Dimensionality Reduction with Image Data

Mónica Benito and Daniel Peña

Universidad Carlos III de Madrid, Spain

Abstract. A common objective in image analysis is dimensionality reduction. The most often used data-exploratory technique with this objective is principal component analysis. We propose a new method based on the projection of the images as matrices after a Procrustes rotation and show that it leads to a better reconstruction of images.

Keywords: Eigenfaces; Multivariate linear regression; Singular value decomposition; Principal component analysis; Generalized proscrustes analysis.

1 Introduction

Exploratory image studies are generally aimed at data inspection and dimensionality reduction. One of the most popular approaches to reduce dimensionality and derive useful compact representations for image data is Principal Component Analysis (PCA). Kirby & Sirovich (1990) proposed using PCA to reduce the dimensionality when representing human faces. Alternative approaches using Independent Component Analysis (ICA) for face representation have been proposed by Barlett and Sejnowski (1997). In the last two decades, PCA has been especially popular in the object recognition community, where it has succesfully been employed by Turk & Pentland (1991)and Valentin et. al (1996). The problem we are interested in is as follows. We have a set of images which represent similar objects, for instance, human faces, temporal images of the same scene, objects in a process of quality control, and so on. Any particular image (say the $n-th$ image) is represented by a matrix X_n of I rows and J columns. We assume that the sample contains the set of N images, $X_1, X_2, ..., X_N$. Each matrix consists of elements x_{ij}, with $i = 1, ..., I$ and $j = 1, ..., J$, that represent the pixel intensities extracted from digitized images. All the elements x_{ij} are in the range between 0 and 255, where the value 0 represents black color, and the value 255 white. Suppose that each matrix is transformed into a vector $\mathbf{x}_n$ by row (or column) concatenation. Therefore, we have a set of N vectors in a high dimensional space, specifically, $\mathbf{x}_n \in \Re^d$ where $d = I \times J$, $n = 1, ..., N$. For convenience, the vectors are assumed to be normalized, so that $\mathbf{x}_n^T \mathbf{x}_n = 1$. Note that this set of vectors can be represented by an $N \times d$ matrix X in which the $n-th$ row is equal to $\mathbf{x}_n$. When dealing with high-dimensional observations, linear mappings are often used to reduce dimensionality of the data by extracting a small (compared to the original dimensionality of the data) number of linear features. Among all linear, orthonormal transformations, principal component

Z.R. Yang et al. (Eds.): IDEAL 2004, LNCS 3177, pp. 326–332, 2004.

analysis is optimal in the sense that it minimizes, in mean square sense, the errors in the reconstruction of the original signal $\mathbf{x}_n$ from its low-dimensional representation, $\widehat{\mathbf{x}}_n$. As is well known, PCA is based on finding directions of maximal variability. In this paper we propose an alternative way of projecting the original data on a subspace of lower dimension. Instead of concatenating rows or columns, we keep the structure of the matrix in the projection. The rest of the paper is organized as follows. In the next section, we propose a new approach which keeps the internal structure of the image and we show that this procedure has important advantages compared to classical PCA. In section 3 we discuss the problems of aligning and scaling images before the dimension reduction is carried out, and introduce a generalized proscrustes rotation to solve this problem. Finally, we present the experimental results of the procedure when applied to a human face data base.

2 An Alternative Approach Based on Matrix Projections

We are interested in a projection method which keep the matrix structure of the image. Yang & Yang (2002) proposed the projection of the rows of the matrix in the context of feature extraction. Here we follow a similar approach. Assume without loss of generality that $I > J$. Then, given $\mathbf{a}$ a unit norm $J \times 1$ vector, we can project the rows of X_n on the $\mathbf{a}$ direction by,

