Principal Component Analysis (PCA)

Dr. Xiqun Lu

College of Computer Science

Zhejiang University

Image Representation

POSTERIOR BASIS FOR NATURAL IMAGES

Basis for Natural Images

- **Priori Basis**
  - Fourier transforms
  - DCT
  - …
- **Posterior Basis** (data-driven)
  - To estimate the linear transform from the data itself, and the transform could be ideally adapted to the kind of data that is being processed.
    - PCA
    - Dictionary based on machine learning

Two Properties of Image Transforms

- Variable decoupling
  - The coefficients of these bases are **less correlated** or become **independent** in ideal cases.
- Dimension reduction
  - The number of bases for approximately reconstructing an image is often much smaller than the number of pixels.

PCA [1]

- Either DFT or DCT has “energy compaction” property.
- What is the **optimal** transform in terms of energy compaction?

- Two basic concepts in statistical analysis: *variance* and *covariance* of random vectors

Variance and Covariance

- Consider two random vectors with zero means $\mathbf{a} = [a_1, a_2, \ldots, a_n]^T$, $\mathbf{b} = [b_1, b_2, \ldots, b_n]^T$.
- The **variance** of **a** and **b** are defined as

$$\sigma_{\mathbf{a}}^2 = \langle a_i \quad a_i \rangle_i = \sum_{i=1}^{n} a_i^2$$

$$\sigma_{\mathbf{b}}^2 = \langle b_i \quad b_i \rangle_i$$

  - where the expectation $\langle\rangle_i$ is the average over n variables.
- The **covariance** between **a** and **b** is a straight-forward generalization:

$$\sigma_{\mathbf{ab}}^2 = \langle a_i \quad b_i \rangle_i$$

Sample Covariance Matrix

- We generalize two random vectors to an arbitrary number. Suppose we have *m* random vectors, and each vector has *n* variables.
- The $n \times m$ observation matrix **X** is defined as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_m \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{mn} \end{bmatrix}$$

