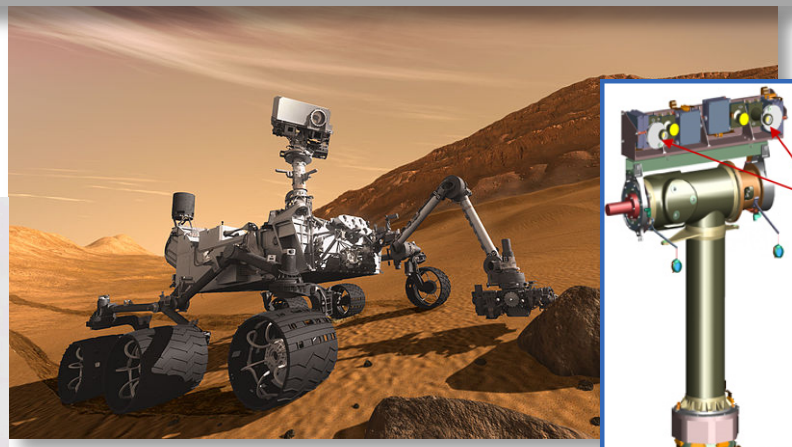
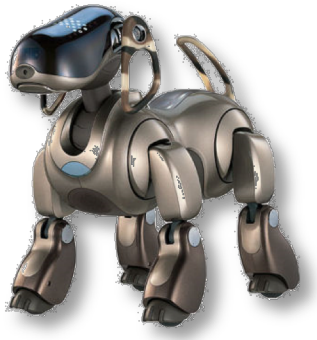
CSC589: Intro to Computer Vision

Instructor: Bei Xiao, American University, Spring, 2019





Today's class

- Introduction
- What is computer vision?
- Why vision is hard?
- Applications of computer vision
- What we will learn in this course

First week's task

- Finish the to-do list on First Steps
- Numpy tutorials
- Set up Python and OpenCV
- Finish the course survey (will send out today)

Instructor

- Bei Xiao (bxiao@american.edu)

- Research interests:
  - Human vision
  - Computer vision and graphics
  - Machine learning for estimation materials of objects
  - Tactile visual integration
  - VR

A bit about me

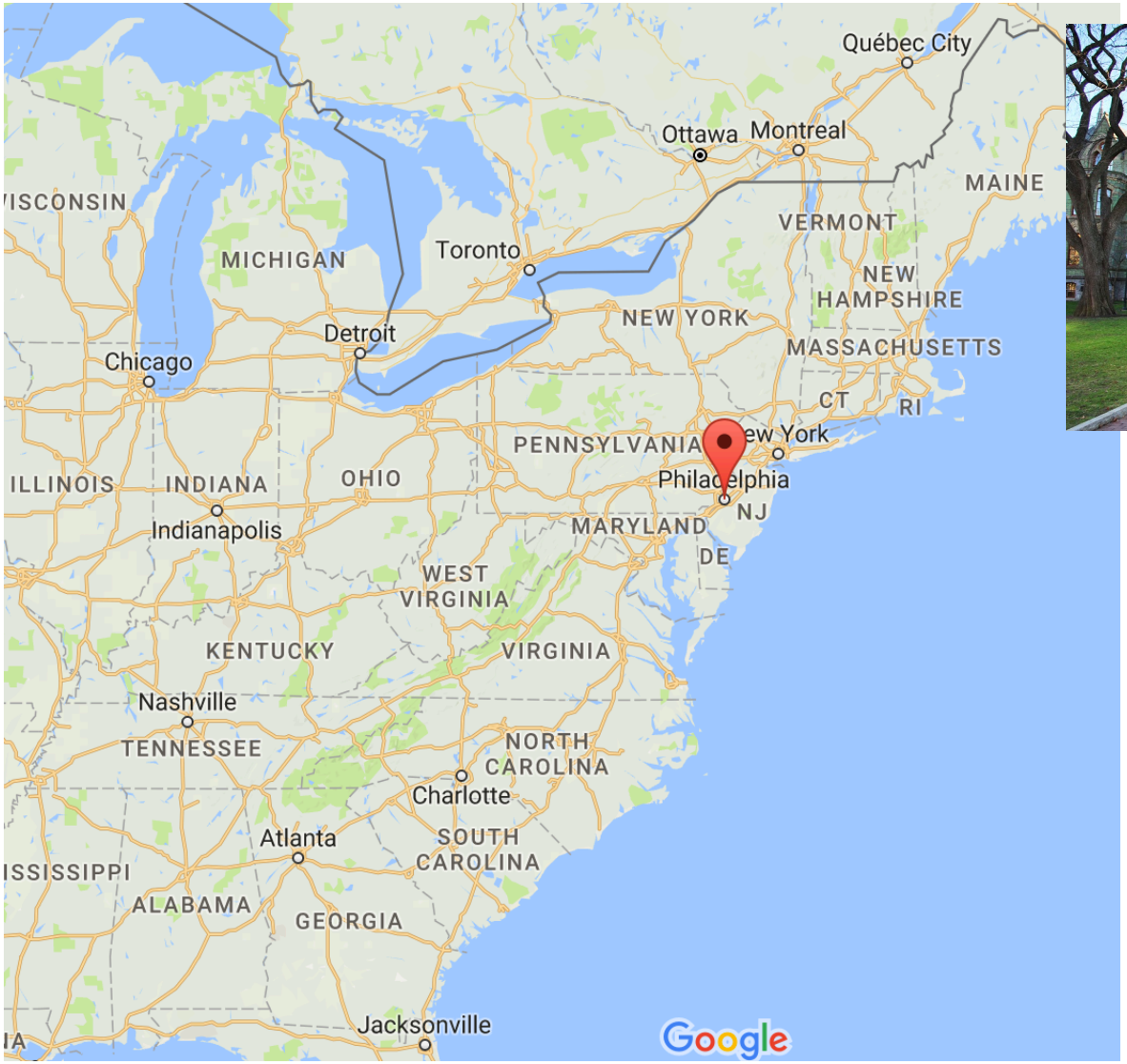
College:

