Introduction

Introduction to Computer Vision
CSE 152
Lecture 1

- We'll start with some introductory material
- …. And end with
  - Syllabus
  - Organizational materials
  - Prequisites

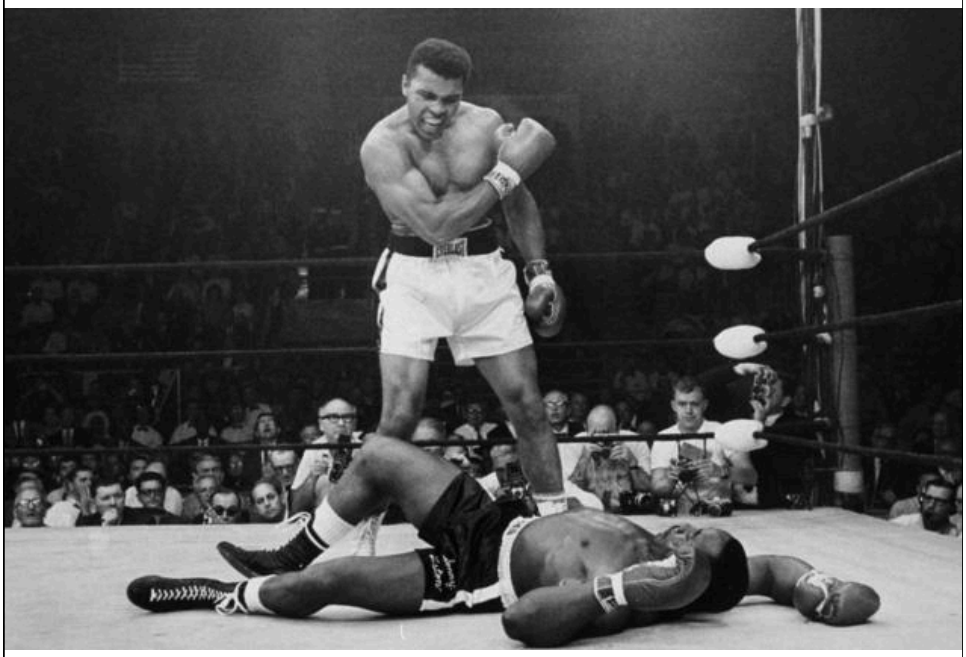
What is computer vision?

Add camera as input device to computer.

Done?

What is Computer Vision?

- Szeliski (Text): "In computer vision, we are trying to...describe the world that we see in one or more images and to reconstruct its properties."
- Trucco and Verri (secondary text): Computing properties of the 3-D world from one or more digital images
- Sockman and Shapiro: To make useful decisions about real physical objects and scenes based on sensed images
- Ballard and Brown: The construction of explicit, meaningful description of physical objects from images.
- Forsyth and Ponce: Extracting descriptions of the world from pictures or sequences of pictures"



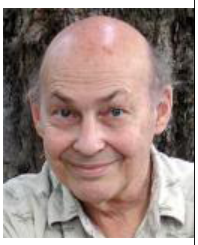
Why is this hard?

What is in this image?

1. A hand holding a man?
2. A hand holding a mirrored sphere?
3. An Escher drawing?
4. A 1935 self portrait of Escher

- Interpretations are ambiguous
- The forward problem (graphics) is well-posed
- The "inverse problem" (vision) is not

The challenge has been underestimated: We all make mistakes.



- "In 1966, Minsky hired a first-year undergraduate student and assigned him a problem to solve over the summer: connect a television camera to a computer and get the machine to describe what it sees."
  - Crevier 1993, pg. 88

- "640K ought to be enough for anybody."
  – Bill Gates, 1981

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

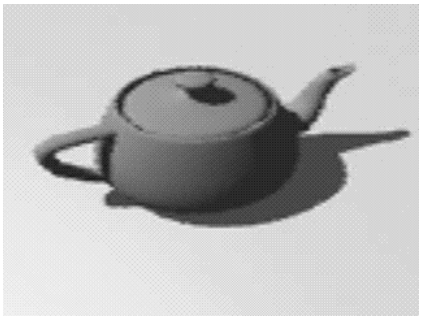
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

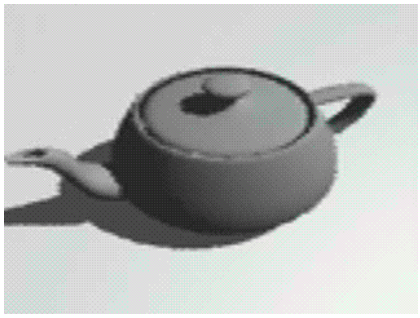
What do you see?

- ❑ Changing viewpoint
- ❑ Moving light source
- ❑ Deforming shape



What was happening

- ❑ Changing viewpoint
- ❑ Moving light source
- ☑ Deforming shape



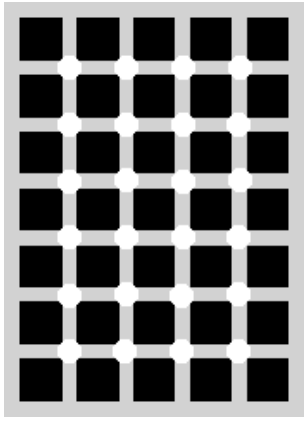
Should computer vision follow from our understanding of human vision?

Yes & No

1. Who would ever be crazy enough to even try creating machine vision?
2. Human vision "works", and copying is easier than creating.
3. Secondary benefit – in trying to mimic human vision, we learn about it.

