Lecture 3: Face Detection

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

Handouts: Course Description
PS1 Assigned

Black or White Video

- ✓ Face Detection
- Face Localization
- Segmentation
- Face Tracking
- Facial features localization
- Facial features tracking
- Morphing

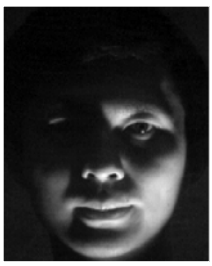
www.youtube.com/watch?v=ZI9OYMRwN1Q

- "…If I look at your face I immediately recognize that I have seen it before. …Yet there is no machine which, with that speed, can take a picture of a face and say even that it is a man; and much less that it is the same man that you showed it before—unless it is exactly the same picture. If the face is changed; if I am closer to the face; if I am further from the face; if the light changes—I recognize it anyway. Now, this little computer I carry in my head is easily able to do that. The computers that we build are not able to do that. …"

Richard P. Feynman, Dec. 29, 1959
There's Plenty of Room at the Bottom
An Invitation to Enter a New Field of Physics

Why is Face Detection Difficult?

- Severe illumination change

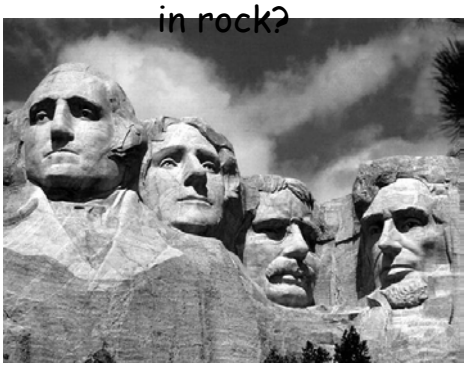


Automated Face Detection Why is it Difficult?

- Varying viewpoint, illumination, etc.

Face Detection

Coincidental appearance of faces in rock?

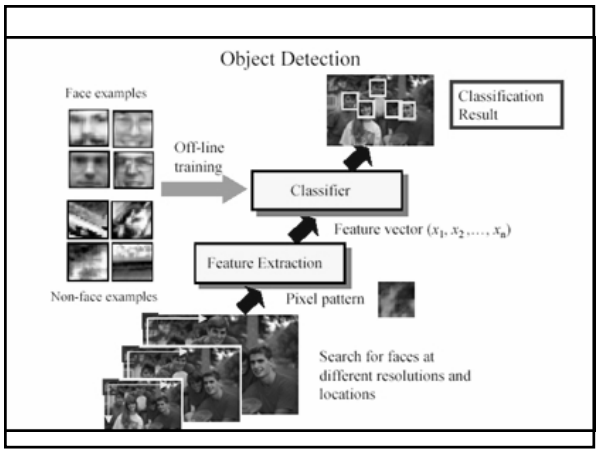


http://bensguide.gpo.gov/3-5/symbols/print/mountrushmore.html

Coincidental appearance of face profile in rock?

http://www.cs.dartmouth.edu/whites/old_man.html

Face Detection



Nearest Neighbor Clasiffier

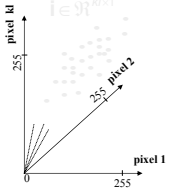
- *Euclidean distance:*

$$d = \|\mathbf{y}_1 - \mathbf{y}_2\|^2 = \sum_{c=1}^{N} (y_{1c} - y_{2c})^2$$

- *Given an input image **y** (also called a probe), 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)*

Linear Models

Images



- An image is a point in dimensional space

Image Representation

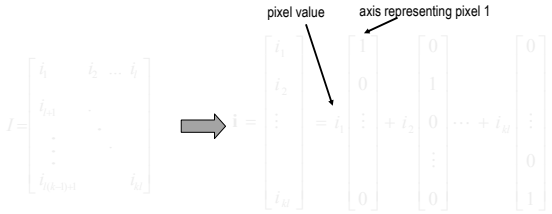
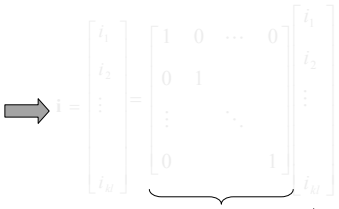


