HO CHI MINH CITY UNIVERSITY OF SCIENCE
FACULTY OF ELECTRONICS AND TELECOMMUNICATIONS
DEPARTMENT OF TELECOMMUNICATIONS AND NETWORKS

Course

DIGITAL SIGNAL PROCESSING

Topic 8

Principal Component Analysis in image processing

Name	ID Student
1. Phạm Hoài An	21207120
2. Đỗ Minh Chương	21207126
3. Trần Thiên Phúc	21207077
4. Ngô Chơn Quang	21207085

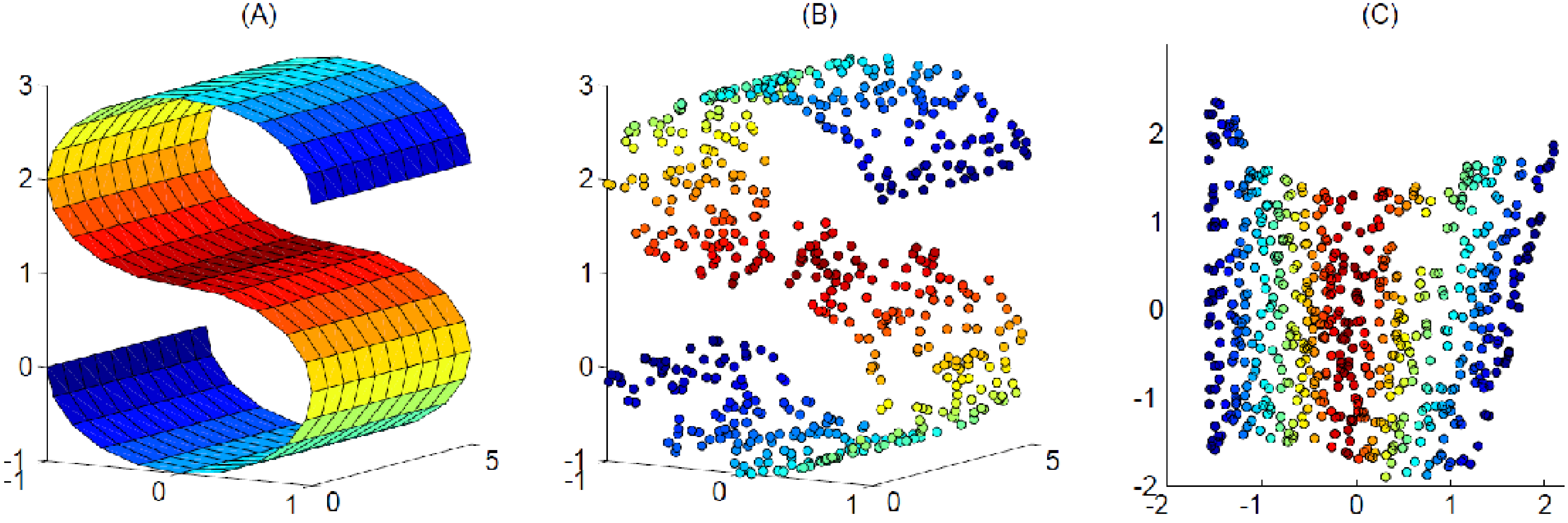
Outline

1. Principal Component Analysis (PCA)
2. Methods of performing PCA in image processing
3. Applications of PCA in image processing
4. PCA for Process Monitoring

1. Principal Component Analysis (PCA)

- An exploratory technique used to reduce the dimensionality of the data set to 2D or 3D
- Can be used to:
  - Reduce number of dimensions in data
  - Find patterns in high-dimensional data
  - Visualize data of high dimensionality
- Example applications:
  - Face recognition
  - Image compression
  - Gene expression analysis

❖ Example for PCA in images processing:



- By using PCA we can reduce the size of the data to make processing easier….
- It can be seen from pic A after PCA -> pic B that the number of samples has been greatly reduced, decreasing from 3D to 2D

2. Methods of performing PCA in image processing

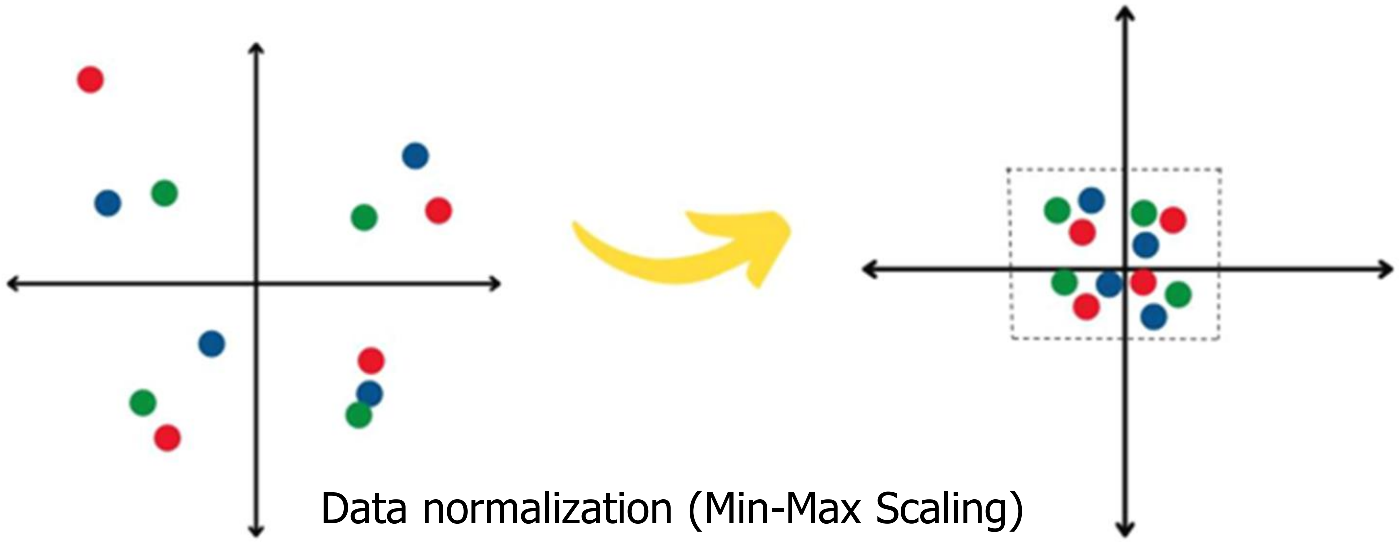
- Data Preprocessing
- Creating the Covariance Matrix
- Calculating Principal Components
- Dimensionality Reduction and Creating New Images

