Lecture 1: CS677

Aug 22, 2017

Introduction

- Course number: CSCI 677 (4 units)
- Course name: Advanced Computer Vision
- Instructor: Prof. Ram Nevatia
  - My background, research interests…
- Today's objective
  - Describe course content
  - Conduct of the class
  - Required work, grading
  - Pre-requisites
  - Intro to potential and problems of vision

Background

- Course web page:
  - http://den.usc.edu
- Office hours
  - Instructor, Tu, Th; 1:30-3:00 P.M., PHE 202; other times by appointment
  - nevatia AT usc DOT edu, 213-740-6427
- TA: Jiyang Gao jiyangga AT usc DOT edu; office hours info will be posted
- Books:
  - Required: "Computer vision" by Forsyth and Ponce, Second Edition, 2012
  - Recommended: "Computer Vision: Algorithms and Applications" by Richard Szeliski; electronic copy available for personal use at http://szeliski.org/Book .
  - Deep Learning: Algorithms and Applications, I. Goodfellow, Y. Bengio and A. Courville, 2017. Free online version available at: http://www.deeplearningbook.org
- Additional reading material will be distributed electronically or available in form of tutorials and research papers online

Class Enrollment

- Sign in waiting list if you are present in class
- We will add a few more seats but may not be able to accommodate everyone
- We have not instituted an exam but students must demonstrate knowledge of pre-requisite material to the instructor
- Adding a class after start of semester requires instructor approval; please do **not** bypass/hack this system.

What is this course about?

- Study of techniques that attempt to create artificial vision capabilities for computers
- What does the term “vision” mean when applied to human vision?
  - Eyes provide images to the brain; perception happens in the brain.
  - Perception consists of computing “good” descriptions of the environment
    - Agree with other measurements
    - Example: class room scene
- Computer Vision
  - Camera is like an eye: provides images
  - Vision is really about perception

Course Objectives

- Understanding key problems of vision
- Alternative approaches to fundamental problems
- Specific applications will be covered only to illustrate the basic techniques
- Provide enough background for further study and for implementation of some practical vision systems
- Vision has become a large field and undergoing revolutionary changes, largely driven by deep learning (DL) techniques
  - Some may argue that we should teach only DL as other parts of CV have become or will soon become obsolete. We will include both the “classical” and “DL” approaches.
  - It is not possible to cover “everything” about “everything” in one semester, nor all of the state-of-art methods as this changes frequently (>1000 papers published each year)

Class Difficulty

- CS 677 is not an “easy” course
  - Other options to satisfy M.S. requirements may be easier
  - Requires skills in several math topics and good programming skills
- Though only fiver years old, textbook is quite out of dae. In particular it lacks any coverage of DL methods.
  - Also, coverage of existing topics is uneven and some parts are very hard to read but still it is the best available
- Course is a mix of traditional, first year graduate class and an advanced, research oriented course
- At this time, there is no plan to offer a more moderately paced course in computer vision in the near future

Pre-requisites: Programming and CS

- This is a graduate course in CS!
  - Proficiency in Python and C++.
  - Must have experience with actual programming, not just concepts of programming.
  - Familiarity with various data structures such as *lists, trees* and *graphs*
  - Algorithms to create and manipulate such structures
  - Basic CS algorithms such as *Sorting, searching, tree/graph traversal etc.*
- Be able to take informal or mathematical descriptions of algorithms and convert them into working programs
- Work with libraries of code with limited documentation

Pre-requistites: Mathematics

- Calculus
  - Derivatives, partial derivatives, Integration
- Geometry
  - Coordinate systems (Cartesian, spherical, cylindrical…)
  - Equations of entities such as points, lines, planes, circles, spheres…;
  - Computing relations between entities (distance, intersection, angles….)
- Linear Algebra
  - Linear transformations
  - Matrix representation, inversion, eigenvectors….
  - Solutions of systems of equations
- Elementary probability theory
  - Discrete/continuous random variables
  - Joint probability distribution/density functions, conditional probabilities
  - Bayes' theorem
- Please take the math requirements seriously!

