Statistical Linear Models: PCA

Reading: Eigenfaces – online paper
FP pgs. 505-512

Last Time

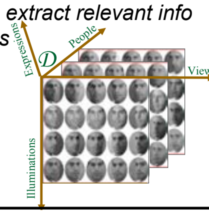
- Radiometry – Radiance and Irradiance
- Color Spaces
  - RGB, nRGB
  - HSV/I/L
  - YCrCb
- Pixel Statistics
  - Color Models
    - Non-parametric – Histogram Table Look-up
    - Parametric – Gaussian Model
  - Classification
    - Maximum Likelihood
  - Skin Color Models

PART I: 2D Vision

- Appearance-Based Methods
  - **Statistical Linear Models:**
    - Today ➔ **PCA**
    - ICA, FLD
    - Non-negative Matrix Factorization, Sparse Matrix Factorization
  - **Statistical Tensor Models:**
    - Multilinear PCA,
    - Multilinear ICA
  - Person and Activity Recognition

Statistical Modeling

- Statistics: the science of collecting, organizing, and interpreting *data.*
  - *Data collection.*
  - *Data analysis - organize & summarize data to bring out main features and clarify their underlying structure.*
  - *Inference and decision theory – extract relevant info from collected data and use it as a guide for further action.*



Data Collection

- **Population:** the entire group of individuals that we want information about.
- **Sample**: a *representative* part of the population that we actually examine in order to gather information.
- **Sample size**: number of observations/individuals in a sample.
- **Statistical inference:** to make an inference about a population based on the information contained in a sample.

Definitions

- **Individuals** (people or things) -- objects described by data.
- Individuals on which an experiment is being performed are known as **experimental units, subjects.**
- **Variables**--describe characteristics of an individual.
  - **Categorical variable** – places an individual into a category such as male/female.
  - **Quantitative variable** – measures some characteristic of the individual, such as height, or pixel values in an image.

Data Analysis

- **Experimental Units:** images
- **Observed Data:** pixel values in images are directly measurable but rarely of direct interest
- **Data Analysis:** extracts the relevant information

  bring out main features and clarify their underlying structure.

Variables

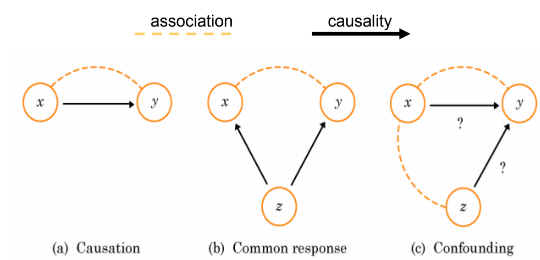
- **Response Variables** – are directly measurable, they measure the outcome of a study.
  - Pixels are response variables that are directly measurable from an image.
- **Explanatory Variables, Factors** – explain or cause changes in the response variable.
  - Pixel values change with scene geometry, illumination location, camera location which are known as the explanatory variables

Response vs. Explanatory Variables

- Pixels (response variables, directly measurable from data) change with changes in view and illumination, the explanatory variables (not directly measurable but of actual interest).

Explaining Association

An association between two variables x and y can reflect many types of relationships



The question of causation

- A strong relationship between two variables does not always mean that changes in one variable causes changes in the other.
- The relationship between two variables is often influenced by other variables which are lurking in the background.
- The best **evidence for causation** comes from *randomized comparative experiments.*
- The observed relationship between two variables may be due to direct causation, common response or confounding.
- ***Common response*** refers to the possibility that a change in a lurking variable is causing changes in both our explanatory variable and our response variable
- ***Confounding*** refers to the possibility that either the change in our explanatory variable is causing changes in the response variable OR that a change in a lurking variable is causing changes in the response variable.

Apperance Based Models

Models based on the appearance of 3D objects in ordinary images.

