

# Introduction

Computer Vision I  
CSE 252A  
Lecture 1

# Announcements

- Class Web Page is up:
  - <http://www.cs.ucsd.edu/classes/wi05/cse252a/>
- Assignment 0: “Getting Started with Matlab” is posted to web page, due 1/13/04
- Read Chapters 1 & 2 of Forsyth & Ponce

# What is Computer Vision?

- Trucco and Verri (Text): Computing properties of the 3-D world from one or more digital images
- Sockman and Shapiro: To make useful decisions about real physical objects and scenes based on sensed images
- Ballard and Brown: The construction of explicit, meaningful description of physical objects from images.
- Forsyth and Ponce: Extracting descriptions of the world from pictures or sequences of pictures”

# Why is this hard?



What is in this image?

1. A hand holding a man?
2. A hand holding a mirrored sphere?
3. An Escher drawing?

- Interpretations are ambiguous
- The forward problem (graphics) is well-posed
- The “inverse problem” (vision) is not

# We all make mistakes

“640K ought to be enough for anybody.” –  
Bill Gates, 1981

“...” – Marvin Minsky

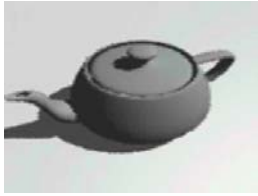
# What do you see?

- ▮ Changing viewpoint
- ▮ Moving light source
- ▮ Deforming shape



## What was happening

- ▮ Changing viewpoint
- ▮ Moving light source
- ▮ Deforming shape



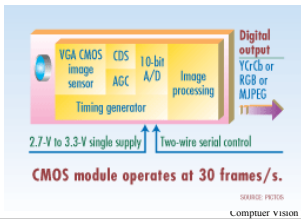
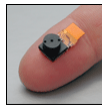
## Should Computer Vision follow from our understanding of Human Vision?

### Yes & No

1. Who would ever be crazy enough to even try creating machine vision?
  2. Human vision “works”, and copying is easier than creating.
  3. Secondary benefit – in trying to mimic human vision, we learn about it.
1. Why limit oneself to human vision when there is even greater diversity in biological vision
  2. Why limit oneself to biological when there may be greater diversity in sensing mechanism?
  3. Biological vision systems evolved to provide functions for “specific” tasks and “specific” environments. These may differ for machine systems
  4. Implementation – hardware is different, and synthetic vision systems may use different techniques/methodologies that are more appropriate to computational mechanisms

## The Near Future: Ubiquitous Vision

- Five years from now, digital cameras will cost 1 cent (sensor cost).
- Digital video will be a widely available commodity component embedded in cell phones, PDA’s, doorbells, bridges, security systems, cars, etc.
- 99.9% of digitized video won’t be seen by a person.
- That doesn’t mean that only 0.1% is important!



## Applications: touching your life

- Football
- Movies
- Surveillance
- HCI – hand gestures, American Sign Language
- Face recognition & Biometrics
- Road monitoring
- Industrial inspection
- Robotic control
- Autonomous driving
- Space: planetary exploration, docking
- Medicine – pathology, surgery, diagnosis
- Microscopy
- Military
- Remote Sensing

## Some Vision Problems

- Segmentation
  - Breaking images and video into meaningful pieces
- Reconstructing the 3D world
  - from multiple views
  - from shading
  - from structural models
- Recognition
  - What are the objects in a scene?
  - What is happening in a video?
- Video
  - Understand movement and change in image sequence.
  - Tracking objects

## Related Fields

- Image Processing
- Computer Graphics
- Pattern Recognition
- Perception
- Robotics
- AI

## Image Interpretation - Cues

- Variation in appearance in multiple views
  - stereo
  - motion
- Shading & highlights
- Shadows
- Contours
- Texture
- Blur
- Geometric constraints
- Prior knowledge

CSE252A, Winter 2005

Computer Vision I

## Shading and lighting

Shading as a result of differences in lighting is

1. A source of information
2. An annoyance

CSE252A, Winter 2005

Computer Vision I

## Illumination Variability An annoyance



“The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity.”