1. Why limit oneself to human vision when there is even greater diversity in biological vision
2. Why limit oneself to biological when there may be greater diversity in sensing mechanism?
3. Biological vision systems evolved to provide functions for "specific" tasks and "specific" environments. These may differ for machine systems
4. Implementation – hardware is different, and synthetic vision systems may use different techniques/methodologies that are more appropriate to computational mechanisms

Hermann Grid



Scan your eyes over the figure. Do you see the gray spots at the intersections? Stare at one of them and it will disappear.

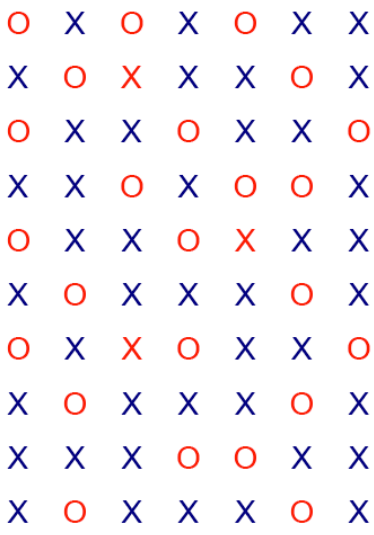
Why does this happen? Is it useful? Eye does not seem be acting like a camera an faith fully capturing scene?

How many red X's are there?

Raise your hand when you know.

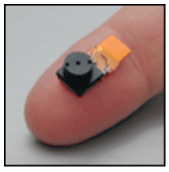
How many red X's are there?

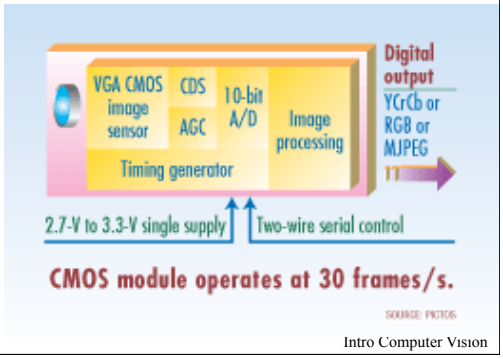
Raise your hand when you know.



Ubiquitous Vision

- Digital video has become really cheap.
- It's widely embedded in cell phones, cars, games, etc.
- 99.9% of digitized video isn't seen by a person.
- That doesn't mean that only 0.1% is important!
- And there's an enormous amount of image and video content on the internet…





Applications: touching your life

- Optical Character Recognition
- Football
- Movies
- Surveillance
- HCI – hand gestures
- Aids to the blind
- Face recognition & biometrics
- Road monitoring
- Industrial inspection
- Virtual Earth; street view
- Robotic control
- Autonomous driving
- Space: planetary exploration, docking
- Medicine – pathology, surgery, diagnosis
- Microscopy
- Military
- Remote Sensing
- Digital photography
- Google Goggles
- Video games

Vision to explore the world

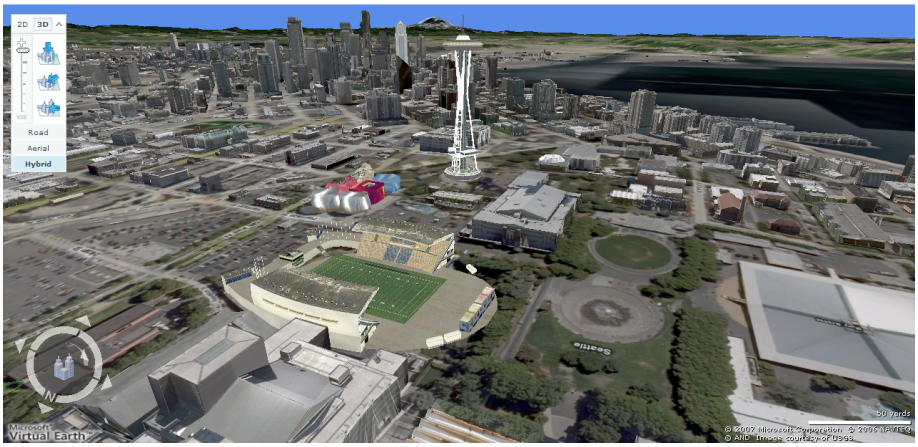


Image from Microsoft Virtual Earth
(see also: Google Earth)

Vision to explore other worlds

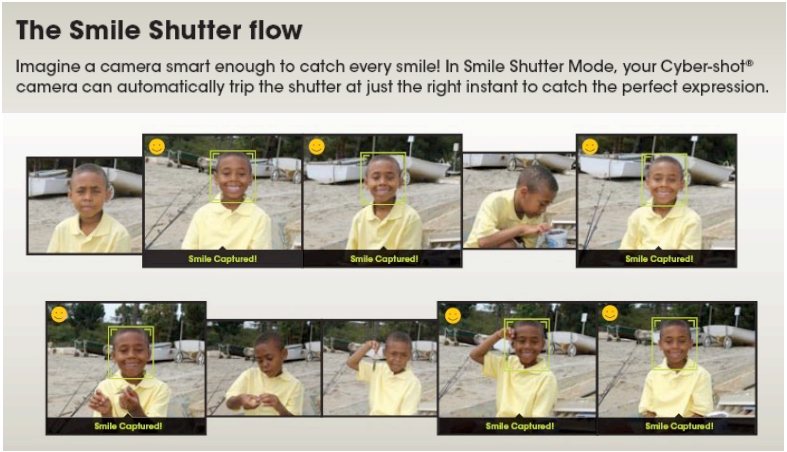
Image from NASA's Mars Exploration Rover Spirit