- Linear Models
  - PCA – Eigenfaces, EigenImages
  - FLD – Fisher Linear Discriminant Analysis
  - ICA – images are a linear combination of multiliple sources
- Multilinear Models
  - Relevant Tensor Math
  - MPCA – TensorFaces
  - MICA

Statistical Linear Models

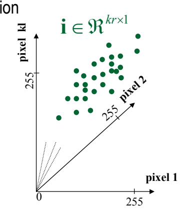
- Generative Models:
  - Second-order methods
    - faithful/accurate data representation - minimal reconstruction (mean-square) error
      - covariance
    - PCA – Principal Component Analysis
    - Factor Analysis
  - Higher Order Methods
    - meaningful representation
      - higher order statistics
    - ICA – Independent Component Analysis
- Descriminant Models:
  - FLD – Fisher Linear Descriminant Analysis

Linear Models

Images

- Image – experimental unit, multivariate function
- Pixel – response variable

$I \in \Re^{k \times r}$



- An image is a point in $\Re^{kr \times 1}$ dimensional space

Image Representation

pixel value

axis representing pixel 1

$$I = \begin{bmatrix} i_1 & i_2 & \dots & i_k \\ i_{k+1} & \ddots & & \\ \vdots & & \ddots & \\ i_{k(r-1)+1} & & & i_{kr} \end{bmatrix} \Rightarrow \mathbf{i} = \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_{kr} \end{bmatrix} = i_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + i_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdots + i_{kr} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Image Representation

$$I = \begin{bmatrix} i_1 & i_2 & \dots & i_k \\ i_{k+1} & \ddots & & \\ \vdots & & \ddots & \\ i_{k(r-1)+1} & & & i_{kr} \end{bmatrix} \Rightarrow \mathbf{i} = \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_{kr} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{bmatrix}}_{\text{Basis Matrix, } \mathbf{B}} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_{kr} \end{bmatrix}$$

vector of coefficients, **c**

$$\mathbf{i} = \mathbf{Bc}$$

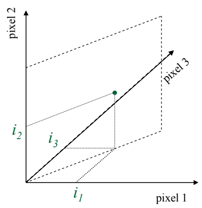
Representation

- Find a new basis matrix that results in a *compact representation useful for face detection/recognition*

Toy Example - Representation Heuristic

- Consider a set of images of N people under the same viewpoint and lighting
- Each image is made up of 3 pixels and pixel 1 has the same value as pixel 3 for all images

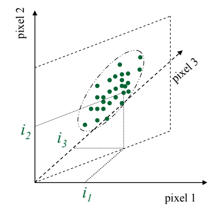
$$\mathbf{i}_n = \begin{bmatrix} i_{1n} \\ i_{2n} \\ i_{3n} \end{bmatrix} \quad \text{s.t.} \quad i_{1n} = i_{3n} \text{ and } 1 \le n \le N$$



Toy Example - Representation Heuristic

- Consider a set of images of N people under the same viewpoint and lighting
- Each image is made up of 3 pixels and pixel 1 has the same value as pixel 3 for all images

$$\mathbf{i}_n = \begin{bmatrix} i_{1n} \\ i_{2n} \\ i_{3n} \end{bmatrix} \quad \text{s.t.} \quad i_{1n} = i_{3n} \text{ and } 1 \le n \le N$$

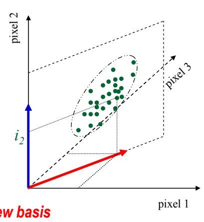


$$\mathbf{i}_n = i_{1n}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + i_{2n}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + i_{3n}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{Basis Matrix, } \mathbf{B}} \begin{bmatrix} i_{1n} \\ i_{2n} \\ i_{3n} \end{bmatrix}$$

