The Contribution of External Features to Face Recognition

Àgata Lapedriza, David Masip, and Jordi Vitrià

Computer Vision Center-Dept. Informàtica
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain
{agata,davidm,jordi}@cvc.uab.es

Abstract. In this paper we propose a face recognition algorithm that combines internal and external information of face images. Most of the previous works dealing with face recognition use only internal face features to classify, not considering the information located at head, chin and ears. Here we propose an adaptation of a top-down segmentation algorithm to extract external features from face images, and then we combine this information with internal features using a modification of the non parametric discriminant analysis technique. In the experimental results we show that the contribution of external features to face classification problems is clearly relevant, specially in presence of occlusions.

1 Introduction

During the past several years, face recognition has received significant attention as one of the most successful applications of image analysis and understanding. Several recognition systems have been recently developed due to their usefulness in a lot of real world applications in different areas such as security, entertainment or user friendly interfaces. However, their success is limited by the conditions imposed by many real applications. Recognition of face images acquired in an outdoor environment with changes in illumination, changes in pose or even occlusions is still an unsolved problem and these methods are still nowadays far away from the capability of the human perception system.

Most of the current face recognition algorithms can be categorized into two classes: geometry feature-based and image template based. The geometry feature-based methods analyze explicit local facial features, and their geometric relationships. The template based methods [1] compute the correlation between a face and one or more model templates to estimate the face identity. Statistical tools such as Support Vector Machines (SVM) [2], Linear Discriminant Analysis (LDA) [3], Principal Component Analysis (PCA) [4, 5], Kernel Methods [6] and Neural Networks [7] have been used to construct a suitable set of face templates that can be viewed as features. These kind of methods have proved to be effective in experiments with large databases.

Face recognition applications related to security need to focus on features difficult to imitate and for this reason the recognition systems tend to use only

J.S. Marques et al. (Eds.): IbPRIA 2005, LNCS 3523, pp. 537–544, 2005.

internal face features. Nevertheless, as technology evolves, it is easier to find electronic devices in our everyday life and in this context new applications dealing with face classification have appeared. In these cases the reasons to use only internal features are not longer valid, and given that the contribution of external features in face recognition is notable [8], their use in automaic systems should be revised. In this paper we propose a method to extract external information in face images and an algorithm to combine the information of external and internal features to solve the recognition problem.

The paper is organized as follows: in section 2 we explain the discriminant analysis algorithm to recognize subjects using only internal features. Section 3 shows how we build a model of known external features, how we use this model for the reconstruction of an unseen image and how we classify the unseen image from its reconstruction. Section 4 describes our experiments and section 5 conclude this work.

2 Extraction of Discriminant Internal Features

The extraction of the internal features in our comparative study has been performed using standard linear techniques found in the literature. One of the most used feature extraction algorithms is Principal Component Analysis (PCA) [4], where we obtain the orthogonal set of basis that preserve the maximum amount of data variance. The original $D-$dimensional data can be reconstructed using M coefficients ($M < D$) minimizing the reconstruction error. Nevertheless, sometimes the features that minimize the reconstruction error are not necessarily the features most suitable for classification [11]. If the class labels of the training sample are taken into account, other linear projections can yield a better classification accuracy even though the reconstruction error is not minimized. In this work we have used a modification of the nonparametric discriminant analysis (NDA) algorithm for this purpose. Below, we briefly describe the classic fisher linear discriminant analysis (FLD) [9] technique and the assumptions performed on the training data, to introduce later the NDA algorithm by Fukunaga et al [10], and the modified version used in this work.

We will assume a Nearest Neighbor classifier in this work, given that the feature extraction performed by the NDA algorithm is specially suitable for the NN rule using euclidean distance.

2.1 Discriminant Analysis

Here we look for a transformation matrix $\mathbf{W}$ which maximizes

$$\mathcal{J} = tr((\mathbf{W}\mathbf{S}_W^{-1}\mathbf{W}^T)(\mathbf{W}\mathbf{S}_B\mathbf{W}^T)) \tag{1}$$

Here $\mathbf{S}_B$ and $\mathbf{S}_W$ are the between-class and the within-class scatter matrix respectively. This problem has an analytical solution [11]. $\mathbf{W}$ is constructed using as its rows the M eigenvectors corresponding to the largest M eigenvalues of $\mathbf{S}_W^{-1}\mathbf{S}_B$.