Tsinghua University

Beijing, China

Degree: Chemistry

A bit about me



Graduate School:

University of Pennsylvania

Philadelphia, PA

Degree: Computational neuroscience

A bit about me

Postdoctoral Fellowship:

MIT

Cambridge, MA

Computer Vision and Human Perception

A bit about me

Research Sabbatical

University of Tuebingen

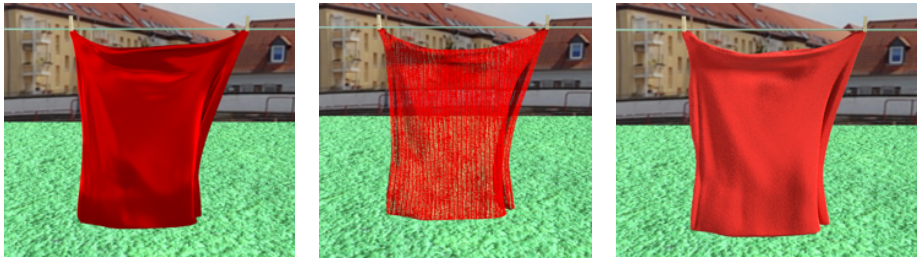
Tubeingen, Germany

Human Perception &
Computer vision

My research: vision and graphics

Physical-realistic rendering of cloth

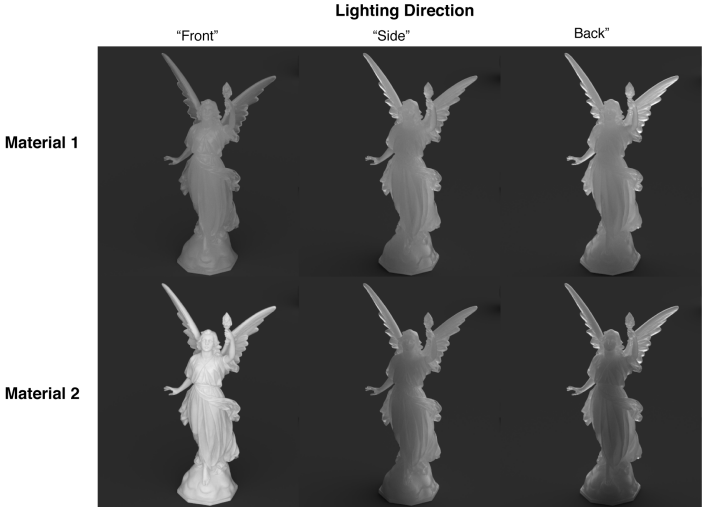
satin knit velvet



Inference of material properties

Visual and tactile material categorization

Understanding translucency



Why vision is important?

Visual cortex occupies about 50% of Macaque brain.

More human brain devoted to vision than anything else.

Visual information is essential in robotic system.

Large percentage of effort in neurobiology is dedicated to the visual system.

Every image tells a story

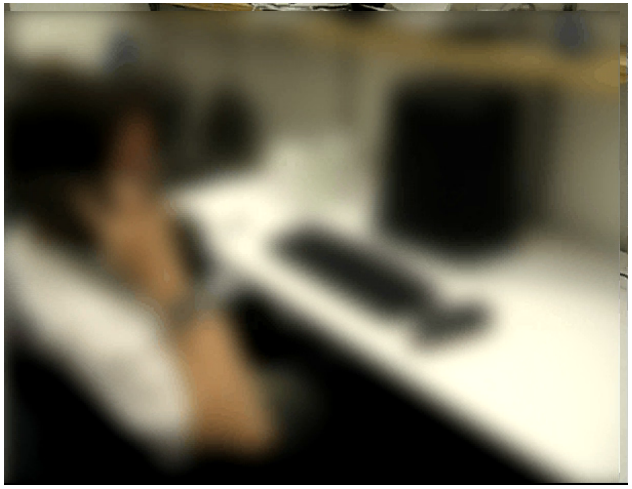
- Goal of computer vision: perceive the “story” behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

The goal of computer vision

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Can the computer match human perception?

- Yes and no (mainly no)
  - computers can be better at “easy” things
  - humans are much better at “hard” things
- But huge progress has been made
  - Accelerating in the last 4 years due to deep learning
  - What is considered “hard” keeps changing



What is computer vision?

What kind of food is it?

What kind of meat?

Are there any vegetables in the image?