Toy Example - Representation Heuristic

- Consider a set of images of N people under the same viewpoint and lighting
- Each image is made up of 3 pixels and pixel 1 has the same value as pixel 3 for all images

$$\mathbf{i}_n = \begin{bmatrix} i_{1n} \\ i_{2n} \\ i_{3n} \end{bmatrix} \quad \text{s.t.} \quad i_{1n} = i_{3n} \text{ and } 1 \le n \le N$$



new basis

$$\mathbf{i}_n = i_{1n}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + i_{2n}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + i_{3n}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{\text{Old Basis}} \begin{bmatrix} i_{1n} \\ i_{2n} \\ i_{3n} \end{bmatrix}$$

$$= i_{1n}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + i_{2n}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\text{New Basis Matrix, } \mathbf{B}} \begin{bmatrix} i_{1n} \\ i_{2n} \end{bmatrix} = \mathbf{Bc}_n$$

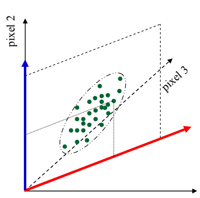
Toy Example-Recognition

- Highly correlated variables were combined
- The new basis (the new axis) are uncorrelated

Toy Example-Recognition

Solve for and store the coefficient matrix **C**:

$$\underbrace{\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{i}_1 & \mathbf{i}_2 & \cdots & \mathbf{i}_N \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}}_{\mathbf{D}\text{, data matrix}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_N \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}}_{\mathbf{C}\text{, coefficient matrix}}$$



$$\mathbf{C} = \mathbf{B}^{-1}\mathbf{D}$$

Given a new image, i_{new} :

$$\mathbf{c}_{new} = \mathbf{B}^{-1}\mathbf{i}_{new} = \begin{bmatrix} .5 & 0 & .5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} i_{1new} \\ i_{2new} \\ i_{3new} \end{bmatrix}$$

• Next, compare $\mathbf{c}_{new}$ a reduced dimensionality representation of i_{new} against all coefficient vectors $\mathbf{c}_n \quad 1 \le n \le N$

•One possible classifier: nearest-neighbor classifier

Nearest Neighbor Classifier

- Given an input *image representation* **y** (input is also called a probe; representation may be the image itself, **i**, or some transformation of the image, ex. **c**), the NN classifier will assign to **y** the label associated with the closest image in the training set.
- So if, it happens to be closest to another face it will be assigned L=1 (face), otherwise it will be assigned L=0 (nonface)
- *Euclidean distance:*

$$d = \left\| \mathbf{y}_L - \mathbf{y} \right\|^2 = \sum_{c=1}^{N} \left(y_{Lc} - y_c \right)^2$$

Principal Component Analysis: Eigenfaces

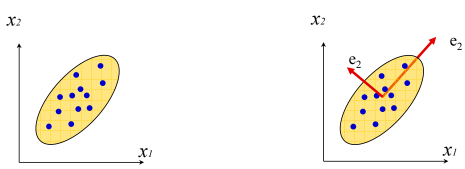
- Employs second order statistics to *compute in a principled way* a new basis matrix

The Principle Behind Principal Component Analysis[1]

- Also called: - Hotteling Transform[2] or the
  - Karhunen-Loeve Method [3].
- *Find an orthogonal coordinate system such that data is approximated best and the correlation between different axis is minimized.*

[1] I.T.Jolliffe; Principle Component Analysis; 1986
[2] R.C.Gonzalas, P.A.Wintz; Digital Image Processing; 1987
[3] K.Karhunen; Uber Lineare Methoden in der Wahrscheinlichkeits Rechnug; 1946
M.M.Loeve; Probability Theory; 1955

PCA: Theory



- Define a new origin as the mean of the data set
- Find the direction of maximum variance in the samples (e_1) and align it with the first axis ,
- Continue this process with orthogonal directions of decreasing variance, aligning each with the next axis
- Thus, we have a rotation which minimizes the covariance

PCA: Goal - Formally Stated

Problem formulation

- Input: $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_N]$ points in d-dimensional space
- Solve for: **B** dxm basis matrix (m≤d)

...