- Panorama stitching
- 3D terrain modeling
- Obstacle detection,
- Position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Vision to look at people

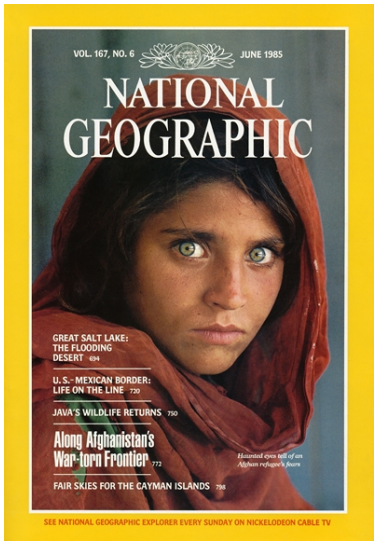
- Digital cameras, smart phones, Facebook, Google Photos, Snapchat etc.

Smile detection



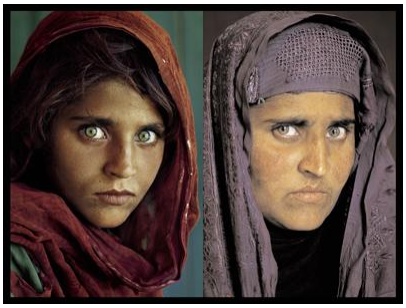
Sony Cyber-shot® T70 Digital Still Camera

Face recognition

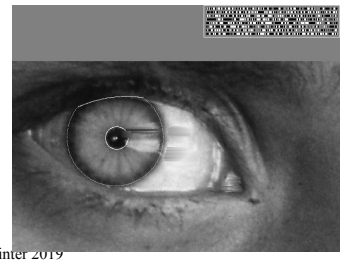


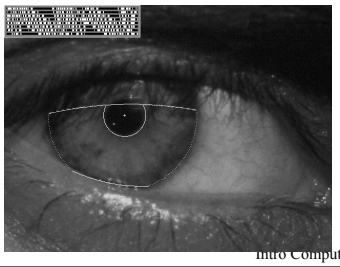
Who is she?

