

Recognition II

Introduction to Computer Vision
CSE 152
Lecture 18

CSE152, Winter 2013

Intro Computer Vision

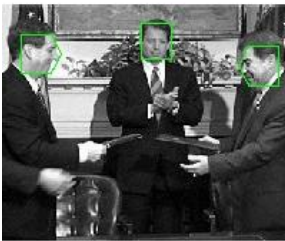
Announcements

- HW 3 due Today.
- HW4 (face recognition) has been posted.

CSE152, Winter 2013

Intro Computer Vision

Recognition

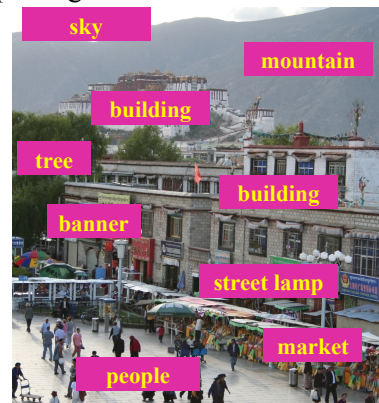


Given a database of objects and an image determine what, if any of the objects are present in the image.

CSE152, Winter 2013

Intro Computer Vision

Image parsing



CSE152, Winter 2013

Intro Computer Vision

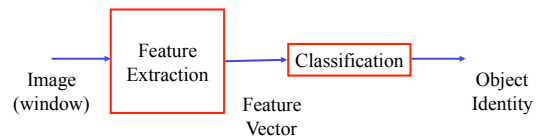
Recognition Challenges

- Within-class variability
 - Different objects within the class have different shapes or different material characteristics
 - Deformable
 - Articulated
 - Compositional
- Pose variability:
 - 2-D Image transformation (translation, rotation, scale)
 - 3-D Pose Variability (perspective, orthographic projection)
- Lighting
 - Direction (multiple sources & type)
 - Color
 - Shadows
- Occlusion – partial
- Clutter in background -> false positives

CSE152, Winter 2013

Intro Computer Vision

Sketch of a Pattern Recognition Architecture



CSE152, Winter 2013

Intro Computer Vision

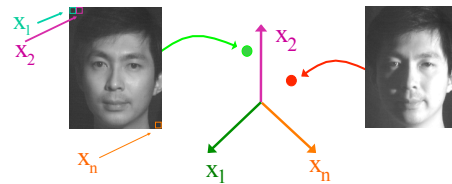
Sliding window approaches



CSE152, Winter 2013

Intro Computer Vision

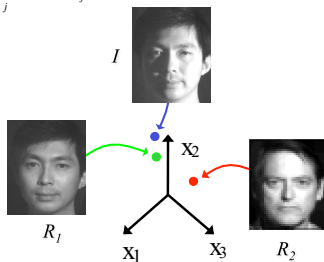
The Space of Images



- We will treat an d -pixel image as a point in an d -dimensional space, $\mathbf{x} \in \mathbb{R}^d$.
- Each pixel value is a coordinate of $\mathbf{x}$.

Nearest Neighbor Classifier

$\{R_j\}$ are set of training images.
 $ID = \arg \min_j \text{dist}(R_j, I)$



CSE152, Winter 2013

Intro Computer Vision

An idea:

Represent the set of images as a linear subspace

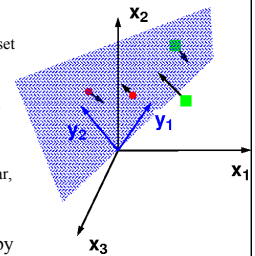
What is a linear subspace?

Let V be a vector space and let W be a subset of V . Then W is a subspace iff:

1. The zero vector, $\mathbf{0}$, is in W .
2. If $\mathbf{u}$ and $\mathbf{v}$ are elements of W , then any linear combination of $\mathbf{u}$ and $\mathbf{v}$ is an element of W ; $a\mathbf{u} + b\mathbf{v} \in W$.
3. If $\mathbf{u}$ is an element of W and c is a scalar, then the scalar product $c\mathbf{u} \in W$.

A k -dimensional subspace is spanned by k linearly independent vectors

It is spanned by a k -dimensional orthogonal basis.



Example: A 2-D linear subspace of $\mathbb{R}^3$ spanned by y_1 and y_2 .

CSE152

Computer Vision I

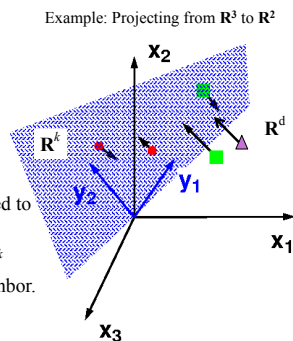
Linear Subspaces & Linear Projection

- A d -pixel image $\mathbf{x} \in \mathbb{R}^d$ can be projected to a low-dimensional feature space $\mathbf{y} \in \mathbb{R}^k$ by

$$\mathbf{y} = W\mathbf{x}$$

where W is an k by d matrix.

