PCA

- Principal Component Analysis(PCA)is a common effective method based on covariance matrix of the information processing, compression and extraction.

- Face recognition process with feature face consists of training and recognition.
- The detail steps as follows:

Face recognition based PCA

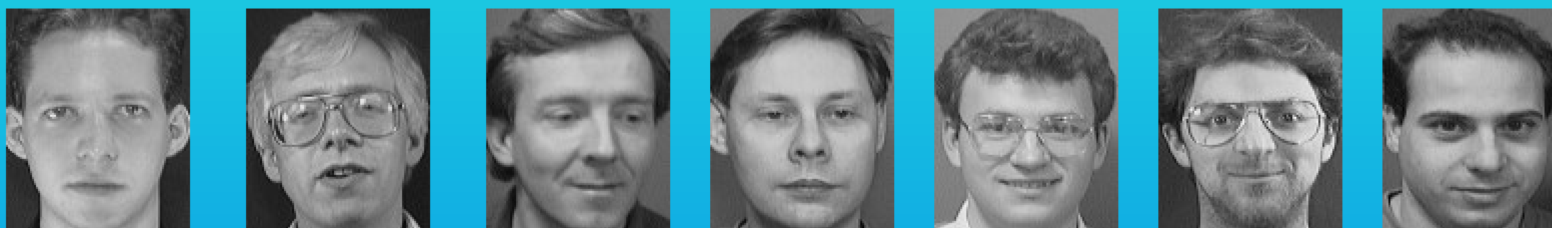
- PCA has been widely used in face recognition because of its effectiveness in the dimension reduction and feature extraction.

- The basic principle of PCA : using the K - L transform to extract the main components of the face, construct feature face, recognition will be tested when the image projected onto this space, a group of projection coefficients will be obtained.

- Recognition need compare with the each face image.

Training

- Step 1 Assume train-dataset has 200samples,gray image, size of sample is M*N



- Write the train-data in matrix：

$$x = \left(x_1, x_2, \ldots, x_{200}\right)^T$$

- Make matrix to vector, as the picture show:

Training

- Matrix of ith image

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

- Xi =

$$\begin{bmatrix} 1 \\ 4 \\ 7 \\ 2 \\ 5 \\ 8 \\ 3 \\ 6 \\ 9 \end{bmatrix}$$

Training

- Step 2 Compute average face values

Compute average face values of train-data :

$$\Psi = \frac{1}{200} \sum_{i=1}^{i=200} x_i$$

Training

- Step 3 Computing difference values of face

Calculate a face with an average face of each difference

$$d_i = x_i - \Psi, i = 1,2,...,200$$

Training

- Step 4 Construct covariance matrix

$$C = \frac{1}{200}\sum_{i=1}^{200} d_i d_i^{T} = \frac{1}{200} AA^{T}$$

$$A = \left(d_1, d_2, \ldots, d_{200}\right)$$

Training

- Step 5 Solve covariance matrix eigenvalue and eigenvector, then construct feature face space.

- The dimensions of the covariance matrix is MN * MN (bigger), using singular value Decomposition (SingularValue Decomposition, SVD) theorem, to get the AA^T eigenvalue and eigenvector by solving $A^T A$ eigenvalue and eigenvector.

v_i

Training

- Solve proper value λ_i of $A^T A$ and its orthogonal normalized vector v_i.
- According to the contribution of the feature value to select top p max eigenvector and their corresponding eigenvectors.
- Contribution rate refers to the selection of feature values with all the eigenvalues and ratio

$$\varphi = \frac{\sum_{i=1}^{i=p} \lambda_i}{\sum_{i=1}^{i=200} \lambda_i} \geq a$$

Training

- Set $a = 99\%$,make sure that train samples projection has 99% energy in top p feature vector on the set .

