Notes on *EE 290-005: Integrated Perception, Learning, and Control* lectures by Professors Yi Ma, Jitendra Malik, Shankar Sastry, and Claire Tomlin, Spring 2021

Jay Monga

Last updated January 27, 2021

Contents

Lecture 1

Introduction

Welcome to EE 290, which can be seen as an advanced robotics course. The faculty has a diverse set of specializations which cover the main themes of the course: perception, planning and control. This is the first time the course is being offered, so it's going to be a fun ride.

Goal: Study the integrated roles of perception, learning, and control in a closed-loop for autonomous robotic systems, under various levels of modeling uncertainty for the environment and of resource constraint for the agent. Many robotics curriculums focus on small parts instead of the integration of perception, learning, and control systems, so this course will focus on the whole, closed-loop autonomous systems.

1.1 Closed-Loop Autonomous System

Model Uncertainty:

- deterministic
- stochastic

Resources: Are we a private researcher at Google or a poor professor at Berkeley?

- cost
- hardware

1.2 Presentation and Project Suggestions

We don't want paper presentations to be a waste of your and our time. Presentations should be engaging and educatiational. Make sure to cover

- Problem formulation and assumptions (what is actually being solved? how is uncertainty modeled?)
- Justification for the proposed method (why not optimal control, learning, etc.)
- Generalizability in the proposed solution (is the solution general? or was a special case solved? how well?)
- Data and computational resourses needed (how easily can these methods be applied? are the results practical?)

In a similar manner, keep these tips in mind to get the most out of your projects.

- Don't stay "cocooned" in your own specific field (ex. I only like vision). This class is about integration.
- Keep physical-world implementation in mind; this makes the work practical.
- Again, integration. Try to combine model-based and data-driven approaches.
- Think out-of-the box to create projects that really capture interest and imagination.

1.3 History of Robotics

Prof. Sastry likes to always start off with a history of robotics.

- "it's important to know whoose shoulders you are standing on"
- Prosthetics was one of the earliest examples of robotics

- Mechatronics became more advanced and even programmable.
- Development of cars and planes provided lots of advancement
- Term "ROBOT" coined in by Karel Capek called "RUR"
- Nyquist, Bode, and Turing driving lots of fundamental theory
- Norbert Wiener did foundational work on cybernetics
- Rudolf Kalman did fundamental work on controls
- Early telemanipulation work done at Argonne national lab

1.4 Integrated Perception and Control

Prof. Malik likes to start out with evolutionary biology and how developed our sensors, acuators, and CPUs.

- Animals developed sight and movement around the same time
  - "Gibson: We see in order to move, and we move in order to see"
- Once we got locomotion down, we started to develop hands
- Use of hands to make tools correlated with growth of intelect
- While vision sensors today are as good as and better than our eyes, nothing really matches our hands yet
- Evolutionary progression came in order of vision and locomotion, manipulation, and then language.
  - Success in AI seems to also come in this order
- Prof. Malik has special intrest in general purpose robots that can work in homes

One field of locomition that has been studied a lot is gaits. Boston Dynamics has developed robots that can emulate animal gates very well, although more complex gaits in real life are left to be effectively realized. In essence, gaits do not encompass all forms of legged locomotion.

Coupled Perception and Action Vision allows a way for a robot to have enough knowledge of its environment to in theory plan robust motion.

Visual landmarks are quite useful for planning in an environment. One classical approach is SLAM - develop a complete map, and then plan in it. However, it may better to utilize visual landmarks rather doing full scene reconstruction.

"Instead of trying to produce a programme to simulate the adult mind, why not try to simulate the child?" - Turing

Six Lessons What we learned about learning by observing people

-

Vision can help us learn from actions and develop a environment model and policy. We also want to incorporate perception in the control loop (Prof. Tomlin's work).

Lecture 2

Learning and Vision

2.1 3D Perception by Humans

Humans rely on many visual cues that provide information about the environment

- Accomodation
  - Change eye focal length
- Convergence
- Convergence angle vs distance
- Motion (optical flow)
  - Estimate derivative
  - Most important signal for depth
  - Can use it to segment occluded objects
- Binocular Stereopsis
  - Stereo vision with our eyes
- Pictorial Cues
  - We can percieve surface normals, not just depth, through a combination of various pictoral cues
- Curvilinear Grouping
  - We can assume depth order since boundaries are smooth in nature
- Texture gradients
  - Different objects appear different in different distances; surface normals get modified
- Shading
  - Changes in shade can make us percieve depth

The main point is that we can't effectively use these features by default; we need to learn how to do so.

2.2

Jitendra likes to emphasize the 3 R's of Vision

- Recognition
  - Classification and categorization of objects in the scene
  - Determining properties of objects as well (density, best grasping points)
- Reorganization
  - Making sense of the objects we know are there
- Reconstruction
  - Recovering 3D scene from image
  - The inverse graphics problem (not just depth!)

People used to view all these We should view these three processes as interacting and each having their own feedback loops.

Deep learning has done a lot of nice work connecting recognition and reorganization

- Determine bounding box, mask, and label

Where do we go from here? 3D reconstruction still needs a lot of work.

- Classical work in 3D vision has given us some nice reconstruction
  - Blocks world
  - Morphable models
  - Shape from shading
  - Structure from motion
- Now deep learning gives us some nice things
  - Mesh R-CNN by Jitendra's gropu
    * Detect object
    * Guess voxels

"Vision can be solved because we do it all the time" - Jitendra

2.3 Robotics and Vision

"Robotics is the killer app for vision" - Yi

- Yi's first project as a PhD student was for self-driving
- In his MASKS book, talks about vision for cars and helicopters
- Need to make vision computational and mathematical

We will kind of be summarizing the MASKS book

- Images and geometry
  - First computer vision pepople were artists
  - Perspective projection ideas
  - A lot happened within the turn of the last century!
  - Ma comes in to help give unifying mathematical perspectie
  - Now we think about not using so many views, because even a single image gives us a lot of information
- Geometry for multiple images
  - "hat" matrix for cross product
  - Homogenous coordinates
  - Perspective projection equation
  - Homogenous representation of 3D line
  - Projection of line onto image is intersection of triangle plane and image plane
  - Specify plane with normal
  - Incidence relation from multiple cameras is the sufficient and necessary constraint for 3D reconstruction
  - We can construct matrices whose rank gives us our ability to reconstruct
  - Special case when points lie on the same plane
  - Iterated solving of depth and perspective
- Geometry for single images
  - Utilize symmetry in scene
  - Windows, walls, edges are straight/orthogonal features
  - Symmetry lets us infer the scene at multiple views
  - Group defintion of symmetric scene
  - Let's us build prceise geometry from few images

We kind of assumed that feature matching or initial views was given - this is where the other R's come in

- Traditional 3D reconstruction bypasses reconstruction problem
  - Suffer from lack of features
  - textureless objects
  - reflection
  - representation patterns (fence)
  - medium/large baseline
  - moving objects
- Deep learning seems to work work well for reconstruction tasks but...
- they are really only doing classification instead of reconstruction, not actually doing a great job
- Problem: we aren't utilizing priors about structure in a scene
- We actually don't care about depth of reconstructing every point in the image
- This can make the problem well defined
- Get wireframe from single image - this gives us good structure of the scene because we used priors and didn't try to get everything
- Learning helps us go from canny edge to wireframe
- Learning to find vanishing points
- Train netural network to identify plane of symmetry and use that to get correspondances
- Datasets are super important for AI projects, but it's kind of lacking for 3D scene
- HoliCity project