- The **sample** covariance matrix is estimated

$$\boxed{C_{\mathbf{X}} = \frac{1}{m-1} \mathbf{X}\mathbf{X}^T}$$

Goal of PCA

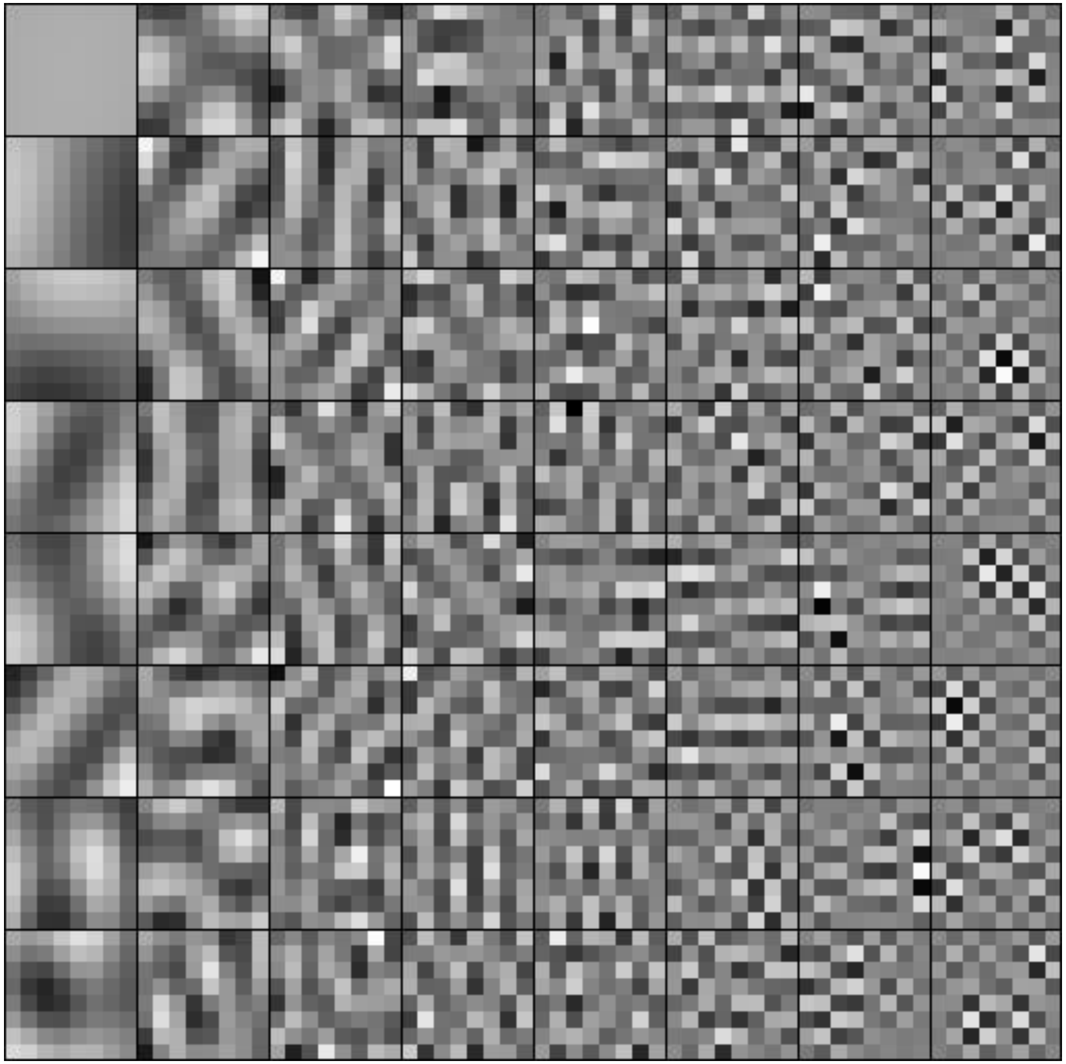
- The goal of PCA is to find *a set of basis* in which the observation matrix **X** is transformed to a new observation matrix **Y**, and *the covariance matrix* of **Y** will be a **diagonal** matrix — to remove correlation among data whereas to preserve the energy as much as possible.
- Assumptions of PCA:
  - Firstly all basis vectors $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ are **orthonormal** (i.e. $\mathbf{p}_i \bullet \mathbf{p}_j = \delta_{ij}$), that is all basis vectors construct an orthonormal matrix **P**.
  - Secondly, *the directions with the largest variances* are more important than those with low variances, so in the transformed space the energy of signals is kept as compact as possible.

Eigen-Value Decomposition

- In the linear algebra, the covariance matrix $C_{\mathbf{x}}$ can be *diagonalized* by an orthogonal matrix of its eigen-vectors.
- Eigen-value decomposition: $C_{\mathrm{x}} = \mathbf{EDE}^{\mathrm{T}}$
- If the orthonormal transform matrix $\mathbf{P}$ is selected as $\mathbf{P} = \mathbf{E}^{\mathrm{T}}$, $\mathbf{Y} = \mathbf{PX} = \mathbf{E}^{\mathrm{T}}\mathbf{X}$

$$\boxed{\begin{aligned}
C_Y &= \frac{1}{m-1}\mathbf{Y}\mathbf{Y}^T \\
&= \frac{1}{m-1}(\mathbf{E}^T\mathbf{X})(\mathbf{E}^T\mathbf{X})^T \\
&= \frac{1}{m-1}\mathbf{E}^T\mathbf{X}\mathbf{X}^T\mathbf{E} \\
&= \mathbf{E}^T C_{\mathbf{x}}\mathbf{E} \\
&= \mathbf{E}^T\,\mathbf{EDE}^T\mathbf{E} \\
&= \mathbf{D}
\end{aligned}}$$

64 8×8 principal components (or basic images) learned from the image “Lena"