-- Moses, Adini, Ullman, ECCV '94

CSE252A, Winter 2005

Computer Vision I

## How do we understand shading

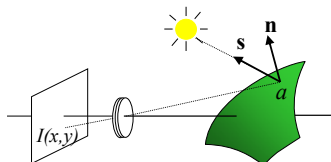
(An idealization of “engineering” research)

1. Construct a model of the domain (usually mathematical, based on physics).
2. Prove properties of that model to better understand the model and opportunities of using it.
3. Develop algorithms to solve a problem that is correct under the model.
4. Implement & evaluate it.
5. Question assumptions of the model & start all over again.

CSE252A, Winter 2005

Computer Vision I

## 1. Image Formation



At image location  $(x,y)$  the intensity of a pixel  $I(x,y)$  is

$$I(x,y) = a(x,y) \mathbf{n}(x,y) \cdot \mathbf{s}$$

where

- $a(x,y)$  is the albedo of the surface projecting to  $(x,y)$ .
- $\mathbf{n}(x,y)$  is the unit surface normal.
- $\mathbf{s}$  is the direction and strength of the light source.

CSE252A, Winter 2005

Computer Vision I

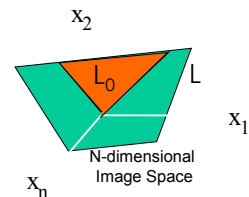
## 2. A property: 3-D Linear subspace

*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]

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

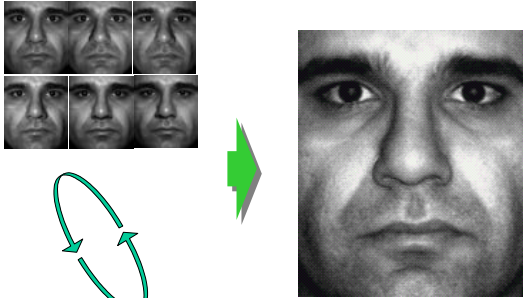
where  $B$  is a  $n$  by 3 matrix whose rows are product of the surface normal and Lambertian albedo



CSE252A, Winter 2005

Computer Vision I

### 3,4 : An implemented algorithm: Relighting

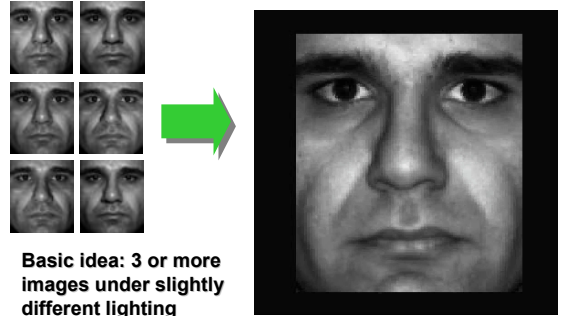


Single Light Source

CSE252A, Winter 2005

Computer Vision I

### 3,4: An implemented algorithm Photometric Stereo



Basic idea: 3 or more  
images under slightly  
different lighting

CSE252A, Winter 2005

Computer Vision I

## 5. Question Assumptions

- Many objects are not Lambertian (specular, complex reflectance functions).



CSE252A, Winter 2005

Computer Vision I

## The course

- Part 1: The Physics of Imaging
- Part 2: Early Vision
- Part 3: Reconstruction
- Part 4: Recognition

CSE252A, Winter 2005

Computer Vision I

## Part I of Course: The Physics of Imaging

- How images are formed
  - Cameras
    - What a camera does
    - How to tell where the camera was located
  - Light
    - How to measure light
    - What light does at surfaces
    - How the brightness values we see in cameras are determined
  - Color
    - The underlying mechanisms of color
- How to describe it and measure it

CSE252A, Winter 2005

Computer Vision I

## Cameras, lenses, and sensors



Figure 1.16 The first photograph on record, *la table servie*, obtained by Nicéphore Niepce in 1822. *Collection Harlinge-Viollet*.

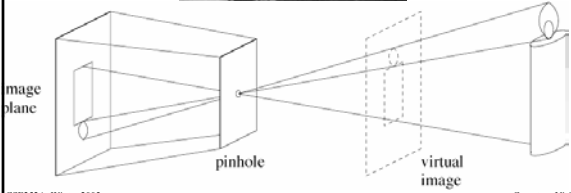
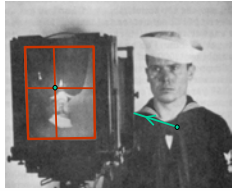
- Pinhole cameras
- Lenses
- Projection models
- Geometric camera parameters

From *Computer Vision*, Forsyth and Ponce, Prentice-Hall, 2002.

