

Recognition I

Introduction to Computer Vision
CSE 152
Lecture 17

CSE152, Spr 12

Intro Computer Vision

Announcements

- HW 3 due 12 noon, tomorrow.
- HW 4 to be posted soon – recognition
- Lecture plan – recognition for next two lectures, then video and motion.

CSE152, Spr 12

Intro Computer Vision

Three Source Photometric stereo:

Offline:

Using source directions & BRDF, construct reflectance map for each light source direction. $R_1(p,q)$, $R_2(p,q)$, $R_3(p,q)$

Online:

1. Acquire three images with known light source directions. $E_1(x,y)$, $E_2(x,y)$, $E_3(x,y)$

Orthographic projection, distant light sources.

2. For each pixel location (x,y) , find (p,q) as the intersection of the three curves

$$R_1(p,q) = E_1(x,y)$$

$$R_2(p,q) = E_2(x,y)$$

$$R_3(p,q) = E_3(x,y)$$

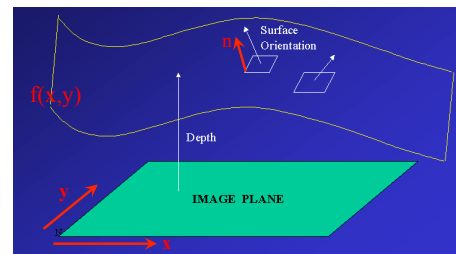
- 3 This is the surface normal at pixel (x,y) . Over image, the normal field is estimated

- 4 Integrate normal field to get depth.

CSE152, Spr 12

Intro Computer Vision

Gradient Space (p,q)



Gradient Space : (p,q)

$$p = \frac{\partial f}{\partial x}, \quad q = \frac{\partial f}{\partial y}$$

Normal vector

$$\mathbf{n} = \frac{\partial s}{\partial x} \times \frac{\partial s}{\partial y} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1 \right)^T$$

$$\hat{\mathbf{n}} = \frac{1}{\sqrt{p^2 + q^2 + 1}} (p, q, -1)^T$$

CSE152, Spr 12

Intro Computer Vision

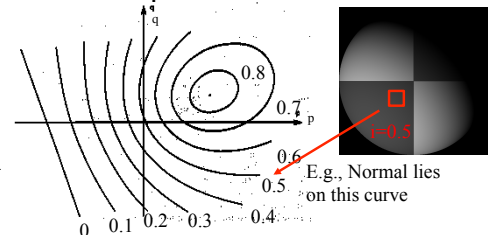
What is the reflectance map?

- Reflectance map is a function $R(p,q)$ – which gives the image irradiance as a function of the surface gradient (p,q) or equivalently the surface normal.
- For a fixed viewing direction, and light source strength and direction
- Can be computed from known BRDF.
- Can be measured using a probe of the same material

CSE152, Spr 12

Intro Computer Vision

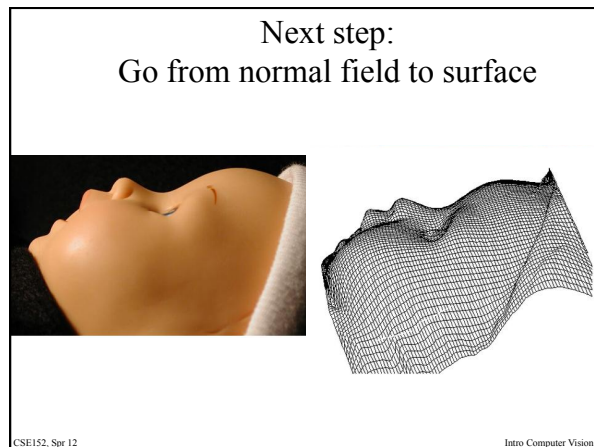
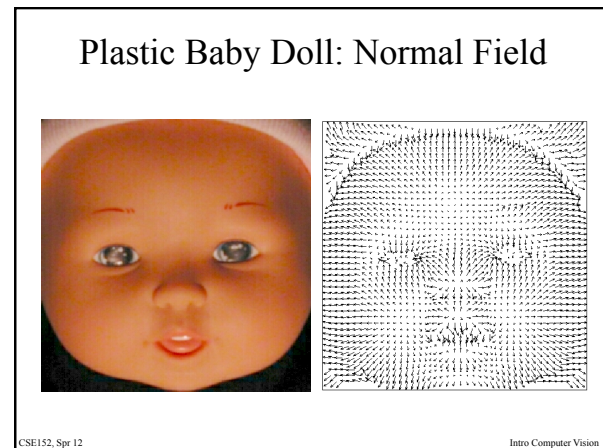
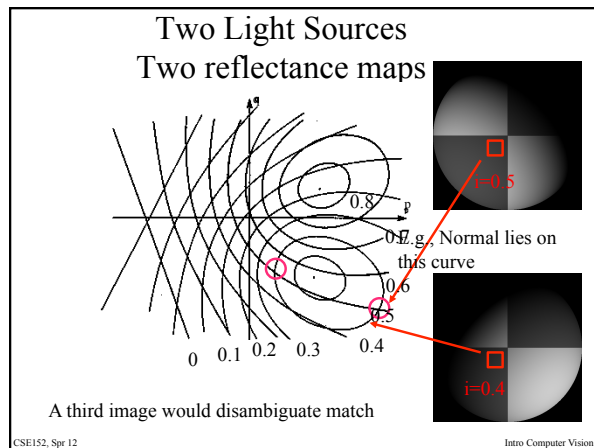
Reflectance Map of Lambertian Surface



