Enhanced face recognition

Alexandre Trilla

Departament de Tecnologies Mèdia
Enginyeria i Arquitectura La Salle, Universitat Ramon Llull
atrilla@salle.url.edu

Abstract

This paper presents the task of face recognition/identification using the Principal Component Analysis tool for dimensionality reduction. Then it extends the functionality of the resulting algorithm through the use of the Singular Value Decomposition method and finally it enhances the identification task applying Bayesian classification.

Index Terms: face recognition, eigenface, principal component analysis, singular value decomposition, bayesian classification, Mahalanobis distance, euclidean distance

1. Introduction

Face recognition is the task associated to identifying an input image as a face, given that the system is trained with a set of face examples, and even spot this face among the training set of examples. In order to accomplish this goal this paper follows the approach taken in [1], in which the identification procedure is based on similarity. The problem is given by a two-dimensional recognition problem, where the images considered are determined by their intensity (luminance/brightness) and the faces that represent are always upright. The face images are projected onto a feature space ("face space") which encodes their most significant features. This space is given by the "eigenfaces", which are the eigenvectors of the set of faces.

It must be taken into consideration that these sort of applications require a lot of computer memory to deal with the resulting huge feature space. In order to surpass this problem, the Principal Component Analysis (PCA) tool for dimensionality reduction is of use.

The experiments have been carried out using the database of faces offered by the Speech, Vision and Robotics Group of the Cambridge University Engineering Department [1] which has also been used in the "example of known face" slide in class. Note that in order to stick to the mathematical development shown in the slides, the images of the database have been previously reduced to fit a 30x25 pixel canvas size so as to avoid a memory overflow problem.

In order to deal with bigger matrices more advanced techniques should be used such as the Singular Value Decomposition. This approach has been followed in order to extend the functionality of the application. But using either the covariance matrix of the training dataset or the matrix of eigenvectors given by the SVD the identification procedure has to be determined with a set of thresholds. This approach may yield some limitations since the values of these thresholds are heuristically determined.

Finally, in order to surpass this possible limitation, the classification process may be enhanced with Bayes' theory. For a sufficient amount of data of the same class, their representation could be given by a gaussian distribution. And then for an input unknown image, assuming a couple of sensible particularities (equiprobable classes and equal covariances), the euclidean distance or the Mahalanobis distance could enable the system determine the most probable class this image may belong to.

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

2. Principal Component Analysis

This dimensionality reduction process begins with the vectorization of the training images, which means that the matrices that represent the images (which in its turn show the intensity values of its pixels) are arranged as column vectors of size NP (N denotes the rows of the face images and P denotes their columns). From now on, each of these vectorized images will be represented by X_i, having a total of M images in the training dataset.

Then the training dataset ($\mathbf{X}$) is built grouping the vectorized images:

$$\mathbf{X} = \{\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_M}\}$$

Afterwards the mean vector is computed like:

$$m = \frac{1}{M}\sum_{i=1}^{M} X_i$$

And then the mean is subtracted from the the training dataset:

$$\mathbf{X'} = \{\mathbf{X_1} - \mathbf{m}, \mathbf{X_2} - \mathbf{m}, ..., \mathbf{X_M} - \mathbf{m}\}$$

Finally the covariance matrix is constructed with $\mathbf{X'}$:

$$\mathbf{C_x} = \mathbf{X'} * \mathbf{X'^T}$$

Note that $\mathbf{C_x}$ is of size $NP \times NP$, which can be extremely high with normal size images. Now the eigenvectors of this covariance matrix are extracted (directions with maximal variation) and they are set in decreasing eigenvalue magnitude (importance) as the rows of a new matrix $\mathbf{A}$ that represents the face space. Each of these eigenvetors are also called "eigenfaces". At this point the dimensionality reduction is applied: a new face space $\mathbf{B}$ is built with a heuristic number of most significant eigenfaces (rows of $\mathbf{C_x}$). This new face space is the one that is tractable to perform the face recognition application. It contains the most significant features used to represent the training corpus.