$$\mathbf{w}_n = X_n \mathbf{a} \tag{1}$$

We will call this $I-$dimensional projected vector $\mathbf{w}_n$ the projected feature vector of X_n. Suppose that we project all the images in this way and obtain a set of vectors, $\mathbf{w}_n$, $n = 1, ..., N$. In order to find a good projection direction, let us call S_r the $I \times I$ covariance matrix for these vectors representing the rows, (the subindex r is due to the projection of the rows). This matrix is given by

$$S_r = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{w}_n - \overline{\mathbf{w}}) (\mathbf{w}_n - \overline{\mathbf{w}})^T , \tag{2}$$

where $\overline{\mathbf{w}}$ is the mean of the projected vectors. The two most often used measures to describe scatter about the mean in multivariate data are the total variation, given by the trace of the covariance matrix, and the generalized variance, given by the determinant of this matrix. For simplicity let us find the direction $\mathbf{a}$ which maximizes the total variation given by the trace of S_r. It follows that vector $\mathbf{a}$ is the eigenvector linked to the largest eigenvalue of the matrix

$$\Sigma_c = \frac{1}{N} \sum_{n=1}^{N} \left(X_n - \overline{X}\right)^T \left(X_n - \overline{X}\right) \quad ; \quad \Sigma_c \in \Re^{J \times J} \tag{3}$$

As we need more than one direction of projection to characterize the sample, we compute the set of eigenvectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_p$, which constitute a basis for

$\Re^p$ from which the data can be estimated using a subspace of lower dimension, $p \leq \min\{I, J\}$. It is easy to prove that the same criterion is obtained if we start projecting the columns instead of the rows. Let $W_n = [X_n \mathbf{a}_1, ..., X_n \mathbf{a}_p] = X_n A_p$, be the feature vectors obtained. We can use these data to predict the matrix X_n by the multivariate regression model

$$X_n = W_n \beta_n + \varepsilon_n \tag{4}$$

where the matrix X_n is predicted from its feature vectors W_n using some parameters $\beta_n = \left[\beta_n^1, ..., \beta_n^J\right] \in \Re^{p \times J}$, which depend on the image. The least squares estimate is given by $\widehat{\beta}_n = \left(W_n^T W_n\right)^{-1} W_n^T X_n$ and the prediction of the matrix X_n with this model is

$$\widehat{X}_n = H_n X_n \tag{5}$$

where $H_n = W_n \left(W_n^T W_n\right)^{-1} W_n^T$ is the perpendicular projection operator onto the column space of W_n .

3 Image Registration

When dealing with a set of homogeneous objects, as in the case of the human face database, the different ilumination and facial expressions greatly increase the difficulty of the reconstruction task. The sample can be seen as a set of shapes with respect to a local 2D coordinates system. We can combine these different local coordinate systems into a common system in order to have a normalized sample of objects before they are analyzed by subspace techniques. This geometric transformation process is known as registration. Depending of the complexity of the object, it may require two or more viewpoints, also called landmarks, to register it appropiately. The most often used procedure in which the landmark points are selected so that these landmarks have the same coordinates in all the images is called Affine transformation. This can be solved easily by,

$$\mathbf{b}_i = D\mathbf{a}_i + \mathbf{s} \quad , \quad i = 1, ..., d \tag{6}$$

where d is the number of pixels, $d = I \times J$ and the vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ belong to $\Re^2$, due the pixel's spatial coordinates. Thus, for any pixel in the image, say the $i - th$, this transformation maps the vector $\mathbf{a}_i$ to $\mathbf{b}_i$. This approach has two main limitations. The first one is that we can select only three points to fix the object normalization. The second is that we are not keeping the relative distances among the landmarks in the transformation. As an alternative, we propose a new procedure to estimate the similarity transformation that avoids these two liminations.

