Principal Component Analysis for Face Images

Daniel L. Swets

January 24, 1996

1 Principal Component Features

A two-dimensional image can be treated as a vector by concatenating the rows of the image together, using each pixel value as a single entry, as shown in figure 1. Thus each $p \times q$ image is considered as a point in a pq-dimensional feature space.

Image instances of a particular face can be represented by an $n = pq$-dimensional vector $\mathbf{X}$. $\mathbf{X}$ can be expanded exactly by

$$\mathbf{X} = V\mathbf{Y},$$

where the columns of the $n \times n$ square matrix V are orthonormal basis vectors. That is,

$$\begin{aligned} x_1 &= v_{11}y_1 + v_{12}y_2 + \cdots + v_{1n}y_n \\ x_2 &= v_{21}y_1 + v_{22}y_2 + \cdots + v_{2n}y_n \\ &\vdots \\ x_n &= v_{n1}y_1 + v_{n2}y_2 + \cdots + v_{nn}y_n \end{aligned}$$

This dimension n of $\mathbf{X}$ is usually very large, on the order of several thousand for even small image sizes. Since we expect that a relatively small number of features are sufficient to characterize a set of images, it is efficient and reasonable to approximate $\mathbf{X}$ using $m < n$ columns of V to give

$$\hat{\mathbf{X}}(m) = \sum_{i=1}^{m} y_i \mathbf{v}_i,$$

where the $\mathbf{v}_i$'s are the column vectors of V.

Let the effectiveness of the approximation be defined as the mean-square error $\|\mathbf{X} - \hat{\mathbf{X}}(m)\|^2$. Then we can use the proven result [2] [4] [6] that the best vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_m$ to use are the unit eigenvectors associated with the m largest eigenvalues of the covariance

$$I = \begin{array}{|c|c|c|c|}\hline f_{1,1} & f_{1,2} & \cdots & f_{1,q} \\ \hline f_{2,1} & f_{2,2} & \cdots & f_{2,q} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline f_{p,1} & f_{p,2} & \cdots & f_{p,q} \\ \hline \end{array}$$

(a) A $p \times q$ image, made up of rows of pixels

$$\mathbf{X} = \begin{bmatrix} f_{1,1} \\ f_{1,2} \\ \vdots \\ f_{1,q} \\ f_{2,1} \\ f_{2,2} \\ \vdots \\ f_{2,q} \\ \vdots \\ f_{p,1} \\ f_{p,2} \\ \vdots \\ f_{p,q} \end{bmatrix}$$

(b) The pq-dimensional vector is formed by concatenating the rows of the image together.

Figure 1. The vectorization of an image. We treat an $p \times q$ image as a point in a pq-dimensional space. The rows of pixels in an image are concatenated together to form a vector.

matrix of $\mathbf{X}$,

$$\Sigma_{\mathbf{X}} = \left[(\mathbf{X} - \mathbf{M}_{\mathbf{X}})(\mathbf{X} - \mathbf{M}_{\mathbf{X}})^t\right],$$

where $\mathbf{M}_{\mathbf{X}}$ is the mean (expected) vector of $\mathbf{X}$. Then the features $y_1, y_2, \cdots, y_m$ can be easily computed from

$$y_i = \mathbf{v}_i^t(\mathbf{X} - \mathbf{M}_{\mathbf{X}}) \ , \ i = 1, 2, \cdots, m.$$

This projection, also called the Karhunen-Loève projection and principal component analysis [3], has been used to represent [5] and recognize [9] [8] face images, for planning the illumination of objects for future recognition tasks [7], and as a component in a lip reading scheme [1], among others.

To determine m, the number of features to use, we first rank the eigenvalues of $\Sigma_{\mathbf{X}}$, $\lambda_1, \lambda_2, \cdots, \lambda_n$, in non-increasing order. The residual mean-square error in using $m < n$

features is simply the sum of the eigenvalues not used,

$$\sum_{i=m+1}^{n} \lambda_i,$$

and is a natural criterion to determine how many features are needed to sufficiently represent a face. We can choose m such that the sum of these unused eigenvalues is less than some fixed percentage P of the sum of the entire set. So we let m satisfy

$$\frac{\sum_{i=m+1}^{n} \lambda_i}{\sum_{i=1}^{n} \lambda_i} < P.$$

If $P = 5\%$, a good reduction in the number of features is obtained while retaining a large proportion of the variance present in the original feature vector [3] [9].

2 Computational Considerations

We can approximate the covariance matrix $\Sigma_{\mathbf{X}}$ with the sample scatter matrix $S = UU^t$, where $U = [\mathbf{U}_1\mathbf{U}_2 \cdots \mathbf{U}_k]$, and $\mathbf{U}_i = \mathbf{X}_i - \overline{\mathbf{X}}$, for k training images. Note that S is $n \times n$. If $k < n$, as is typically the case when dealing with a small number of training samples relative to the image dimension, S is degenerate. When this happens, however, we can find the eigensystem of the $k \times k$ matrix U^tU. This means that

$$U^tU\mathbf{w}_i = \lambda_i \mathbf{w}_i,$$

with eigenvalue λ_i and associated eigenvector $\mathbf{w}_i$. Pre-multiplying by U gives

$$UU^tU\mathbf{w}_i = \lambda_i U\mathbf{w}_i.$$

Then $\mathbf{v}_i = U\mathbf{w}_i$ is the eigenvector of $S = UU^t$ with eigenvalue λ_i. If the number of samples available is more than the image dimensions, then the eigensystem of UU^t can be computed directly.

3 Algorithm

1. Given k $p \times q$ input images, view each as an $n = pq$-dimensional vector, $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_k$.