In order to perform the face identification procedure, the test images have to follow the same modifications as the training images: they have to be vectorized X_t, subtracted the mean X'_t (the mean vector of the training dataset) and then they can be projected into the reduced face space:

$$\hat{y} = \mathbf{B} * \mathbf{X'_t}$$

Then the resulting projection is converted backwards (reconstructed) into the vector image that would have resulted in the projection:

$$\hat{X}'_t = B^T * \hat{y}$$

And finally the distance between the two vector images is computed with their scalar product:

$$S = \hat{X}'^T_t \cdot X'_t$$

S yields the overall similarity between the original face image the the face image reconstructed with the features of the face space. Also, the scalar product between the projections of the test face image and the training face images should yield the identification of the test face as one of the training faces:

$$S = \hat{y}^T \cdot \hat{y_k} \quad k = 1, 2, ..., M$$

One must bear in mind that in this similarity comparison procedure the images being compared $\hat{X}'_t$ and X_t (multiplied with the dot product) must be normalized in energy:

$$newX_t = \frac{X_t}{||X_t||}$$

If the energy normalization is not performed the similarity results are utterly worthless since there is a direct relationship between the energy of the images and the magnitude of the operation (then a plain white image would be more likely to be similar to any image regardless of what it depicted).

Once the similarity score is obtained some heuristic thresholds can be introduced to enable the system to differentiate between a face and a no-face, and in the case of determining a face, be able to discern if it belongs to a face observed in the training dataset or not.

3. Singular Value Decomposition

The main problem that is faced when dealing with the PCA procedure is the production of the covariance matrix because this matrix ens up being computationally intractable even for small pictures. As seen in the previous section, the size of the covariance matrix is $NP \times NP$. Then the Singular Value Decomposition (SVD) is used in favor of the computer resources and capabilities.

The SVD is closely related to PCA and eigendecomposition. In the end it is based on finding the eigenvalues and eigenvectors of AA^T and $A^T A$. Moreover, the SVD is less restrictive in that it can be performed on any $m \times n$ matrix. The form is:

$$X = USV^T$$

where $X \in \Re^{m \times n}$ with $m \geq n$, $U \in \Re^{m \times n}$, $V \in \Re^{n \times n}$ and S being a diagonal matrix of size $\Re^{n \times n}$. Both U and V are orthogonal.

One the SVD is computed, matrix B is directly obtained:

$$B = U^T$$

Note that following this process the number of resulting eigenvectors to build the reduced face space is given by the number of examples in the training dataset. The evaluation of the similarity is performed in the same way as explained in the previous section.

Anyway, still some heuristics are of use to determine if the system is observing a face or a no-face, and whether if it is a familiar face or not (belonging to the training dataset).

4. Bayesian classification

None of the previous approaches based on similarity scores could do without the heuristic thresholds to determine if a given observation belongs to a face or not. This additional degree of knowledge must be introduced by hand in all situations.

In this section this statement is no longer valid since these decisions are taken based on conditional probabilities. Here the aim is to score for a determined image X the probability of belonging to a determined class α_i:

$$X \in \alpha_i \;\; iff \;\; p(\alpha_i|X) > p(\alpha_j|X) \;\; \forall i \neq j$$

Here the classes considered are, on one hand, the faces, and on the other hand, any other group of similar images, say beaches. Given an input image, the probability of belonging to one class should be greater than the probability of belonging to the other class. Applying Bayes' theorem:

$$p(\alpha_i|X) = \frac{p(X|\alpha_i)p(\alpha_i)}{p(X)}$$

In the equation above, $p(\alpha_i|X)$ is identified as the probability *a posteriori*, which is not sensible to compute due to the amount of different values X can take. $p(X|\alpha_i)$ is identified as the probability *a priori*, which represents the feature model given a class and $p(\alpha_i)$ is the probability of observing a class.