This approach for calculating the within- and between-class scatter matrices makes use of only up to second order statistics of the data. This was proposed in the classic paper by Fisher [9] and the technique is referred to as Fisher Linear Discriminant Analysis (FLD). In FLD the within class scatter matrix is computed as the weighted sum of the class-conditional sample covariance matrices. If equal priors are assumed for the classes C_k, $k = 1, \ldots, K$, then

$$\mathbf{S}_W = \frac{1}{K}\sum_{k=1}^{K}\mathbf{S}_k \tag{2}$$

where $\mathbf{S}_k$ is the class-conditional covariance matrix for C_k, estimated from the data. The between class-scatter matrix is defined as,

$$\mathbf{S}_B = \frac{1}{K}\sum_{k=1}^{K}(\mathbf{m}_k - \mathbf{m}_0)(\mathbf{m}_k - \mathbf{m}_0)^T \tag{3}$$

where $\mathbf{m}_k$ is the class-conditional sample mean and $\mathbf{m}_0$ is the unconditional (global) sample mean.

The following two limitations of FLD have to be noted: the rank of $\mathbf{S}_B$ is $K-1$, so the number of extracted features can be, at most one in a gender recognition problem (with only two classes). Second, the scatter matrices are calculated assuming Gaussian classes. The solution provided by FLD is blind beyond second-order statistics, so this method may be inaccurate for complex classification structures.

2.2 Non-parametric Discriminant Analysis

Fukunaga and Mantock [10] propose a nonparametric discriminant analysis method as an attempt to overcome the two limitations of FLD noted above. In NDA the between-class scatter matrix $\mathbf{S}_B$ is calculated without the assumption of Gaussian classes. This scatter matrix is generally full rank, thus loosening the bound on the extracted feature dimensionality. Below we briefly expose this technique, extensively detailed in [11].

In NDA, the between-class scatter matrix is obtained as an average of N local covariance matrices, one for each point in the data set. This is done as follows. Let $\mathbf{x}$ be a data point in $\mathbf{X}$ with class label C_j. Denote by $x^{\text{different}}$ the subset of the k nearest neighbors of $\mathbf{x}$ among the data points in $\mathbf{X}$ with class labels different from C_j. We calculate the "local" between-class matrix for $\mathbf{x}$ as

$$\Delta_B^{\mathbf{x}} = \frac{1}{k-1}\sum_{\mathbf{z}\in x^{\text{different}}}(\mathbf{z}-\mathbf{x})(\mathbf{z}-\mathbf{x})^T \tag{4}$$

The estimate of the between-class scatter matrix $\mathbf{S}_B$ is found as the average of the local matrices

$$\mathbf{S}_B = \frac{1}{N}\sum_{\mathbf{z}\in X}\Delta_B^{\mathbf{z}} \tag{5}$$

We use $k = 1$ in this study, hence $x^{\text{different}}$ contains only one element, $\mathbf{z}_{\mathbf{x}}^{\text{different}}$, and

$$\mathbf{S}_B = \frac{1}{N} \sum_{\mathbf{x} \in X} (\mathbf{x} - \mathbf{z}_{\mathbf{x}}^{\text{different}})(\mathbf{x} - \mathbf{z}_{\mathbf{x}}^{\text{different}})^T. \tag{6}$$

The M eigenvectors corresponding to the largest M eigenvalues of $\mathbf{S}_W^{-1}\mathbf{S}_B$ define the projection matrix $\mathbf{W}$. In [12] they introduced also a non parametric form of the within-class scatter matrix $\mathbf{S}_W$, to extract features more suitable for the nearest neighbor classification. They propose to use

$$\mathbf{S}_W = \frac{1}{N} \sum_{\mathbf{z} \in X} \Delta_W^{\mathbf{z}} \tag{7}$$

where $\Delta_W^{\mathbf{x}}$ is calculated from the set of k nearest neighbors of $\mathbf{x}$ from the same class label, C_j, x^{same}

$$\Delta_W^{\mathbf{x}} = \frac{1}{k-1} \sum_{\mathbf{z} \in x^{\text{same}}} (\mathbf{z} - \mathbf{x})(\mathbf{z} - \mathbf{x})^T \tag{8}$$

For $k = 1$,

$$\mathbf{S}_W = \frac{1}{N} \sum_{\mathbf{x} \in X} (\mathbf{x} - \mathbf{z}_{\mathbf{x}}^{\text{same}})(\mathbf{x} - \mathbf{z}_{\mathbf{x}}^{\text{same}})^T. \tag{9}$$

In this paper we use the modified NDA algorithm as was proposed in [12] (using the local approximations of $\mathbf{S}_B$ and $\mathbf{S}_W$).

3 Extraction of the External Features

To extract the external features from the face images we have adapted a Top-Down Segmentation algorithm [13] to build a model based on a selection of parts from the object, faces in our case. In fact, we consider their segmentation as a reconstruction of the original image. Then, given a new unseen image, we find the subset of these parts that best reconstruct the image. The information of the matching between the parts-based model and the unseen image is used to classify the sample in classes. The set of pieces of the learned model are called in this paper *Building Blocks*.

