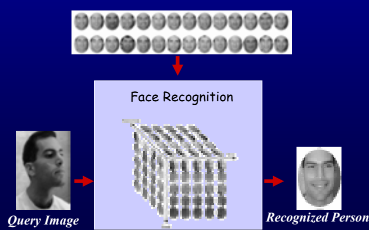
Face Recognition



- "…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 Recognition Difficult?

- Severe illumination change

Automated Face Recognition Why is it Difficult?

- Varying viewpoint, illumination, etc.

Applications of Face Biometrics

- financial transactions
- check-in or boarding planes
- crossing borders
- casting votes
- security or surveillance
- identity fraud
- criminal justice & law enforcement
- access to facilities, databases or privileged information, etc

Face Recognition

- Definition:
  - Given a database of labeled facial images
  - Recognize an individual from an image formed from new and varying conditions (pose, expression, lighting etc.)
- Sub-Problems:
  - Representation:
    - How do we represent images of faces?
    - What information do we store?
  - Classification:
    - How do we compare stored information to a new sample?
  - Search

Representation

Goal:

Compact, descriptive object representation for recognition

Representations:

- Shape Representation:
  - Generalized cylinders, Superquadrics …
- Apperace Based Representation for Recognition:
  - Ordinary images
  - statistics

Today: Apperance Based Recognition

Appearance based recognition refers to the recognition of 3D objects from ordinary images.

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

Linear Algebra

- The algebra of vectors and matrices
  - Traditionally of great value in image science
    - Fourier transform
    - Karhunen-Loeve transform (PCA)
    - Eigenfaces
  - Linear methods model:
    - Linear operators over a vector space
    - Single-factor linear variation in image formation
    - The linear combination of multiple sources (ICA)

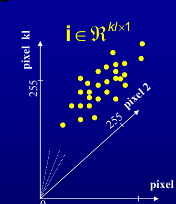
Multilinear Algebra

- The algebra of higher-order (>2) tensors
  - A unifying mathematical framework for image science
  - Natural images result from the interaction of multiple factors related to
    - scene geometry
    - Illumination
    - Imaging
  - Multilinear algebra can explicitly represent multiple factors
    - Multilinear operators over a **set** of vector spaces
  - Multilinear algebra subsumes linear algebra as a special case

Linear Models

Images

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



$\mathbf{i} \in \Re^{kl \times 1}$

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

Image Representation

pixel value   axis representing pixel 1

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

Image Representation

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

vector of coefficients, **c**

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

© 2002 by M. Alex O. Vasilescu

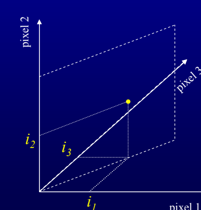
Representation

- 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

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

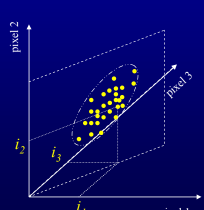


© 2002 by M. Alex O. Vasilescu

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

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



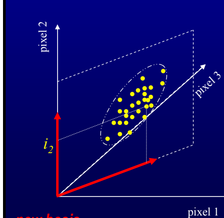
$$\mathbf{i}_n = i_{1n}\begin{bmatrix}1\\0\\0\end{bmatrix} + i_{2n}\begin{bmatrix}0\\1\\0\end{bmatrix} + i_{3n}\begin{bmatrix}0\\0\\1\end{bmatrix} = \underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{\text{Basis Matrix, } \mathbf{B}}\begin{bmatrix} i_{1n}\\ i_{2n}\\ i_{3n}\end{bmatrix}$$

© 2002 by M. Alex O. Vasilescu

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

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



$$\mathbf{i}_n = i_{1n}\begin{bmatrix}1\\0\\0\end{bmatrix} + i_{2n}\begin{bmatrix}0\\1\\0\end{bmatrix} + i_{3n}\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix} i_{1n}\\ i_{2n}\\ i_{3n}\end{bmatrix}$$

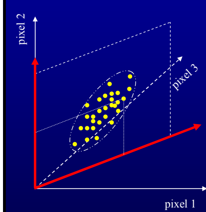
$$= i_{1n}\begin{bmatrix}1\\0\\1\end{bmatrix} + i_{2n}\begin{bmatrix}0\\1\\0\end{bmatrix} = \underbrace{\begin{bmatrix}1&0\\0&1\\1&0\end{bmatrix}}_{\text{New Basis Matrix, } \mathbf{B}}\begin{bmatrix} i_{1n}\\ i_{2n}\end{bmatrix} = \mathbf{Bc}_n$$

© 2002 by M. Alex O. Vasilescu

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} :

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

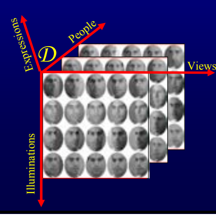
- Next, compare $\mathbf{c}_{new}$ a reduced dimensionality representation of i_{new} against all coefficient vectors $\mathbf{c}_n \quad 1 \le n \le N$
- One possible classifier: nearest-neighbor classifier