- Each training image is projected to the subspace
- Recognition is performed in $\mathbb{R}^k$ using, for example, nearest neighbor.
- How do we choose a good W ?



CSE152, Winter 2013

Intro Computer Vision

Linear Subspaces & Recognition

1. Eigenfaces: Approximate all training images as a single linear subspace
2. Distance to subspace 1: Represent lighting variation w/o shadowing for a single individual as a 3-D linear subspace. n individuals are modeled as n 3-D linear subspaces..
3. Fisherfaces: Project all training images to a single subspace that enhances discriminability

CSE152, Winter 2013

Intro Computer Vision

Eigenfaces: Principal Component Analysis (PCA)

Assume we have a set of n feature vectors x_i ($i = 1, \dots, n$) in $\mathbb{R}^d$. Write

$$\mu = \frac{1}{n} \sum_i x_i$$

$$\Sigma = \frac{1}{n-1} \sum_i (x_i - \mu)(x_i - \mu)^T$$

The unit eigenvectors of Σ — which we write as $v_1, v_2, \dots, v_d$, where the order is given by the size of the eigenvalue and v_1 has the largest eigenvalue — give a set of features with the following properties:

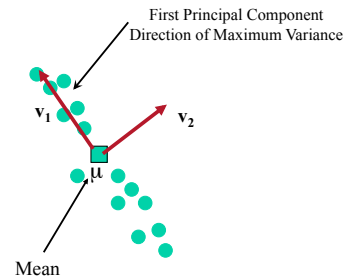
- They are independent.
- Projection onto the basis $\{v_1, \dots, v_k\}$ gives the k -dimensional set of linear features that preserves the most variance.

Algorithm 22.5: Principal components analysis identifies a collection of linear features that are independent, and capture as much variance as possible from a dataset.

CSE152, Winter 2013

Intro Computer Vision

PCA Example



CSE152, Winter 2013

Intro Computer Vision

Eigenfaces

- Modeling
 1. Given a collection of n training images x_i , represent each one as d -dimensional column vector
 2. Compute the mean image and covariance matrix.
 3. Compute k Eigenvectors of the covariance matrix corresponding to the k largest Eigenvalues and form matrix $W^T = [v_1, v_2, \dots, v_k]$ (Or perform using SVD!!) (note that the Eigenvectors are images)
 4. Project the training images to the k -dimensional Eigenspace. $y_i = Wx_i$
- Recognition
 1. Given a test image x , project the vectorized image to the Eigenspace by $y = Wx$
 2. Perform classification to the projected training images.

CSE152, Winter 2013

Intro Computer Vision

Why is W a good projection?

- The linear subspace spanned by W maximizes the variance (i.e., the spread) of the projected data.
- W spans a subspace that is the best approximation to the data in a least squares sense. E.g., W is the subspace that minimizes the the sum of the squared distances from each datapoint to the the subspace.

CSE152, Winter 2013

Intro Computer Vision

Eigenfaces: Training Images



[Turk, Pentland 91]

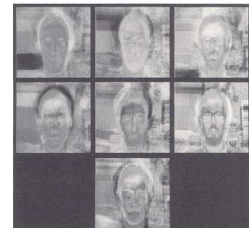
CSE152, Winter 2013

Intro Computer Vision

Eigenfaces



Mean Image

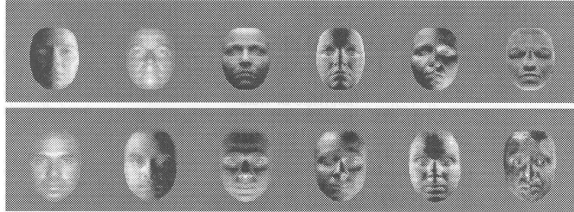


Basis Images

CSE152, Winter 2013

Intro Computer Vision

Basis Images for Variable Lighting



CSE152, Winter 2013

Intro Computer Vision

Distance to Linear Subspace

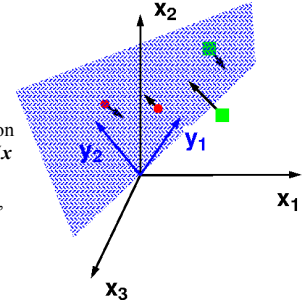
- An n -pixel image $x \in \mathbb{R}^d$ can be projected to a low-dimensional feature space $y \in \mathbb{R}^k$ by

$$y = Wx$$

- From $y \in \mathbb{R}^k$, the reconstruction of the point in $\mathbb{R}^d$ is $W^T y = W^T W x$

- The error of the reconstruction, or the distance from x to the subspace spanned by W is:

$$\|x - W^T W x\|$$



CSE152, Winter 2013

Intro Computer Vision

Distance to Affine Subspace (i.e., Distance to Face Space)