CSE252A, Winter 2005

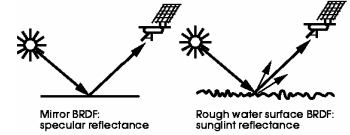
Computer Vision I

### A real camera ... and its model



## Lighting & Photometry

- How does measurement relate to light energy?
- Sensor response
- Light sources
- Reflectance



## Color



## Part II: Early Vision in One Image

- Representing small patches of image
  - For three reasons
    - We wish to establish correspondence between (say) points in different images, so we need to describe the neighborhood of the points
    - Sharp changes are important in practice --- known as “edges”
    - Representing texture by giving some statistics of the different kinds of small patch present in the texture.
      - Tigers have lots of bars, few spots
      - Leopards are the other way

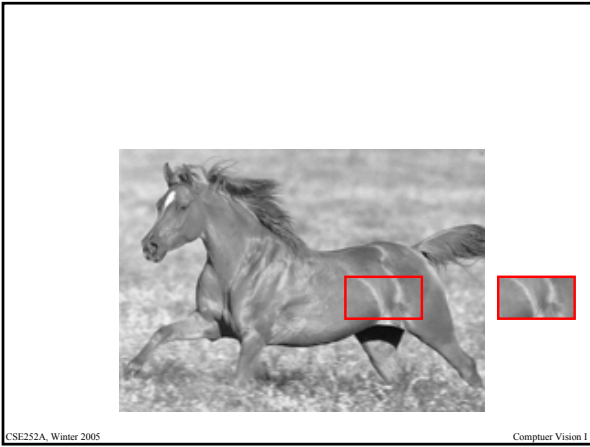
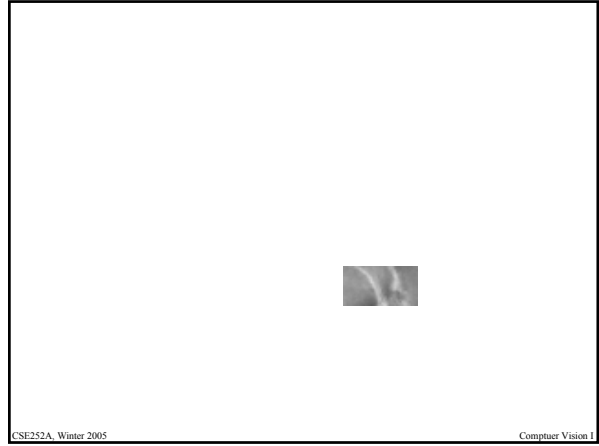
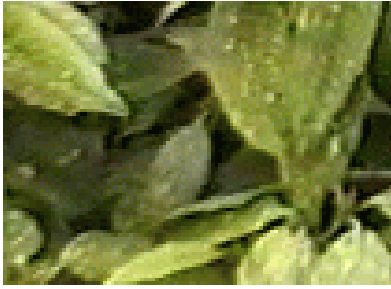
## Segmentation

- Which image components “belong together”?
- Belong together=lie on the same object
- Cues
  - similar color
  - similar texture
  - not separated by contour
  - form a suggestive shape when assembled

## Boundary Detection: Local cues



# Boundary Detection: Local cues



# Boundary Detection



# Gradients



# Boundary Detection



Finding the Corpus Callosum

(G. Hamarneh, T. McInerney, D. Terzopoulos)

## Part 3: Reconstruction from Multiple Images

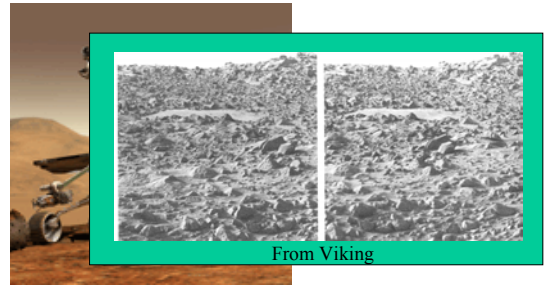
- Photometric Stereo
    - What we know about the world from lighting changes.
  - The geometry of multiple views
  - Stereopsis
    - What we know about the world from having 2 eyes
  - Structure from motion
    - What we know about the world from having many eyes
- or, more commonly, our eyes moving.