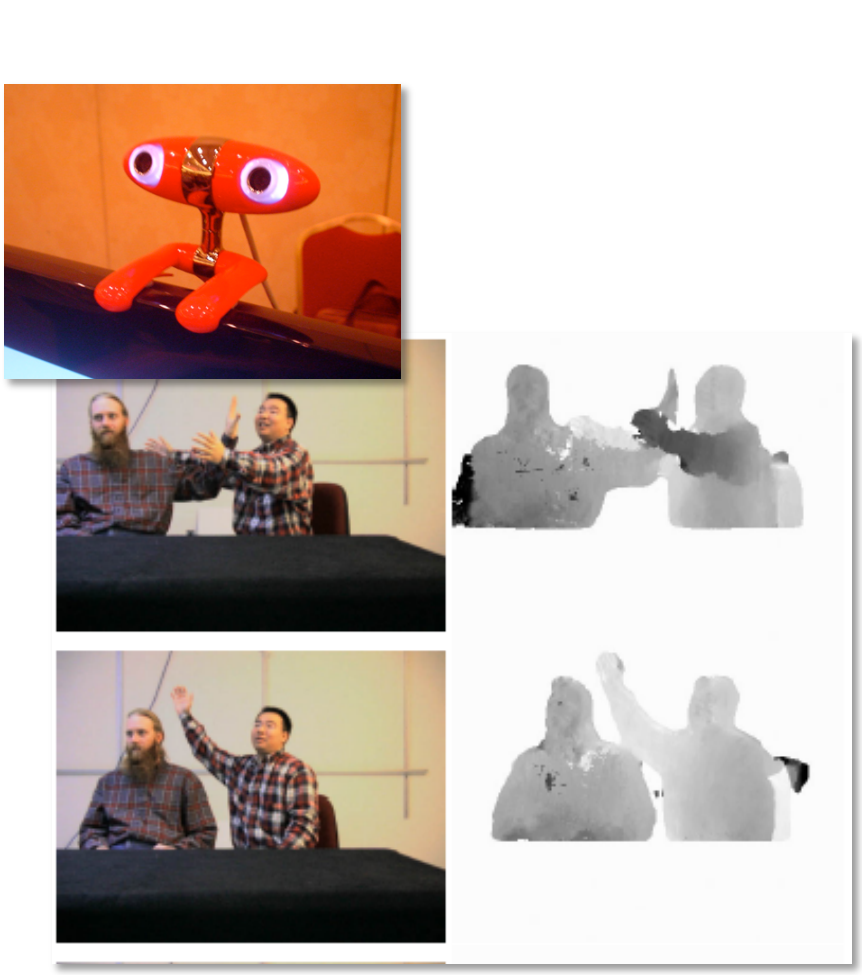
What kind of cuisine is it?

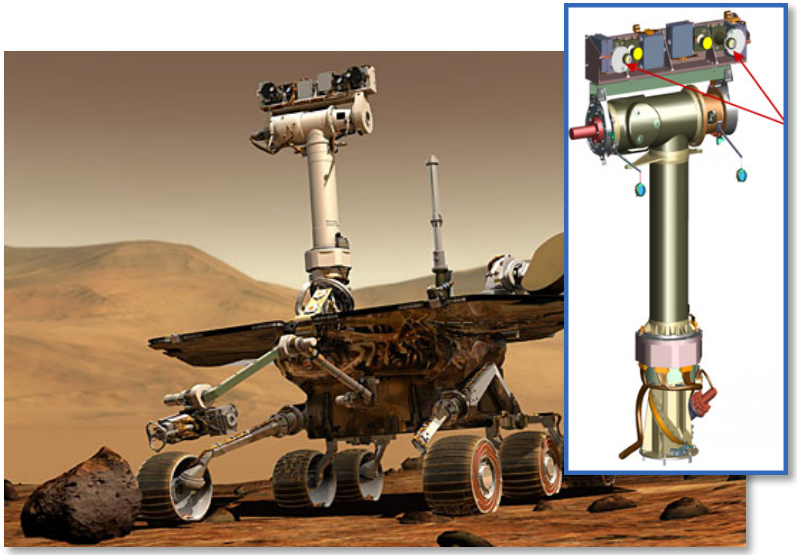
What are the most useful utensils to use to eat the food? Knife and fork?