- Represented by mean vector μ and basis images W

- An n -pixel image $x \in \mathbb{R}^d$ can be projected to a low-dimensional feature space $y \in \mathbb{R}^k$ by

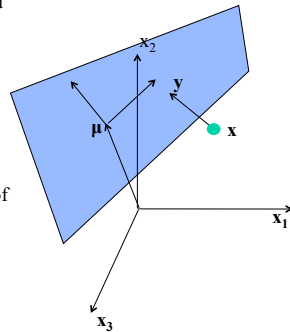
$$y = W(x - \mu)$$

- From $y \in \mathbb{R}^k$, the reconstruction of the point in $\mathbb{R}^d$ is

$$W^T y + \mu = W^T W(x - \mu) + \mu$$

- The error of the reconstruction, or the distance from x to the affine is:

$$\|x - W^T W(x - \mu) - \mu\| = \|(I - W^T W)(x - \mu)\|$$



CSE152, Winter 2013

Intro Computer Vision

An important footnote:

We don't really implement PCA by constructing a covariance matrix!

Why?

- How big is Σ ?
 - d by d where d is the number of pixels in an image!!
- You only need the first k Eigenvectors

Assume we have a set of n feature vectors x_i ($i = 1, \dots, n$) in $\mathbb{R}^d$. Write

$$\mu = \frac{1}{n} \sum_i x_i$$

$$\Sigma = \frac{1}{n-1} \sum_i (x_i - \mu)(x_i - \mu)^T$$

The unit eigenvectors of Σ — which we write as $v_1, v_2, \dots, v_d$, where the order is given by the size of the eigenvalue and v_1 has the largest eigenvalue — give a set of features with the following properties:

CSE152, Winter 2013

Intro Computer Vision

Singular Value Decomposition

- Any m by n matrix A may be factored such that

$$A = U \Sigma V^T$$

$$[m \times n] = [m \times m][m \times n][n \times n]$$

- U : m by m , orthogonal matrix
 - Columns of U are the eigenvectors of AA^T
- V : n by n , orthogonal matrix,
 - columns are the eigenvectors of $A^T A$
- Σ : m by n , diagonal with non-negative entries ($\sigma_1, \sigma_2, \dots, \sigma_s$) with $s = \min(m, n)$ are called the singular values. SVD algorithm produces sorted singular values: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s$

Important property

- Singular values are the square roots of Eigenvalues of both AA^T and $A^T A$ & Columns of U are corresponding Eigenvectors!!

CSE152, Winter 2013

Intro Computer Vision

SVD Properties

- In Matlab $[u \ s \ v] = \text{svd}(A)$, and you can verify that: $A = U \Sigma V^T$
- $r = \text{Rank}(A) = \#$ of non-zero singular values.
- U, V give an orthonormal bases for the subspaces of A :
 - 1st r columns of U : Column space of A
 - Last $m - r$ columns of U : Left nullspace of A
 - 1st r columns of V : Row space of A
 - Last $n - r$ columns of V : Nullspace of A
- For some d where $d \leq r$, the first d column of U provide the best d -dimensional basis for columns of A in least squares sense.

CSE152, Winter 2013

Intro Computer Vision

Performing PCA with SVD

- Singular values of A are the square roots of eigenvalues of both AA^T and $A^T A$ & Columns of U are corresponding Eigenvectors
- And $\sum_{i=1}^n a_i a_i^T = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^T = AA^T$
- Covariance matrix is:

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (\tilde{x}_i - \bar{\mu})(\tilde{x}_i - \bar{\mu})^T$$

- So, ignoring $1/n$ subtract mean image μ from each input image, create a $d \times n$ data matrix, and perform thin SVD on the data matrix. $D = [x_1 - \mu \mid x_2 - \mu \mid \dots \mid x_n - \mu]$

CSE152, Winter 2013

Intro Computer Vision

Thin SVD

- Any m by n matrix A may be factored such that

$$A = U \Sigma V^T$$

$$[m \times n] = [m \times m][m \times n][n \times n]$$

- If $m > n$, then one can view Σ as: (i.e., more pixels than images)

$$\begin{bmatrix} \Sigma' \\ 0 \end{bmatrix}$$

- Where $\Sigma' = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_s)$ with $s = \min(m, n)$, and lower matrix is $(n-m \text{ by } m)$ of zeros.

- Alternatively, you can write:

$$A = U' \Sigma' V^T$$

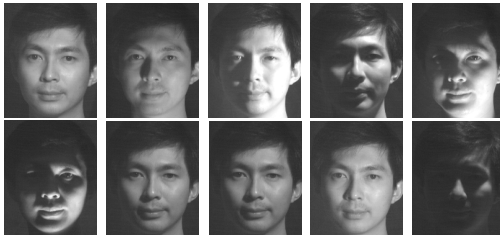
This is what you should use!!

