CS 652

Computer Vision

Joachim Stahl
Assistant Professor
Dept. of Computer Science

Outline

- What is Computer Vision?
- Course Overview
  - Administrative stuff
- Basic Concepts

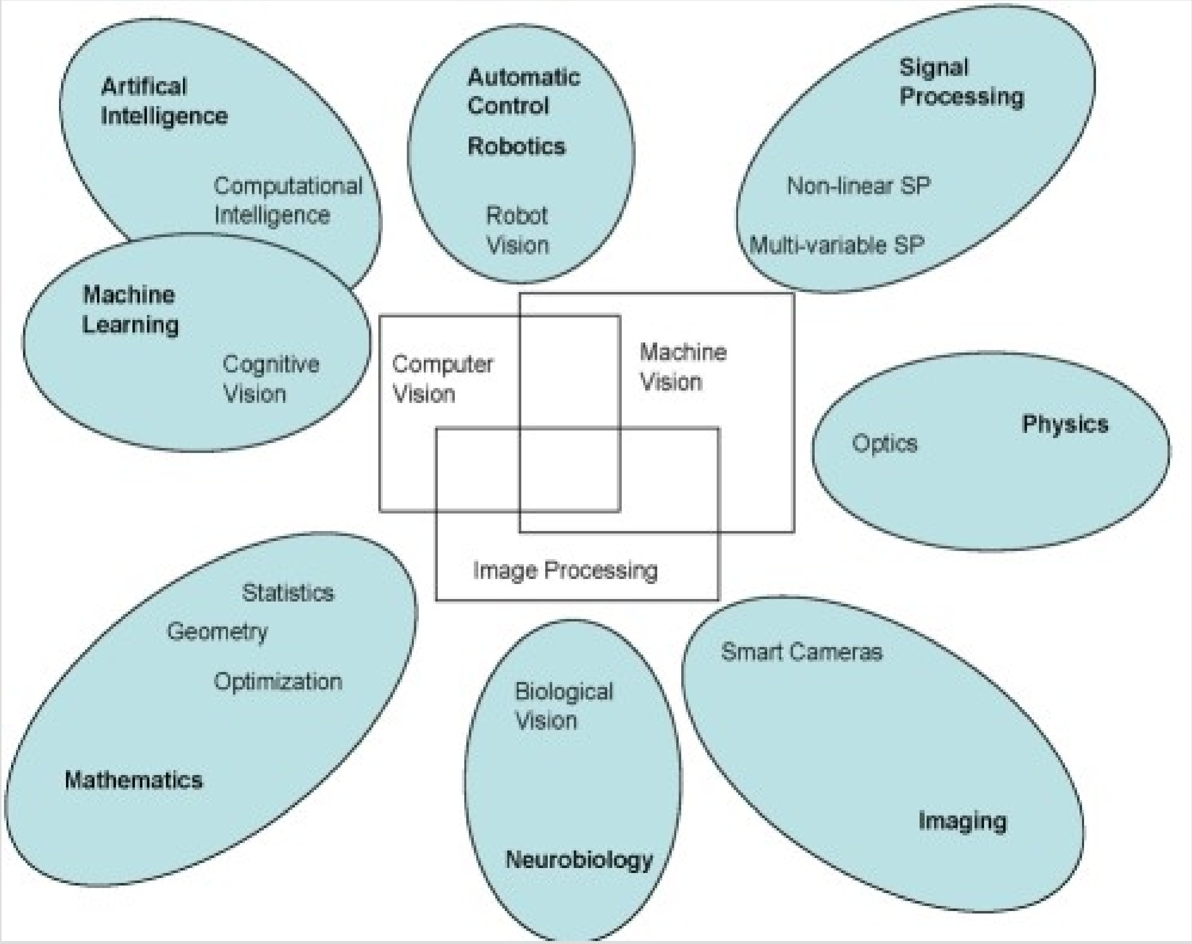
What is Computer Vision?