Data Preprocessing

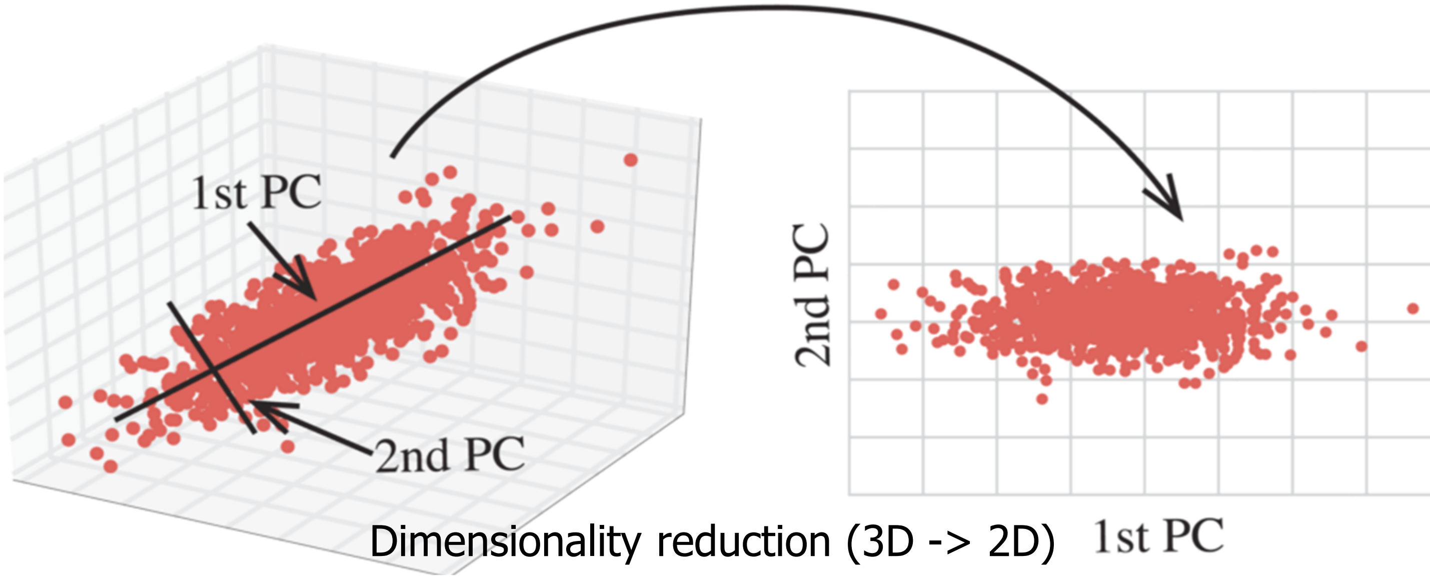
- Data Preprocessing is the process of preparing data before feeding it into machine learning models or data analysis.

- Objective: Normalize data, reduce noise, handle missing data, and prepare data for analysis.
  - Data normalization (Min-Max Scaling, Standardization)
  - Handling missing data
  - Discrete data processing (DFT, FFT,…)
  - Dimensionality reduction (PCA)

Data Preprocessing



Data normalization (Min-Max Scaling)



Dimensionality reduction (3D -> 2D)

Creating the Covariance Matrix

- The covariance matrix is a square matrix with a size equal to the number of variables in the image

- From the matrix covariance, we calculate the eigenvectors and eigenvalues for use in the next steps of PCA.

- Covariance: measures the correlation between X and Y
  - $Cov(X, Y) = 0$: **Independent**
  - $Cov(X, Y) > 0$: **Move Same Direction**
  - $Cov(X, Y) < 0$: **Move Opposite Direction**

Creating the Covariance Matrix

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

- X_i, Y_i are values at position i in the two respective columns of the data matrix.
- $\bar{X}$, $\bar{Y}$ are the mean values of columns X and Y.
- n is the number of samples (in this case, 7).

$$\bar{X} = \frac{40+40+40+30+15+12+12}{7} = 27$$

$$\bar{Y} = \frac{90+90+90+87+70+70+70}{7} = 81$$

Example:

X=Temperature	Y=Humidity
40	90
40	90
40	90
30	87
15	70
12	70
12	70

$$\Rightarrow Cov(X,Y) = \frac{(40-27)*(90-81)+ \ldots + (12-27)*(70-81)}{7-1} = 138.5 \longrightarrow$$ Move Same Direction

Calculating Principal Components

- To perform PCA, we need to calculate the covariance matrix to understand the relationships between pixels in the image.
- The covariance matrix stores information about how pixel values change together.
- Eigenvalues & eigenvectors:
  - Vectors **x** having same direction as C**x** are called *eigenvectors* of C (C is an n by n matrix).
  - In the equation C**x**=λ**x**, λ is called an *eigenvalue* of C.

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} x \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4x \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Calculating Principal Components

- Let's illustrate these steps with an example:
- Assume we have the covariance matrix C after computing it from the standardized data X:

$$C = \begin{pmatrix} 3/2 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{pmatrix}$$

- Calculate det(C-λI):

$$\det = \begin{pmatrix} \frac{3}{2} - \lambda & -1/2 & -1/2 \\ -1/2 & 1 - \lambda & -1/2 \\ -1/2 & -1/2 & 1 - \lambda \end{pmatrix}$$

- Determine roots to det(C-λI)=0 $\Rightarrow$ $\begin{matrix} \lambda 1 = 2 \\ \lambda 2 = 1/2 \\ \lambda 3 = 1/2 \end{matrix}$