Requirements

- Assignments (~6-8)
  - 1-2 "written" (mathematical) assignments
  - 5-6 programming assignments
    - Will make extensive use of functions in the Intel **OpenCV** library, Caffe and TensorFlow
    - Preferred language for programming is Python though much can be done using C++ or MATLAB.
- Grading:
  - Assignments 30%
  - Exam1: 30%, Exam 2: 30%; note Exam 2 is not a "final" exam
    - Exam 1, 7th or 8th week of classes (will be announced >1 week in advance); Exam 2, Nov 30, last class day (tentative date)
  - Class attendance and participation 10% (not applicable to DEN students or students with special exemption; their others scores will be scaled accordingly).
- All assignments and exams are required, missing any will result in a grade of "F". Late submissions will be accepted with prior permission only.

Academic Integrity

- Assignments and Exams are to be completed ***individually***, unless otherwise specified.
- We encourage discussions among students but ***not copying*** of the answers.
- The class will be conducted on the basis of mutual trust and respect; for USC ethics code, see

  http://web-app.usc.edu/scampus/principles-of-community/
- Unfortunately, there are occasional instances of gross abuse; these instances will be referred for action according to the USC Student Conduct Code, which can be found at this link:
  - http://web-app.usc.edu/scampus/university-student-conduct-code/
- Grades can only be assigned based on performance, not need.

Related Courses

- Artificial Intelligence (cs561)
  - Many AI techniques are used in vision but this course is not a pre-req
- Machine Learning (cs576)
  - Modern computer vision makes heavy use of these techniques; again it may be helpful but is not a a pre-requisite
- Deep Learning (CS599)
  - CV is now dominated by DL (and DL by CV)
  - DL599 is a new course, syllabus is not fully defined yet
  - Possible significant overlap with 677; we will focus only on use of DL, not a broader study of DL.
- Mathematical Pattern Recognition (ee559)
  - Overlaps with 573 and 576 to some extent
- Image Processing (ee 569)
  - In image processing, goal is usual to enhance, compress or modify images for better human perception or transmission
  - Recent versions of the course may material that overlaps 677.
- Robotics and Graphics
  - Cover some major applications of vision

Why Study Computer Vision (or not)?

- Required course for some tracks
  - Easier alternatives may exist
- Many important and exciting applications
  - Number of applications has been growing rapidly
  - Good job prospects
- Exciting intellectually
  - Vision is one of few remaining mysteries of natural world
  - Vision problems seem fundamentally unsolvable, great intellectual challenge to discover methods for solving them
  - Computer vision is a fast changing field, what is exciting today maybe obsolete in just a few years
  - Many fundamental problems remain unsolved, even difficult to formulate precisely
  - Theories are not always coherent
  - Many advanced mathematical tools are required

Two Major Components of the Field

- Infer 3-D scene geometry
  - Needed for navigation and manipulation
  - May be helpful for object/activity recognition
  - How can we infer 3-D info from a single 2-D image?
  - Can we use multiple images to simplify the problem?
  - Can we measure 3-D directly (and bypass some basic vision problems)?
  - Above problems relatively well understood, many working systems
- Semantic understanding
  - Recognition of objects, relations, activities….
  - Difficult to formulate mathematically
  - Very active area of research: methods have changed from "intuitive" to "statistical" to "deep learning"

Traditional (classic) vs Current Approaches

- Field of vision is < 50 years old
- Work older than five years is being called "classic"
- Major changes have come from machine learning, most recently deep learning
- Deep learning has proved very effective for recognition, less important for 3-D geometry topics
- We could just teach a course based on deep learning but geometry topics are useful for many important applications, e.g. vehicle navigation, augmented reality, movie making…
- Deep learning experiments require use of GPU
  - This will limit the range of assignments for the course.

What is Vision useful for?