- In a nutshell:

- Sub-field of Artificial Intelligence
  - Give computer ability to see and understand images
- Two goals:
  - Simulate human vision
  - Enhance human vision

What is Computer Vision?

- Is a multidisciplinary field



Overview

- Relatively new field
  - Intellectual frontier = disorganized
  - Many useful ideas with no theoretical ground
  - Many interesting theories useless in practice
- Relies on understanding of the physical process of image formation
  - Both human and artificial
- Acquire knowledge from image(s)
  - Done in steps
  - Early (or low-level) vision, mid-level vision, high-level vision

Overview

- Why study vision?
  - Extracting information from the world from pictures is unequivocally useful
  - Process is non-destructive, and can be discreet
  - Also cheap (nowadays)
- Many important applications
  - Robot navigation
  - Image Enhancement (e.g. military applications)
  - Medical Imaging (diagnostic, screening, etc.)
  - Image Database organization
  - etc...

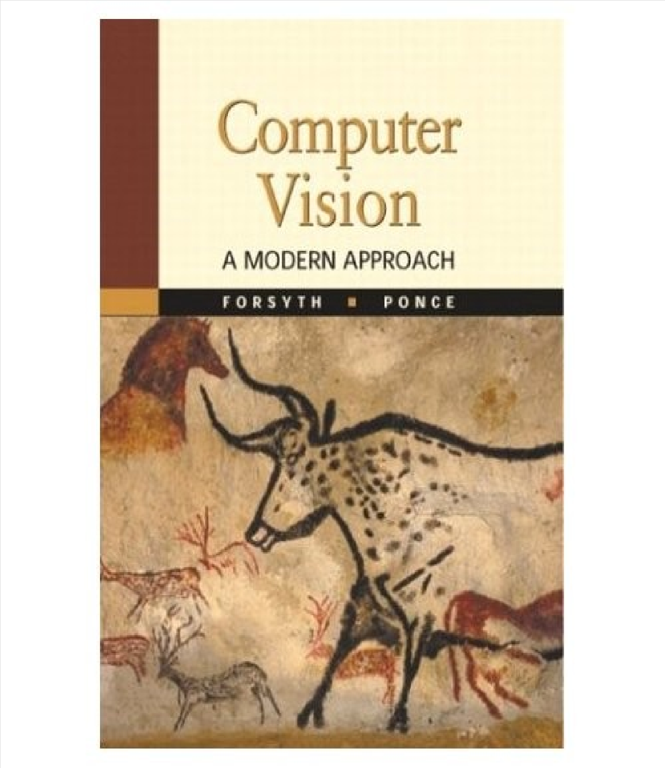
Overview

- Computer Vision started in 1960's
- Many theories developed at the time could not be applied
  - Lack of powerful enough equipment
  - Available capable hardware too expensive
    - Not long ago, color digital camera = $10,000
  - Some technologies didn't exist
- These are becoming possible now
  - Many research groups going back to these “old school” theories
- Still a very young field, lots of problems to be solved

Course Details

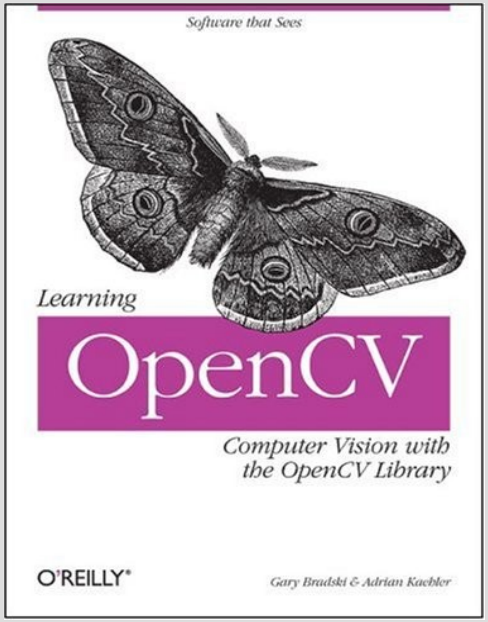
- Objectives:
  - become familiar with the field of computer vision
  - gain the necessary understanding of the field to be able to critically analyze research papers
- Evaluation criteria
  - be able to present research papers, and their critiques and findings, in a seminar setting
  - be able to solve practical computer vision problems