Calculating Principal Components

- Solve (C- λ I) x=0 for each λ to obtain eigenvectors x
  - For λ1 = 2: $\begin{pmatrix} \frac{3}{2}-2 & -1/2 & -1/2 \\ -1/2 & 1-2 & -1/2 \\ -1/2 & -1/2 & 1-2 \end{pmatrix} \times x_1 = 0, \Rightarrow \boldsymbol{x_1}$ **= [1, -1, -1]**
  - For λ2 = ½: $\begin{pmatrix} \frac{3}{2}-1/2 & -1/2 & -1/2 \\ -1/2 & 1-1/2 & -1/2 \\ -1/2 & -1/2 & 1-1/2 \end{pmatrix} \times x_2 = 0, \Rightarrow \boldsymbol{x_2}$ **= [-1, 0, 1]**
  - For λ3 = ½: ⇒ $\boldsymbol{x_3}$ **= [-1, 0, 1]**
- Sort the Eigenvalues:
  - Arrange the eigenvalues in descending order: **λ1 > λ2 > λ3, (2 > ½ > ½)**
- Select the Principal Components:
  - $\boldsymbol{x_1}$ **= [1, -1, -1]**
  - $\boldsymbol{x_2}$ **= [-1, 0, 1]**

✓ *These eigenvectors* $\boldsymbol{x_1}$ *and* $\boldsymbol{x_2}$ *are the principal components of the data. We can use them to reduce the dimensionality of the data in PCA.*

Dimensionality Reduction and Creating New Images

- Dimensionality Reduction
  - Reduce color space dimensionality using selected eigenvectors.
  - Transform original pixel values into a new, lower-dimensional representation.
- Creating New Images
  - Use the reduced representation to generate new images.
  - Combine the selected principal components to form the new image.
- Benefits of Dimensionality Reduction
  - Compact data representation: Reduced storage space.
  - Noise reduction: Remove less significant features.
  - Efficient computation: Faster image processing.
- Application Examples
  - Image compression: Reduce image size while preserving essential features.
  - Image reconstruction: Recover original image from reduced representation.

3. Applications of PCA in image processing

- Eigenfaces are

  the eigenvectors of
  the covariance matrix of
  the probability distribution of
  the vector space of
  human faces

- Eigenfaces are the ‘standardized face ingredients’ derived from the statistical analysis of many pictures of human faces
- A human face may be considered to be a combination of these standard faces

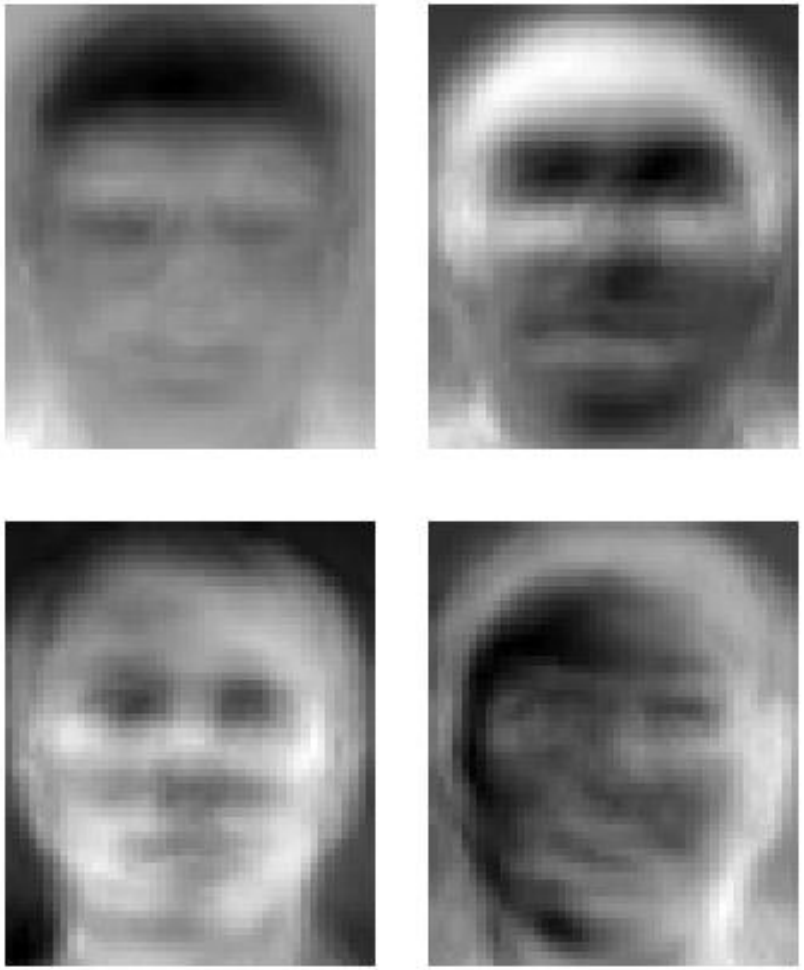
PCA applications -Eigenfaces

To generate a **set of eigenfaces:**

1. Large set of digitized images of human faces is taken under the same lighting conditions.
2. The images are normalized to line up the eyes and mouths.
3. The eigenvectors of the covariance matrix of the statistical distribution of face image vectors are then extracted.
4. These eigenvectors are called eigenfaces.

PCA applications -Eigenfaces

- The principal eigenface looks like a bland androgynous average human face



http://en.wikipedia.org/wiki/Image:Eigenfaces.png

Eigenfaces – Face Recognition

- When properly weighted, eigenfaces can be summed together to create an approximate gray-scale rendering of a human face.
- Remarkably few eigenvector terms are needed to give a fair likeness of most people's faces
- Hence eigenfaces provide a means of applying data compression to faces for identification purposes.
- Similarly, Expert Object Recognition in Video

Eigenfaces

- <u>Experiment and Results</u>

  Data used here are from the ORL database of faces. Facial images of 16 persons each with 10 views are used. - Training set contains 16×7 images.

