Introduction

Computer Vision I
CSE 252A
Lecture 1

• We’ll start with some introductory material
• …. And end with
  – Syllabus
  – Organizational materials
  – Wait list

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?

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

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
Underestimates

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

• “... in three to eight years we will have a machine with the general intelligence of an average human being ... The machine will begin to educate itself with fantastic speed. In a few months it will be at genius level and a few months after that its powers will be incalculable.”
– Marvin Minsky, Life Magazine, 1970

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. 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.

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!
- And there’s an enormous amount of image and video content on the internet…

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)

Image from Microsoft Virtual Earth
See also: Google Earth

Photosynth

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

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

Digit recognition, AT&T labs
http://research.att.com/~yann/
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection

• Digital cameras, smart phones, face book, etc.

Smile detection

Sony Cyber-shot T70 Digital Still Camera

Face recognition

Who is she?
Vision-based biometrics

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

Object recognition (in supermarkets)

LaneHawk by EvolutionRobotics

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk you are assured to get paid for it...”

Login without a password

Fingerprint scanners on laptops, mice, other devices

Face recognition systems now beginning to appear more widely

http://www.sensiblevision.com/

Object recognition (in mobile phones)

– Point & Find, Nokia
– SnapTell.com (now Amazon)
– Mobile Acuity
– Google Photos
– Apple Photos

Leafsnap.com -> Dogsnap -> Birdsnap
Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

- Vicon

Sports

Spacevision first down line
Nice explanation on www.howstuffworks.com

Smart cars

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

http://www.youtube.com/watch?v=cdIpq1pUUE

Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in.

Digimask: put your face on a 3D avatar.

Xbox Kinect

Playmotion game a Disney Epic

3D sensors

Vision in space

NASA's Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

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

NASA's Mars Spirit Rover


Robotics

First person vision

Oracam

Google Glass

http://www.robocup.org/

http://www.roborise.org/
Medical imaging

Image guided surgery

Molecular Reconstruction from Cryo-electron Microscope Images

Coralnet.ucsd.edu

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 applications in the next 5 years

• To learn more about vision applications and companies
  – David Lowe maintained a list of vision companies, until 2015 …
    • http://www.cs.ubc.ca/spider/lowe/vision.html

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
\[ I(x,y) = a(x,y)n(x,y)s \]
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.

\[ L = \{ x | x = Bs, \forall s \in \mathbb{R}^3 \} \]

where \(B\) is a \(n\) by 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).

Or can we learn about these variations from data

• Google FaceNet trained on hundreds of millions of cropped face images
• 140 Million Parameters
• 1.6B Flops

[Schroff, Kalenichenko, Philbin, 2015]

Related disciplines

Deep Learning

Cognitive science
Algorithms
Image processing

Artificial intelligence
Robotics

Related Fields

Deep Learning

Robotics
Control

Signal processing
Mathematics
Neurobiology

Four Rs of computer vision

• Reprojection
  – Rendering a scene and features from a different view, under different illumination, under different surface properties, etc.
• Reconstruction
  – Multiple view geometry, structure from motion, shape from X (where X is texture, shading, contour, etc.), etc.
• Registration
  – Tracking, alignment, optical flow, correspondence, etc.
• Recognition
  – Recognizing objects, scenes, events, etc.

Rudiments: The implied fifth R

• image filtering
• interest point detection
• edge detection
• probability
• statistics
• linear algebra
• projective geometry
• optics
• Fourier analysis
• sampling
• algorithms
• photometry
• physics of color
• human vision
• psychophysics
• performance evaluation

From Serge Belongie
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

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/~vgd/dynamics.html

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.

Mars Rover

Curiosity

From Viking Lander, 1976

Video-Motion Analysis

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

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

Why is Face Recognition Hard?

Many faces of Madona

Scene Interpretation

“The Swing” Fragonard, 1766

About the class
Text


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

Primary Text

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

Secondary Text

- Class Web Page and syllabus is at:
  - http://cseweb.ucsd.edu/classes/fa17/cse252A-a/
  - HW0: “Image manipulation in Python” to be posted to web page by tomorrow, due next Wed 10/11.

About the class

- Read:
  - Chapters 1 & 2 of Forsyth & Ponce
  - Chapter 1 of Szeliski (Optional)

Academic Integrity Policy

Integrity of scholarship is essential for an academic community. The University expects that both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

Collaboration Policy

It is expected that you complete your academic assignments on your own and in your own words and code. The assignments have been developed by the instructor to facilitate your learning and to provide a method for fairly evaluating your knowledge and abilities (not the knowledge and abilities of others). So, to facilitate learning, you are authorized to discuss assignments with others; however, to ensure fair evaluations, you are not authorized to use the answers developed by another, copy the work completed by others in the past or present, or write your academic assignments in collaboration with another person. If the work you submit is determined to be other than your own, you will be reported to the Academic Integrity Office for violating UCSD’s Policy on Integrity of Scholarship.
Wait List

- Number of enrolled students is limited by
  - Size of room
- General advice
  - Wait for as long as you can
- Concurrent enrollment (Extension) students have lowest priority

- And, if you are going to drop the class, please officially drop it to make room for others