Textbook

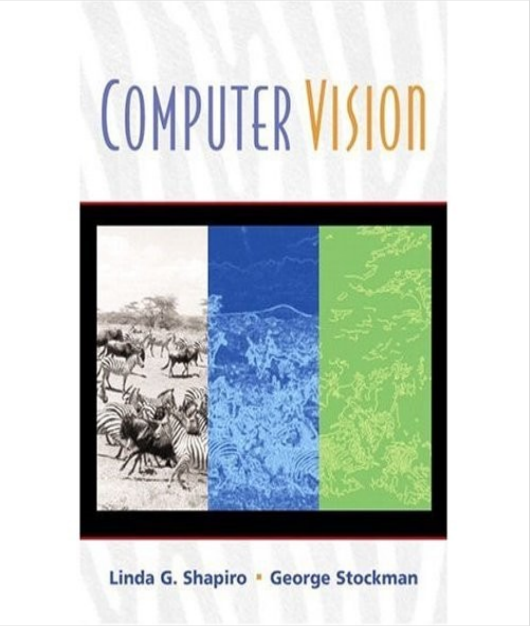
- Computer Vision: A Modern Approach, by David Forsyth and Jean Ponce, Prentice Hall, 2002 (ISBN 0130851981)
- 

Other

- Recommended texts (optional)
- Learning OpenCV by Bradksi and Kaehler



- Computer Vision by Shapiro and Stockman



Evaluation method

- The final grade will be computed on the total score weighted as follows:
  - Attendance, Participation 10%
  - Midterm 15%
  - Presentations 50%
  - Homework projects 25%

- The most important component of evaluation will be two individual presentations.
- Study relevant papers in computer vision and introduce them to the class.

Administrative

- Read Syllabus for different policies
  - Assignment submission
  - Deadlines
  - Academic integrity
  - Grading scheme
  - Disclaimer
  - etc.

Contact

- Instructor: Joachim Stahl
- Email: jstahl@clarkson.edu

- Office Hours: T 9am to 12pm
  Th 10am to 12pm

- Office Location: SC 383
- Office Phone: ext 2387

Basic Concepts

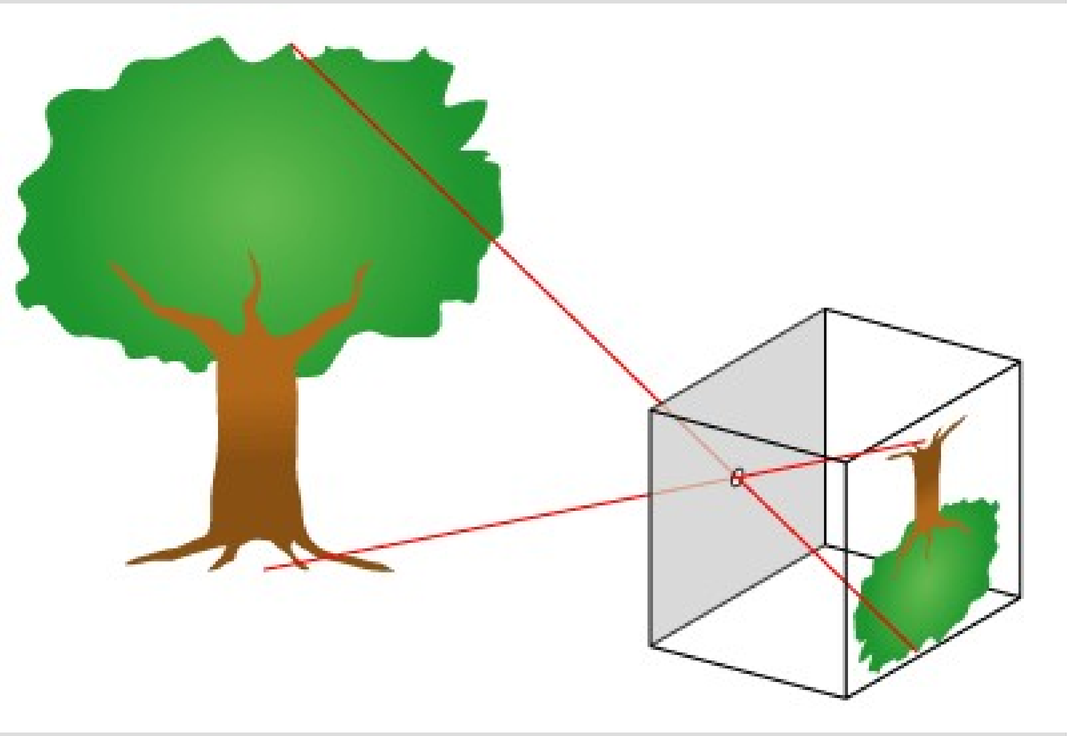
- Early Vision
- Mid-level Vision
- High-level Vision