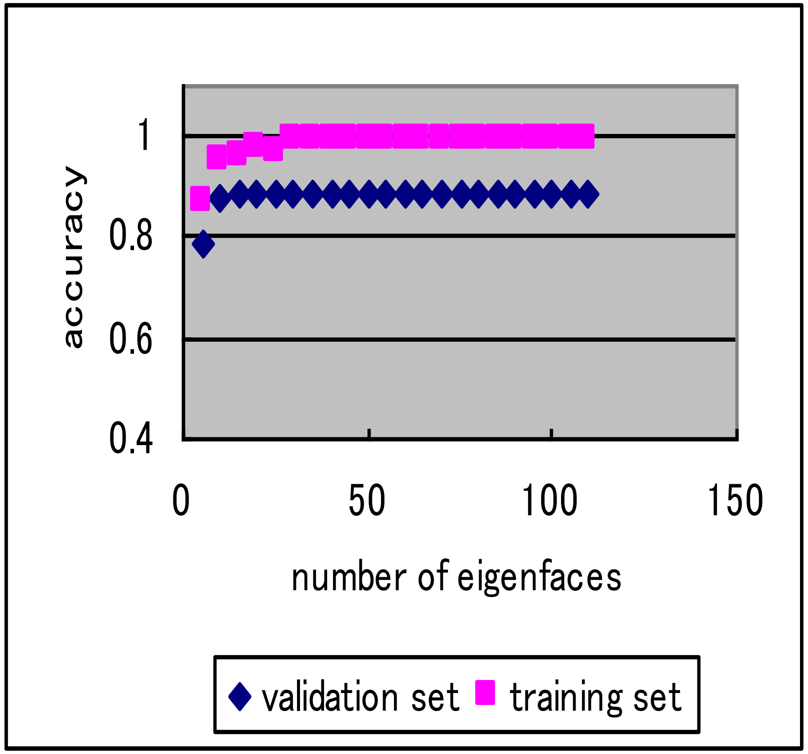
  Test set contains 16×3 images.

First three eigenfaces :

Classification Using Nearest Neighbor

- Save average coefficients for each person. Classify new face as the person with the closest average.
- Recognition accuracy increases with number of eigenfaces till 15.

Later eigenfaces do not help much with recognition.



Best recognition rates

Training set 99%

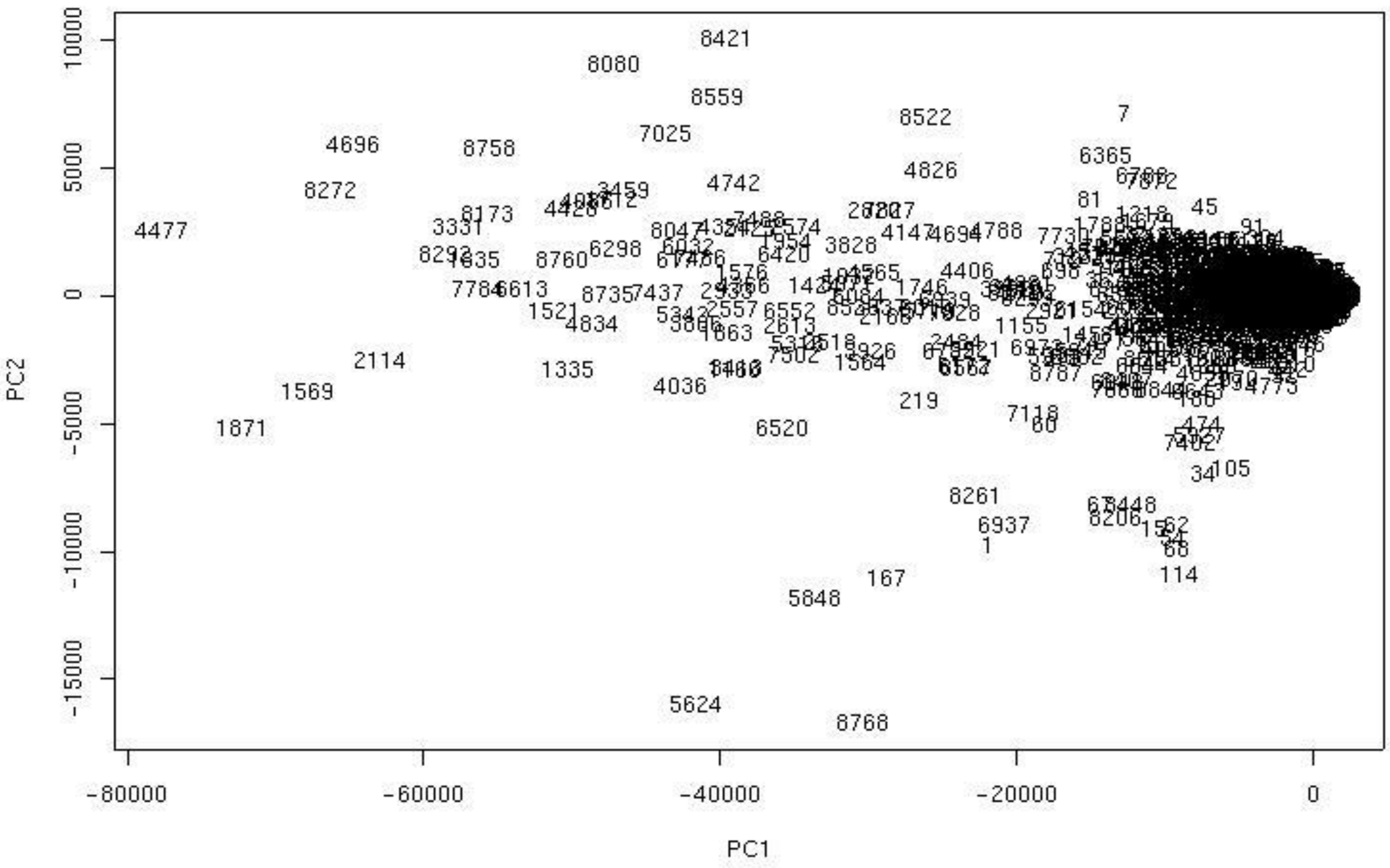
Test set 89%

PCA of Genes (Leukemia data, precursor B and T cells)