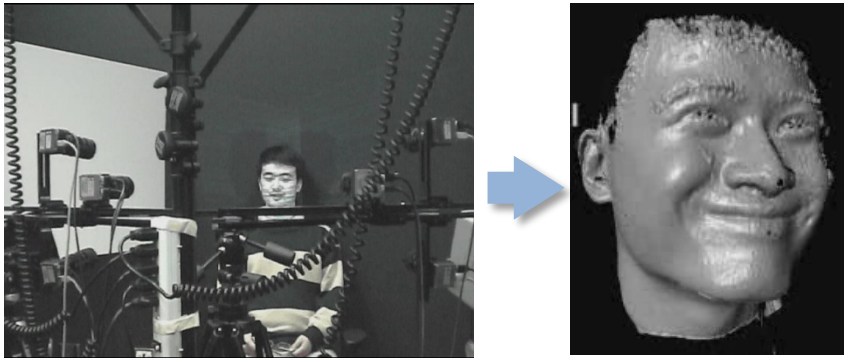
The goal of computer vision

The goal of computer vision

- Compute the 3D shape of the world





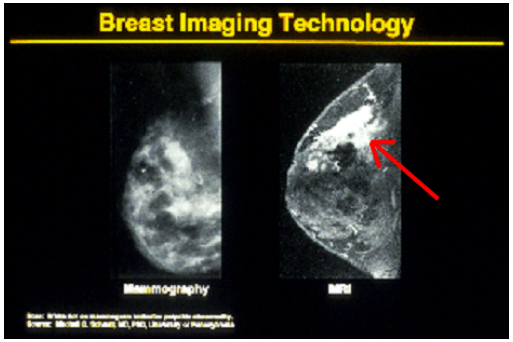


Slide courtesy from Noah Snavely

Slide courtesy from Noah Snavely

Computer Vision Applications

Safety

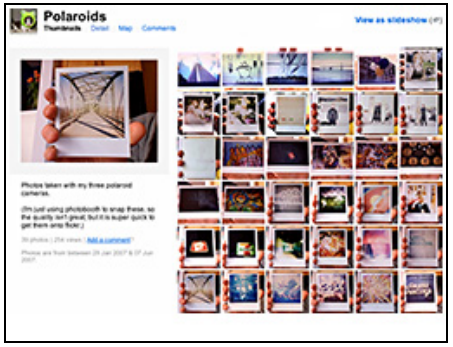


Health

Security

Comfort

Fun

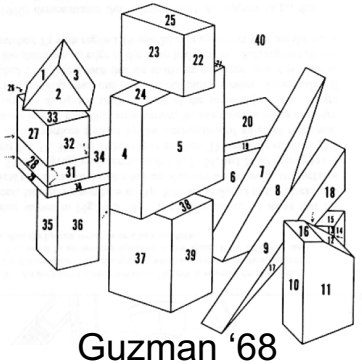


Access

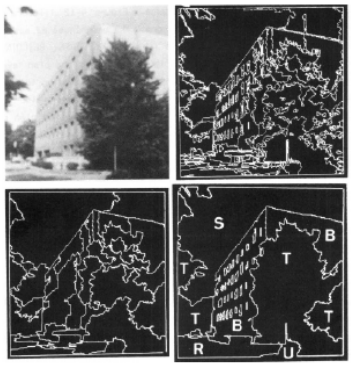
Slides from James Hayes

Ridiculously short history of CV

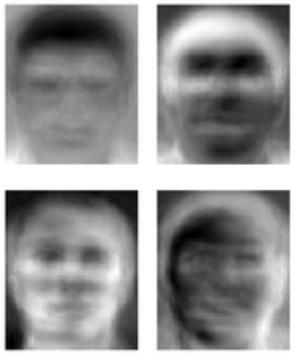
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: Deep learning with ConvNets
- 2030's: robot uprising?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

Self driving Cars

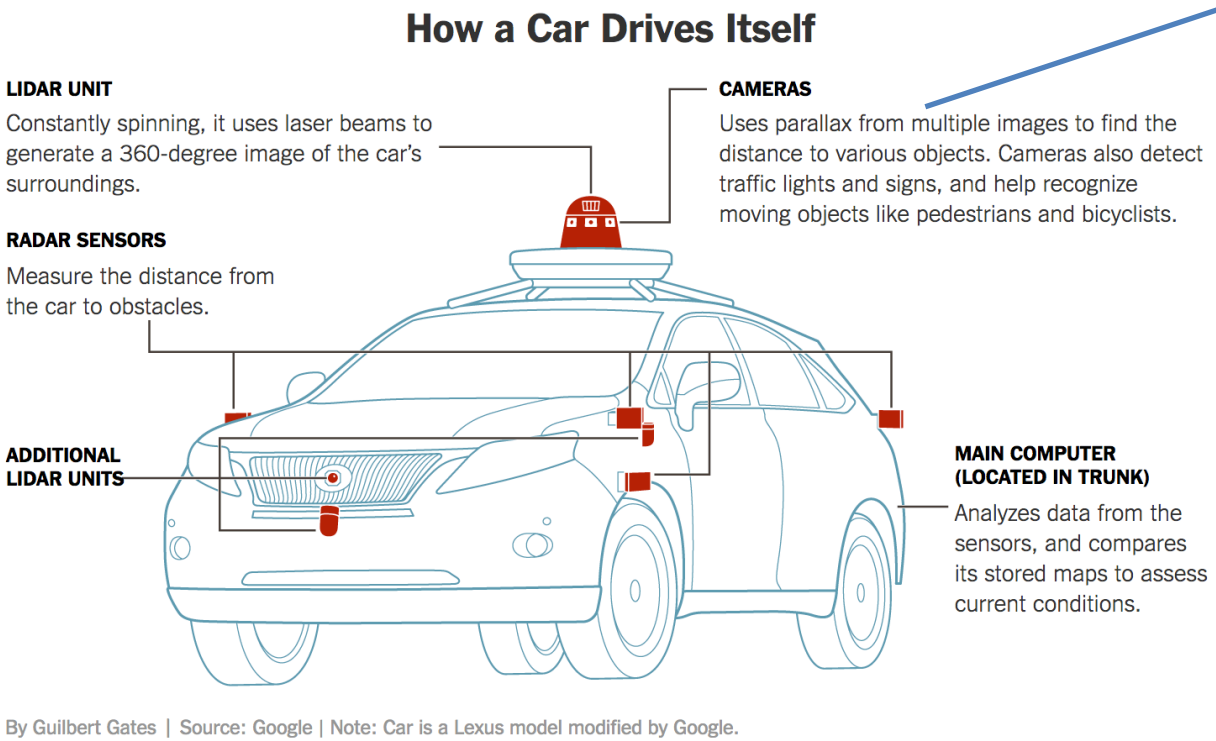
Oct 9, 2010. "Google Cars Drive Themselves, in Traffic". *The New York Times*. John Markoff

June 24, 2011. "Nevada state law paves the way for driverless cars". *Financial Post*. Christine Dobby