Image Formation

- First photograph on record, “La Table Servie”, 1822

Image Formation

- First step in Computer Vision is to understand how images are formed
- First cameras worked using the concept of “pinhole camera”
- Image captured on paper treated with silver chloride
- Later cameras added use of lenses to help focus and capture more light



Early Vision

- Once we obtain an image, we want to extract information from it
- Early vision works at the lowest level, looking at single images and individual pixels to extract information
- This may include grouping pixels with similar characteristics, and generate higher level structures
  - E.g. find a line, a corner, or homogenous regions

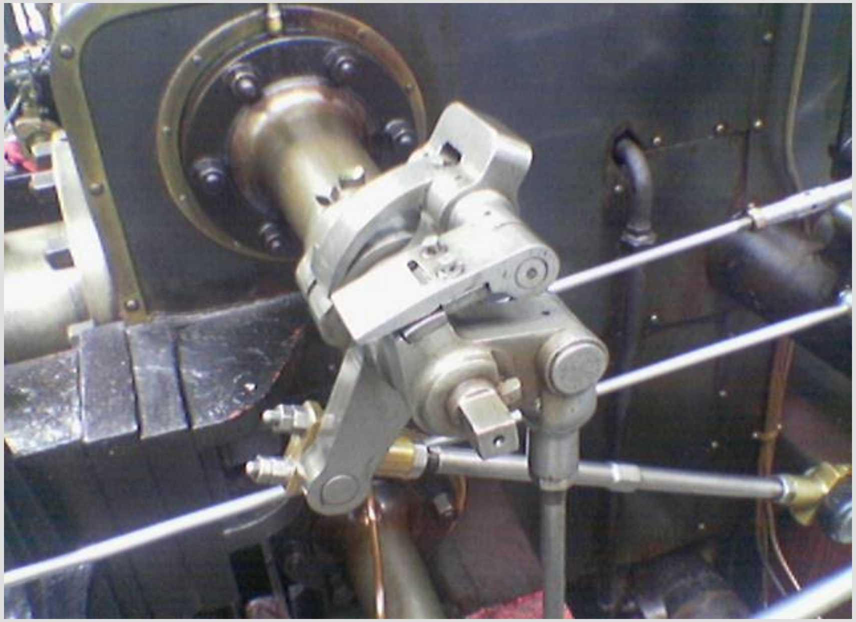
Early Vision

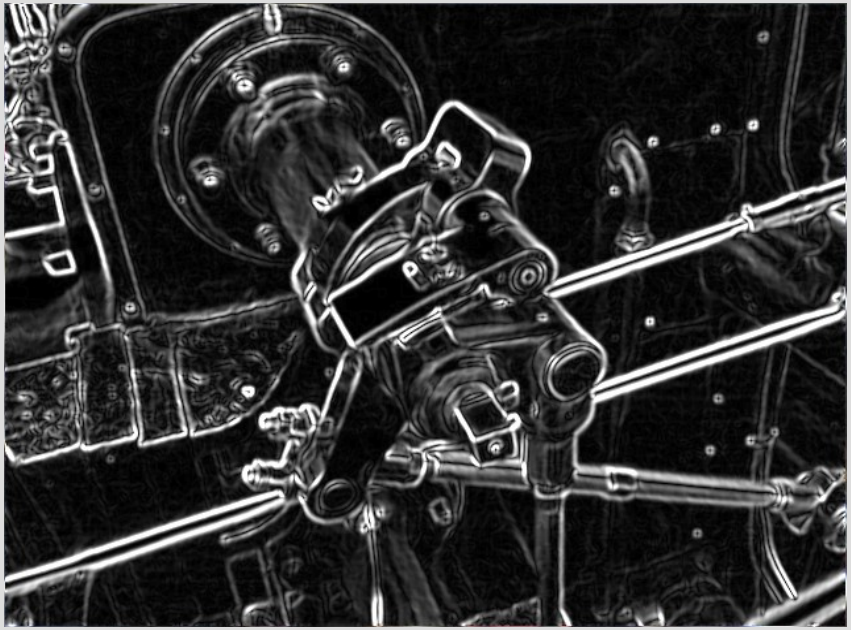
- Several methods are based on psychological studies on how the human brain processes low level information
  - points, lines, etc.
- Example: Detecting Edges
  - An edge can be defined as a fast change in brightness in an image
  - We can create a mathematical model of this change:
    - Represent image as matrix
    - Find derivatives

Edge Detection

- Find 2D derivative of image I for the x- and y-axis: $\partial I/\partial x$ and $\partial I \partial y$
- These can be obtained by doing a convolution between image and Laplacian operator
- With derivatives we can calculate an image gradient magnitude
  - Points of high gradient magnitude correspond to edge pixels

Edge Detection





Mid-Level Vision

- Takes as input results from low-level vision
- Combines results from one or more images
- Generates meaningful structures
  - Set of edges from a single structure (boundary)
  - Groups of pixels corresponding to a structure or shape
- Many detection methods fall into this category