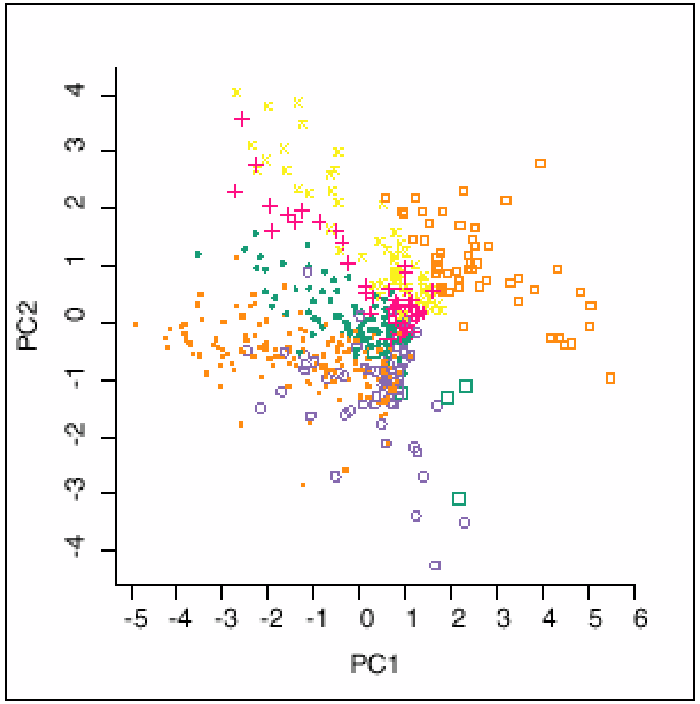
- 34 patients, dimension of 8973 genes reduced to 2

PCA of genes (Leukemia data)

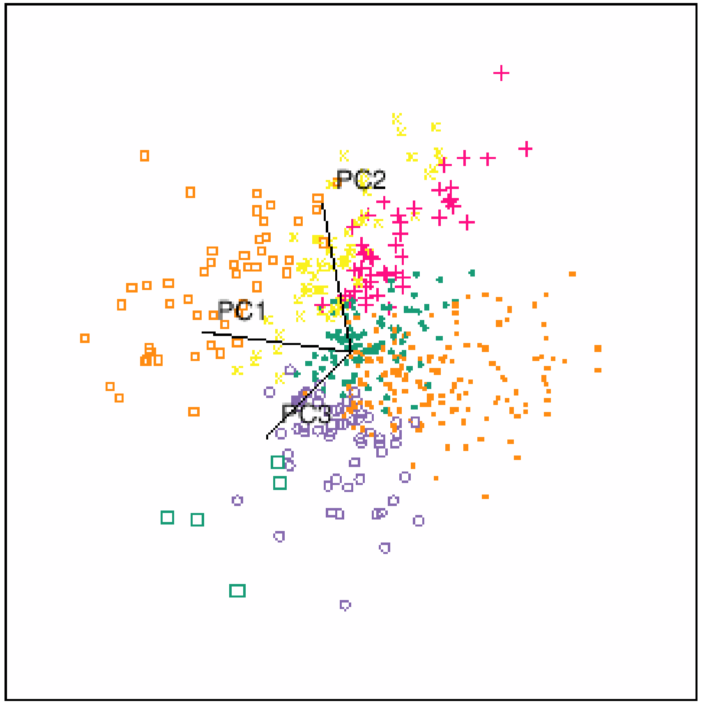
- Plot of 8973 genes, dimension of 34 patients reduced to 2



Sporulation Data



(a) In the subspace of the first 2 PC's



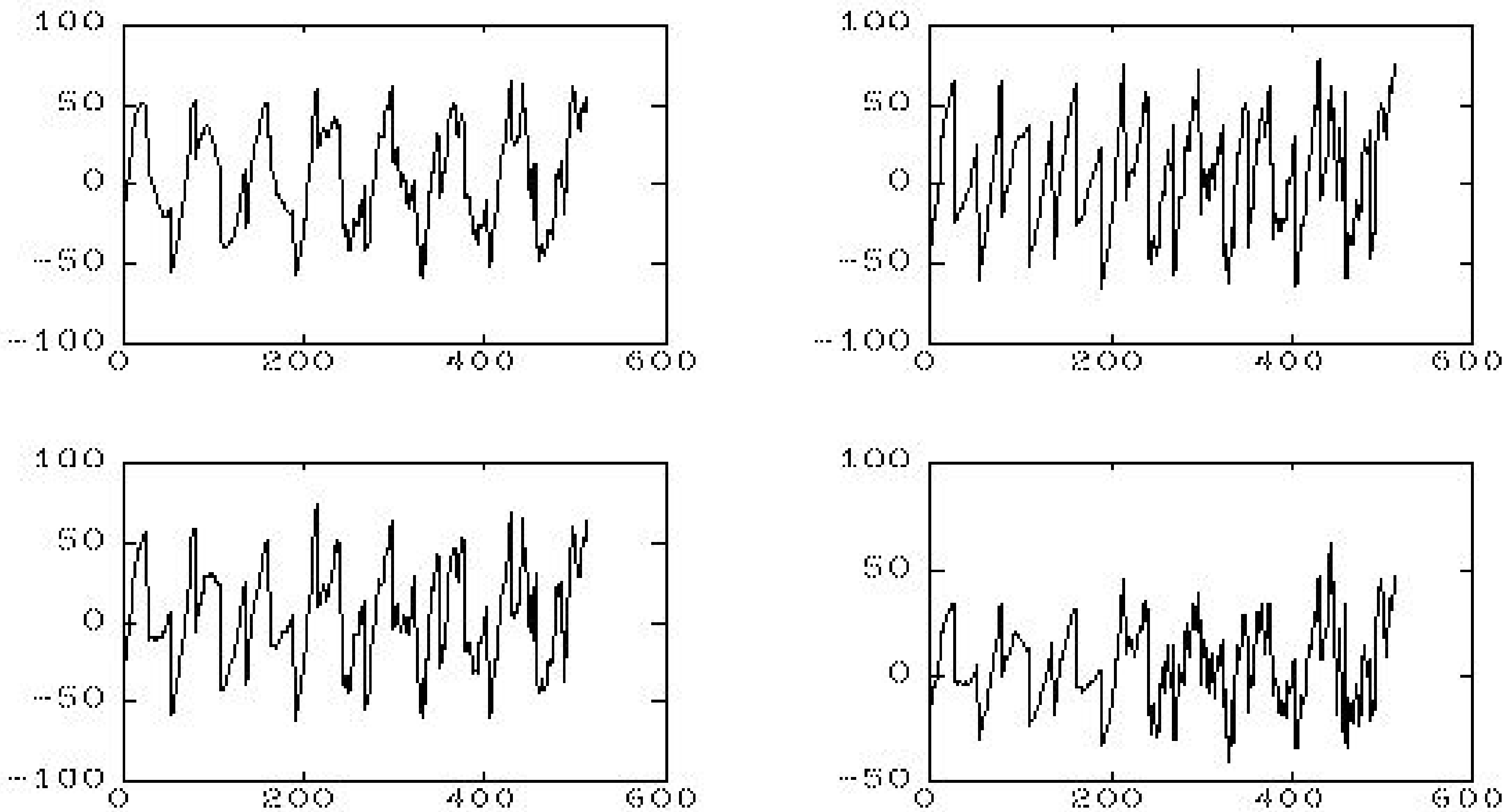
(b) In the subspace of the first 3 PC's

Fig. 1. Visualization of a subset of the sporulation data. (a) In the subspace of the first two PCs. (b) In the subspace of the first three PCs.