$$\mathbf{C} = [\mathbf{c}_1 \ \cdots \ \mathbf{c}_N] = \mathbf{B}^T [\mathbf{x}_1 \ \cdots \ \mathbf{x}_N]$$

and correlation is minimized
(or cov. is diagonalized)

Recall:

- Correlation: $$\mathrm{cor}(\mathbf{x},\mathbf{y}) = \frac{\mathrm{cov}(\mathbf{x},\mathbf{y})}{\sigma_\mathbf{x}\sigma_\mathbf{y}}$$
- Sample Covariance: $$\mathrm{cov}(\mathbf{x},\mathbf{y}) = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^T$$

The Sample Covariance Matrix

- Define the covariance (scatter) matrix of the input samples:

$$\mathbf{S}_T = \frac{1}{N-1}\sum_{n=1}^{N} (\mathbf{i}_n - \boldsymbol{\mu})(\mathbf{i}_n - \boldsymbol{\mu})^{\mathrm{T}}$$

(where μ is the sample mean)

$$\mathbf{S}_T = \frac{1}{N-1}(\mathbf{D}-\mathbf{M})(\mathbf{D}-\mathbf{M})^{\mathrm{T}} \quad \text{where} \quad \mathbf{M} = [\boldsymbol{\mu} \ \cdots \ \boldsymbol{\mu}]$$

$$\mathbf{S}_T = \frac{1}{N-1}\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{i}_1-\boldsymbol{\mu} & \mathbf{i}_2-\boldsymbol{\mu} & \cdots & \mathbf{i}_N-\boldsymbol{\mu} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & \mathbf{i}_1-\boldsymbol{\mu} & \rightarrow \\ \leftarrow & \mathbf{i}_2-\boldsymbol{\mu} & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{i}_N-\boldsymbol{\mu} & \rightarrow \end{bmatrix}$$

PCA: Some Properties of the Covariance/Scatter Matrix

- The covariance matrix $\mathbf{S}_T$ is symmetric
- The diagonal contains the variance of each parameter (i.e. element $\mathbf{S}_{T,ii}$ is the variance in the i'th direction).
- Each element $S_{T,ij}$ is the co-variance between the two directions i and j, represents the level of correlation (i.e. a value of zero indicates that the two dimensions are uncorrelated).

PCA: Goal Revisited

- Look for: **B**
- Such that:

$$[\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_N] = \mathbf{B}^T[\mathbf{i}_1 - \boldsymbol{\mu} \quad \cdots \quad \mathbf{i}_N - \boldsymbol{\mu}]$$

  - correlation is minimized ⟶ cov(**C**) is diagonal

Note that Cov(**C**) can be expressed via Cov(**D**) and **B** :

$$\mathbf{C}\mathbf{C}^T = \mathbf{B}^T(\mathbf{D}-\mathbf{M})(\mathbf{D}-\mathbf{M})^T\mathbf{B} = \mathbf{B}^T\mathbf{S}_T\mathbf{B}$$

Algebraic definition of PCs

- Given a sample of *N* observations on a vector of *d* variables

$$\mathbf{x} = [x_1 \quad \cdots \quad x_N]^T$$

- Define the k^{th} principal coefficient of the sample by the linear transformation

$$c_k = \mathbf{b}_k^T\mathbf{x} = \sum_{i=1}^{d} b_{ik}x_i$$

- where the vector $\mathbf{b}_k = [b_{1k} \quad \cdots \quad b_{dk}]^T$
- Chosen such that $\mathrm{var}[c_k]$ is maximal
- Subject to $\mathrm{cov}[c_k, c_l] = 0, \; k > l \geq 1$ and to $\mathbf{b}_k^T\mathbf{b}_k = 1$

Algebraic Derivation of b_1

To find b1 maximize var[c1] subject to $\mathbf{b}_k^T\mathbf{b}_k = 1$

- Maximize objective function:

$$L = \mathbf{b}_1^T\mathbf{S}\mathbf{b}_1 - \lambda(\mathbf{b}_1^T\mathbf{b}_1 - 1)$$