Vision-based biometrics

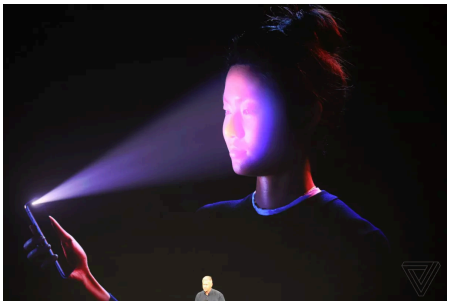


"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story

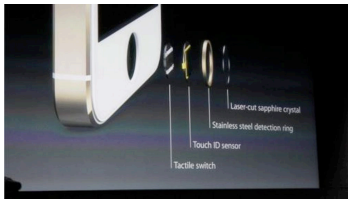




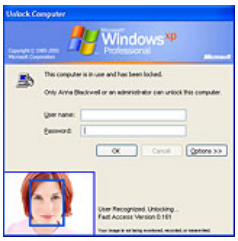
Login without a password…



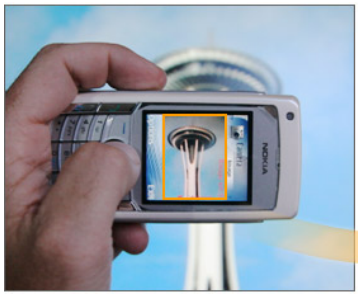
iPhone X



Fingerprint scanners on
laptops,
mice, other devices

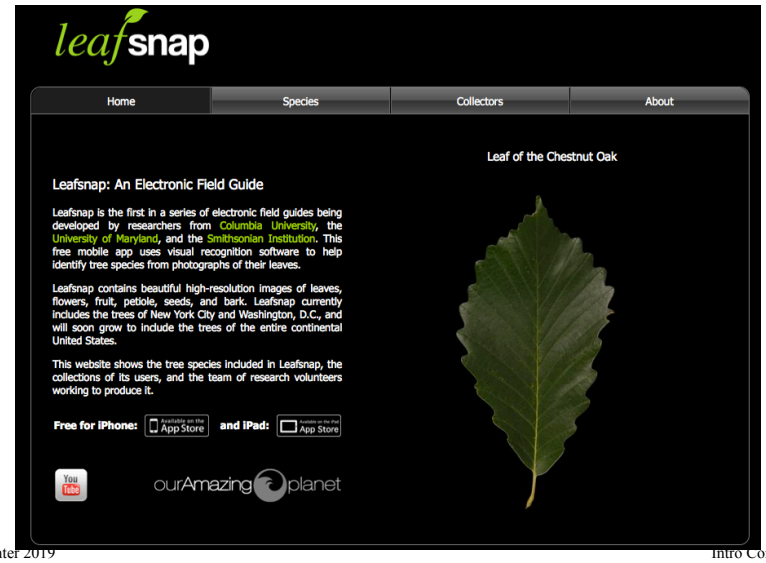


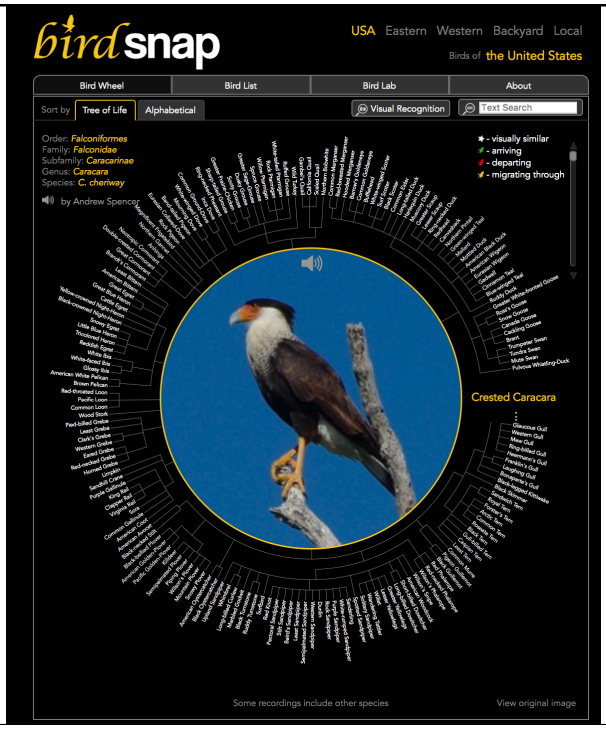
Vision to look at stuff: Object recognition



- Point & Find, Nokia
- SnapTell.com (now Amazon)
- Mobile Acuity
- Google Photos
- Apple Photos

Visual Field Guides:
Leafsnap.com -> Dogsnap -> Birdsnap -> Merlin, iNaturalist, etc



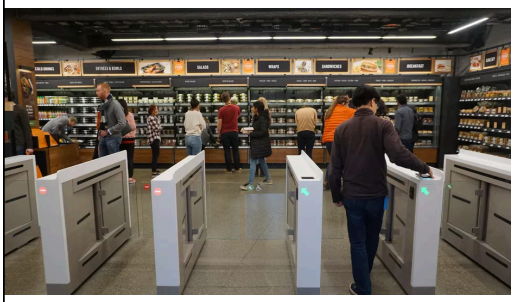


Object recognition (in supermarkets)

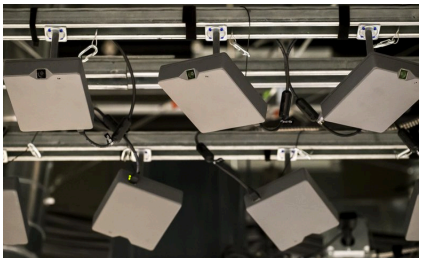
LaneHawk by EvolutionRobotics

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

Amazon Go



1. Turn-style entry. Consumer scans in with Amazon App on smartphone
2. Consumer goes around the store, picks up items, adds to bag, shops like normal
3. Consumer exits



Vision to create special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Vision to create special effects: motion capture

Facial Motion capture

Vision for Entertainment

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

Vision for augmented reality

- AR Toolkit
- Blippar
- Magic Leap
- Microsoft HoloLens

Augmented reality

- Text detection, localization, and translation, then render with similar font

Augmented reality

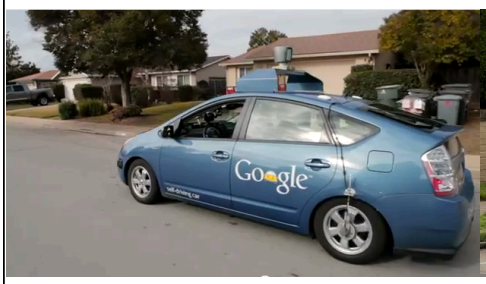
- Pokémon Go

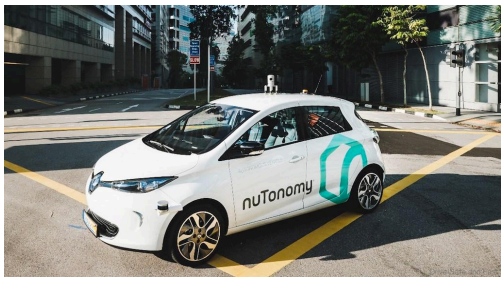
Vision for mart cars

- Mobileye
  - Vision systems currently in high-end BMW

Slide content courtesy of Amnon Shashua

Autonomous Cars





http://www.youtube.com/watch?v=cdgQpa1pUUE

Vision-based interaction (and games)

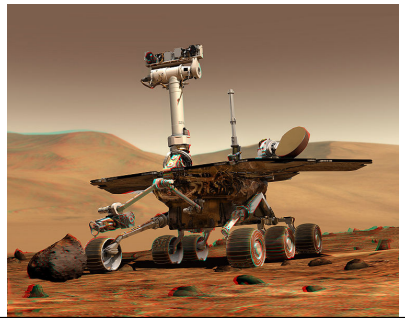
Nintendo Wii has camera-based IR tracking built in.

Digimask: put your face on a 3D avatar.