- The patterns overlap around the origin in (**1a**).
- The patterns are much more separated in (**1b**).

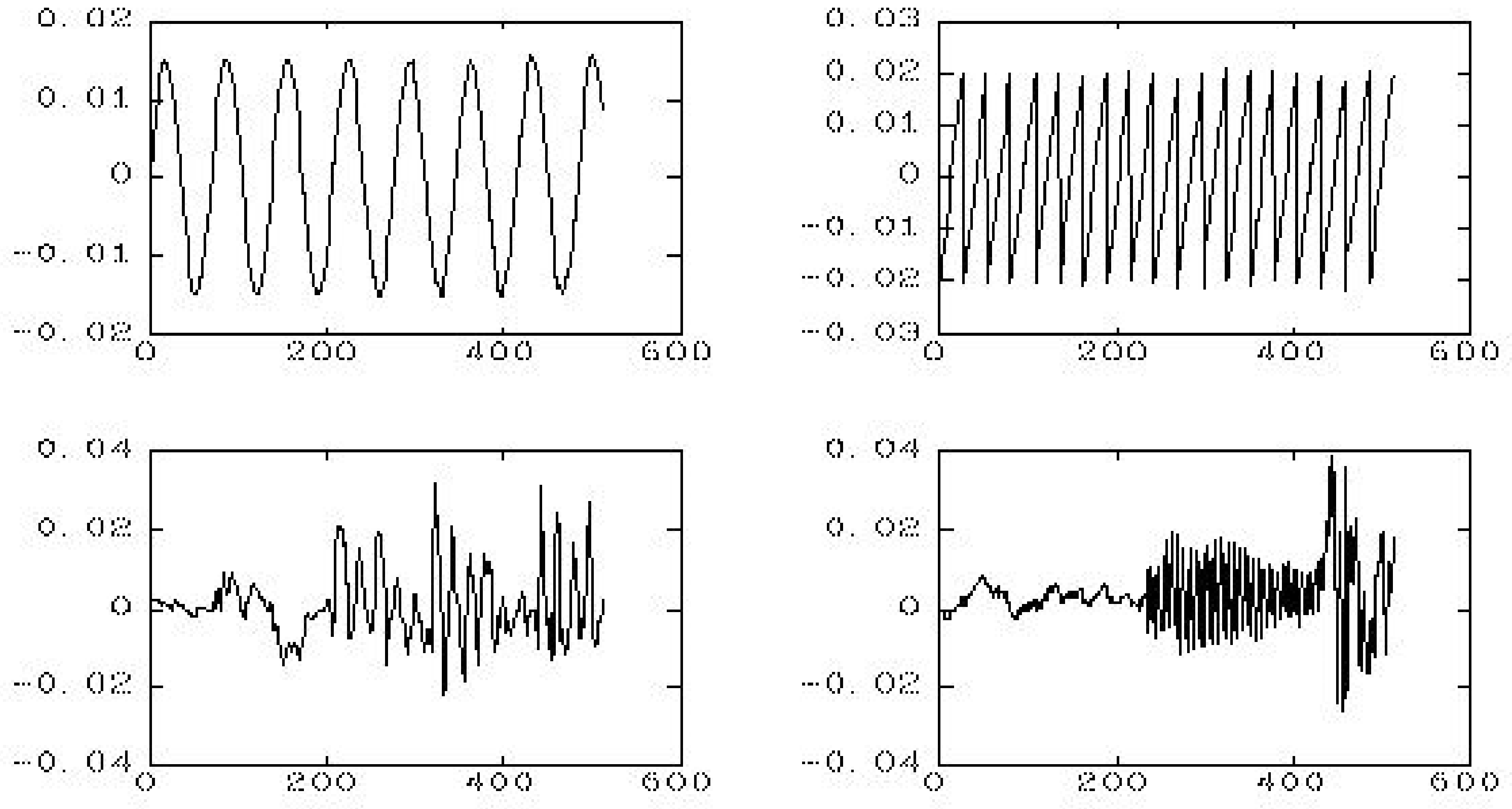
Variation in processes:

- Chance
  - Natural variation inherent in a process. Cumulative effect of many small, unavoidable causes.
- Assignable
  - Variations in raw material, machine tools, mechanical failure and human error. These are accountable circumstances and are normally larger.



- This is recorded by the microphones: a linear mixture of the sources
- $x_i(t) = a_{i1}*s_1(t) + a_{i2}*s_2(t) + a_{i3}*s_3(t) + a_{i4}*s_4(t)$