- Differentiate and set to 0:

$$\frac{\partial L}{\partial \mathbf{b}_1} = \mathbf{S}\mathbf{b}_1 - \lambda\mathbf{b}_1 = 0 \quad \Rightarrow \quad (\mathbf{S} - \lambda\mathbf{I})\mathbf{b}_1 = 0$$

- Therefore, $\mathbf{b}_1$ is an eigenvector of $\mathbf{S}$ corresponding to eigenvalue $\lambda = \lambda_1$

Algebraic Derivation of b_1

- We have maximized

$$\mathrm{var}[c_1] = \mathbf{b}_1^T\mathbf{S}\mathbf{b}_1 = \mathbf{b}_1^T\lambda_1\mathbf{b}_1 = \lambda_1$$

- So, λ_1 is the largest eigenvalue of S

Algebraic Derivation of b_2

To find the next principal direction maximize var[c_2] subject to cov[c_2,c_1]=0 and $\mathbf{b}_2^T\mathbf{b}_2 = 1$

- Maximize objective function:

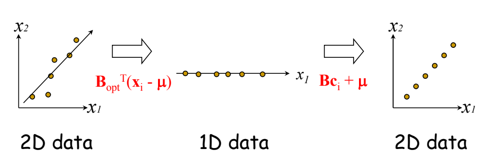
$$L = \mathbf{b}_2^T\mathbf{S}\mathbf{b}_2 - \lambda(\mathbf{b}_2^T\mathbf{b}_2 - 1) - \delta(\mathbf{b}_2^T\mathbf{b}_1 - 0)$$

- Differentiate and set to 0:

$$\frac{\partial L}{\partial \mathbf{b}_2} = \mathbf{S}\mathbf{b}_2 - \lambda\mathbf{b}_2 - \delta\mathbf{b}_1 = 0$$

Data Loss

- Sample points can be projected via the new m×d projection matrix $\mathbf{B}_{opt}$ and can still be reconstructed, but some information will be lost.



Data Loss (cont.)

- It can be shown that the mean square error between $\mathbf{x}_i$ and its reconstruction using only m principle eigenvectors is given by the expression :

$$\sum_{j=1}^{N} \lambda_j - \sum_{j=1}^{m} \lambda_j = \sum_{j=m+1}^{N} \lambda_j$$

Data Reduction: Theory

- Each eigenvalue represents the the total variance in its dimension.
- Throwing away the least significant eigenvectors in B_{opt} means throwing away the least significant variance information

Singular Value Decomposition

- For a square matrix

$$\mathbf{C}_x = \mathbf{U}\mathbf{C}_y\mathbf{U}^T$$

  - Remember that: $\mathbf{C}_x = (\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T \equiv \mathbf{D}\mathbf{D}^T$
  - then: $\mathbf{D} = \mathbf{U}\tilde{\mathbf{C}}_y\mathbf{V}^T$
  - where $\mathbf{D} \in IR^{d \times N}$ is non-square

SVD: definition

Any real $\mathbf{D}$ matrix $d \times N$

Can be decomposed: $\mathbf{D} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$

where

$$\mathbf{U}^T\mathbf{U} = \mathbf{V}\mathbf{V}^T = \mathbf{I}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_q \end{bmatrix}_{q \times q} \qquad q = \min(N, d)$$

The λ 's are called singular values

EVD vs. SVD

- And: Square $\mathbf{C}_x = \mathbf{U}\mathbf{C}_y\mathbf{U}^T$ Non-square $\mathbf{D} = \mathbf{U}\tilde{\mathbf{C}}_y\mathbf{V}^T$

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1^2 & & & \\ & \ddots & & \\ & & \lambda_N^2 & \\ & & & \ddots \\ & & & & 0 \end{bmatrix}_{d \times d} \qquad \tilde{\mathbf{C}}_y = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}_{N \times N}$$