The algorithm can be divided in two parts: learning the model from sample images, and the reconstruction of a new image using the set of *building blocks*. In the first step the optimal set of fragments from the object are learned, and the later step yields the features useful for classification of each new image.

3.1 Learning the Model

In this step the best an optimal set of fragments from the face images is computed. Usually this is the most time consuming part of the algorithm, although it is performed only once, off line, and using a generic face training set.

Given a training set consisting on face images with only the external characteristics to analyze (set C), and non face images acquired in natural environments (set $\overline{C}$), we generate subimages at sizes ranging from 12×12 to 24×24 from the training set. Each subimage will be a candidate fragment F_i for the final model. For each F_i the maximum values of the normalized correlation NC_i between F_i and each image from C and $\overline{C}$ are computed. The model is built storing the K fragments with best probability to describe the elements of the class C and not $\overline{C}$ $p(NC_i > \theta_i \mid C)$. The value of the threshold θ_i is computed taking into account a predefined number of false positive that can be tolerated $p(NC_i > \theta_i \mid \overline{C}) \leq \alpha$. For each fragment we also store a mask where only the pixels of the object are active.

For the construction of the model we guarantee that the set of fragments that we keep is able to reconstruct the external features of a generic face image. So additional restrictions to the relative position of the fragments are imposed. We discard similar pieces from the same relative position on a face, trying to achieve enough diversity in the fragments that compose the model for the external features. To perform it, we separately compute the model for different parts of the face images: forehead part, left side, right side, and chin part. Although there is some overlapping in these parts we obtain enough variety of pieces from each part in the final model.

3.2 Extraction of the External Features from Unseen Images

Suppose now that we have learned a model of external features. Given an unseen image our goal is to select the set of fragments of the building blocks that best reconstruct the image, and use this reconstruction to recognize or classify the subject. In figure 1 we show an example where it seems reasonable the use of the reconstruction for classification, given that the obtained reconstruction is not affected by image artifacts.

In this study we have proceeded as follows: we have computed the normalized correlation between the new image and each of the fragments of the building blocks and we have encoded the new image as a vector with the correlation values. We have considered this vector as a representation of the external features of the subject in the image, and we have used it to classify.

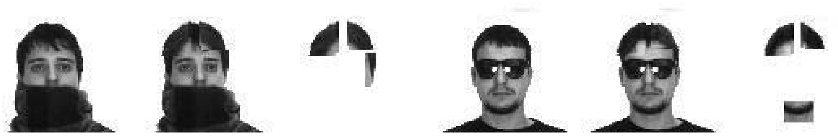

Fig. 1. Example of face images from the AR Face database [14] where different fragments are analyzed at each position. The first image shows the original face, in the second image we plot only the 5 most similar fragments according to the normalized correlation on the face. The third image shows the fragments alone.

3.3 Combination of the Internal and External Information

Once we have obtained the internal and external features for each new unseen image is needed to combine this information in a classification rule. Nevertheless it is not easy to understand the role that the different facial features play in a judgment of identity. In this work we have joined both sets of internal and external features in a single matrix and we have applied standard discriminant analysis techniques such as NDA to select a linear projection that performs the feature extraction from the whole set.

4 Experiments

To show the performance of our purpose in a face recognition problem, we have used the AR Face database [14], which is composed by 26 samples from 126 different subjects. Images were acquired in two different sessions, and for each session there is a frontal face, 3 samples with faces gesturing, 3 samples with illumination changes (frontal and lateral illumination), 3 samples with occlusions produced by the use of glasses (also with frontal and lateral illumination), and 3 samples with occlusions produced by the use of a scarf (with the 3 kinds of illumination). One sample from each type is plotted in table 1 above the results.

We have set the configuration parameters of the experiment as done in [15]. The image set has been split in non overlapping training and test sets, randomly selecting half of the subjects as a generic training data. The testing has been performed on the remaining subjects. A previous preprocess has been preformed on the face database, consisting on resizing each face according to the inter eye distance, and aligning the central position of both eyes. Also the mean has been subtracted from each image. The internal features have been extracted using different linear feature extraction algorithms: PCA and NDA. In all the cases we have selected the central part (33×33 pixels) of each image to be used as a input for each internal feature extractor. In the figure 2 we show some examples of the training set.

The external features have been extracted using the algorithm shown in section 3. We have randomly selected 40 generic faces and 40 images with no faces from natural environments. Using this set we have learned the fragments based model, setting the default threshold α to 0.01. Only the 400 fragments with maximum probability have been preserved. We also guarantee that there are fragments from each part of the face in the final set (frontal, chin zone, and both laterals). In the table 1 we compare the results using just internal features with PCA and NDA, with the use of our purpose with external information. We have used the nearest neighbor classifier on each case, with euclidean distance. The optimal dimensionality reduction on each case as been selected cross validating the training data. We show on table 1 the accuracies on each type of image for each case. As can be seen, the NDA algorithm outperforms the PCA in almost all the cases, given that NDA focuses the feature extraction on finding the most discriminative features while PCA just tries to preserve the maximum amount of data variance.