PCA for Process Monitoring

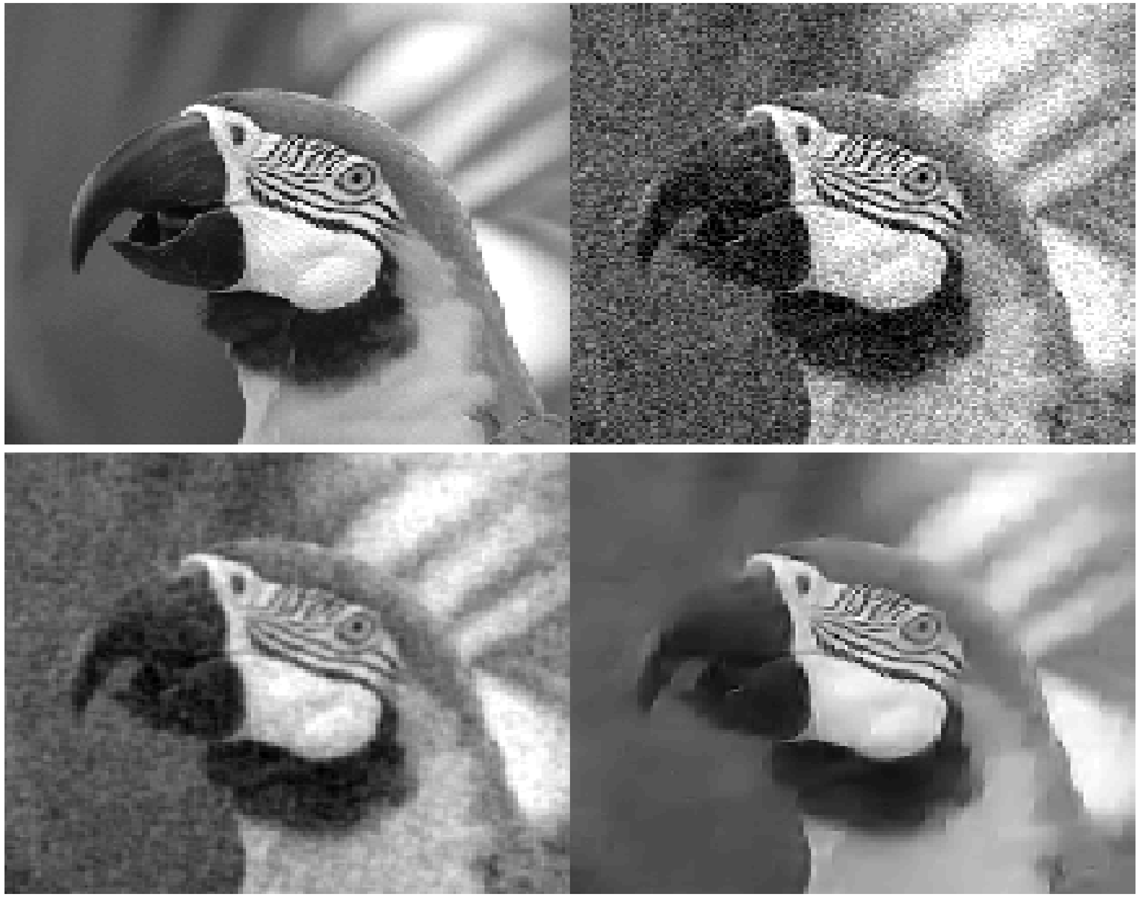


Recovered signals

Image denoising

Original image

Noisy image

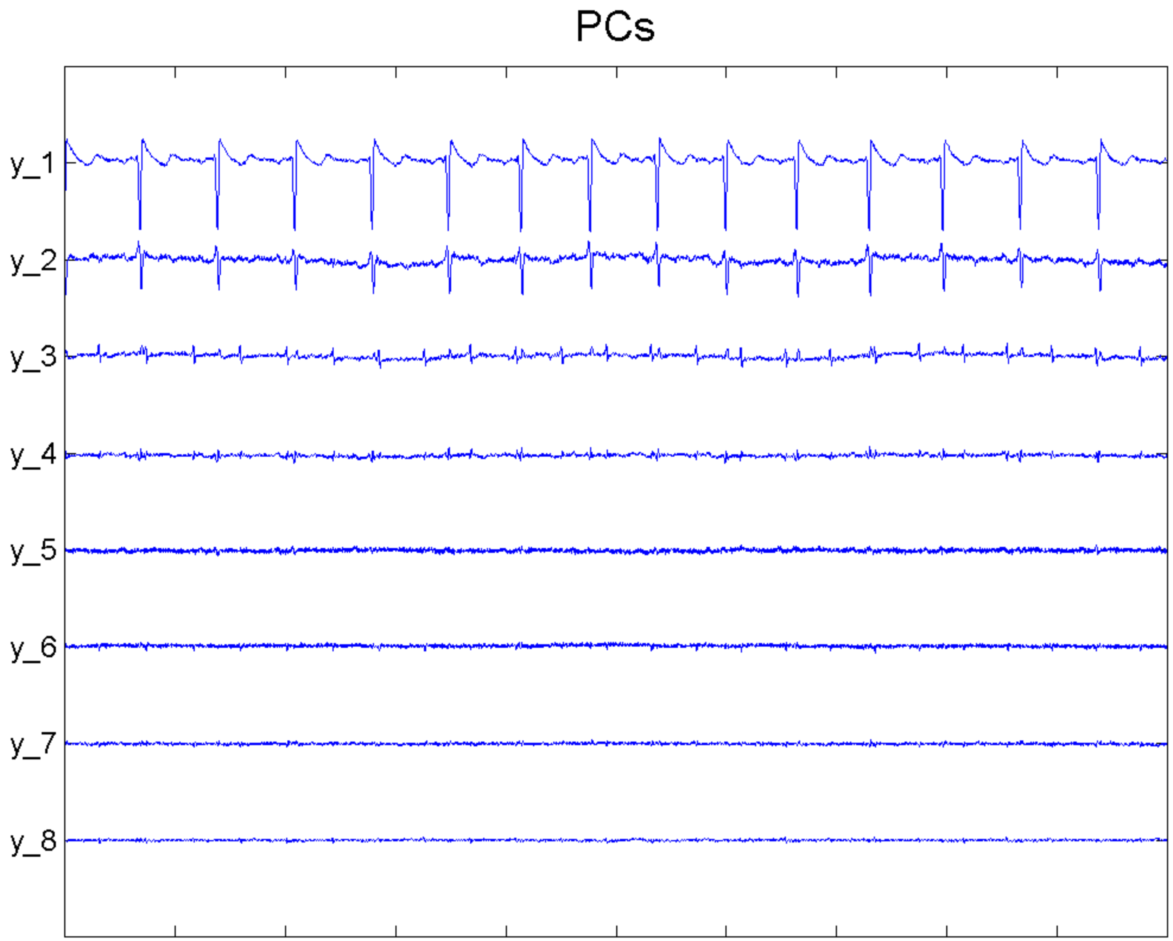


Wiener filtering

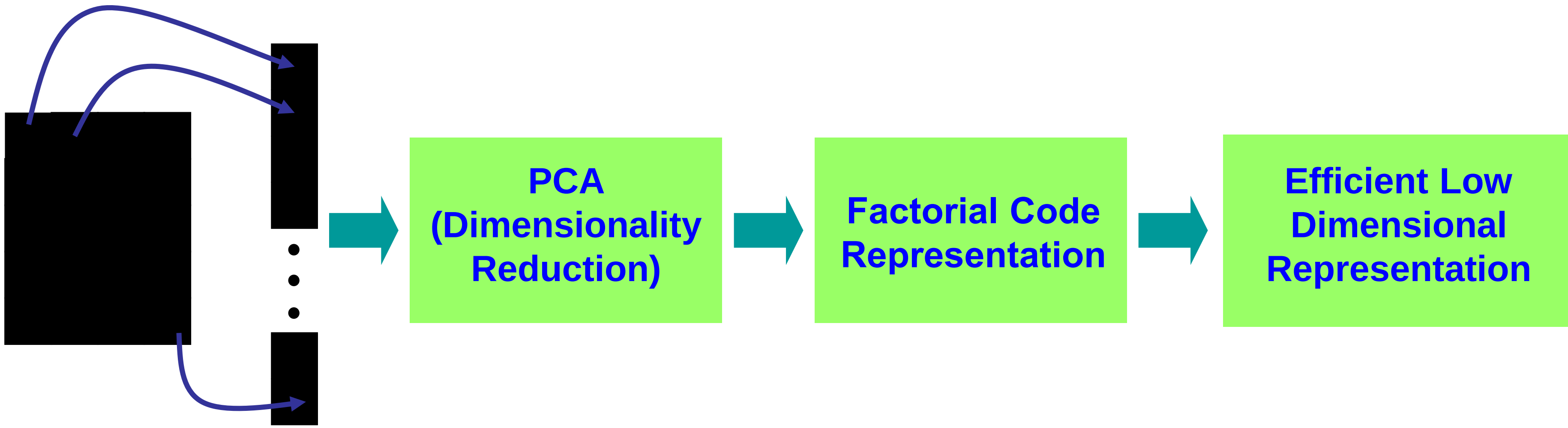
ICA filtering

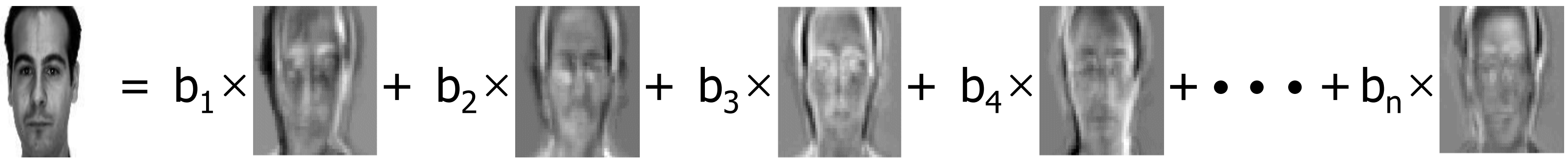
PCA for Process Monitoring

- Feature Extraction in ECG data (PCA)



- ***Factorial Code Representation:***
  Factorial Faces





Experimental Results

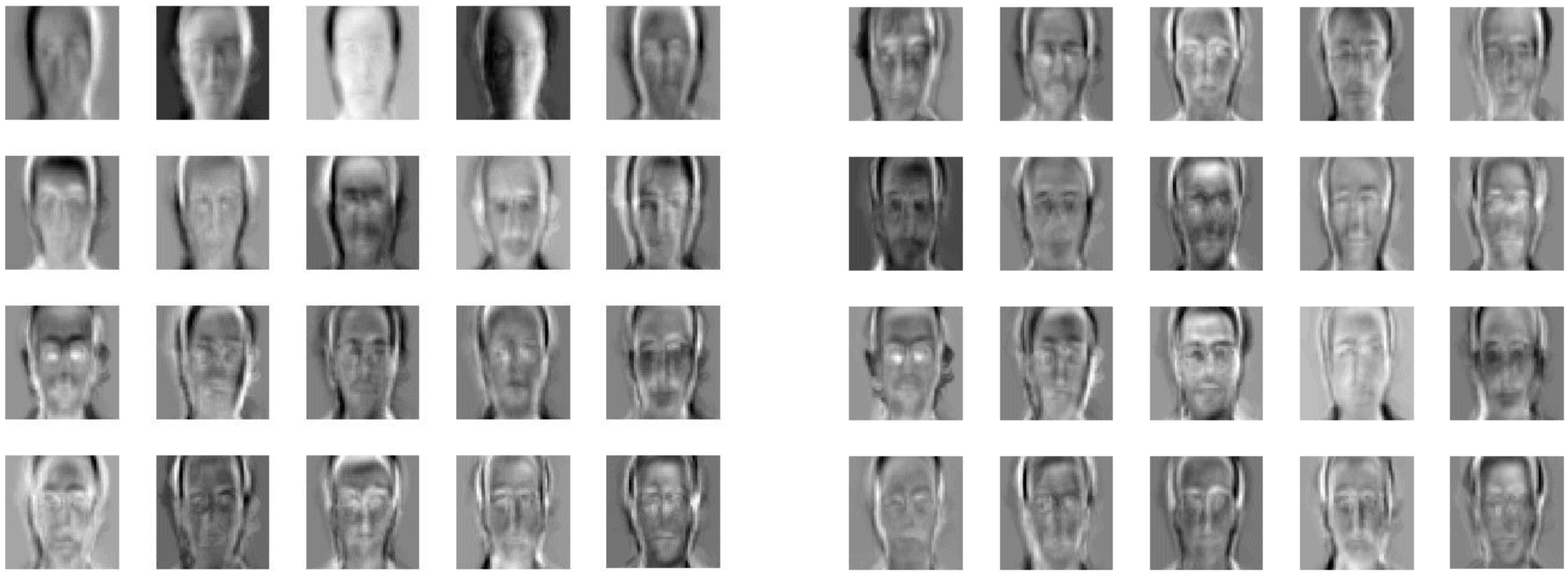
Figure 1. Sample images in the training set.

(neutral expression, anger, and right-light-on from first session; smile and left-light-on from second session)

Figure 2. Sample images in the test set.

(smile and left-light-on from first session; neutral expression, anger, and right-light-on from second session)

Eigenfaces vs Factorial Faces



(a) (b)

Figure 3. First 20 basis images: (a) in eigenface method; (b) factorial code. They are ordered by column, then, by row.