Image Representation

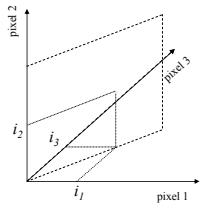


Representation

- Find a new basis matrix that results in a compact representation

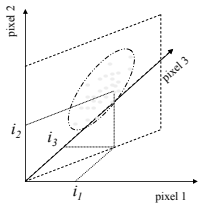
Toy Example - Representation Heuristic

- 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



Toy Example - Representation Heuristic

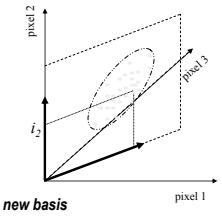
- Consider a set of images of N people under the same viewpoint and lighting
- Each image is made up of 3 pixels and pixel 1 has the same value as pixel 3 for all images



Basis Matrix, **B**

Toy Example - Representation Heuristic

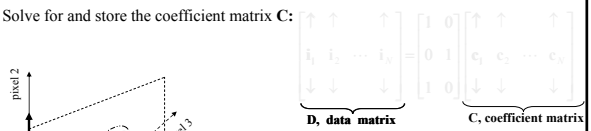
- Consider a set of images of N people under the same viewpoint and lighting
- Each image is made up of 3 pixels and pixel 1 has the same value as pixel 3 for all images



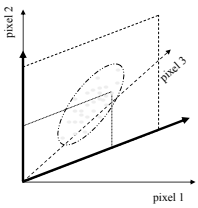
New Basis Matrix, **B**

Toy Example-Recognition

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



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



Given a new image, i_{new} :

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

- 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

Principal Component Analysis: Eigenfaces

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

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).

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.*

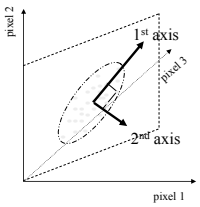
1 I.T.Jolliffe; Principle Component Analysis; 1986
2 R.C.Gonzalas, P.A.Wintz; Digital Image Processing; 1987
3 K.Karhunen; Uber Lineare Methoden in der Wahrscheinlichkeits Rechnug; 1946
M.M.Loeve; Probability Theory; 1955

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 (y_1),
- 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-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 $i_n = [i_{n1} \;\; i_{n2} \;\; i_{n3}] \quad \text{s.t.} \quad i_{n1} = i_{n3}$ and $1 \le n \le N$
- PCA chooses axis in the direction of highest variability of the data, maximum scatter



$$\underbrace{\begin{bmatrix} | & | & & | \\ i_1 & i_2 & \cdots & i_N \\ | & | & & | \end{bmatrix}}_{\text{data matrix, D}} = B \begin{bmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_N \\ | & | & & | \end{bmatrix}$$

- Each image i is now represented by a vector of coefficients c_n in a reduced dimensionality space.

$D = USV^T$ (svd of D) ⇒ set $B = U$

- B minimize the following function

$E = B^T S_T B$ such that $B^T B = I$ identity

The Covariance Matrix

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

$$S_T = \sum_{n=1}^{N} (i_n - \mu)(i_n - \mu)^T$$

(where μ is the sample mean)

$$S_T = (D - M)(D - M)^T \quad \text{where} \quad M = [\mu \;\; \cdots \;\; \mu]$$

$$S_T = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ i_1 - \mu & i_2 - \mu & \cdots & i_N - \mu \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & i_1 - \mu & \rightarrow \\ \leftarrow & i_2 - \mu & \rightarrow \\ & \vdots & \\ \leftarrow & i_N - \mu & \rightarrow \end{bmatrix}$$

PCA: Some Properties of the Covariance/Scatter Matrix

- The 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).