  - Face detection
  - Tumor detection (Medical Imaging)
- Note difference between detection and recognition

Mid-Level Vision

- Most methods at this level are some form of image segmentation
  - Segment into regions that convey some meaning
- Segmentation helps summarize the information in the image
- Its more compact and meaningful than pixels
  - Pixels are an artificial artifact product of digitalization
  - Helps speed up further processing
    - Millions of pixels vs hundreds of regions

Example: Statistical Region Merging

- Assume each pixel is a region
- Create a list of all pairs of neighboring pixels
  - For (x,y): ((x,y),(x,y+1)), ((x,y),(x+1,y)), ((x,y),(x,y-1)), ((x,y),(x-1,y))
- Sort this list according to function f(p,p')
  - e.g. $f(p,p') = |p_a - p_a'|$
- Go through list, comparing regions where p and p' belong to and merge according to:

$$P(R,R') = \begin{cases} true & |R'_a - R_a| \leq \sqrt{b^2(R) + b^2(R')} \\ false & otherwise \end{cases}, \forall a \in \{R, G, B\}$$

Statistical Region Merging

- Where function b in P(R,R') is

$$b(R) = g\sqrt{(1/(2Q|R|))\ln|R_{|R|}|/\delta}$$

- And $|R_{|R|}|$ is the number of regions of size |R| and Q is an image complexity factor

High-level Vision

- This level of vision uses information from lower levels to recognize objects
- Recognition can be achieved by:
- Fitting a model (requires thorough understanding of problem)
  - Geometric information
  - Shape matching
- Through probabilistic techniques (requires learning mechanisms)
  - Classifiers
  - Artificial neural networks
- Could combine both

High-level Vision

- Example: Face Recognition
- Fitting a model to extract face features
  - Face has two eyes, a nose, a mouth, etc.
- Use learned DB to find best match