As it can be observed, $p(X)$ is irrelevant to the comparison, thus it can be omitted, resulting in:

$$X \in \alpha_i \;\; iff \;\; p(X|\alpha_i)p(\alpha_i) > p(X|\alpha_j)p(\alpha_j) \;\; \forall i \neq j$$

From now on, three assumptions will be considered in order to obtain a feasible procedure to achieve the classes identification objective. The first one assumes that the feature model probability has a gaussian distribution (this assumption is sensible for an abundant population). Then, for a n-dimensional space:

$$p(X|\alpha_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|C_i|^{\frac{n}{2}}} e^{-\frac{1}{2}(X-m_i)^T C_i^{-1}(X-m_i)}$$

The second assumption states that the probability of observing a particular class is equiprobable. This is sensible for balanced datasets. Then:

$$p(\alpha_i) = p(\alpha_j) \;\; \forall i, j$$

$$X \in \alpha_i \;\; iff \;\; p(X|\alpha_i) > p(X|\alpha_j) \;\; \forall i \neq j$$

And the third and last assumption states that the covariance matrices are equal for all classes. With this last assumption, the identification equation results:

$$X \in \alpha_i \;\; iff \;\; (X-m_i)C(X-m_i) < (X-m_j)C(X-m_j) \;\; \forall i \neq j$$

The last equation above arises some discussions according to the covariance matrix C of use. These finer details will be discussed in the following subsections.

4.1. Euclidean distance

In the case that the covariance matrix C is equal to the identity matrix, the identification equation computes the euclidean distance.

$$C = I$$

The euclidean distance is the geometric distance between two vectors given by the square root of the sum of the squared differences in each dimension (variable). This distance measure has a straightforward geometric interpretation, it is easy to code and it is fast to calculate, but it has two basic drawbacks:

1. It is extremely sensitive to the scales of the variables involved. The brightness/contrast of the images may bias the results towards false conclusions.
2. It is blind to correlated variables. In the case that these dimensions are related, the euclidean distance has no no means of taking into account that they don't bring new information, and it will just weigh this information more heavily.

4.2. Mahalanobis distance

In the case that the covariance matrix C is equal to the covariance matrix between the two mean vectors the identification equation computes the Mahalanobis distance.

$$M = [m_i \;\; m_j]$$
$$C = M * M^T$$

The Mahalanobis distance takes into account the covariance among the dimensions in calculating distances. With this measure, the problems of scale and correlation inherent in the euclidean distance are no longer an issue. Intuitively, when using the euclidean distance, the set of points equidistant from a given location is a sphere. The Mahalanobis distance stretches this sphere to correct for the respective scales of the different dimensions and to account for their correlation. The Mahalanobis distance can be applied directly to modeling problems as a replacement for the euclidean distance.

Another important use of the Mahalanobis distance is the detection of outliers. Therefore it is useful to increase the distance measure when different elements are to be compared.

5. Practical issues

The training dataset has been extracted from the database of faces of the Speech, Vision and Robotics Group of the Cambridge University Engineering Department. Five faces have been selected for training, two additional faces for testing and two extra images (a percentage symbol and a capital A letter) that in principle have nothing to do with faces. They are saved in the 'myfaces' folder.

This dataset of faces has been prepared in a small size (30x25) for the face recognition experiments dealing with the covariance matrix directly. The dataset has also been produced in a double size (61x50) in order to test the effectiveness of the face recognition experiments using the SVD. These bigger images are saved in the 'myfaces/big' folder.

Additionally a second dataset of beaches has been produced with the two size variants in order to build a new class of images to identify with the Bayesian classification procedure.

The code to recognize face images has been implemented in Scilab[2] with the Image Processing toolbox[3]. The code (scripts) and results of the programs are available in appendices being:

1. Appendix A: Face recognition using the direct covariance matrix
2. Appendix B: Face recognition using the SVD
3. Appendix C: Face/beach identification using Bayesian classification

The code includes several comments showing the different parts involved in the processing chain. Also, when the programs are executed, several information messages are issued on the console, as well as the obtained results.

[2] http://www.scilab.org/

[3] http://siptoolbox.sourceforge.net/

6. Results

As it has been mentioned above, the results of the experiments are included in the appendices of this work.

