CHALLENGES OF VISUAL RECOGNITION

• Appearance
  • DOF: texture, illumination, material, shading, ...
• Shape
  • DOF: object category, geometric pose, viewpoint, ...
Challenges of Visual Recognition

• Appearance
  • DOF: texture, Illumination, material, shading, ...

• Shape
  • DOF: object category, geometric pose, viewpoint, ...

[Images of people and objects in a park setting]
Space of Appearance (Fixed Shape)

$x \in \mathbb{R}^D$

Template

High dimension (D)
e.g., D: 10,000 = 100 x 100
**SPACE OF APPEARANCE (FIXED SHAPE)**

Template

High dimension (D)

\( \text{e.g., } D: 10,000 = 100 \times 100 \)

\( x \in \mathbb{R}^D \)

Naïve face detection algorithm:

Use NCC or SSD to measure similarity.

\[
\begin{align*}
\text{maximize} & \quad \text{corr}(x, y) \\
\text{minimize} & \quad \|x - y\|^2
\end{align*}
\]

Why not working?
SPACE OF FACE APPEARANCE
SPACE OF FACE APPEARANCE
Miss Korea Contestants

Observation: not all pixels are equally informative to detect a face
**Miss Korea Contestants**

Observation: not all pixels are equally informative to detect a face

Average image
Structured Appearance

Idea: face images are highly correlated and can be represented in a low-dimensional subspace.
**Linear Basis**

\[
\begin{align*}
\text{Face} & = \text{Mean face} + \alpha_1 + \alpha_2 + \alpha_3 + \cdots \\
y & = m + \alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3 + \cdots
\end{align*}
\]
**Linear Basis**

\[
y = m + \alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3 + \cdots
\]
**Linear Basis**

\[
\text{Face} = \text{Mean face} + \alpha_1 + \alpha_2 + \alpha_3 + \cdots
\]

\[
y = m + \alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3 + \cdots
\]

\[
y \approx m + B \alpha
\]
Reconstruction from Linear Basis

$$y = m + \alpha_1 b_1 + \alpha_2 b_2 + \alpha_3 b_3 + \cdots$$

$$\alpha^* = \min_{\alpha} \|y - m - B\alpha\|^2$$

Cf) $$\min_{x} \|x - y\|^2$$ Template matching
RECONSTRUCTION EXPRESSIBILITY
Reconstruction Expressibility
Reconstruction Expressibility
Reconstruction Expressibility
Reconstruction Expressibility

Reconstruction

# basis: 5
Reconstruction Expressibility
**How to Compute Mean and Basis from Database?**

\[ m^*, B^*, \alpha^* = \minimize_{m, B, \alpha} \| y - m - B\alpha \|^2 \]
**How to Compute Mean?**

Consider a dataset $\{x_i\}_{i=1}^n$ consisting of $n$ images, where each image is represented as a vector $x_i$. The goal is to find the mean image $m$ that best represents the dataset.

Mathematically, we aim to find $m^*, B^*, \alpha^*$ that minimize the distance between the mean and the dataset:

$$m^*, B^*, \alpha^* = \min_{m, B, \alpha} \| y - m - B\alpha \|^2$$

The mean $m$ is calculated as follows:

$$m = \frac{1}{n} \sum_{i=1}^{n} x_i$$
How to Compute Mean?

$m^*, B^*, \alpha^* = \underset{m, B, \alpha}{\text{minimize}} \| y - m - B\alpha \|^2$

$m = \frac{1}{n} \sum_{i} x_i$
Mean Subtraction

\[
m^*, B^*, \alpha^* = \text{minimize} \| y - m - B\alpha \|^2_{m,B,\alpha}
\]

\[
m = \frac{1}{n} \sum_{i} x_i
\]
How to Compute Basis?

\[
m^*, B^*, \alpha^* = \text{minimize} \| y - m - B\alpha \|^2
\]

\[
m = \frac{1}{n} \sum_{i=1}^{n} x_i
\]
Principal Axis
**Principal Axis**

Basis is the axis that represents the maximum data covariance.
**Principal Axis**

Basis is the axis that represents the maximum data covariance.