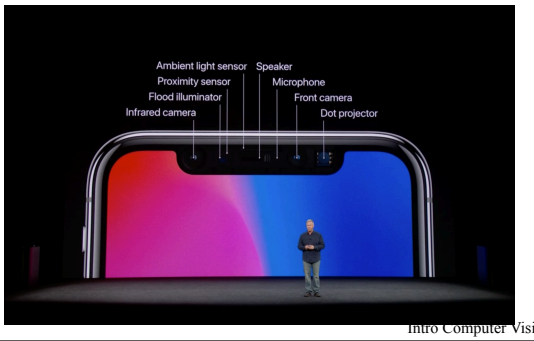
Playmotion game a Disney Epcot

Xbox Kinnect

Vision for Robotics

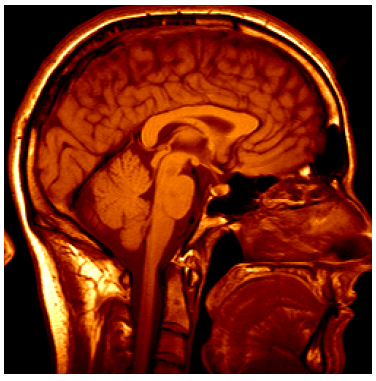


3D sensors



Vision from your persective: First person vision

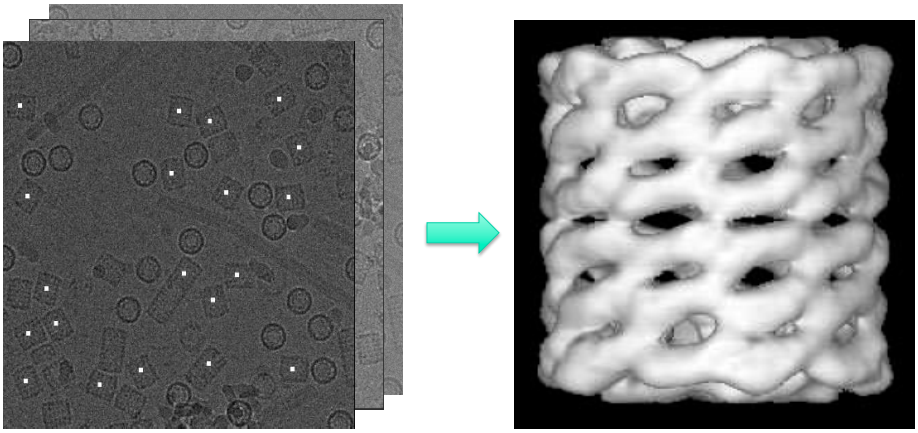
Vision for medicine



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

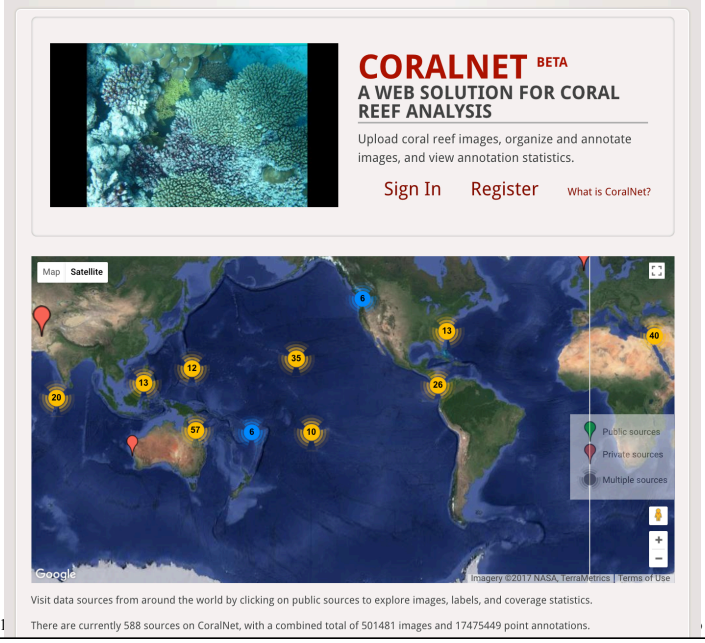
Vision for Science



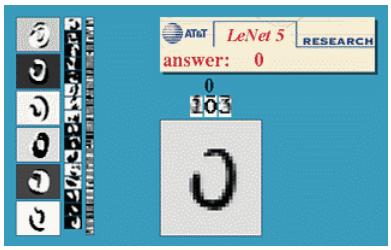
Molecular Reconstruction from
Cryo-electron Microscope Images

[Mallick, Zhu, Kriegman]

Coralnet.ucsd.edu



Vision to bridge communication between physical and digital worlds: Optical Character Recogntion (OCR)