6.1. Face recognition using the direct covariance matrix

In this experiment, six images are tested: two pertaining to the faces training dataset, two faces not seen before and two additional nonsense images (a percentage and a capital 'A'). Note that in this situation the pictures used have been previously reduced in order to be able to compute and operate with the resulting covariance matrix.

The decision thresholds for this application have been heuristically set to 0.57 for face identification and 0.7 for face recognition among the faces observed in the training dataset.

The system can well identify the training images as faces and discern them among the rest of the training faces. The similarity between the images and their face space reconstruction are equal to the similarity between the projections of the test image and the training image.

Then, the first test image is identified as a face, and surprisingly has a high similarity score against the last of the training faces. The second test image has a low similarity with the face space and none of the training faces is similar to it. It must be considered that this second face image corresponds to a woman face while the whole training dataset is composed of male faces. Possibly features like long hair and cute face characteristics are hard to find between the male images of the training dataset.

Finally, in order to test the robustness of the system against images that have nothing to do with faces, a couple of pictures corresponding to a percentage symbol drawing and a capital 'A' letter are tested. None of these pictures is identified as a face, which empirically demonstrates that the overall system performs quite successfully with the dataset at hand.

6.2. Face recognition using the SVD

The main deficiency found in the direct usage of the covariance matrix is the preprocessing procedure applied to the images to be used in order to shrink them. Since the computation burden is the main restriction when using the covariance matrix directly, the SVD method is tested.

The decision thresholds for this application have been heuristically set to 0.50 for face identification and 0.7 for face recognition among the faces observed in the training dataset.

The SVD permits the application to use bigger images, now their size is reasonable (61x50) and they can be interpreted (or represented) when needed (in order to see, for example, what constitutes the face space). The obtained results are though very similar to the ones obtained with the covariance matrix directly. By adjusting the decision thresholds appropriately the resulting performance is the same.

6.3. Face/beach identification using Bayesian classification

In order to surpass the usage of heuristic thresholds, the Bayesian conditional probabilities determine whether an input image is more likely to pertain to the class of faces or to the class of beaches.

In this experiment, eight pictures are used: four faces and four beaches. And within each subgroup of four elements, two pictures from the training dataset and two new pictures to test.

As it has been described above, the identification equation may be implemented with the euclidean distance or the Mahalanobis distance according to the form of the covariance matrix

C. With the euclidean distance the performance of the system results in 75% of accuracy while with the Mahalanobis distance the accuracy gets 100%. This fact shows that the scaling and correlation robustness of the Mahalanobis distance is far better than simple projections with the euclidean distance.

7. Conclusions

The face recognition system identifies face images quite well, especially when the faces that compose the training dataset are presented to the system as test images. When this situation happens, the similarity score is equal to the similarity between the projection of the test image and the projection of the training face image for the algorithms that make use of some sort of eigendecomposition.

When using the covariance matrix directly the computational load is very high and the images to be used have to be previously reduced with an image processing software. In order to enhance the system with face discrimination capabilities, this knowledge is transferred with two thresholds set heuristically. Then the resulting application performs rather well with pictures reasonably similar or different from the facial space model.

The usage of more advanced decomposition techniques such as the SVD enhances the application enabling it to use bigger pictures. Since the SVD operates directly on the training dataset, without operating on the covariance matrix directly, the computational load is not so heavy and the processing of bigger images is feasible on a personal computer. The resulting accuracy is though not better than the previous approach, at least with the dataset at hand. And the heuristic tuning of the thresholds prevents it from inferring knowledge from the data.

Finally, the usage of a Bayesian classifier with a balanced dataset (assuming equal class observation probability), assuming that the distribution of the features for a given class is a gaussian and assuming equal covariance matrices among the classes permits the omission of the heuristic thresholds. With the Mahalanobis distance the obtained results have been the best of all: the resulting system has not failed in any case. With this configuration the system cannot though determine which face is presently recognizing. It would need a dataset of faces for each person, model each person's face with a gaussian, make the same assumptions as before and then follow the same procedure described above.

8. References

[1] M. Turk and A. Pentland, "Face recognition using eigenfaces," Jun 1991, pp. 586–591.