Coefficient \( \alpha^i = \frac{b \cdot \bar{x}_i}{\|b\|} \)
**Principal Axis**

Basis is the axis that represents the maximum data covariance.

Coefficient \( \alpha^i = \frac{b \cdot \bar{x}_i}{\|b\|} \)

\( b^* = \max_b \sum_{i=1}^{n} (\alpha^i)^2 \)
**Principal Axis**

Basis is the axis that represents the maximum data covariance.

Coefficient  \( \alpha^i = \frac{b \cdot \bar{x}_i}{\|b\|} \)

\( b^* = \max_b \sum_{i=1}^n (\alpha^i)^2 \)

\( = \max_b \sum_{i=1}^n \left( \frac{b \cdot \bar{x}_i}{\|b\|} \right)^2 \)
**Principal Axis**

Basis is the axis that represents the maximum data covariance.

Coefficient \( \alpha^i = \frac{b \cdot \bar{x}_i}{\|b\|} \)

\( b^* = \max_b \sum_{i=1}^n (\alpha^i)^2 \)

\( = \max_b \sum_{i=1}^n \left( \frac{b \cdot \bar{x}_i}{\|b\|} \right)^2 \)

\( = \max_b b^T X^T X b \)

Covariance matrix

where \( x = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \)
**Principal Axis**

Basis is the axis that represents the maximum data covariance.

Coefficient \( \alpha^i = \frac{b \cdot \bar{x}_i}{\|b\|} \)

\[ b^* = \underset{b}{\text{maximize}} \sum_{i=1}^{n} (\alpha^i)^2 \]

\[ = \underset{b}{\text{maximize}} \sum_{i=1}^{n} \left( \frac{b \cdot \bar{x}_i}{\|b\|} \right)^2 \]

\[ = \underset{b}{\text{maximize}} b^T X^T X b \]

Covariance matrix

Solution is the eigenvector corresponding to the largest eigenvalue: \( b^* = \lambda_{\text{max}} (X^T X) \)
Principal Axis

Basis is the axis that represents the maximum data covariance.

Coefficient $\alpha^i = \frac{b \cdot \bar{x}_i}{\|b\|}$

$$\begin{bmatrix} \alpha^1 \\ \vdots \\ \alpha^n \end{bmatrix} = \begin{bmatrix} \bar{x}^T_1 \\ \vdots \\ \bar{x}^T_n \end{bmatrix} \begin{bmatrix} b \end{bmatrix}$$

 nx1  nxD  Dx1
**Principal Axes**

Basis is the axis that represents the maximum data covariance.

Coefficient

\[ \alpha_{i}^{j} = \frac{b_{j} \cdot \bar{x}_{i}}{\|b_{j}\|} \]

Orthogonal principal axes: first \( d \) largest eigenvectors

\[
\begin{bmatrix}
\alpha_{1}^{1} & \alpha_{d}^{1} \\
\vdots & \vdots \\
\alpha_{1}^{d} & \alpha_{d}^{d}
\end{bmatrix}
\begin{bmatrix}
\bar{x}_{1}^T \\
\vdots \\
\bar{x}_{n}^T
\end{bmatrix}
= 
\begin{bmatrix}
b_{1} \\
\vdots \\
b_{d}
\end{bmatrix}
\]

\( d \ll D \)
**PCA: Dimensional Reduction**

\[ A = X B \]

\[ n \times d \quad n \times D \quad D \times d \]

\( d \ll D \)
How to Compute Basis?

Set of basis vectors
How to Choose # of Basis Vectors?
 FACE DETECTION

Residual

\[ \| \bar{x} - BB^T \bar{x} \| \]

Face subspace
FACE DETECTION

Residual

\[ \| \bar{x} - \frac{\bar{x} \cdot b}{\|b\|^2} b \| \]
LIMITATION

Object distribution does not follow Guassian!