2. Compute the mean of these training images, $\bar{\mathbf{X}}$. Note that

$$\bar{\mathbf{X}} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{X}_i.$$

3. Compute a deviation matrix $U = [\mathbf{U}_1 \mathbf{U}_2 \cdots \mathbf{U}_k]$ of size $n \times k$, where $\mathbf{U}_i = \mathbf{X}_i - \bar{\mathbf{X}}$.

4. If $k < n$ compute $M = U^t U$. In this case, M is $k \times k$; call the dimension $d = k$. Otherwise, compute $M = UU^t$. In this case M is $n \times n$; call the dimension $d = n$.

5. Perform the eigenvalue decomposition of M. This will produce d eigenvectors $\mathbf{w}_i$ and their associated eigenvalues λ_i, $i = 1, 2, \cdots d$.

6. Sort the eigenvalue/eigenvector pairs in non-increasing order.

7. Choose m such that

$$\frac{\sum_{i=m+1}^{d} \lambda_i}{\sum_{i=1}^{d} \lambda_i} < P.$$

for a given percentage of the total variance P.

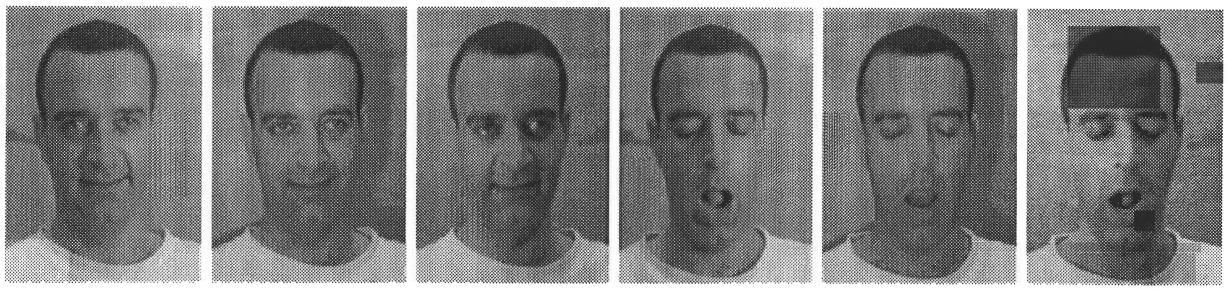
8. If $k < n$, compute $\mathbf{v}_i = U\mathbf{w}_i$, $i = 1, 2, \cdots m$.
Otherwise, $\mathbf{v}_i = \mathbf{w}_i$, $i = 1, 2, \cdots, m$.

9. These $\mathbf{v}_i$'s are the eigenvectors ($n \times 1$ dimensionality) associated with the largest eigenvalues for the k input images. These vectors are what Pentland calls "Eigenfaces" [9]. These vectors can now be "unstacked" and viewed as $p \times q$ images.

10. Each input image can be approximated by a weighted sum of these $\mathbf{v}_i$'s, as $\hat{\mathbf{X}}_j = \sum_{i=1}^{m} y_{j,i} \mathbf{v}_i$

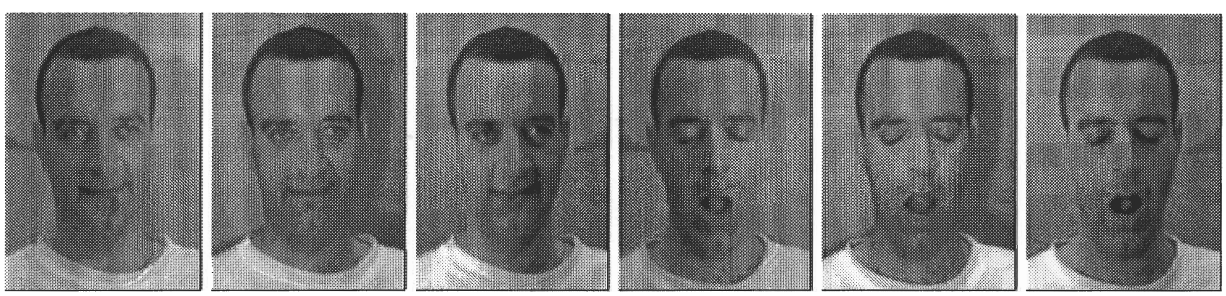
11. $y_{j,i} = \mathbf{v}_i^t \mathbf{U}_j, i = 1, 2, \cdots, m$. This vector $\mathbf{Y}_j = [y_{j,1}, y_{j,2}, \cdots y_{j,m}]$ is a feature vector for input image $\mathbf{X}_j$.

4 Example



Set of input images from the Weizmann Institute. Each image is 88×64 pixels.

Mean MEF$_1$ MEF$_2$ MEF$_3$



Approximations of input images using a weighted sum of the first three eigenvectors. Note that the three largest eigenvalues satisfied the 95% criterion.

References

[1] C. Bregler and S. M. Omohundro, "Nonlinear manifold learning for visual speech recognition," in *Proc. International Conference on Computer Vision*, pp. 494–499, 1995.

[2] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Academic Press, New York, 2nd ed., 1990.

[3] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data.* Prentice-Hall, New Jersey, 1988.

[4] I. T. Jolliffe, *Principal Component Analysis.* Springer-Verlag, New York, 1986.

[5] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, January 1990.

[6] M. M. Loève, *Probability Theory.* Princeton, NJ: Van Nostrand, 1955.

[7] H. Murase and S. K. Nayar, "Illumination planning for object recognition in structured environments," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 31–38, June 1994. Seattle, Washington.

[8] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 84–91, June 1994. Seattle, Washington.

[9] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.