Lecture 2: CS677

Aug 24, 2017

1

Review

- Previous class
  - Course requirements
  - Assignments, grading
  - Adding more students to the class
  - Topics to be studied in class
  - Some problems of vision
- **TODAY ONLY: office hours 1-2PM**
- Today's objective
  - Some example state-of-art apps
  - Human visual system (very briefly)
  - Image formation

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

Find Objects in this Image

- Where is the object of interest? (Figure-ground problem)
- Do we need to know we are looking for a bicyle?
- How do we know if the object is a bicycle?
  - Do we need to know bikes have two wheels, handlebar etc
  - If so, how do we find the wheels and the other parts?

Find Objects

- What is figure, what is ground?
- Different shape of bicycle, with a rider
- What color is the backpack of the rider?
- How far is the fence from the biker?

Additional Complexities

- Harder to segment figure from ground
- If we draw a box around bicycle, image will also have a car in it. Do we need to separate the two before we can recognize or do we recognize first and then separate?
- How far is the car from the bicycle?

More Problems of Vision

- Recovery of 3-D
- Variations in pose, illumination, camera properties..
- Dealing with occlusion
- Inference of surface properties (material)
- Dynamic scene analysis
- …

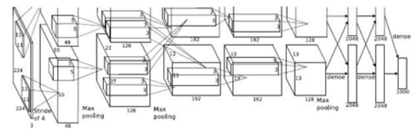
Mathematics or Machine_Learning?

- All vision problems can be stated as learning a function between input and output, say $\mathbf{y} = f(\mathbf{x})$
- If *f* can be described (or well approximated) by an analytical function, say a polynomial in case of scalar values, the task reduces to find the parameters of the function
- If the form of *f* can not be derived by analysis, then we can try to fit a complex, generic function with many degrees of freedom
  - Illustrate by example (fit curves to set of points)
  - This is the approach taken by machine learning, in particular deep learning
- Which is better?
  - If *f* is indeed a simple, derivable function, we can be confident of the solution; otherwise, it may "underfit" the data
  - Deep learning is susceptible to "overfitting" and requires huge amounts of training data
  - Transparency, ease of human interaction

Evolution of Computer Vision Approaches

- Early methods used representations based on intuition
  - "Hand-designed" descriptors and classification rules
- Later methods incorporated sophisticated mathematical models
  - These turned out to be very effective for recovering 3-D geometry from multiple images as problem is well posed mathematically
  - Less effective for semantic analysis such as object segmentation and recognition
    - Trend was to use hand-designed features but machine learned classifiers
- Current trend
  - Let machine learn the complete pipeline though structure of the pipe is still defined by designers
  - Achieves much higher accuracies when sufficient training data is available but methods are not transparent; hard to find source of errors

"Alexnet"



- First deep learning network that achieved high object classification performance (2012)
- Large number of parameters (~100M)
- Intermediate layers are "hidden" (we don't know what the right values are, they may not represent any recognizable entities such as parts)

What are we going to Study?

- A combination of mathematical and learning methods
- More emphasis on mathematical methods in first part of the course as the geometry problems are relatively well-defined
- More emphasis on machine learning (deep learning) in second half as problems are not easy to describe in precise math terms
- Anticipation that future systems will use a combination of techniques so best to learn basic principles of both.

Current state of the art

- The following slides show some examples of what current vision systems can do
  - Many taken from class page of Prof. Seitz/Szeliski

