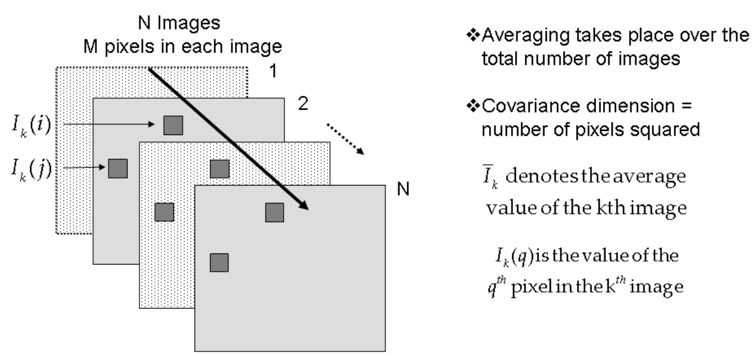
Principal component analysis (PCA) on ensembles of digital images

We can generally carry out PCA in **two senses** –

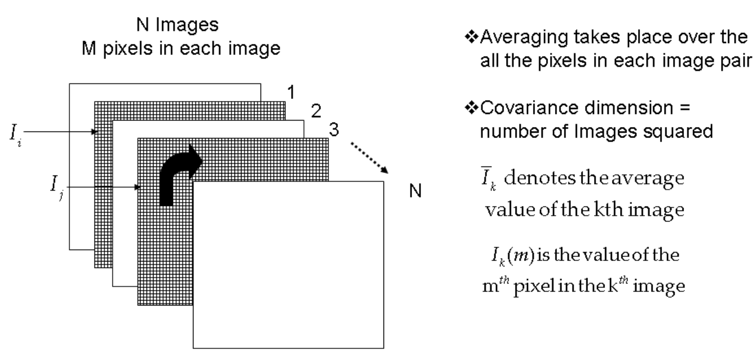
i) By taking our statistical average over the ensemble of vectors.
ii) By taking our statistical average over the elements of the vectors themselves.

This point has rarely been made explicit in existing image processing textbooks. The one which we choose in a given instance will naturally depend on what we are trying to do but often **only one approach will allow a computationally viable solution to the problem**. Such is the case with an ensemble of digital images.

Consider figure 1 in which we depict a stack of M images, each of which contains N pixels. In the first approach to PCA the covariance is calculated on a *pixel-to-pixel* basis. That is to say that the ijth element of the covariance matrix is calculated by examining the values of the ith and jth pixels in each image and averaging over the whole sample of M images. Unfortunately, such an approach leads to the calculation of a covariance matrix of quite unmanageable size (consider that even a modest image size of 256^2 would then require calculation of a covariance matrix of 256^2 x 256^2 ~ 4295 million elements).



$$\text{Covariance} \quad C_{ij} = \frac{1}{N-1}\sum_{n=1}^{N}\left(I_i(n)-\bar{I}_i\right)\left(I_j(n)-\bar{I}_j\right)$$



$$\text{Covariance} \quad C_{ij} = \frac{1}{M-1}\sum_{m=1}^{M}\left(I_i(m)-\bar{I}_i\right)\left(I_j(m)-\bar{I}_j\right)$$

Figure 1: Two ways of carrying out PCA on image ensembles. Only the second approach in which the statistical averaging takes place over the image pixels is computationally viable.

The second approach in which the ensemble averaging takes place over the pixels of the images generally results in a much more tractable problem – the resulting covariance matrix has a square dimension equal to the number of images in the ensemble and can be diagonalised in a straightforward fashion.

Consider then a sample of N images ($I_1, I_{2,\ldots} I_N$) each containing M pixels so that the mean pixel value of each image given by $\bar{I}_i = \frac{1}{M}\sum_{i=1}^{M} I_i(i)$. We order the M pixels in each mean-subtracted image[1] $\mathbf{I}_k = I_k - \bar{I}_k$ and assign them as column vectors of the overall data matrix **X**. Thus,

$$\mathbf{X} = \begin{bmatrix} \mathbf{I}_1(1) & \mathbf{I}_2(1) & \ldots & \mathbf{I}_N(1) \\ \mathbf{I}_1(2) & \mathbf{I}_2(2) & \cdots & \mathbf{I}_N(2) \\ \vdots & & \ddots & \vdots \\ \mathbf{I}_1(M) & \mathbf{I}_2(M) & \cdots & \mathbf{I}_N(M) \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{I}_1 & \mathbf{I}_2 & \ldots & \mathbf{I}_N \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} \qquad (1)$$

