COMP 776: Computer Vision
Today

- Introduction to computer vision
- Course overview
- Course requirements
The goal of computer vision

- To extract “meaning” from pixels

What we see

What a computer sees

Source: S. Narasimhan
The goal of computer vision

- To extract "meaning" from pixels

Humans are remarkably good at this…

Source: "80 million tiny images" by Torralba et al.
What kind of information can be extracted from an image?

- Metric 3D information
- Semantic information
Vision as measurement device

Real-time stereo

Structure from motion

Reconstruction from Internet photo collections

Pollefeys et al.

Goesele et al.
Vision as a source of semantic information
Scene and context categorization

- outdoor
- city
- traffic
- ...
Qualitative spatial information

- slanted
- vertical
- rigid moving object
- non-rigid moving object
- horizontal

slide credit: Fei-Fei, Fergus & Torralba
Why study computer vision?

- Vision is useful: Images and video are everywhere!
Why study computer vision?

- Vision is useful
- Vision is interesting
- Vision is difficult
  - Half of primate cerebral cortex is devoted to visual processing
  - Achieving human-level visual perception is probably “AI-complete”
Why is computer vision difficult?
Challenges: viewpoint variation

Michelangelo 1475-1564
Challenges: illumination

image credit: J. Koenderink
Challenges: scale
Challenges: deformation

Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba
Challenges: occlusion

Magritte, 1957

slide credit: Fei-Fei, Fergus & Torralba
Challenges: background clutter
Challenges: Motion
Challenges: object intra-class variation

slide credit: Fei-Fei, Fergus & Torralba
Challenges: local ambiguity
Challenges: local ambiguity

Source: Rob Fergus and Antonio Torralba
Challenges: local ambiguity

Source: Rob Fergus and Antonio Torralba
Challenges or opportunities?

- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Image source: J. Koenderink
Depth cues: Linear perspective
Depth cues: Aerial perspective
Depth ordering cues: Occlusion

Source: J. Koenderink
Shape cues: Texture gradient
Shape and lighting cues: Shading

Source: J. Koenderink
Position and lighting cues: Cast shadows

Source: J. Koenderink
Grouping cues: Similarity (color, texture, proximity)
Grouping cues: "Common fate"
Inherent ambiguity of the problem

- Many different 3D scenes could have given rise to a particular 2D picture
Inherent ambiguity of the problem

• Many different 3D scenes could have given rise to a particular 2D picture

• Possible solutions
  – Bring in more constraints (more images)
  – Use prior knowledge about the structure of the world

• Need a combination of geometric and statistical methods
Connections to other disciplines

- Artificial Intelligence
- Robotics
- Machine Learning
- Computer Graphics
- Computer Vision
- Image Processing
- Cognitive science
  - Neuroscience
Origins of computer vision

Successes of computer vision to date
Optical character recognition (OCR)

Digit recognition
yann.lecun.com

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Automatic check processing

Sudoku grabber
http://sudokugrab.blogspot.com/

Source: S. Seitz, N. Snavely
Biometrics

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Source: S. Seitz
How the Afghan Girl was Identified by Her Iris Patterns

Source: S. Seitz
Mobile visual search: **Google Goggles**

Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.

- Landmark
- Book
- Contact Info
- Artwork
- Places
- Wine
- Logo
Face detection

Many new digital cameras now detect faces

- Canon, Sony, Fuji, …

Source: S. Seitz
Smile detection

The Smile Shutter flow
Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.

Sony Cyber-shot® T70 Digital Still Camera

Source: S. Seitz
Face recognition: Apple iPhoto software

http://www.apple.com/ilife/iphoto/
Automotive safety

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

- Pedestrian collision warning
- Forward collision warning
- Lane departure warning
- Headway monitoring and warning

Source: A. Shashua, S. Seitz
Vision-based interaction: Xbox Kinect


http://electronics.howstuffworks.com/microsoft-kinect.htm

Special effects: shape and motion capture

Source: S. Seitz
3D visualization: Microsoft Photosynth

http://photosynth.net

Source: S. Seitz
Vision for robotics, space exploration

*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 Matthies et al.

Source: S. Seitz
The computer vision industry

• A list of companies here:

http://www.cs.ubc.ca/spider/lowe/vision.html
Basic Info

• Instructor: Svetlana Lazebnik (lazebnik@cs.unc.edu)
• Office hours: By appointment, FB 244
• Class webpage: http://www.cs.unc.edu/~lazebnik/spring11

• Textbooks (suggested):
  Forsyth & Ponce, *Computer Vision: A Modern Approach*
  Richard Szeliski, *Computer Vision: Algorithms and Applications* (available online)
Course requirements

• Philosophy: computer vision is best experienced hands-on

• Programming assignments: 50%
  – About four assignments
  – Expect the first one in the next couple of classes
  – Brush up on your MATLAB skills (see web page for tutorial)

• Final assignment: 30%
  – Recognition competition
  – Winner gets a prize!

• Participation: 20%
  – Come to class regularly
  – Ask questions
  – Answer questions
Collaboration policy

• Feel free to discuss assignments with each other, but coding must be done individually

• Feel free to incorporate code or tips you find on the Web, provided this doesn’t make the assignment trivial and you explicitly acknowledge your sources

• Remember: I can Google too (and I have the copies of everybody’s assignments from the last three years this class was offered)
Course overview

I. Early vision: Image formation and processing
II. Mid-level vision: Grouping and fitting
III. Multi-view geometry
IV. Recognition
V. Advanced topics
I. Early vision

- Basic image formation and processing

Cameras and sensors
Light and color

* Linear filtering
Edge detection

Feature extraction: corner and blob detection
II. “Mid-level vision”

- Fitting and grouping

Fitting: Least squares
Hough transform
RANSAC
III. Multi-view geometry

- Stereo
- Epipolar geometry
- Tomasi & Kanade (1993)
- Affine structure from motion
- Projective structure from motion

Tomasi & Kanade (1993)
IV. Recognition

Patch description and matching

Clustering and visual vocabularies

Bag-of-features models

Classification

Sources: D. Lowe, L. Fei-Fei
V. Advanced Topics

- Time permitting…

Segmentation

Face detection

Articulated models

Motion and tracking