CSE252A, Winter 2005

Computer Vision I

## Mars Rover

Spirit



CSE252A, Winter 2005

Computer Vision I

Façade (Debevec, Taylor and Malik, 1996)

Reconstruction from multiple views, constraints, rendering



Architectural modeling:

- photogrammetry;
- view-dependent texture mapping;
- model-based stereopsis.

Reprinted from "Modeling and Rendering Architecture from Photographic: A Hybrid Geometry- and Image-Based Approach,"  
By P. Debevec, C.J. Taylor, and J. Malik,  
Proc. SIGGRAPH (1996). © 1996 ACM, Inc.  
Included here by permission.

CSE252A, Winter 2005

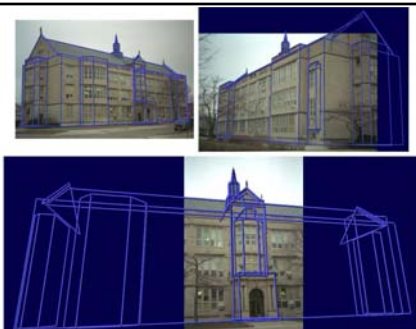
Computer Vision I

## Images with marked features



CSE252A, Winter 2005

Computer Vision I



Recovered model edges reprojected through recovered camera positions into the three original images

CSE252A, Winter 2005

Computer Vision I

## Resulting model & Camera Positions

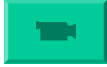


CSE252A, Winter 2005

Computer Vision I

## Façade

- The Camponile Movie

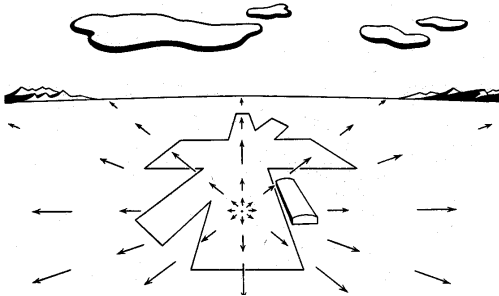


## Video-Motion Analysis



- Where “things” are moving in image – segmentation.
- Determining observer motion (egomotion)
- Determining scene structure
- Tracking objects
- Understanding activities & actions

## Forward Translation & Focus of Expansion [Gibson, 1950]



## Visual Tracking



### Main Challenges

1. 3-D Pose Variation
2. Occlusion of the target
3. Illumination variation
4. Camera jitter
5. Expression variation etc.

[ Ho, Lee, Kriegman ]

## Visual Tracking

- State: usually a finite number of parameters (a vector) that characterizes the “state” (e.g., location, size, pose, deformation of thing being tracked).
- Dynamics: How does the state change over time? How is that changed constrained?
- Representation: How do you represent the thing being tracked
- Prediction: Given the state at time  $t-1$ , what is an estimate of the state at time  $t$ ?
- Correction: Given the predicted state at time  $t$ , and a measurement at time  $t$ , update the state.

## Tracking



(www.brickstream.com)