What does the intensity (Irradiance) of one pixel in one im
It constrains the surface normal projecting to that point to

CSE152, Spr 12

Intro Computer Vision



Lambertian Photometric stereo

- If the light sources s_1 , s_2 , and s_3 are **known**, then we **can** recover b from as few as three images. (Photometric Stereo: Silver 80, Woodham81).

$$[e_1 \ e_2 \ e_3] = b^T [s_1 \ s_2 \ s_3]$$

- i.e., we measure e_1 , e_2 , and e_3 for each pixel location and we know s_1 , s_2 , and s_3 . We can then solve for b by solving a linear system.

$$b^T = [e_1 \ e_2 \ e_3] [s_1 \ s_2 \ s_3]^{-1}$$

- Normal is: $\hat{n} = b/|b|$, albedo is: $|b|$

CSE152, Spr 12

Intro Computer Vision

Uncalibrated photometric stereo

- Take n images as input without knowledge of light directions or strengths
- Perform SVD to compute B^* .
- Find some A such that B^*A is close to integrable.
- Integrate resulting gradient field to obtain height function $f^*(x,y)$.

Comments:

- $f^*(x,y)$ differs from $f(x,y)$ by a GBR.
- Can use specularities to resolve GBR for non-Lambertian surface.

CSE152, Spr 12

Intro Computer Vision

Generalized Bas-Relief Transformations

Without knowledge of light source location, one can only recover surfaces up to GBR transformations.

$$\tilde{f}(x,y) = \lambda f(x,y) + \mu x + \nu y$$

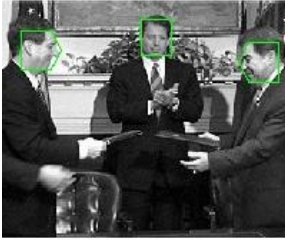
f : true depth

$\tilde{f}$: GBR transform of depth

CSE152, Spr 12

Intro Computer Vision

Recognition

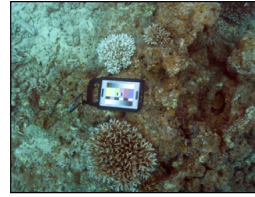


Given a database of objects and an image determine what, if any of the objects are present in the image.

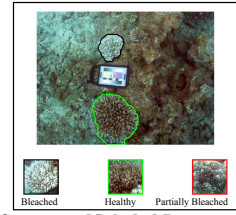
CSE152, Spr 12

Intro Computer Vision

Where are the coral heads and which ones are healthy and which are bleached?



Input Image

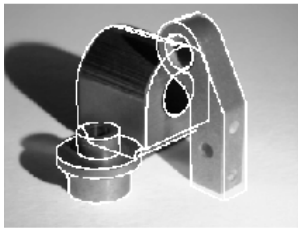


Segmented/labeled Image

CSE152, Spr 12

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, Spr 12

Intro Computer Vision

Recognition

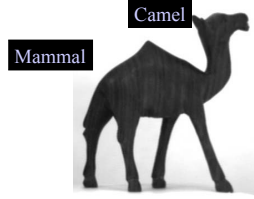


Given a database of objects and an image determine what, if any of the objects are present in the image.

CSE152, Spr 12

Intro Computer Vision

Categories vs. Instances



Two challenges:
Recognizing instances
Recognizing categories

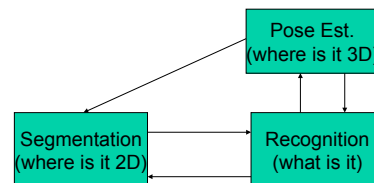
CSE152, Spr 12

Intro Computer Vision

Object Recognition: The Problem

Given: A database D of "known" objects and an image I :

1. Determine which (if any) objects in D appear in I
2. Determine the pose (rotation and translation) of the object



WHAT AND WHERE!!!

