Introduction
Computer Vision I
CSE 252A
Lecture 1

What is computer vision?

What is Computer Vision?
• Trucco and Verri: 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 (Text): 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
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 vision 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

Hermann Grid

Scan your eyes over the figure. Do you see the gray spots at the intersections? Stare at one of them and it will disappear.

How many red X’s are there?
Raise your hand when you know.

How many red X’s are there?
Raise your hand when you know.
The Near Future: Ubiquitous Vision

- Digital video has become really cheap.
- It’s widely embedded in cell phones, cars, games, etc.
- 99.9% of digitized video isn’t seen by a person.
- That doesn’t mean that only 0.1% is important!

Applications: touching your life

- Optical Character Recognition
- Football
- Movies
- Surveillance
- HCI – hand gestures
- Aids to the blind
- Face recognition & biometrics
- Road monitoring
- Industrial inspection
- Virtual Earth, street view
- Robotic control
- Autonomous driving
- Space: planetary exploration, docking
- Medicine – pathology, surgery, diagnosis
- Microscopy
- Military
- Remote Sensing
- Digital photography
- Google Goggles
- Video games

Earth viewers (3D modeling)

Photosynth

http://photosynth.net
Based on Photo Tourism technology developed by Noah Snavely, Steve Seitz, and Rick Szeliski
Photo Tourism overview

System for interactive browsing and exploring large collections of photos of a scene. Computes viewpoint of each photo as well as a sparse 3d model of the scene.

Optical character recognition (OCR)
Technology to convert scanned docs to text
- If you have a scanner, it probably came with OCR software

Face detection
- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Smile detection?

Object recognition (in supermarkets)

Face recognition
Vision-based biometrics

“How the Afghan Girl was Identified by Her Iris Patterns” Read the story

Object recognition (in mobile phones)

- This is becoming real:
  - Point & Find, Nokia
  - SnapTell.com (now Amazon)
  - Google Goggles

Special effects: shape capture

Leafsnap.com

Sports

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely

Login without a password…

Face recognition systems now beginning to appear more widely

Login without a password…

Fingerprint scanners on many new laptops, other devices
Smart cars

- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models

Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in.

Vision systems (JPL) used for several tasks
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Robotics

NASA’s Mars Spirit Rover

http://www.youtube.com/watch?v=cdQps1tJUE

Autonomous Vehicles


http://www.youtube.com/watch?v=cdQps1tJUE
Current state of the art

- You just saw examples of current systems.
  - Many of these are less than 5 years old

- This is a very active research area, and rapidly changing
  - Many new apps in the next 5 years

- To learn more about vision applications and companies
  - David Lowe maintains an excellent overview of vision companies

Image Interpretation - Cues

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

An example of a cue:
Shading and lighting
Shading as a result of differences in lighting is

1. A source of information
2. An annoyance

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
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.

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

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

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

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.

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

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

3,4: An implemented algorithm:
Relighting

3,4: An implemented algorithm
Photometric Stereo

Basic idea: 3 or more images under slightly different lighting

5. Question Assumptions
- Many objects are not Lambertian (specular, complex reflectance functions).
The course

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

Part I of Course: The Physics of Imaging

• How images are formed
  – Cameras
    • What a camera does
    • Projection Models (Projective spaces, etc.)
    • How to tell where the camera was located
  – Light
    • How to measure light
    • What happens to light at surfaces
    • How the brightness values we see in images are determined
  – Color
    • The underlying mechanisms of color
    • How to describe it and measure it

Cameras, lenses, and sensors

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

Lighting & Photometry

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

A real camera ... and its model

Color
Part II: Early Vision in One Image

• Representing small patches of image
• Noise
• Filtering
• Edge Detection
• Corner Detection
• Texture
• Segmentation

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

http://www.robots.ox.ac.uk/~vgg/dynamics.html

Boundary Detection: Local cues

(Sharon, Balun, Brandt, Basri)
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 two eyes
- Structure from motion
  - What we know about the world from having many eyes
  - or, more commonly, our eyes moving.

Façade (Debevec, Taylor and Malik, 1996)
Reconstruction from multiple views, constraints, rendering

Architectural modeling:
- photogrammetry;
- view-dependent texture mapping;
- model-based stereopsis.

Images with marked features
Resulting model & Camera Positions

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

UNI High 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 change 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.
Part 4: Recognition

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

Recognition Challenges

- Within-class variability
  - Different objects within the class have different shapes or different material characteristics
  - 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

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

Why is Face Recognition Hard?
Many faces of Madona

Scene Interpretation

“The Swing”
Fragonard, 1766

The Syllabus
The primary course text is:

The secondary text is: Rick Szeliski’s book *Computer Vision: Algorithms and Applications*; Printed copy available, also softcopy online http://szeliski.org/Book/

**A note about subsequent courses**
- CSE 252B – to be offered in spring
- CSE252C – not offered this year
- CSE190 Biometrics - not offered this year

**Announcements**
- Class Web Page is at:
  - http://cseweb.ucsd.edu/classes/fa12/cse252A-a
  - HW0: “Getting Started with Matlab” to be posted to web page by tomorrow, due next Thursday 10/4.
- Read:
  - Chapters 1 & 2 of Forsyth & Ponce
  - Chapter 1 of Szeliski (Optional)

**About working together**
- You may work together on homework assignments to discuss ideas and methods, however what you turn in should be your own work and any code should be your own coding. Copying is not permitted.