The $(j,k)^{th}$ element of the covariance matrix gives the covariance between the j^{th} and k^{th} images in the ensemble and is given by

$$\mathbf{C}_\mathbf{I}(j,k) = \frac{1}{M-1}\mathbf{I}_j^\mathbf{T}\mathbf{I}_k \quad = \frac{1}{M-1}(I_j - \bar{I}_j)^\mathbf{T}(I_k - \bar{I}_k)$$

it follows that the sample covariance matrix is simply given by the matrix multiplication –

$$\mathbf{C}_\mathbf{I} = \frac{1}{M-1}\mathbf{X}^\mathbf{T}\mathbf{X} \qquad (2)$$

In general, we expect $\mathbf{C}_\mathbf{I}(j,k) \neq 0$ - that is, that the sample images are *correlated*. Our aim is to effect a linear transformation, i.e. derive a new set of uncorrelated images which are linear combinations of the original sample, which will *remove* the correlation. The linear transform is given by $\mathbf{P}_k = \sum_{i=1}^{N} u_k(i)\mathbf{I}_i$ with the weights $u_k(i)$ to be determined. In matrix form, we can express this compactly as –

$$\mathbf{P} = \mathbf{X}\mathbf{U} \qquad (3)$$

[1] Conventionally, we strip the pixels for each column in the image and stack them to form one long vector – this is sometimes referred to as applying the stacking operator.

It is instructive to write this in tabular form -

$$\begin{bmatrix} \mathbf{P}_1(1) & \mathbf{P}_2(1) & \dots & \mathbf{P}_N(1) \\ \mathbf{P}_1(2) & \mathbf{P}_2(2) & \cdots & \mathbf{P}_N(2) \\ \vdots & & \ddots & \vdots \\ \mathbf{P}_1(M) & \mathbf{P}_2(M) & \cdots & \mathbf{P}_N(M) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1(1) & \mathbf{I}_2(1) & \dots & \mathbf{I}_N(1) \\ \mathbf{I}_1(2) & \mathbf{I}_2(2) & \cdots & \mathbf{I}_N(2) \\ \vdots & & \ddots & \vdots \\ \mathbf{I}_1(M) & \mathbf{I}_2(M) & \cdots & \mathbf{I}_N(M) \end{bmatrix} \begin{bmatrix} u_1(1) & u_2(1) & \dots & u_N(1) \\ u_1(2) & u_2(2) & \cdots & u\ (2) \\ \vdots & & \ddots & \vdots \\ u_1(N) & u_2(N) & \cdots & u_N(N) \end{bmatrix} \quad (4)$$

Note the new set of images form the *columns* of the matrix **P**, the original sample images form the *columns* of **X** and the matrix of weighting coefficients **U** is to be determined**.**

We must enforce the condition that the new set of images be uncorrelated. [2] This requires us to find a matrix **U** such that P is *orthogonal* matrix. We thus demand that -

$$\frac{1}{M-1}\mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{U}^T \left\{ \frac{1}{M-1} \quad \mathbf{X}^T\mathbf{X} \right\} \mathbf{U} \;=\; \Lambda \quad \textit{with } \Lambda \textit{ a diagonal matrix.}$$

Note that demanding Λ is diagonal ensures that the inner product of any two columns of **P** will be zero unless the columns are the same – precisely the condition required if the new vectors are to be uncorrelated. Note that we have included the (optional) scaling factor of $\frac{1}{M-1}$ so that the sample covariance matrix ($\mathbf{C}_{\mathbf{I}} = \frac{1}{M-1}\mathbf{X}^{\mathbf{T}}\mathbf{X}$) appears precisely on the RHS. The problem thus fundamentally reduces to solving for a matrix **U** which diagonalises the sample covariance matrix -

$$\mathbf{U}^T\mathbf{C}_{\mathbf{I}}\mathbf{U} = \Lambda \quad \textit{with } \Lambda \textit{ a diagonalmatrix.} \qquad (5)$$

This is the eigenvalue/eigenvector problem which yields to standard numerical procedures. Once **U** has been found, the principal components are calculated using eq. (3).

Note also that since **U** is a matrix of orthonormal eigenvectors, the original (mean-subtracted) data can be expressed in terms of the principal components by inverting eq. 14.11. Thus we can write –

$$\mathbf{X} = \mathbf{P}\mathbf{U}^T \qquad (6)$$

[2] i.e. that $\mathbf{P^T}_k\mathbf{P}_j = \sum_{i=1}^{N} u_k(i)\mathbf{I^T}_i \;\; \sum_{l=1}^{N} u_j(l)\mathbf{I}_l = \lambda_k \;\; \delta_{jk}$.