Fig. 2. Central part of some training images with only internal features.

Once we have fixed the NDA technique as the most suitable for the face recognition task, we have combined the external features and the internal ones using this algorithm, in order to obtain higher accuracies combining both approaches (COM in the table 1). The results obtained show that the external features help significantly the face recognition task. Actually, the technique is specially suitable when there are occlusions (sets A08 to A13), as can be seen, the contribution of the external features is notable in this cases, due to the large presence of fragments on the non occluded parts of the face. This allows new features for the NDA algorithm that have not been affected by the occlusions. In frontal and gesture images, the use of external features also improves the NDA using just internal features. In images with strong changes in the illumination, the external features also help when the light is focused on a lateral, giving more importance to the non illuminated side (more fragments). In Images with strong light changes on both sides the external features do not contribute to the global accuracy of the NDA given that the correlation values are mislead in all the cases due to the light.

Table 1. Results using Principal Component Analysis (PCA), Non Parametric Discriminant Analysis (NDA), and our purpose combining internal and external features (COM) for each AR Face image subset (percentage rates).

	AR01	AR02	AR03	AR04	AR05	AR06	AR07	AR08	AR09	AR10	AR11	AR12	AR13
PCA	76.7	69.7	74.4	51.1	67.4	60.4	60.4	34.9	39.5	34.9	44.2	37.2	23.2
NDA	74.4	79.1	76.7	51.2	76.7	69.8	62.8	34.9	46.5	37.2	62.8	44.2	46.5
COM	81.4	86.0	76.7	55.80	72.1	69.8	74.4	44.2	53.5	41.9	67.4	60.5	51.2

5 Conclusions

In this paper we have proposed a method for face recognition using internal and external features. We have adapted a Parts Based Segmentation algorithm to learn a model to extract the external features of new unseen images. These external features have been combined with the internal ones using classic discriminant analysis techniques. In this work the NDA algorithm has been used for this purpose. The experimental results show that external features improve the ones obtained using only internal information and this indicates that the contribution of external features is relevant for classification purposes, specially when occlusions are present.

As a future work we plan to get better reconstruction methods to represent the external features of each subject using the learned model. Alternative information to the normalized correlation could be used (derivatives, edge detectors). Also the reconstruction could be more reliable if the geometric relationships between the fragments are taken into account. Other algorithms could be used to combine external and internal information, weighting its relative significance.

Acknowledgments

This work has been partially supported by MCYT grant TIC2003-00654 and FP2000-4960 Ministerio de Ciencia y Tecnologia, Spain.

References

1. R. J. Baron. Mechanisms of human facial recognition. International Journal of Man-Machine Studies 15(2):137–178, 1981.
2. V. N. Vapnik. The nature of statistical learning theory. Springer Verlag, Heidelberg, DE, 1995.
3. R. J. Baron. Mechanisms of human facial recognition. International Journal of Man-Machine Studies, 15(2):137–178, 1981.
4. L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. Journal of Optical Society of America, 4(3):519–524, March 1987.
5. M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71–86, 1991.
6. Y. Li, S. Gong and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. In IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.
7. A. J. Howell, C. Lee Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural network approach. IEEE Transactions on Neural Networks, 8(1):98–113, 1998.
8. I. N. Jarudi and P. Sinha. Relative Contributions of Internal and External Features to Face Recognition. Massachusetts Institute of technology, 2003.
9. R. Fisher.The use of multiple measurements in taxonomic problems. Ann. Eugenics, vol. 7, pp. 179–188, 1936.
10. K. Fukunaga and J. Mantock. Nonparametric discriminant analysis. IEEE Trans. Pattern Anal. Machine Intell., vol. 5, no. 6, pp. 671–678, nov 1983.
11. K. Fukunaga. Introduction to Statistical Pattern Recognition, 2nd ed. Boston, MA: Academic Press, 1990.
12. M. Bressan and J. Vitria. Nonparametric discriminant analysis and nearest neighbor classification. Pattern Recognition Letters, vol. 24, no. 15, pp. 2743–2749, nov 2003.
13. E. Borenstein, S. Ullman. Class-Specific, Top-Down Segmentation. Proceedings of the 7th European Conference on Computer Vision-Part II, 2002, Springer-Verlag.
14. A. Martinez and R. Benavente. The AR face database. Computer Vision Center. Technical Report #24. June 1998.
15. R. Gross, J. Shi and J.Cohn. Quo vadis Face Recognition?, Carnegie Mellon University. Technical report CMU-RI-TR-01-17