- In Matlab, thin SVD is: $[U \Sigma V] = \text{svds}(A)$

CSE152, Winter 2013

Intro Computer Vision

Illumination Variability



- How does the set of images of an object in fixed pose vary as lighting changes?
- How can we recognize people across all lighting conditions without having to see the person every way?

CSE152, Winter 2013

Intro Computer Vision

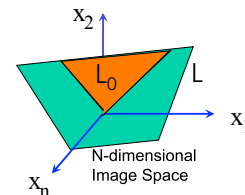
3-D Linear subspace

The set of images of a Lambertian surface with no shadowing is a subset of 3-D linear subspace.

[Moses 93], [Nayar, Murase 96], [Shashua 97]

$$L = \{x \mid x = Bs, \forall s \in \mathbb{R}^3\}$$

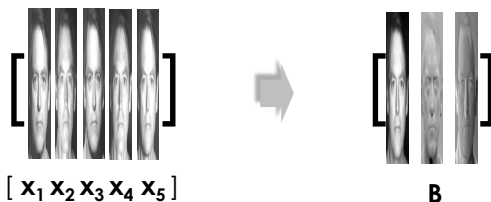
where B is a n by 3



CSE152, Winter 2013

Intro Computer Vision

How do you construct the 3-D subspace?



With more than three images, perform least squares estimate of B using PCA or Singular Value Decomposition (SVD)

CSE152, Winter 2013

Intro Computer Vision

Face Basis

Original Images

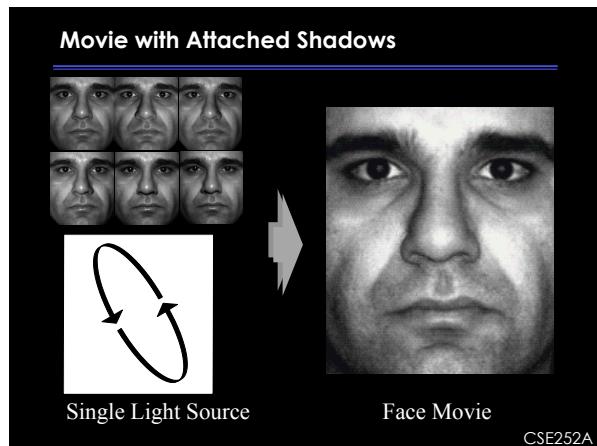


Basis Images



Span B

CSE252A



Distance to Linear Subspace

- An n -pixel image $x \in \mathbb{R}^d$ can be projected to a low-dimensional feature space $y \in \mathbb{R}^k$ by

$$y = Wx$$
- From $y \in \mathbb{R}^k$, the reconstruction of the point in $\mathbb{R}^d$ is $W^T y = W^T W x$
- The error of the reconstruction, or the distance from x to the subspace spanned by W is:

$$\|x - W^T W x\|$$

CSE152, Winter 2013

Intro Computer Vision

Distance to Linear Subspace

- From a collection of images of person i under variable lighting, compute basis B_i images of a 3-D linear subspace using PCA or SVD.
- Recognize individual by finding subspace with shortest distance from a test image to the subspace.

$$d_i = \|x - B_i B_i^T x\|$$

CSE152, Winter 2013

Intro Computer Vision

Alternative projections

CSE152, Winter 2013

Intro Computer Vision

Fisherfaces: Class specific linear projection

P. Belhumeur, J. Hespanha, D. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, PAMI, July 1997, pp. 711--720.

- An n -pixel image $x \in \mathbb{R}^d$ can be projected to a low-dimensional feature space $y \in \mathbb{R}^k$ by

$$y = Wx$$
- where W is an k by d matrix.
- Recognition is performed using nearest neighbor in $\mathbb{R}^k$.
- How do we choose a good W ?

CSE152, Winter 2013

Intro Computer Vision

PCA & Fisher's Linear Discriminant

- Between-class scatter

$$S_B = \sum_{i=1}^c |\chi_i| (\mu_i - \mu)(\mu_i - \mu)^T$$
- Within-class scatter

$$S_W = \sum_{i=1}^c \sum_{x_i \in \chi_i} (x_i - \mu_i)(x_i - \mu_i)^T$$
- Total scatter

$$S_T = \sum_{i=1}^c \sum_{x_i \in \chi_i} (x_i - \mu)(x_i - \mu)^T = S_B + S_W$$
- Where
 - c is the number of classes
 - μ_i is the mean of class χ_i
 - $|\chi_i|$ is number of samples of χ_i .

If the data points x_i are projected by $y_i = Wx_i$ and the scatter of x_i is S , then the scatter of the projected points y_i is $W^T S W$

CSE152, Winter 2013

Intro Computer Vision