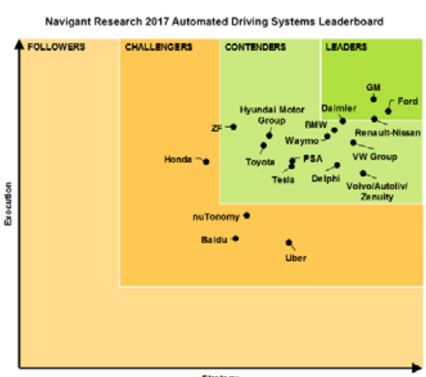
Driving Scene

From Mobileye

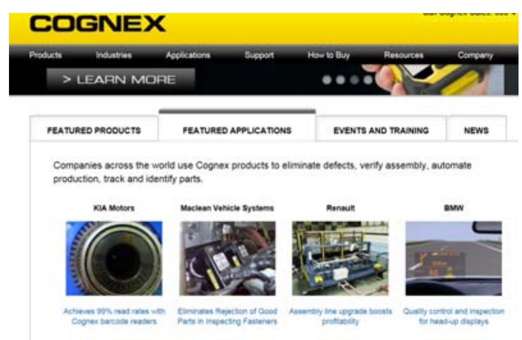
Self-Driving Cars

- A short video showing some visual needs and capabilities
  - https://www.youtube.com/watch?v=42rmGs0Rvtw
- A long talk on status of self-driving cars (watch on your own)
  - https://www.youtube.com/watch?v=GJ82mk99Agw
- A business analysis of participants in self-driving technology
  - http://www.businessinsider.com/the-companies-most-likely-to-get-driverless-cars-on-the-road-first-2017-4/#1-ford-18

Autonomous Driving Leaders



Manufacturing and Inspection



Earth viewers (3D modeling)

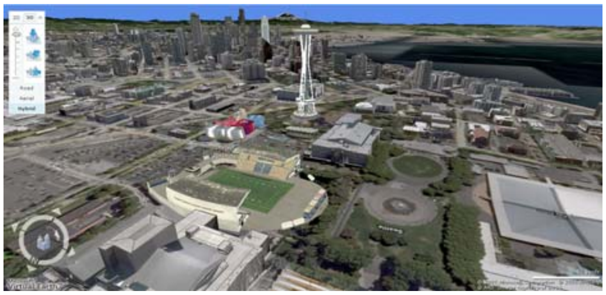


Image from Microsoft's Virtual Earth
(see also: Google Earth)

iPhone PANO Images

http://labs.live.com/photosynth/

Based on Photo Tourism technology

Optical character recognition (OCR)

- Technology to convert scanned docs to text
  - If you have a scanner, it probably came with OCR software

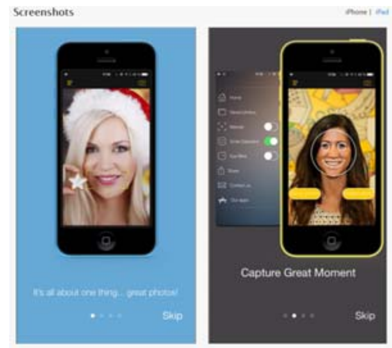
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

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

Face detection

- Most new digital cameras now detect faces

Smile Detector (from Quanticapps)



SensibleVision

Cognitec Face Recognition

A9.com

Visual Search also develops computer vision solutions that support Amazon initiatives along the entire product delivery pipeline: from the time a new product is photographed and added to our catalog to the time an item is bought and shipped to the customer.

Amazon Go

- https://www.youtube.com/watch?v=NrmMk1Myrxc

Special Effects

- From movie “Avatar”; image from www.rockying.com

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in.

Digimask: put your face on a 3D avatar.

"Game turns moviegoers into Human Joysticks", CNET
Camera tracking a crowd, based on this work.

Microsoft Kinect
(no images here)

Microsoft Kinect

Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

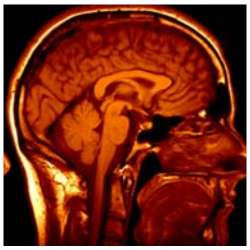
- Vision systems (JPL) used for several tasks
  - Panorama stitching
  - 3D terrain modeling
  - Obstacle detection, position tracking

Robotics

NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

http://www.robocup.org/

Medical imaging



3D imaging
MRI, CT

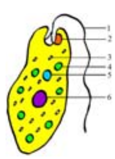
Image guided surgery
Grimson et al., MIT

33

5-Minute Break

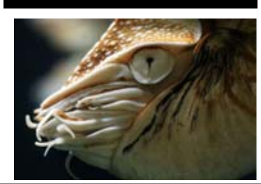
34

Simple Eyes



Single cell organism, can sense presence/absence of light only
http://www.wikiwand.com/en/Evolution_of_the_eye

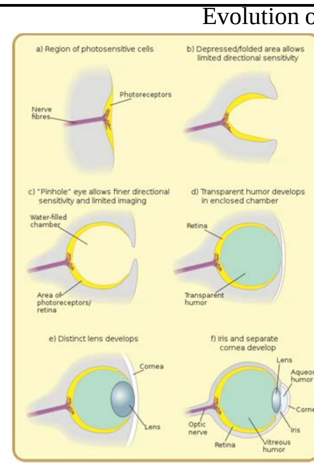
Holes provide some directional sensitivity.
From Alessandro: wikipedia