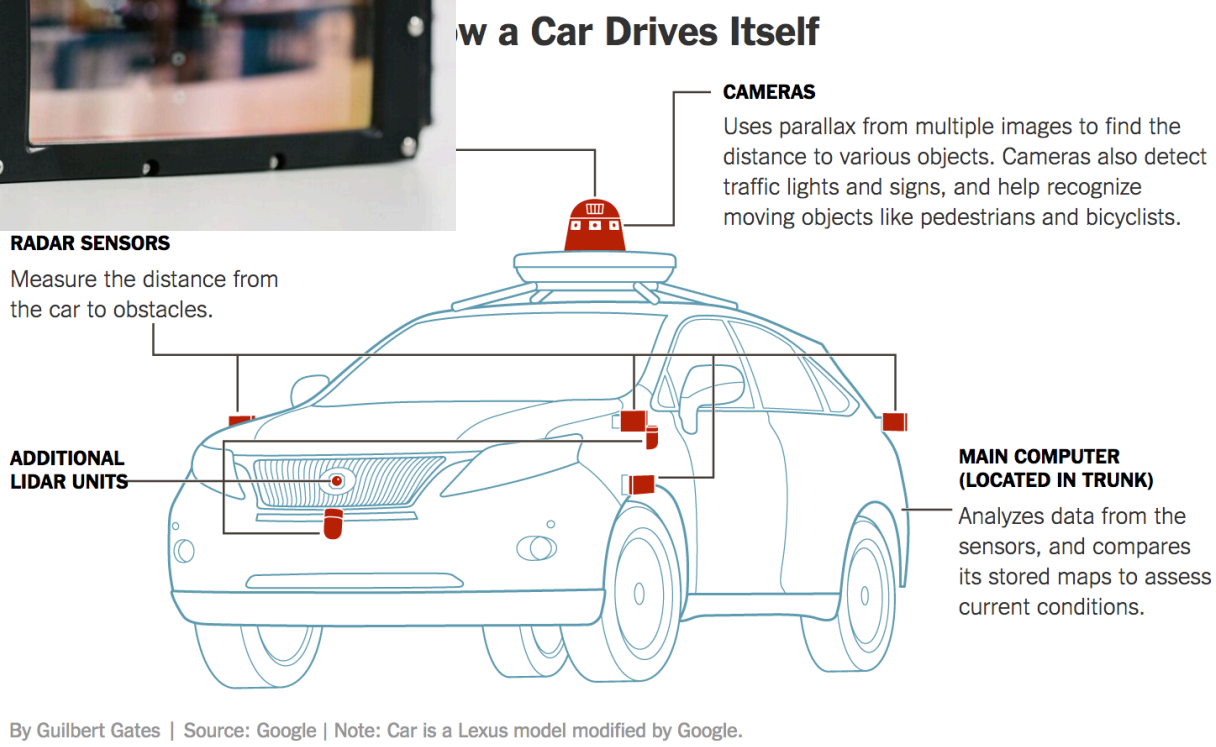
Aug 9, 2011, "Human error blamed after Google's driverless car sparks five-vehicle crash". *The Star* (Toronto)

How a car drives itself?