Data Reduction and SVD

- Set to 0 redundant singular values

$$\mathbf{D}' = \mathbf{U}\begin{bmatrix} \lambda_1^2 & & & & \\ & \ddots & & & \\ & & \lambda_m^2 & & \\ & & & \ddots & \\ & & & & 0 \end{bmatrix}\mathbf{V}^T$$

- Given the data dimension is m we can solve for the first m vectors of $\mathbf{U}$

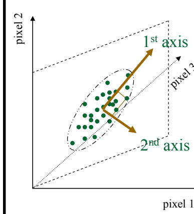
(No need to find all of them)

PCA : Conclusion

- A multi-variant analysis method.
- Finds a more "natural" coordinate system for the sample data.
- Allows for data to be removed with minimum loss in reconstruction ability.

PCA-Dimensionality Reduction

- Consider a set of images, & each image is made up of 3 pixels and pixel 1 has the same value as pixel 3 for all images $\mathbf{i}_n = [i_{1n} \;\; i_{2n} \;\; i_{3n}]^T$ s.t. $i_{1n} = i_{3n}$ and $1 \le n \le N$
- PCA chooses axis in the direction of highest variability of the data, maximum scatter



$$\underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{i}_1 & \mathbf{i}_2 & \cdots & \mathbf{i}_N \\ | & | & & | \end{bmatrix}}_{\text{data matrix, } \mathbf{D}} = \mathbf{B} \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_N \\ | & | & & | \end{bmatrix}$$

- Each image $\mathbf{i}_n$ is now represented by a vector of coefficients $\mathbf{c}_n$ in a reduced dimensionality space.

$\mathbf{D} = \mathbf{USV}^T$ (svd of D) ⇨ set $\mathbf{B} = \mathbf{U}$

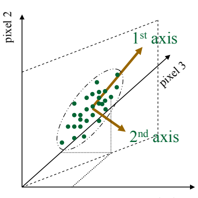
- B minimize the following function

$$E = \mathbf{B}^T \mathbf{S}_T \mathbf{B} \text{ such that } \mathbf{B}^T\mathbf{B} = \text{Identity}$$

PCA for Recognition

- Consider the set of images $\mathbf{i}_n = [i_{1n} \;\; i_{2n} \;\; i_{3n}]^T$ s.t. $i_{1n} = i_{3n}$ and $1 \le n \le N$
- PCA chooses axis in the direction of highest variability of the data
- Given a new image, $\mathbf{i}_{new}$, compute the vector of coefficients $\mathbf{c}_{new}$ associated with the new basis, **B**

$$\mathbf{c}_{new} = \mathbf{B}^T \mathbf{i}_{new} \qquad \mathbf{B}^{-1} = \mathbf{B}^T$$



- Next, compare $\mathbf{c}_{new}$ a reduced dimensionality representation of i_{new} against all coefficient vectors $\mathbf{c}_n$ $1 \le n \le N$
- One possible classifier: nearest-neighbor classifier

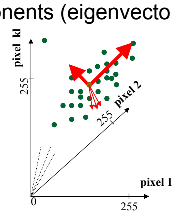
Data and Eigenfaces

- Data is composed of 28 faces photographed under same lighting and viewing conditions

- Each image below is a column vector in the basis matrix B

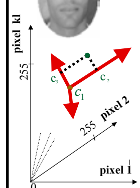
Eigenimages

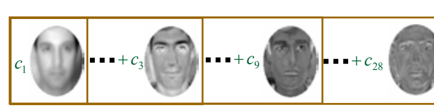
- Principal components (eigenvectors) of image ensemble



- Eigenvectors are typically computed using the Singular Value Decomposition (SVD) algorithm