Principal Component Analysis: Eigenfaces

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

Statistical Learning

- 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

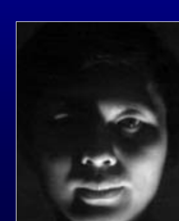
- **Population:** the entire group of individuals that we want information about.
- **Sample**: a *representative* part of the population that we actually examine in order to gather information.
- **Sample size**: number of observations/individuals in a sample.
- **Statistical inference:** to make an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



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]

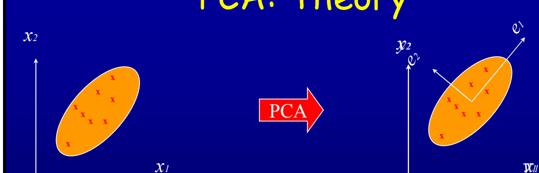
- Also called:- Hotteling Transform[2] or the
  - Karhunen Loeve Method [3].
- *Find an orthogonal coordinate system such that data is approximated best and the correlation between different axis is minimized.*

[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 for Recognition Eigenimages

- PCA / Eigenimages:
  - Sirovich & Kirby 1987
    "Low Dimensional Procedure for the Characterization of Human Faces"
  - Turk & Pentland 1991
    "Face Recognition Using Eigenfaces"
  - Murase & Nayar 1995
    "Visual learning and recognition of 3D objects from appearance"

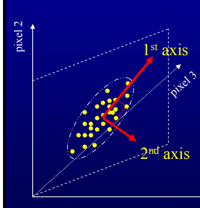
PCA: Theory



- 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 $\mathbf{i}_n = [i_{1n} \quad i_{2n} \quad i_{3n}]^T$ 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



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

- 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

$$E = \mathbf{B}^T \mathbf{S}_T \mathbf{B} \quad \text{such that} \quad \mathbf{B}^T\mathbf{B} = \mathbf{I}\text{dentity}$$

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)

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

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

PCA: Some Properties of the Covariance/Scatter Matrix

- The 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: $\mathbf{S}_T = (\mathbf{D} - \mathbf{M})(\mathbf{D} - \mathbf{M})^T$

$(\mathbf{D} - \mathbf{M}) = \mathbf{U}\Sigma\mathbf{V}^T$ by svd of $(\mathbf{D}-\mathbf{M})$ ➡ set $\mathbf{B} = \mathbf{U}$

$(\mathbf{D} - \mathbf{M})(\mathbf{D} - \mathbf{M})^T = \mathbf{U}\Sigma^2\mathbf{U}^T$ (svd of $\mathbf{S}_T$) ➡ set $\mathbf{B} = \mathbf{U}$

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 :

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

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

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

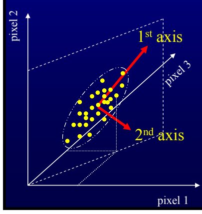
Data Reduction: Theory

- 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. $i_{1n} = i_{3n}$ 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, $\mathbf{B}$

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



- 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

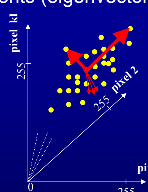
Data and Eigenfaces

- 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

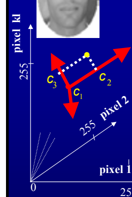
- Principal components (eigenvectors) of image ensemble

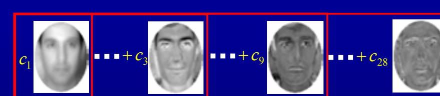


.....

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

Linear Representation:





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

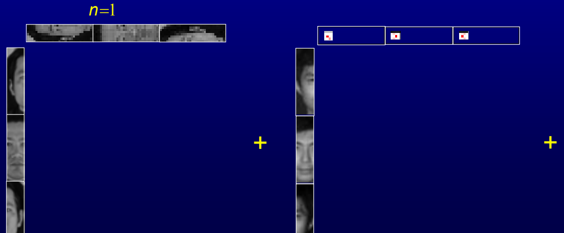
Running Sum: 1 term 3 terms 9 terms 28 terms

The Covariance Matrix

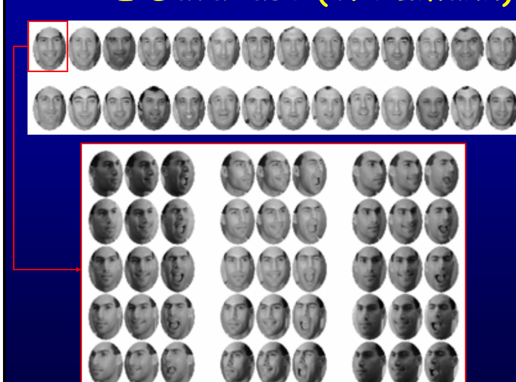
- Define the covariance (scatter) matrix of the input samples:

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

(where μ is the sample mean)



PIE Database (Weizmann)

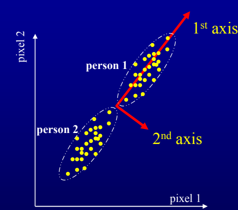


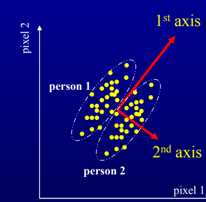
EigenImages-Basis Vectors

- 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 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