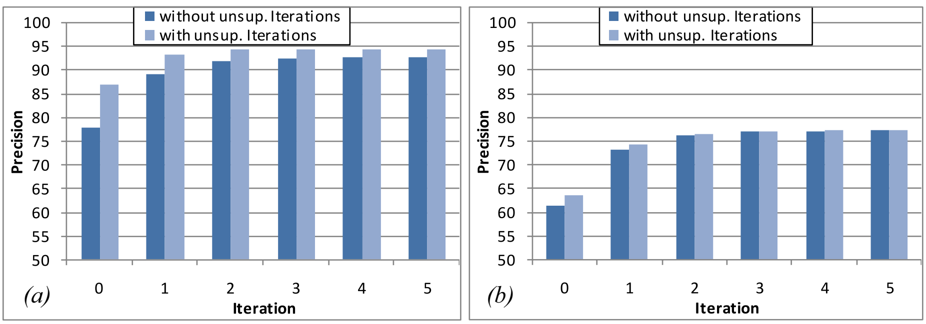

Fig. 3. Precision of the *supervised* search approach as a function of the number of iterations for the ORL (*a*) and Faces96 (*b*) databases. The result obtained with and without previous *unsupervised* iterations is given.

of comparison for the results of the *unsupervised* search is given by the performance measured for the *supervised* approach reported in Fig. 3.

In the second set of experiments the results obtained with the supervised feedback procedure are analyzed. To obtain objective and reproducible results, the supervised feedback has been simulated by automatically assigning a positive feedback to the images belonging to the same class of the query image. The final aim of our work is to reduce the effort required to the user to retrieve a satisfactory set of images; the advantage of the proposed approach is that a few unsupervised iterations can be performed before the user starts its interaction with the system. If the initial set of images is better, the number of supervised iterations needed to obtain a satisfactory result is lower. This evaluation is reported in Fig. 3. where the results obtained with a totally supervised search are compared to those obtained with a supervised search performed after 5 unsupervised iterations. The number of images to retrieve *r* was fixed to 15 in this test, and the number *r1* of images to score in the second level queries is 10. Analogous behaviors have been observed with other parameters setup.

The graph shows that the supervised search clearly benefits from some previous unsupervised iteration since the precision reached is higher, particularly in the first supervised iterations. This result is encouraging since it shows that the effort required to the user can be significantly reduced. The overall performance measured in the experiments are clearly higher for the ORL database where the images have been acquired under more controlled conditions; the Faces96 database contains very complex images and in some cases the automatic face detection provides inaccurate results. In both cases, however, the results confirm the efficacy of the proposed unsupervised feedback technique.

Experiments with the mental face retrieval (search without query image) require an extensive experimentation with several real users to provide reliable results, and will be thus performed in a future study.

4 Conclusions

In this paper a new approach for similarity searches in face database has been presented. The method exploits relevance feedback to gradually improve the quality of the result obtained. The feedback iterations can either be supervised by the user or be carried out automatically by the systems without requiring the interaction with the user. The two operative modalities can be combined to reduce the effort required to the user, obtaining at the same time satisfactory results.

This is only a preliminary work and the approach must be reinforced to better deal with typical changes in the face aspect that can make the search process more difficult. Moreover, the framework will be extended with the possibility of performing searches starting from sketches (as typically occurs when an identikit of a suspect is available).

References

[1] Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)

[2] Baker, E., Seltzer, M.: The Mug-Shot Search Problem. In: Proc. Vis. Int., pp. 421–430 (1998)

[3] Essex facial images database, `http://cswww.essex.ac.uk/mv/allfaces/index.html`

[4] Fang, Y., Geman, D., Boujemaa, N.: An Interactive System for Mental Face Retrieval. In: Proc. International Workshop on Multimedia information retrieval, pp. 193–200 (2005)

[5] Franco, A., Lumini, A.: Mixture of KL Subspaces for relevance feedback. Multimedia Tools and Applications 37(2), 189–209 (2008)

[6] Franco, A., Maio, D., Lumini, A.: A new approach for relevance feedback through positive and negative samples. In: Proc. ICPR, vol. 4, pp. 905–908 (2004)

[7] Fukunaga, K.: Statistical Pattern Recognition. Academic Press, San Diego (1990)

[8] MacQueen, J.B.: Some Methods for classification and Analysis of Multivariate Observations. In: Proc. Symp. on Math. Stat. and Probability, vol. 1, pp. 281–297 (1967)

[9] Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1), 51–59 (1996)
[10] ORL database of faces, `http://www.uk.research.att.com/facedatabase.html`
[11] Viola, P., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE ICCVPR, vol. 1, pp. 511–518 (2001)
[12] Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Computing Surveys 35(4), 399–458 (2003)
[13] Zhou, X.S., Huang, T.S.: Relevance feedback for image retrieval: a comprehensive review. Multimedia Systems 8(6), 536–544 (2003)