3.1 Procrustes Analysis

Procrustes analysis theory is a set of mathematical tools to directly estimate and perform simulteneous similarity transformations among the objects landmarks up to their maximal agreement. Based on this idea, we can focus on a goodness of fit measure used to compare N configurations of points. The basic procedure is as follows. Let A_n be the $r \times 2$ matrix of coordinates of r landmarks in the $n-th$ image, $n = 1, ..., N$. We wish to find simultaneous translations, rotations, and scale factors of these N sets of points into positions of best fit with respect to each other. The functional model of the transformation is stated as follows,

$$\widehat{A}_n = c_n A_n T_n + \mathbf{1}\mathbf{t}_n^T \quad , \quad n = 1, ..., N \tag{7}$$

where c_n is the scale factor, T_n is 2×2 orthogonal rotation matrix, $\mathbf{t}_n$ is a 2×1 traslation vector, and $\mathbf{1}$ is a 2×1 unit vector. According to Goodall (1991), there is a matrix B, also called consensus matrix, which contains the true coordinates of the r points defined in a mean and common coordinate system. The N matched configurations are measured by means of the residual sum of squares between each point of each configuration and the corresponding point of the average configuration or common coordinate system. For this task, Generalized Orthogonal Procrustes Analysis (Gower, 1975) provides least-squares correspondence of more than two point matrices. To obtain the initial centroid C, we should define one of the coordinates matrices A_n as fixed, and sequently link the others by means of the Extended Orthogonal Procrustes (EOP) algorithm (Beinat and Crosilla, 2001). Defining $C = \frac{1}{N}\sum_{n=1}^{N} \widehat{A}_n$, as the geometrical centroid of the transformed matrices, the solution of the registration problem is achieved by using the following minimum condition

$$\sum_{n=1}^{N} tr\left\{\left[\widehat{A}_n - C\right]^T \left[\widehat{A}_n - C\right]\right\} \tag{8}$$

in an iterative computation scheme of centroid C until global convergence. Hence, the final solution of the centroid corresponds to the least squares estimation $\widehat{B}$ and shows the final coordinates of r points in the maximal agreement with respect to least squares objetive function. Finally, the unknown similarity transformation parameters $(T_n, \mathbf{t}_n, c_n)$, $n = 1, ..., N$, are then determined using the procrustes algorithm procedure for fitting two given sets of points, A_n and $\widehat{B}$ (Schoenemann and Carroll, 1970).

4 Experiments

In the first example the method proposed in (5) for dimension reduction is compared to the standard eigenface technique on a gray-level database. We compare the dimensionality reduction performance when a frontal view face database is used, showing that the new technique leads to a better result for the data analyzed. In the second example we show that the proposed Procrustes analysis works well for the image registration problem.

4.1 Example 1

We use a gray-level frontal view face database that comprises 114 full-face pictures, 56 males and 58 females ($N = 114$). Each image is digitized in a gray-scale, with a resolution of 248×186, i.e. 248 rows and 186 columns ($I = 248$, $J = 186$). We compare the reconstruction performance of the traditional method with the new one when the number of singular values used (i.e. dimension of the subspace) increase gradually. The quality of the reconstruction, as the efficiency of representing the data by the subspace, is measured by the mean square error (MSE). In Figure 1 we plot the average reconstruction error ($AMSE$) for the training sample when the number of estimated parameters k increase as a function of the number of singular values used, p, in the reconstruction by the standard method and the new one. For simplicity, we only consider $p = 1, ..., 40$. Figure 1 is a 3D graph, in which each point has three coordinates, $(x, y, z) = (k, AMSE, p)$. Thus, when the number of singular values are fixed, the x-axis represents the amount of parameters needed to reconstruct the image, and the average mean square error (AMSE) in the reconstruction is computed (y-axis). The upper plotted points correspond to the singular values used by the standard method, and the lower points are the ones used by the proposed method. This graph demostrates that the quality of the reconstruction by the new procedure is better than the traditional one. To visualize in more detail the performance of the reconstruction by both methods, Figure 2 gradually shows the reconstruction of one individual of the sample when the number of singular values is $p = 5, 10, 20$ and 50. Its reconstruction accuracy is measured by the MSE.