SVD of a Matrix

Scatter of matrix: $S_T = (D - M)(D - M)^T$

$(D - M) = U \Sigma V^T$ by svd of $(D - M)$ ⇒ set $B = U$

$(D - M)(D - M)^T = U \Sigma^2 U^T$ (svd of S_T) ⇒ set $B = U$

PCA: Goal Revisited

- Look for: - B
- Such that:
  - $[c_1 \ldots c_N] = B^T [i_1 \ldots i_N]$
  - correlation is mininmized ⇒ Cov(C) is diagonal

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

$$CC^T = B^T (D - M)(D - M)^T B$$
$$= B^T S_T B$$

Selecting the Optimal **B**

How do we find such **B** ?

$$(\mathbf{D}-\boldsymbol{\mu})(\mathbf{D}-\boldsymbol{\mu})^T\mathbf{b}_i = \lambda_i \mathbf{b}_i$$

$$\mathbf{S}_T\mathbf{B} = \Lambda\mathbf{B}$$

$\mathbf{B}_{opt}$ contains the eigenvectors of the covariance of D

$$\mathbf{B}_{opt} = [\mathbf{b}_1|\ldots|\mathbf{b}_d]$$

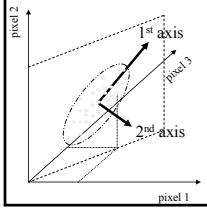
Data Reduction: Theory

- 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

PCA for Recognition

- Consider the set of images $\mathbf{i}_n = [i_{1n} \;\; i_{2n} \;\; i_{3n}]^T$ s.t. $i_{in} = i_{jn}$ 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**

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



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

•One possible classifier: nearest-neighbor classifier

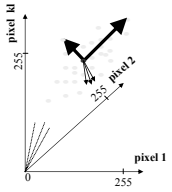
Data and Eigenfaces

- 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

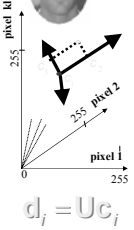
Eigenimages

- Principal components (eigenvectors) of image ensemble

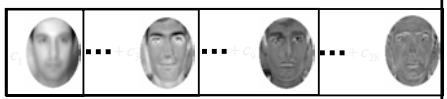


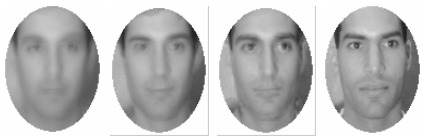
- Eigenvectors are typically computed using the Singular Value Decomposition (SVD) algorithm

Linear Representation:



$$\mathbf{d}_i = \mathbf{U}\mathbf{c}_i$$





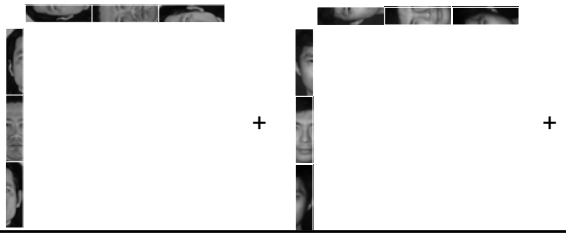
Running Sum: 1 term 3 terms 9 terms 28 terms

The Covariance Matrix

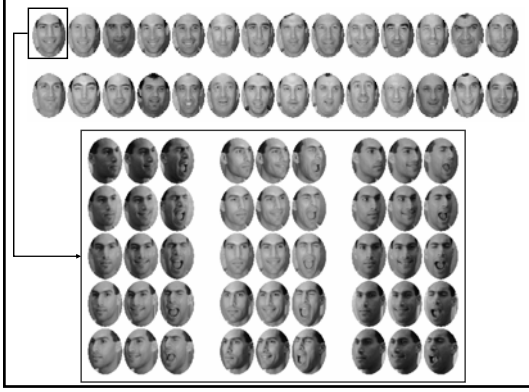
- Define the covariance (scatter) matrix of the input samples:

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

(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:

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