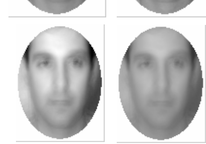
Linear Representation:





$$\mathbf{d}_{ij} = \mathbf{U}\mathbf{c}_{ij}$$

Running Sum: 1 term 3 terms

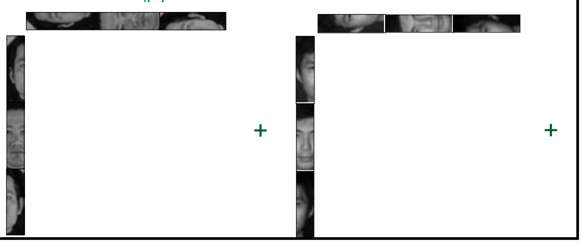


The Covariance Matrix

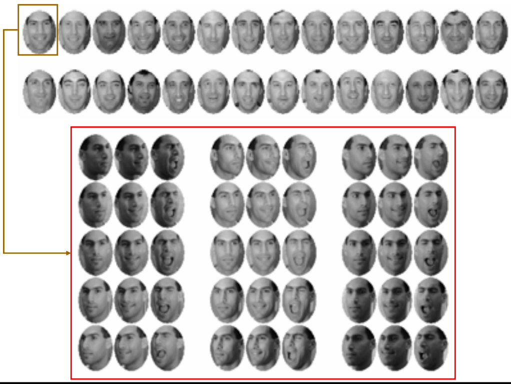
- Define the covariance (scatter) matrix of the input samples:

$$\mathbf{S}_T = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{i}_n - \boldsymbol{\mu})(\mathbf{i}_n - \boldsymbol{\mu})^{\mathrm{T}}$$

(where μ is the sample mean)



PIE Database (Weizmann)



EigenImages-Basis Vectors

- Each image bellow is a column vector in the basis matrix B
- PCA encodes encodes the variability across images without distinguishing between variability in people, viewpoints and illumination

PCA Classifier

- Distance to Face Subspace:

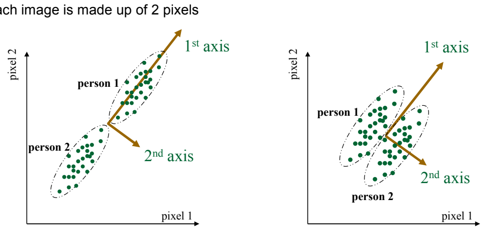
$$d_f(\mathbf{y}) = \left\| \mathbf{y} - \mathbf{U}_f \mathbf{U}_f^T \mathbf{y} \right\|^2$$

- Likelihood ratio (LR) test to classify a probe **y** as face or nonface. Intuitively, we expect $d_n(\mathbf{y}) > d_f(\mathbf{y})$ to suggest that **y** is a face.
- The LR for PCA is defined as:

$$\Delta_d = \frac{d_n(\mathbf{y})}{d_f(\mathbf{y})} \underset{L=0}{\overset{L=1}{\gtrless}} \eta$$

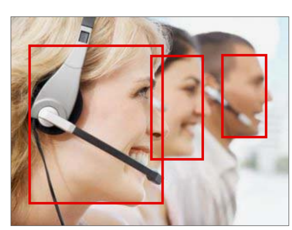
PCA for Recognition - EigenImages

- Consider a set of images of 2 people under fixed viewpoint & N lighting condition
- Each image is made up of 2 pixels



- Reduce dimensionality by throwing away the axis along which the data varies the least
- The coefficient vector associated with the 1st basis vector is used for classifiction
- Possible classifier: Mahalanobis distance
- Each image is represented by one coefficient vector
- Each person is displayed in N images and therefore has N coefficient vectors

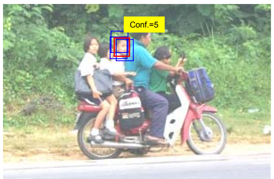
Face Detection/Recognition



location and scale in an image

Face Localization

- Scan and classify using image windows at different positions and scales



- Cluster detections in the space-scale space
- Assign cluster size to the detection confidence

Face Detection and Localization