Detect
Moving
Targets
Traffic Lights

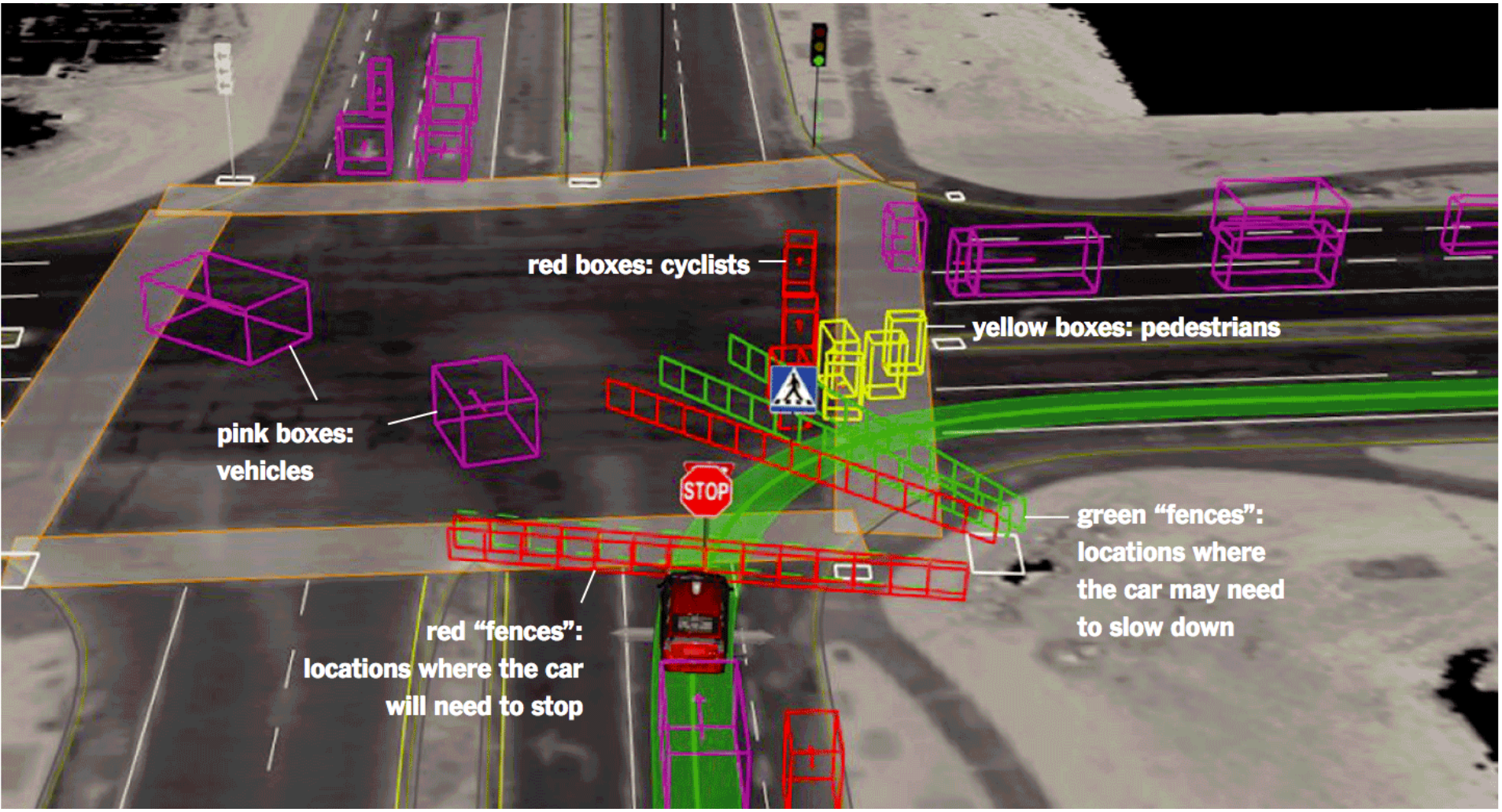


How a car drives itself?



For an excellent coverages:
https://www.nytimes.com/interactive/2016/12/14/technology/how-self-driving-cars-work.html
https://www.nytimes.com/2017/05/25/automobiles/wheels/lidar-self-driving-cars.html

What self-driving cars can see?



https://www.nytimes.com/2017/05/25/automobiles/wheels/lidar-self-driving-cars.html?_r=0

Image-based search

Query image

Visually similar images

Report images

Label: Residenzplatz, Salzburg, Austria

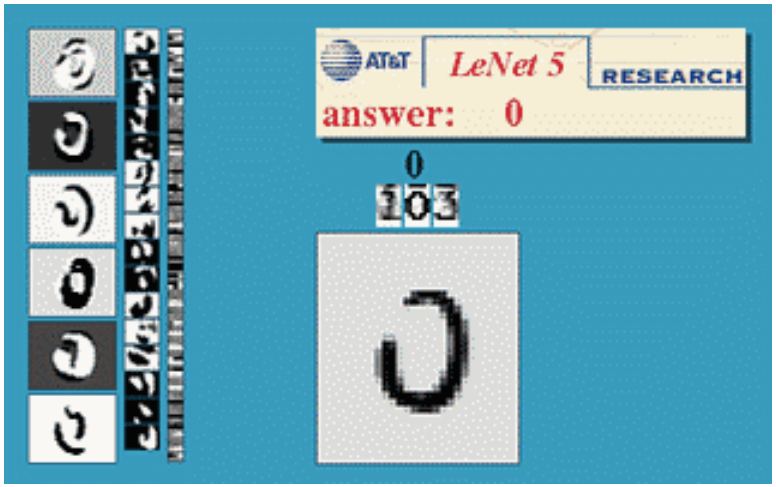
Usually compares image similarity distance such as similarity in color, texture, shape, etc.

Read about content-based image retrieval

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

Slide from James Hayes

Face Detection

Face Detection function keeps subjects' faces in **sharp focus**

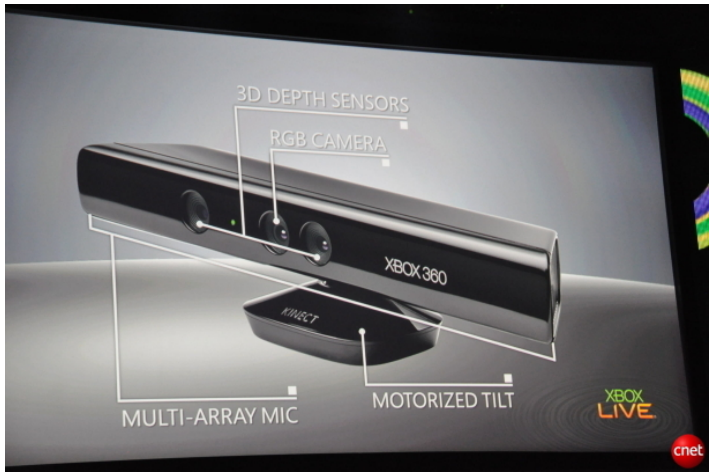
Object recognition (in supermarkets)

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… "