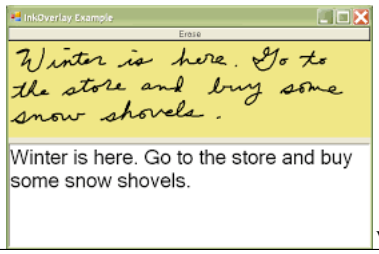


Digit recognition, AT&T labs
http://www.research.att.com/~yann/

Or more recent, see blog post about Dropbox OCR

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

Handwriting recognition

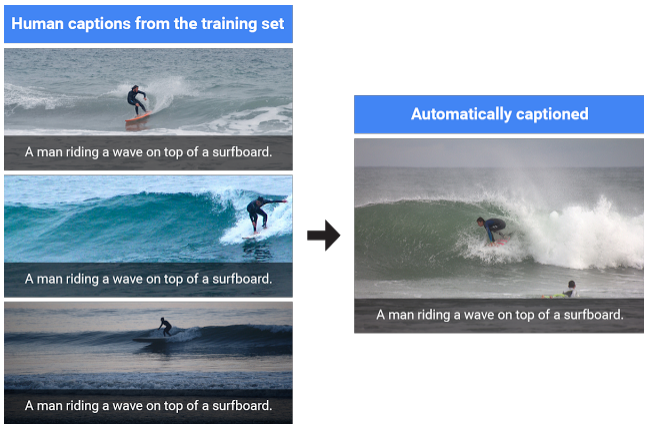


Scene Text: Text Recognition in the Wild

COCO-Text
A large-Scale Scene Text Dataset
https://bgshih.github.io/cocotext/

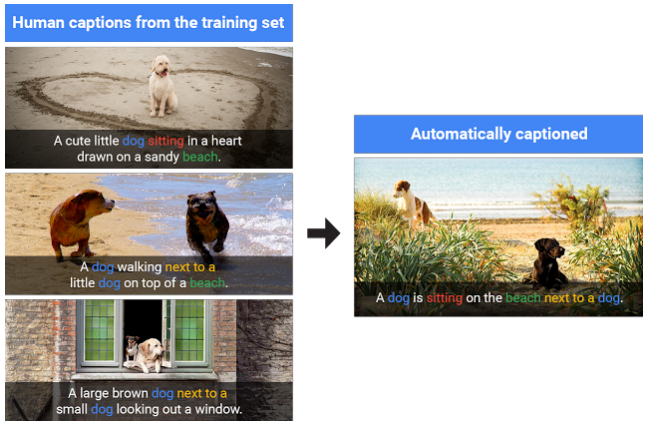
Automatic image captioning

- Deep learning



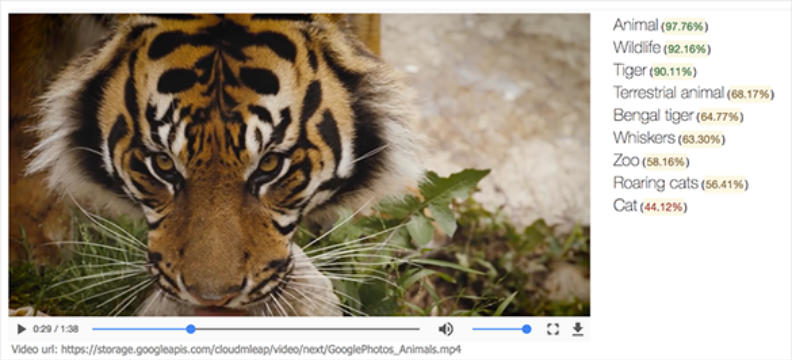
Automatic image captioning

- Deep learning

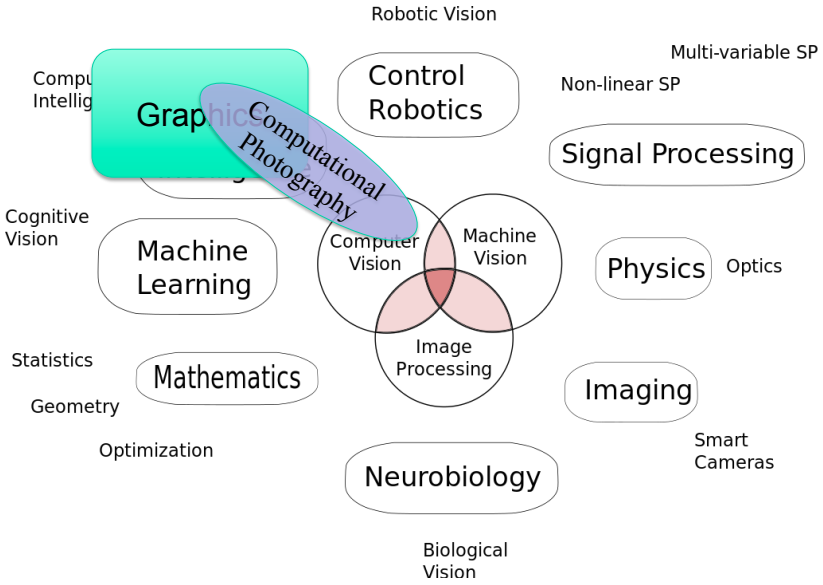


Video understanding

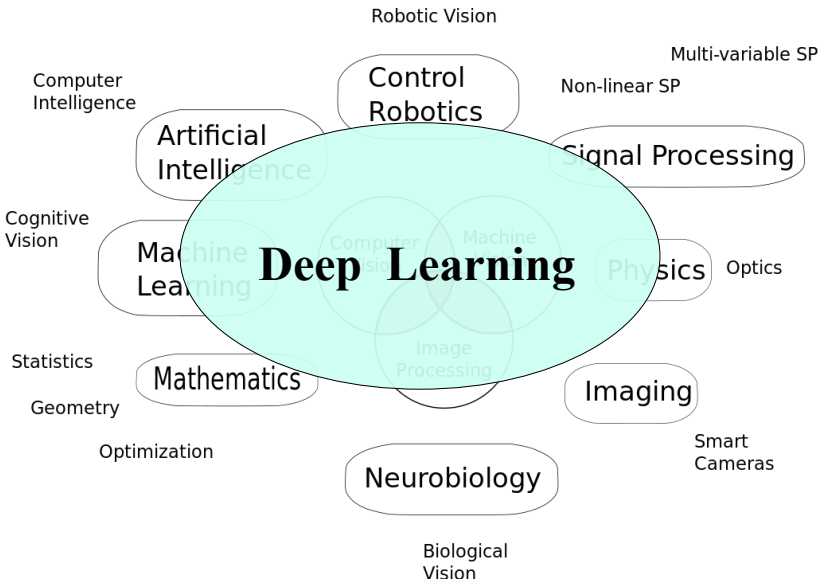
- Deep learning



Related Fields



Related Fields



Four Rs of computer vision

- Reprojection
  - Rendering a scene and features from a different view, under different illumination, under different surface properties, etc.
- Reconstruction
  - Multiple view geometry, structure from motion, shape from X (where X is texture, shading, contour, etc.), etc.
- Registration
  - Tracking, alignment, optical flow, correspondence, etc.
- Recognition
  - Recognizing objects, scenes, events, etc.