CSE152, Spr 12

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, Spr 12

Intro Computer Vision

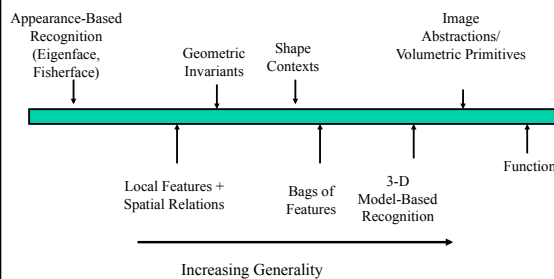
Further Recognition Issues

- How general is the problem?
 - 2D vs. 3D
 - range of viewing conditions
 - available context
 - segmentation cues
- What sort of data is best suited to the problem?
 - Whole images
 - Local 2D features (color, texture,
 - 3D (range) data
- What information do we have in the database?
 - Collection of images?
 - 3-D models?
 - Learned representation?
 - Learned classifiers?
- How many objects are involved?
 - small: brute force search
 - large: ??

CSE152, Spr 12

Intro Computer Vision

A Rough Recognition Spectrum



CSE152, Spr 12

Intro Computer Vision

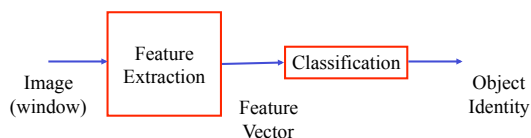
Appearance-Based Vision: A Pattern Classification Viewpoint

1. Feature Space + Nearest Neighbor
2. Dimensionality Reduction
3. Bayesian Classification
4. Appearance Manifolds

CSE152, Spr 12

Intro Computer Vision

Sketch of a Pattern Recognition Architecture



CSE152, Spr 12

Intro Computer Vision

Example: Face Detection

- Scan window over image.
- Search over position & scale.
- Classify window as either:
 - Face
 - Non-face



CSE152, Spr 12

Intro Computer Vision

- So, what are the features?
- So, what is the classifier

CSE152, Spr 12

Intro Computer Vision

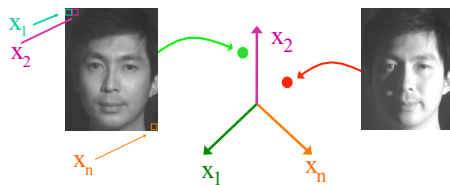
Pattern Classification Summary

- Supervised vs. Unsupervised: Do we have labels?
- Supervised
 - Nearest Neighbor
 - Bayesian
 - Plug in classifier
 - Distribution-based
 - Projection Methods (Fisher's, LDA)
 - Neural Network
 - Support Vector Machine
 - Kernel methods
- Unsupervised
 - Clustering
 - Reinforcement learning

CSE152, Spr 12

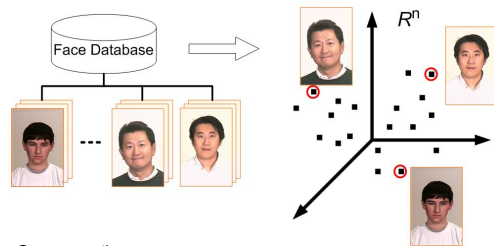
Intro Computer Vision

The Space of Images



- Consider an n -pixel image to be a point in an n -dimensional space, $\mathbf{x} \in \mathbb{R}^n$.
- Each pixel value is a coordinate of $\mathbf{x}$.

- Face Space:
 - A set of face images construct a *face space* in $\mathbb{R}^n$
 - Appearance-based methods analyze the distributions of individual faces in face space



Some questions:

1. How are images of an individual, under all conditions, distributed in this space?
2. How are the images of all individuals distributed in this space?

CSE152

Computer Vision I

More features

- Filtered image
- Filter with multiple filters (bank of filters)
- Histogram of colors
- Histogram of Gradients (HOG)
- Haar wavelets
- Scale Invariant Feature Transform (SIFT)
- Speeded Up Robust Feature (SURF)

CSE152

Computer Vision I

- So, what are the features?
- So, what is the classifier

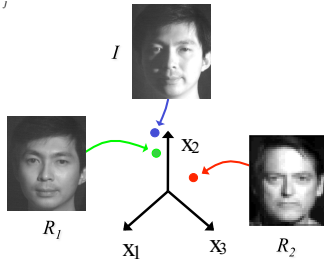
CSE152

Computer Vision I

Nearest Neighbor Classifier

$\{R_j\}$ are set of training images.

$$ID = \arg \min_j \text{dist}(R_j, I)$$



CSE152, Spr 12

Intro Computer Vision

Comments on Nearest Neighbor

- Sometimes called “Template Matching”
- Variations on distance function (e.g. L_1 , robust distances)
- Multiple templates per class- perhaps many training images per class.
- Expensive to compute k distances, especially when each image is big (N dimensional).
- May not generalize well to unseen examples of class.
- No worse than twice the error rate of the optimal classifier -- if enough training samples
- Some solutions:
 - Bayesian classification
 - Dimensionality reduction

CSE152, Spr 12

Intro Computer Vision

Do features vectors have structure in the image space?

