**Admistrivia**

This course requires a CS account. Apparently (new for 2010), if you have a UA email and are registered, an account is automatically created for you.

Course page is now up: http://www.cs.arizona.edu/classes/cs477/spring10
(Linked from instructor’s home page http://kobus.ca)

Lectures and assignments will require either connecting from a UA machine, OR a login id (“me”) and password (“vision4fun”).

Office hours (TBA) will be via electronic signup (http://kobus.ca/calendar). Signup MUST occur by 6pm the day before.

Significant communication for the course will happen using the class mail list (cs477@listserv.arizona.edu).

Eight machines in 9th floor lab (gr01-gr08) will be available with priority for this course.

“…, vision is the process of discovering from images what is present in the world, and where it is.

… our brains must be capable of representing this information …”

Marr 82, page 3.

**What is (computer) vision?**

Visual Representation $\rightarrow$ Semantic Representation $\rightarrow$ A tiger lying in the grass
Is vision computational?

Why study Computer Vision?

- Images and movies are everywhere
- Fast-growing collection of useful applications
  - building representations of the 3D world from pictures
  - automated surveillance (who’s doing what)
  - movie post-processing
  - automated analysis of scientific data
- Various deep and attractive scientific mysteries
  - how does object recognition work?
- Greater understanding of human vision

More Applications

Image and video retrieval and data mining
Robotics
  - Defect spotting
  - Driving aids, autonomous flight
  - Surveillance, identification
Graphics, Virtual Reality, Printing

Computer Vision in Context

Part of Artificial Intelligence

Connected to cognitive psychology perceptual psychology robotics databases imaging science

Key methods
- math
- stats
- programming
- empirical science
Graphics versus Vision

Graphics
model of the world --> images

Vision
images --> model of the world

Vision Systems

Biological
eye + brain

Man made
camera + computer

Computer Vision History

(Nope, it does not work yet)
Properties of Vision

- One can “see the future”
  - You can avoid getting hit by an approaching object

Properties of Vision

- 3D representations are easily constructed
  - Useful
    - to humans (avoid bumping into things; planning a grasp; etc.)
    - for applications (build models for movies).
  - Many different cues including
    - multiple views (motion, stereopsis)
    - texture
    - shading

Shading Cues

Shape from texture
• People draw distinctions between what is seen
  – “Object recognition”
  – This could mean “is this a fish or a bicycle?”
  – It could mean “is this George Washington?”
  – It could mean “is this poisonous or not?”
  – It could mean “is this slippery or not?”
  – It could mean “will this support my weight?”

• Great mystery
  – How does it all work?

• Great challenge
  – Build programs that can infer useful information about the world from image data.
Part I: How images are formed

- Abstraction level is a single pixel
- Cameras
  - What a camera does
- Light
  - How to measure light
  - What light does at surfaces
  - How the brightness values we see in cameras are determined
- Color

* Here "part" refers to the book section. We cannot do them all in detail.

Part II: Early Vision in One Image

- Single pixels are not very informative
- Representing small patches of image
  - Often want to match points in different images, so we need to describe the neighborhood of the points (e.g. for stereo)
  - Sharp changes are important in practice --- known as "edges"
  - Representing texture by statistics of local structure
    - Zebras have lots of bars, few spots, leopards are the other way
Filters as templates
Filters respond to structures that “look” like the filter

Part III: Early Vision in Multiple Images
- (Limited coverage in 2010)
- The geometry of multiple views
  - Two views of the same patch reveal depth if you can match them
- Stereo vision, structure from motion

Part IV: Mid-Level Vision
- Finding coherent structure to break the image or movie into bigger units of semantic significance
  - Segmentation
    - Breaking images or videos into useful pieces
      - finding video sequences that correspond to one shot
      - finding image regions of consistent color and texture
    - Tracking (limited coverage in 2010)
      - Keeping track of a moving object through a sequence of views
      - Provides an effective way to segment whole objects

Grouping
- Which image components “belong together’’?
- “belong together” typically means “lie on the same object”
- Cues
  - similar colour
  - similar texture
  - not separated by contour
  - form a suggestive shape when assembled
  - move together
Part V & VI: High Level Vision

- Visual information → semantics
- Object recognition
  - Specific object (my car)
  - Object category (a car)
- Object / scene understanding
  - What are the pose (orientation) and location of scene objects
  - What about the object structure (very leafy plant?)
  - What might an object be good for (a step?)
- Activity recognition

High Level Vision

- Many applications
  - Digital libraries
    - Find me the picture of JFK and Marilyn Monroe embracing
  - Surveillance
    - Warn me if there is a mugging in the grove
  - HCI
    - Do what I show you
    - Military
      - Shoot this, not that
- Variable appearance
  - Objects appear different due to pose, illumination, occlusion, etc.
- Which bits of image should be recognized together?
  - Segmentation (issue: segment first or recognize first)
- How can objects be recognized without focusing on detail?
  - Levels of abstraction (object categories vs individuals)

What are the problems in recognition?
Part V: High Level Vision (Geometry)

- The relations between object geometry and image geometry
- One difficulty in recognition is pose change
  - One approach is to use descriptors which are invariant to (small) pose changes
  - If we know the object well, then we can use the fact that valid features must be consistent in pose.

Part VI: High Level Vision (Probabilistic)

- Using probabilistic models and classifiers to recognize objects
  - Probabilistic methodology
    - Bayes rule: \( p(\text{object} | \text{data}) \sim p(\text{data} | \text{object}) p(\text{object}) \)
  - Templates and classifiers
    - Find objects from a canonical view (e.g. frontal face) by matching a template
      - best when combined with a probabilistic classifier
    - Transformed views require looking with more templates (e.g. rotated, scaled, faces)
  - Relations
    - Find the parts with a classifier, and then reason about the relationships between the parts to find the object.

Matching templates

- Some objects are 2D patterns
  - e.g. faces
- Build an explicit pattern matcher
  - need to discount changes in illumination
  - changes in background are hard
  - changes in pose are hard

http://www.ri.cmu.edu/projects/project_271.html
Relations between templates

- e.g. find faces by
  - finding eyes, nose, mouth
  - finding an assembly of the three that has the “right” relations