Slide from James Hayes

Motion Sensing: Kinect

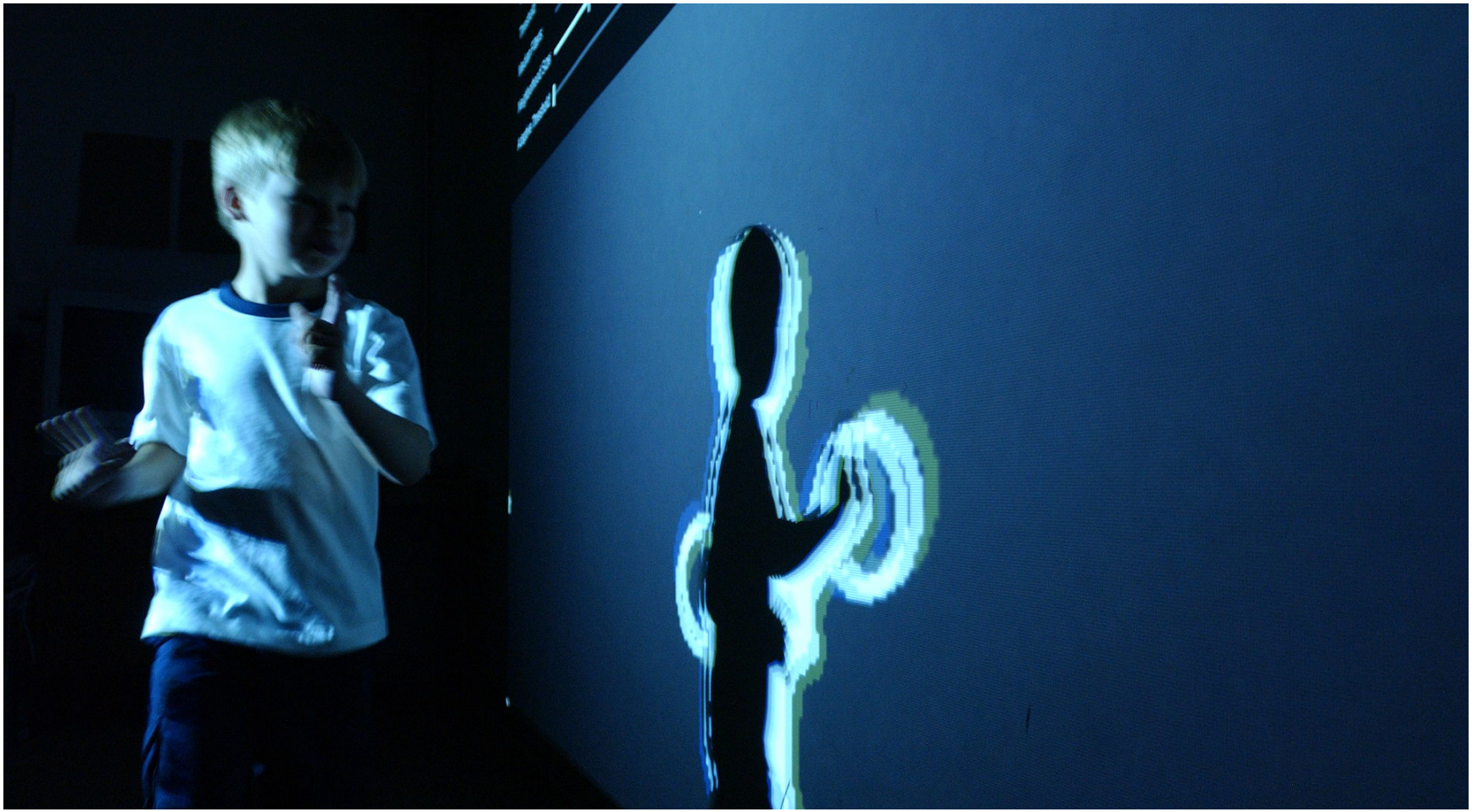


Robot : https://www.youtube.com/watch?v=w8BmgtMKFbY

Object recognition:
https://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o

Slide from James Hayes

Gesture Recognition



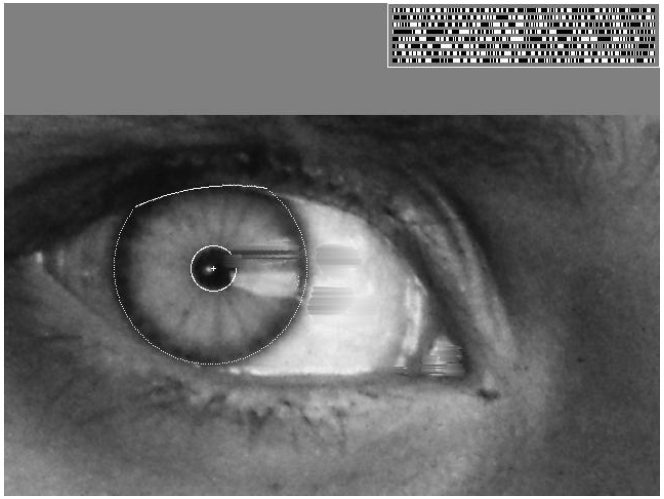
http://www.webopedia.com/TERM/G/gesture_recognition.html

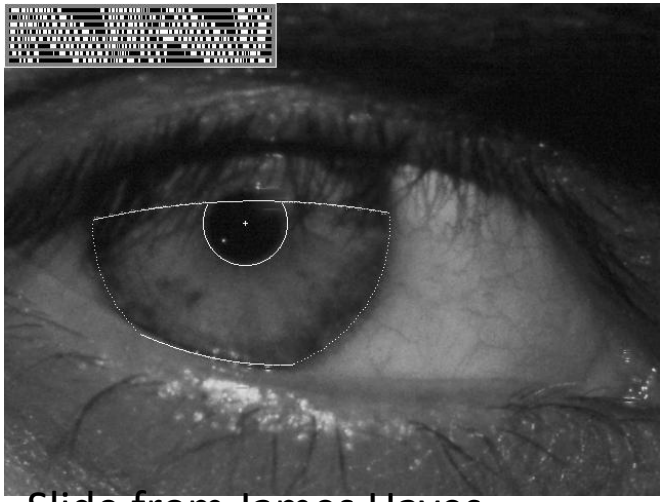
Special effects: motion capture

Dawn Of The Planet of The Apes, Director talks about the visual effects
https://www.youtube.com/watch?v=hlWyAePmAYM