- Faces of individuals cluster in the image space. (Not true).
- Faces of individuals are confined to a linear or affine subspace of $\mathbf{R}^n$
- Faces of an individual are approximated by a linear subspace
- Faces and objects lie on or near a manifold in the space of images.

CSE152

Computer Vision I

An idea:

Represent the set of images as a linear subspace

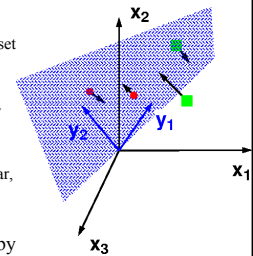
What is a linear subspace?

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

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

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

It is spanned by a d -dimensional orthogonal basis.



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

CSE152

Computer Vision I

Linear Subspaces & Linear Projection

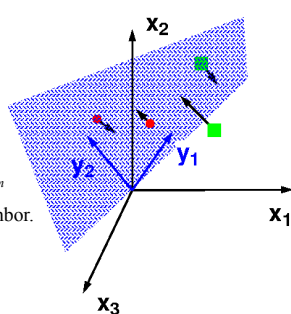
- An n -pixel image $x \in \mathbf{R}^n$ can be projected to a low-dimensional feature space $y \in \mathbf{R}^m$ by

$$y = Wx$$

where W is an m by n matrix.

- Recognition is performed in $\mathbf{R}^m$ using, for example, nearest neighbor.
- How do we choose a good W ?

Example: Projecting from $\mathbf{R}^3$ to $\mathbf{R}^2$



CSE152, Spr 12

Intro Computer Vision

Linear Subspaces & Recognition

1. Approximate all training images as a single linear subspace (Eigenfaces).
2. Represent lighting variation w/o shadowing for a single individual as a 3-D linear subspace. A collection of n individuals is modeled as n 3-D linear subspaces.
3. Represent lighting variation w/ shadowing for a single individual as a 9-D linear subspace. A collection of n individuals is modeled as n 9-D linear subspaces.
4. Project all training images to a single subspace that enhances discriminability (Fisherfaces).

CSE152, Spr 12

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:

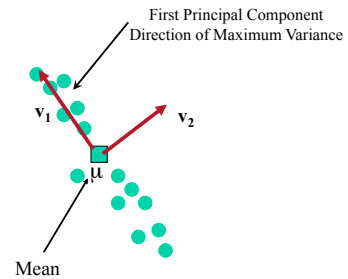
- They are independent.
- Projection onto the basis $\{v_1, \dots, v_k\}$ gives the k -dimensional set of linear features that preserves the most variance.

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, Spr 12

Intro Computer Vision

PCA Example



CSE152, Spr 12

Intro Computer Vision

Eigenfaces

- Modeling
 1. Given a collection of n labeled k by l training images, represent each one as $k \times l$ dimensional column vector
 2. Compute mean image and covariance matrix.
 3. Compute k Eigenvectors (note that these are images) of covariance matrix corresponding to k largest Eigenvalues. (Or perform using SVD!!)
 4. Project the training images to the k -dimensional Eigenspace.
- Recognition
 1. Given a test image, project vectorized image to Eigenspace.
 2. Perform classification to the projected training images.

CSE152, Spr 12

Intro Computer Vision

Eigenfaces: Training Images



[Turk, Pentland 91]

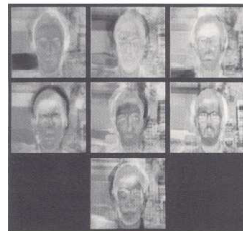
CSE152, Spr 12

Intro Computer Vision

Eigenfaces



Mean Image

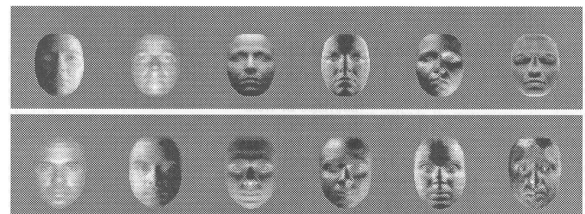


Basis Images

CSE152, Spr 12

Intro Computer Vision

Basis Images for Variable Lighting



CSE152, Spr 12

Intro Computer Vision