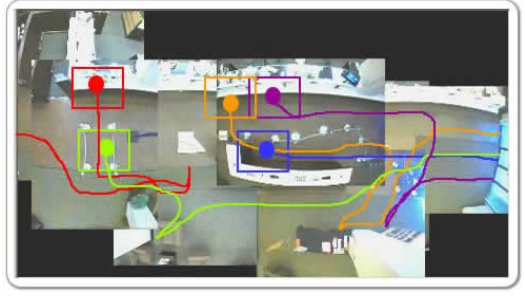
## Tracking



CSE252A, Winter 2005

Computer Vision I

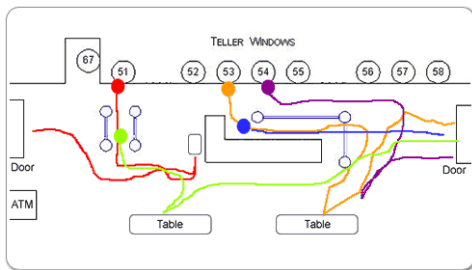
## Tracking



CSE252A, Winter 2005

Computer Vision I

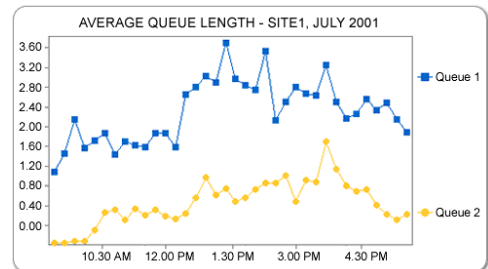
## Tracking



CSE252A, Winter 2005

Computer Vision I

## Tracking



CSE252A, Winter 2005

Computer Vision I

## Part 4: Recognition



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

CSE252A, Winter 2005

Computer Vision I

## 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

CSE252A, Winter 2005

Computer Vision I

## Object 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: ??

CSE252A, Winter 2005

Computer Vision I

## Recognition Example: Face Detection: Classify face vs. non-face



CSE252A, Winter 2005

Computer Vision I

## Why is Face Recognition Hard?

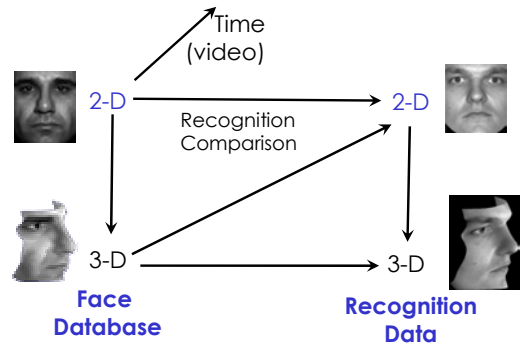
### Many faces of Madonna



CSE252A, Winter 2005

Computer Vision I

## Face Recognition: 2-D and 3-D



CSE252A, Winter 2005

Computer Vision I

## Yale Face Database B



64 Lighting Conditions  
9 Poses  
=> 576 Images per Person

CSE252A, Winter 2005

Computer Vision I

## Real vs. Synthetic



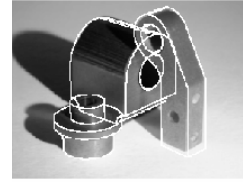
CSE252A, Winter 2005

Computer Vision I



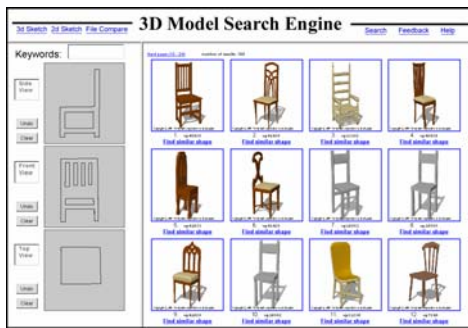
[http://www.ri.cmu.edu/projects/project\\_271.html](http://www.ri.cmu.edu/projects/project_271.html)

## Model-Based Vision



- Given 3-D models of each object
- Detect image features (often edges, line segments, conic sections)
- Establish correspondence between model & image features
- Estimate pose
- Consistency of projected model with image.

## Object Classes: Chairs



(Funkhauser, Min, Kazhdan, Chen, Halderman, Dobkin, Jacobs)

## Scene Interpretation



“The Swing”  
Fragonard, 1766

## The Syllabus

## cse252a: early vision and recognition

- Cameras
- Human Vision
- Photometry (radiance, irradiance, BRDF)
- Illumination cones
- Shape from Shading, Photometric Stereo
- Curves & Surfaces
- Color
- Filtering
- Edges & Features
- Stereo Matching
- Optical Flow and Motion
- Tracking
- Statistical pattern recognition (Bayes, SVM, Kernel methods)
- Object Recognition
- Behavior Recognition (HMM's)

## cse252b: Multiview Geometry & Segmentation

- Multiview Geometry
- Affine Structure from Motion
- Projective Structure from Motion
- Robust F-matrix estimation
- Image Segmentation
- Texture: Synthesis, Recognition, Shape-from
- Motion Segmentation
- Object Detection
- Image Registration
- Image Based Rendering

## Announcements

- Class Web Page is up:
  - <http://www.cs.ucsd.edu/classes/wi05/cse252a/>
- Assignment 0: “**Getting Started with Matlab**” is posted to web page, due 1/13/04
- Read Chapters 1 & 2 of Forsyth & Ponce