The eigenvectors of the covariance matrix

$$u_i = \frac{1}{\sqrt{\lambda_i}} A v_i (i = 1,2,\ldots, p)$$

Feature face space：

$$w = \left(u_1, u_2, \ldots, \quad u_p\right)$$

Training

- Step 6 Difference vector between face and average face projects to feature face space ,then

$$\Omega_i = w^T d_i \left(i = 1,2,\ldots,200 \right)$$

Recognition

- Step 1 Difference between face image Γ prepare for recognition and average face image project to feature space, and then get the feature vectors:

$$\Omega^{\Gamma} = w^{T}\left(\Gamma - \Psi\right)$$

Recognition

- Step 2 Define threshold

$$\theta = \frac{1}{2}\max_{i,j}\left\{\left\|\Omega_i - \Omega_j\right\|\right\}, i, j = 1,2,\ldots,200$$

Recognition

- Step 3 Compute distance between Ω^{Γ} and each face image ε_i with Euclidean distance

$$\varepsilon_i^{\,2} = \left\| \Omega_i - \Omega^{\Gamma} \right\|^2 \left(i = 1,2,\ldots,200 \right)$$

Recognition

- In order to distinguish face and non face, we need to calculate distance ε between the original image Γ and the reconstruction image Γ_f with feature face space

$$\varepsilon^2 = \left\| \Gamma - \Gamma_f \right\|^2$$

- And：

$$\Gamma_f = w\Omega^{\Gamma} + \Psi$$

Recognition

- According to the following rules to classify faces
- 1) if $\varepsilon \geq \theta$, input image is not face image;
- 2) if $\varepsilon < \theta$, && $\forall i$, $\varepsilon_i \geq \theta$,input image include unknown face images;
- 3) if $\varepsilon < \theta$, && $\forall i$,$\varepsilon_i < \theta$,input image is kth face image in face library.

2D-PCA

- 2 D - PCA is the improvement on the basic PCA algorithm.
- The main difference is the method to construct covariance matrix, the selection of top P maximum eigenvalue and eigenvector is also different.

Training

- 1 Assuming train samples set is:：

$$\left\{s_j^{\,i} \in R^{m\cdot n}, i = 1,2,\ldots, N, J = 1,2,\ldots, K\right\}$$

i means ith person, number of labels;

j means jth image of ith person;

N means numbers of recognition face;

K means each person has K images;

M means total samples and M=NK.

Training

- 2 Compute the average image of all train samples

$$S = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{K} S_j^i$$

Training

- 3 Compute covariance matrix of samples：

$$G = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{K}\left(S_j^i - S\right)^T\left(S_j^i - S\right)$$

Training

- 4 Solve eigenvalue of covariance matrix ,selecte the Orthogonal eigenvectors $X_1 \ldots X_p$ corresponding to max eigenvalue $u_1 \ldots u_p$ as projection space.

Let the projection matrix Y total discrete degree as the criterion function J (U) to measure the pros and cons of projection space U: $J(U) = tr(S_u)$

Training

- S_u is covariance matrix of projection matrixY，
  $tr(S_u)$ is trace of matrix S_u

$$S_U = U^T E\{[x - E(x)]^T [x - E(x)]\}U$$

- Feature vector that be selected :

$$U = (X_1, X_2, ..., X_p) = \arg\max[J(U)],$$

$$X_i^T X_j = 0; i \neq j; i, j = 1,2,..., p$$

Training

- 5 train-samples $\left\{s_j^{\,i}, i = 1,2,\ldots, N, j = 1,2,\ldots, K\right\}$ project to $X_1 \ldots X_p$:

$$Y_j^{\,i} = [S_j^{\,i} X_i, \ldots, S_j^{\,i} X_p] = [Y_j^{\,i}(1), \ldots, Y_j^{\,i}(p)] \in R^{m*p}$$

Recognition

- Step 1 Test sample $W \in R^{m*n}$ to $X_1 \ldots X_p$ projection and obtain feature vector Y_t and principle component $Y_j^i(1),\ldots,Y_j^i(p)$ of sample W

$$Y_t = [Y_j^i(1),\ldots,Y_j^i\left(p\right)] = \left[WX_1,\ldots,WX_p\right]$$

Recognition

- Step 2 According to the minimum distance between test sample projection feature matrix and all the training sample projection feature matrix to determine test sample category.
- Define the distance measurement rule as follows:

$$p\left(Y_j^{\,i}, Y_t\right) = \sum_{n=1}^{p} \left\| Y_j^{\,i}(n) - Y_t(n) \right\|^2$$

- $\left\| Y_j^{\,i}(n) - Y_t(n) \right\|^2$ means Euclidean distance between two feature vector.

Recognition

- Step 3 if

$$p\left(Y_d{}^q, Y_t\right) = \min_i \min_j p\left(Y_j{}^i, Y_t\right)$$

then Y_t belongs to qth sample