Others may have slightly different Rs

Rudiments: The implied fifth R

- image filtering
- interest point detection
- edge detection
- probability
- statistics
- linear algebra
- projective geometry
- optics
- Fourier analysis
- sampling
- algorithms
- photometry
- physics of color
- human vision
- psychophysics
- performance evaluation

How are images understood? Visual cues

- Variation in appearance in multiple views
  - stereo
  - motion
- Shading & highlights
- Shadows
- Contours
- Texture
- Blur
- Geometric constraints
- Prior knowledge

How do we solve computer vision problems?

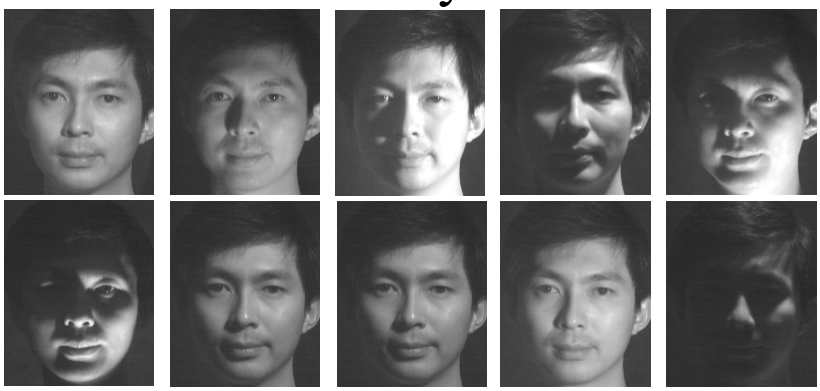
1. Build mathematical/physical model of the problem and implement algorithm with provably correct properties.
2. Craft a solution using software libraries of established methods and tailor them to the particulars of the problem.
3. Gather image data, potentially label it, and use machine learning to provide solution.

An example problem:

Shading as a result of differences in lighting is

1. A source of information
2. An annoyance

Illumination variability as an annoyance



"The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity."

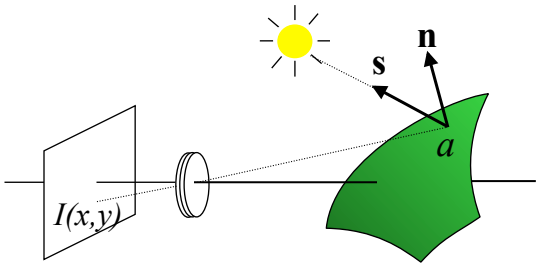
-- Moses, Adini, Ullman, ECCV '94

How do we understand shading?

(An idealization of "engineering" research)

1. Construct a model of the domain (usually mathematical, based on physics).
2. Prove properties of that model to better understand the model and opportunities of using it.
3. Develop algorithms to solve a problem that is correct under the model.
4. Implement & evaluate it.
5. Question assumptions of the model & start all over again.

Step 1. Build an Model: Image Formation Model



At image location *(x,y)* the intensity of a pixel *I(x,y)* is

$$I(x,y) = a(x,y)\ \mathbf{n}(x,y) \bullet\ \mathbf{s}$$

where

- *a(x,y)* is the albedo of the surface projecting to *(x,y)*.
- **n***(x,y)* is the unit surface normal.
- **s** is the direction and strength of the light source.

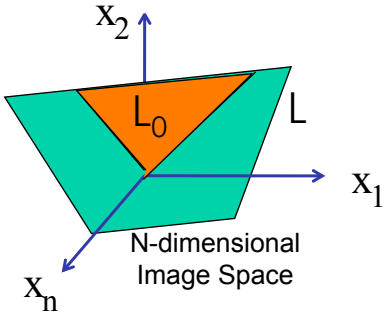
Step 2. Prove a property

The set of images of a Lambertian surface with no shadowing is a subset of 3-D linear subspace.

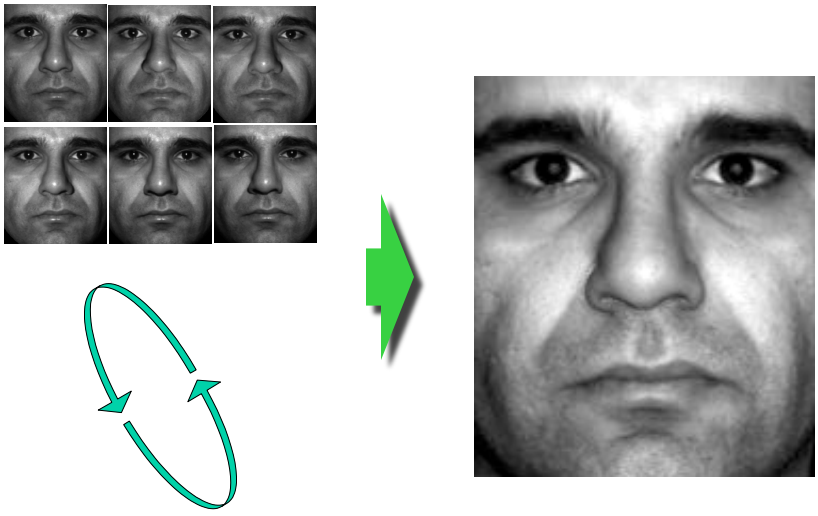
[Moses 93], [Nayar, Murase 96], [Shashua 97]