A toy example of PCA

1) Divide an image with the size of 512*512 into 8*8 blocks

2) Observation Matrix $\mathbf{X}$: 64 * 4096

3) The Sampling Covariance Matrix $\mathbf{C}_{\mathbf{X}}$: 64*64

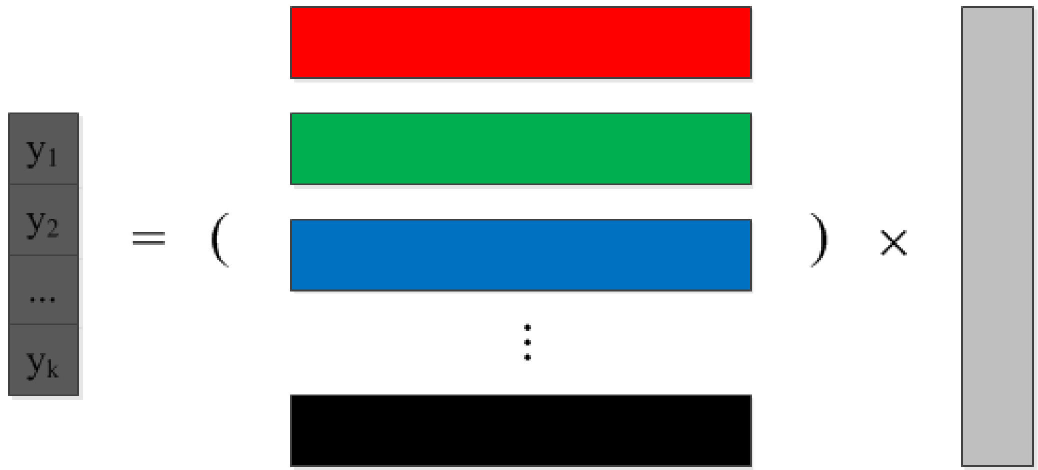
4) Eigenvalue Decomposition of $\mathbf{C}_{\mathbf{X}}$: $\mathbf{C}_{\mathbf{X}} = \mathrm{EDE}^{\mathrm{T}}$

5) We select the eigenvectors with the first k largest eigenvalues, and form a PCA transform matrix $\mathbf{P}$ ($\mathbf{P}$: 64 * k, k <<64)

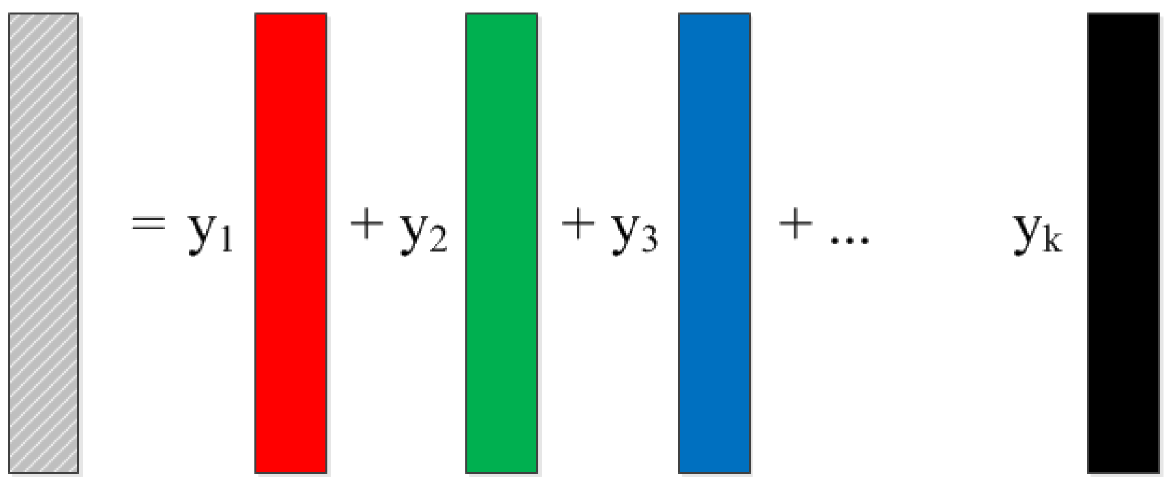
6) Now we can transform the observation matrix X into another space by the PCA matrix $\mathbf{P}$ ($\mathbf{Y} = \mathbf{P}^{\mathrm{T}}\mathbf{X}$ and Y: k * 4096)

7) Reconstruction: $\mathbf{X}_{\mathrm{R}} = \mathbf{PY}$ (Note $\mathbf{X}_{\mathrm{R}} \sim= \mathbf{X}$)

PCA Forward Transform: $y = \mathbf{P}^T x$



PCA Inverse Transform: $x = \mathbf{P}y$



Reconstructed images with different number of PCs

20

10

5

References

- [1] M. J. T. Smith and A. Docef, A Study Guide for Digital Image Processing, Scientific Publishers, Inc. Riverdale, Georgia, 1999.

Thank You!

Dr. Xiqun Lu

xqlu@zju.edu.cn