Nautilus Eye: like a pin hole camera
From Hillewaert: wikipedia

35

Evolution of Eyes

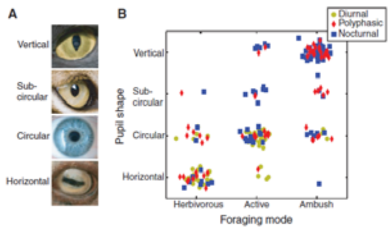


From Matticus: Wikipedia

Also,
http://www.wikiwand.com/en/Evolution_of_the_eye

36

Pupil Shape

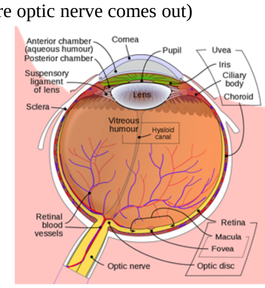


From Banks *et al*: “Why do animal eyes have pupils of different shapes?”, in Science Advances, August 2015

Human Eye

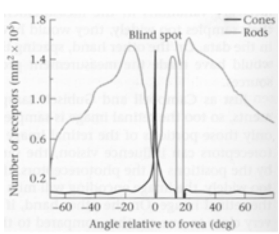
- Like a camera
  - Lens, pupil (iris), focus by *accommodation*
- Image formed on back of eye (retina)
- Optic nerve sends *data* to brain (cortex)
  - Blind spot (where optic nerve comes out)

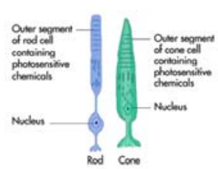
From Wikipedia



Retina

- Two types of photoreceptors
  - Rods: highly sensitive to light, not used for color vision, ~ 100M rods
  - Cones: 3 different types with different spectral sensitivities, less sensitive to light, ~ 5M cones
    - Explains why *color* is not seen at night
- Distribution is not uniform
  - High concentration of cones in fovea (0.5 minute visual angle)
  - Fixation (*foveation*) to get high resolution everywhere

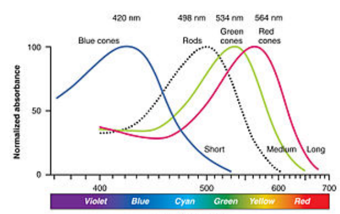




http://ionabio.weebly.com/

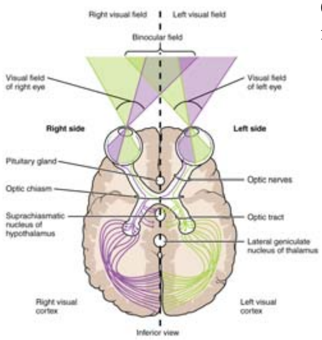
Color Sensor Response

- Eyes do not have built in color spectrometer
- Rather, we have 3 sensors with different responses to lights of different color
- Perceived color depends on relative responses of three sensors



https://en.wikipedia.org/wiki/Photoreceptor_cell

Cortex schematic



Optical nerve carries signals from retina to cortex

~100:1 ratio of nerve fibers to receptors: some processing performed at this level

Optical *chiasma*: optic nerve fibers split to two halves of the brain

Many functional areas (V1, V2,….): knowledge about them is limited

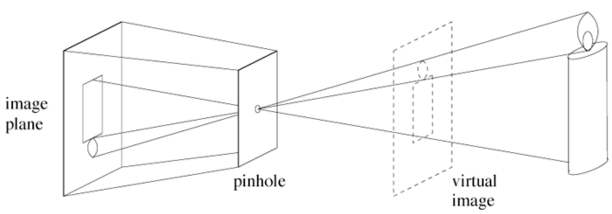
From lindsayoptometric.com

Image Formation

- Geometry
  - Where is the image of a point formed?
- Photometry/Colorimetry
  - How bright is the point?
  - What is its *color*?
- Ideal camera models
- Real lenses

Pinhole cameras

- Abstract camera model - box with a small hole in it
- Note inverted image
- Pinhole cameras work in practice, ignoring diffraction



Next Class

- FP: Sections 1.3, 2.1, 2.3.4, 2.4