$$L = \{\mathbf{x} \mid \mathbf{x} = B\mathbf{s},\ \forall \mathbf{s} \in R^3\}$$

where B is a *n* by 3 matrix whose rows are product of the surface normal and Lambertian albedo



Steps 3,4 : Develop a provable algorithms and implement it: Relighting



Single Light Source

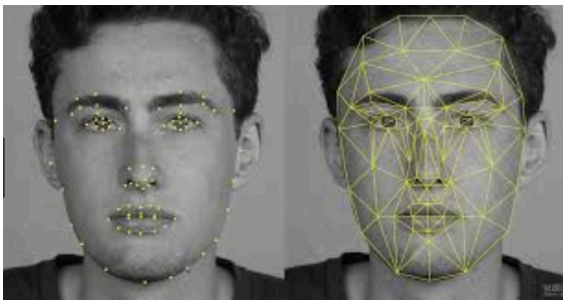
Steps 3,4 : Develop a provable algorithms and implement it: Photometric Stereo

Basic idea: 3 or more images under slightly different lighting

Step 5. Question Assumptions

- What about cast shaodws?
- Many objects are glossy or have more complex reflectance functions

Crafting a solution



- Distance between facial features (eye center, tip of nose, lip corners, eyebrows) should be good for recognition.
- Feature descriptors like SIFT or ORB are pretty insensetive to lighting and available in OpenCV
- Let's detect those features, build a Delaunay triangulation, and use an SVM for classification.

Use lots of data and machine learning

- Google FaceNet trained on hundreds of millions of cropped face images
- 140 Million Parameters
- 1.6B Flops

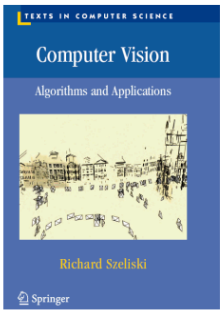
[Schroff, Kalenichenko, Philbin, 2015]

About the Course

Primary Text 1

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research



Welcome to the Web site (http://szeliski.org/Book) for my computer vision textbook, which you can now purchase at a variety of locations, including Springer (SpringerLink, DOI), Amazon, and Barnes & Noble.

This book is largely based on the computer vision courses that I have co-taught at the University of Washington (2008, 2005, 2001) and Stanford (2003) with Steve Seitz and David Fleet.

You are welcome to download the PDF from this Web site for personal use, but **not** to repost it on any other Web site. Please post a link to this URL (http://szeliski.org/Book) instead. An electronic version of this manuscript will continue to be available even after the book is published. Note, however, that while the content of the electronic and hardcopy versions are the same, the page layout (pagination) is different, since the electronic

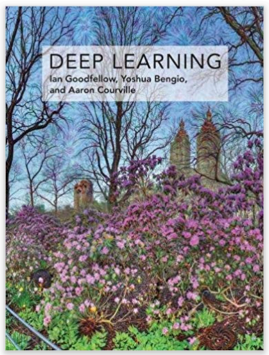
Primary Text 2

Deep Learning (Adaptive Computation and Machine Learning series)

by Ian Goodfellow (Author), Yoshua Bengio (Author), Aaron Courville (Author)

193 customer reviews

Look inside



eTextbook $51.57 | Hardcover $24.96 - $28.80 | Other Sellers See all 3 versions

Buy used — prime $24.96

Buy new — prime $28.80

Get FREE Delivery **Tomorrow** if you order within 19 hrs 47 mins and choose this date at checkout. Details

In Stock.

Sold by Hardin Stores and Fulfilled by Amazon. Gift-wrap available.

prime

FREE One-Day Pickup. Details

List Price: $80.00

Save: $51.20 (64%)

39 New from $28.80

Select delivery location

Qty: 1

Add to Cart

Buy Now

Secondary Text

Available on Ereserves

The Syllabus

http://cseweb.ucsd.edu/classes/sp19/cse152-a/

Background

- Linear algebra
- Multivariable calculus
- Probability
- Programming (we'll use python)

Announcements

- HW0 to be posted on Tuesday
- Read Chapter 1, Szeliski

Academic Integrity Policy

Integrity of scholarship is essential for an academic community. The University expects that both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

Collaboration Policy

It is expected that you complete your academic assignments on your own and in your own words and code. The assignments have been developed by the instructor to facilitate your learning and to provide a method for fairly evaluating your knowledge and abilities (not the knowledge and abilities of others). So, to facilitate learning, you are authorized to discuss assignments with others; however, to ensure fair evaluations, you are not authorized to use the answers developed by another, copy the work completed by others in the past or present, or write your academic assignments in collaboration with another person. If the work you submit is determined to be other than your own, you will be reported to the Academic Integrity Office for violating UCSD's Policy on Integrity of Scholarship.