- Biological organisms
  - Navigate: avoid obstacles, don't get lost…
  - Recognize objects of interest: predators, food, friends…
  - Estimate motion, observe activities, predict…
  - Visual communication
  - Entertainment
  - …..
- Note: capabilities of biological vision span a broad spectrum: there are single cell organisms that can just sense presence/absence of light but still find it useful for survival
- Artificial mechanisms
  - Capabilities depend on the goals and the environment
  - Human level generality is difficult to achieve and not required in all cases

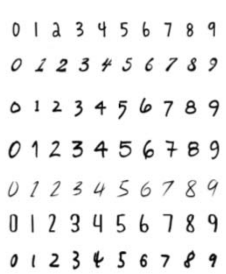
Application Areas

- **Mobile robots**
  - Self-driving cars, vehicles for hazardous environments, military vehicles..
- **Visual surveillance, security**
- **Manufacturing:**
  - Inspection, assembly...
- **Biometrics**
    - Face/Iris, fingerprint recognition
- **Visual aids**
  - Lane warnings, aid for the blind, …
- **Multi-media**
  - Visual communication, enhanced reality, models for virtual environments, content based retrieval...
- **Remote Sensing**
  - Maps (2-D and 3-D), change detection, crop/weather assessment…
- **Biomedical**
  - Large population screening, assisted surgery….

Why is Vision Hard?

- Seems easy to us, no conscious effort is needed by human viewers
- Small variations in human population’s ability to see/perceive
  - Does not require training/education for everyday tasks
- Can’t we just recognize objects based on “how they look”?
  - Isn’t a pen (a chair) a pen (chair) because it looks like a pen (chair)?
  - What does a pen (chair) look like?
  - Do we memorize images of pens or extract some more abstract representations (such as thin, mostly cylindrical objects with a conical section narrowing to a small circle at the end)?
  - We also need to detect/segment objects from others

From MNIST Database



From MNIST Database

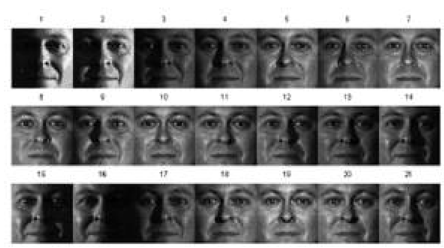
Same Object Class?

Some Issues: Representation

- What is representation of an object
- Objects of same class can have large variations in shape, size, color, material and other properties
  - Think about every day objects, such as chairs, coffee mugs, telephones…
- What is representation of an action (say throw an object)?
- Same action can be performed in different ways by different actors or even the same actor at different times or in different contexts

Viewpoint Change Examples

Illumination Change Examples



Depth Ambiguity and Occlusions

- World is 3-D, images are 2-D
  - There is an inherent loss of information; process is not truly invertible
    - Many 3-D environments could produce the same 2-D images
  - Our perception of 3-D from single 2-D images must take advantage of some regularities of the natural world
    - How do we isolate and exploit these regularities?
- Occlusion is (almost) ever-present
  - Objects occlude one another
  - Self-occlusion

Multiple Objects in a Scene

Complexity

How many objects are in this image?

What can we say about each?

What can we say about this scene?

Video Analysis

- Adds difficulties of detecting and tracking moving objects
  - If camera also moves, we need to distinguish between object and background motion
- We want to detect not only objects but also events/activities taking place in the environment
- Make inferences about the intentions/plans of actors
- With expansion of the course to 4 units, we expect to have 1-2 weeks coverage of video analysis; this is a new component of the course.

Topics to be studied in this class

- **Introduction (1 week)**
  Background, requirements and issues, human vision.
- **Image formation: geometry and photometry (2 weeks)**
  Geometry, brightness, quantization, camera calibration, photometry (brightness and color)
- **Image segmentation (1 week)**
  Region segmentation, Edge and line finding
- **Multi-view Geometry (3 weeks)**
  Shape from stereo and motion, feature matching, surface fitting, Active ranging
- **Object Recognition: Traditional Methods (2 weeks)**
  HoG/SIFT features, Bayes classifiers, SVM classifiers
- **Object Recognition: Deep Learning Methods (3 weeks)**
  Deep neural networks, classification networks, object proposal networks
- **Video analysis (1 week)**
  Activity Recognition, Detection and Tracking of objects
- **Language and Vision (1 Week)**

Next Class

- Read ch. 1 of Forsyth/Ponce book
  - Sections 1.1, 1.2.