These figures clearly demostrate that when the dimensionality of the subspace is the same, the new method always perform better than the standard eigenface technique.

4.2 Example 2

In this example, we will show that the proposed image registration procedure is more effective than the affine transformation. For this purpose, we will register

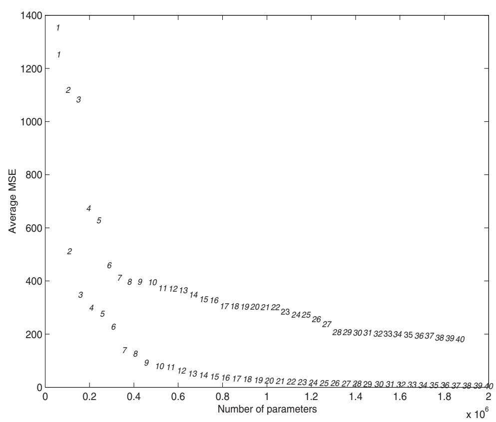

Fig. 1. Comparision of the average mean square error between eigenface method (upper points) and the proposed method (lower points) when the number of singular values used increases from 1 to 40

Fig. 2. Image Reconstruction by means of the standard method (left panels) and by the new method (right panels) using $p = 5, 10, 20$ and 50 singular values

Fig. 3. Image Registration of one individual in the sample

the face database used in example 1 in order to work with normalized objects. We choose as control points (landmarks) the coordinates associated to the left and right eyes and the end point of the chin. As an illustration, Figure 3 shows the solution of the registration problem for the $10-th$ image in the sample. The left panel in Figure 3 shows the original image. The middle panel shows the image registration by means of the affine transformation and the right panel by means of the procrustes analysis. Notice that while in the middle panel the classical affine transformation procedure deforms the original image, in the left image the procrustes algorithm perfectly reproduces the image.

References

[Barlett and Sejnowski (1997)] Barlett, M.S. and Sejnowski, T.J. (1997). Independent components of face images: a representation for face recognition. *Proceedings of the Fourth Annual Joint Symposium on Neural Computation, CA.*

[Beinat and Crosilla (2001)] Beinat, A. and Crosilla, F. (2001). Generalized Procrustes analysis for size and shape 3-D object reconstruction. *Optical 3-D Measurements TechniquesV, Viena*, pp. 345-353.

[Christensen (1991)] Christensen, R. (1991). *Linear Models for Multivariate, Time Series and Spatial Data.* New York: Springer-Verlag.

[Goodall (1991)] Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal Royal Statistical Society Series B*, **53**, pp.285-339

[Gower (1975)] Gower, J.C. (1975). Generalized Procrustes analysis. *Psychometrika*, **40**, pp.33-51

[Kirby and Sirovich (1990)] Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Machine Intell.*, **12**, pp.103-108

[Krzanowski and Marriot (1994).] Krzanowski, W.J. and Marriot, F.H.C. (1994). *Multivariate Analysis: Distributions, ordination and inference.* UK: Edward Arnold.

[Mardia, Kent and Bibby (1992).] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1992). *Multivariate Analysis.* CA: Academic Press.

[Schoenemann and Carroll (1970)] Schoenemann, P.H. and Carroll, R. (1970). Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika*, **35**, pp.245-255

[Turk and Pentland (1999)] Turk,M. and Pentland, A. (1999). Face recognition using Eigenfaces. *Proceedings of the IEEE Conference in Computer Vision and Pattern Recognition*, pp.586-591.

[Valentin, Abdi and O'Toole (1996)] Valentin, D., Abdi, H. and O'Toole, A. (1996). Principal Component and Neural Network Analysis of face images. *Progress in Mathematical Psycology.*

[Yang and Yang (2002)] Yang, J. and Yang, J. (2002). From Image vector to matrix: a straightforward image projection technique. *Pattern Recognition*, **35**, pp.1997-1999