Vision-based biometrics

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

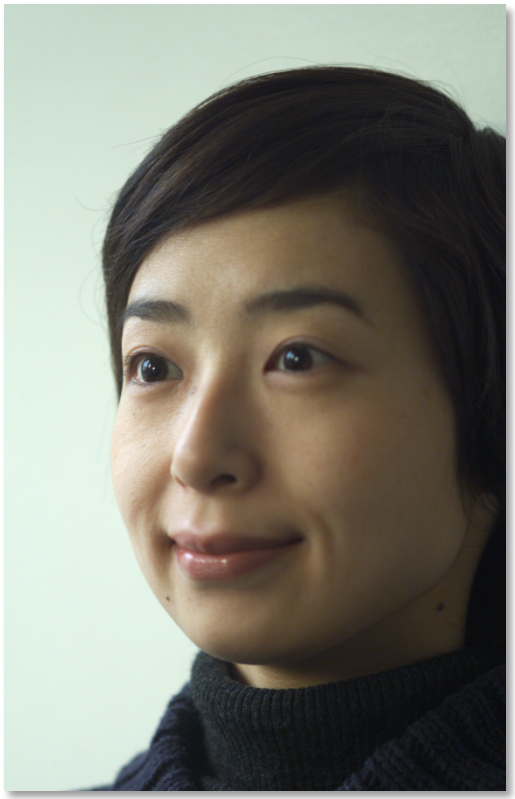




Slide from James Hayes

The goal of computer vision

- Forensics



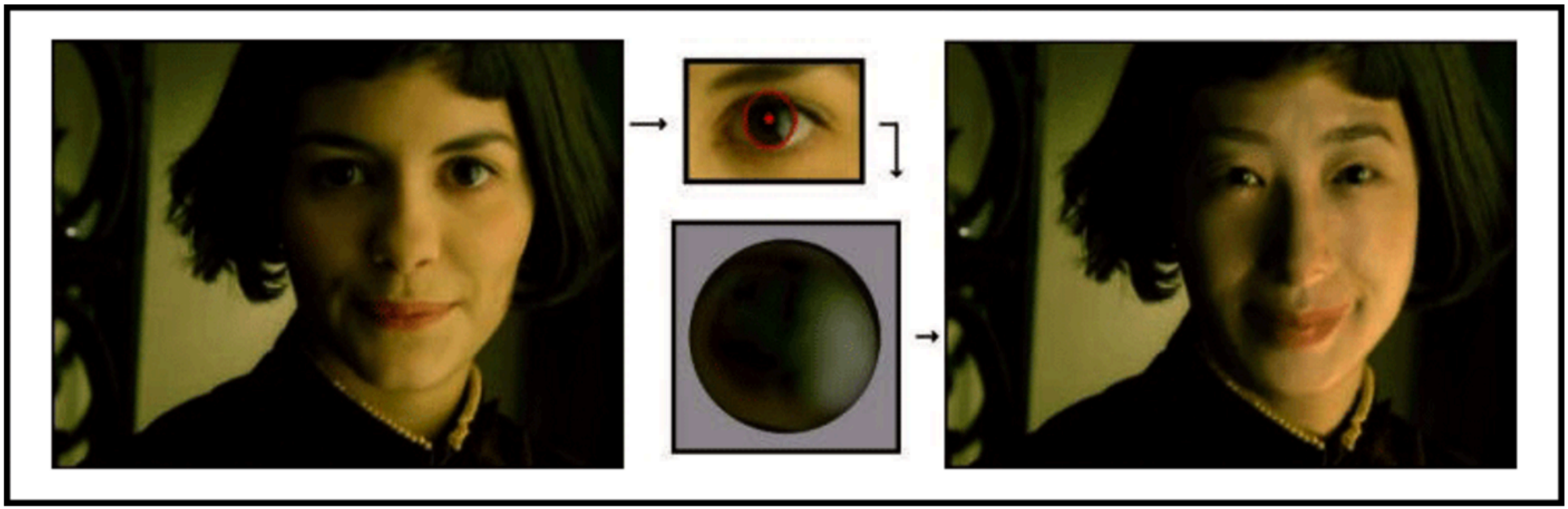
Source: Nayar and Nishino, “Eyes for Re

Source: Nayar and Nishino, “Eyes for Re

http://www1.cs.columbia.edu/CAVE/projects/eyes_relight/

Source: Nayar and Nishino, “Eyes for Relighting”

Eyes for Relighting



http://www1.cs.columbia.edu/CAVE/projects/eyes_relight/

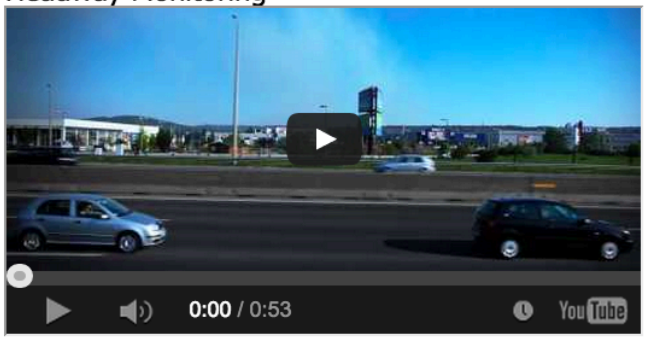
Mobile Eye for driving safety

Pedestrian Collision Warning

Lane Departure

Forward Collision Warning

Headway Monitoring



Mobileye Intelligent High-Beam

Speed Limit Indication

http://www.mobileye.com/technology/

leafsnap

Leaf of the Bottlebrush Buckeye

Leafsnap: An Electronic Field Guide

Leafsnap is the first in a series of electronic field guides being developed by researchers from Columbia University, the University of Maryland, and the Smithsonian Institution. This free mobile app uses visual recognition software to help identify tree species from photographs of their leaves.

Leafsnap contains beautiful high-resolution images of leaves, flowers, fruit, petiole, seeds, and bark. Leafsnap currently includes the trees of the Northeast and will soon grow to include the trees of the entire continental United States.

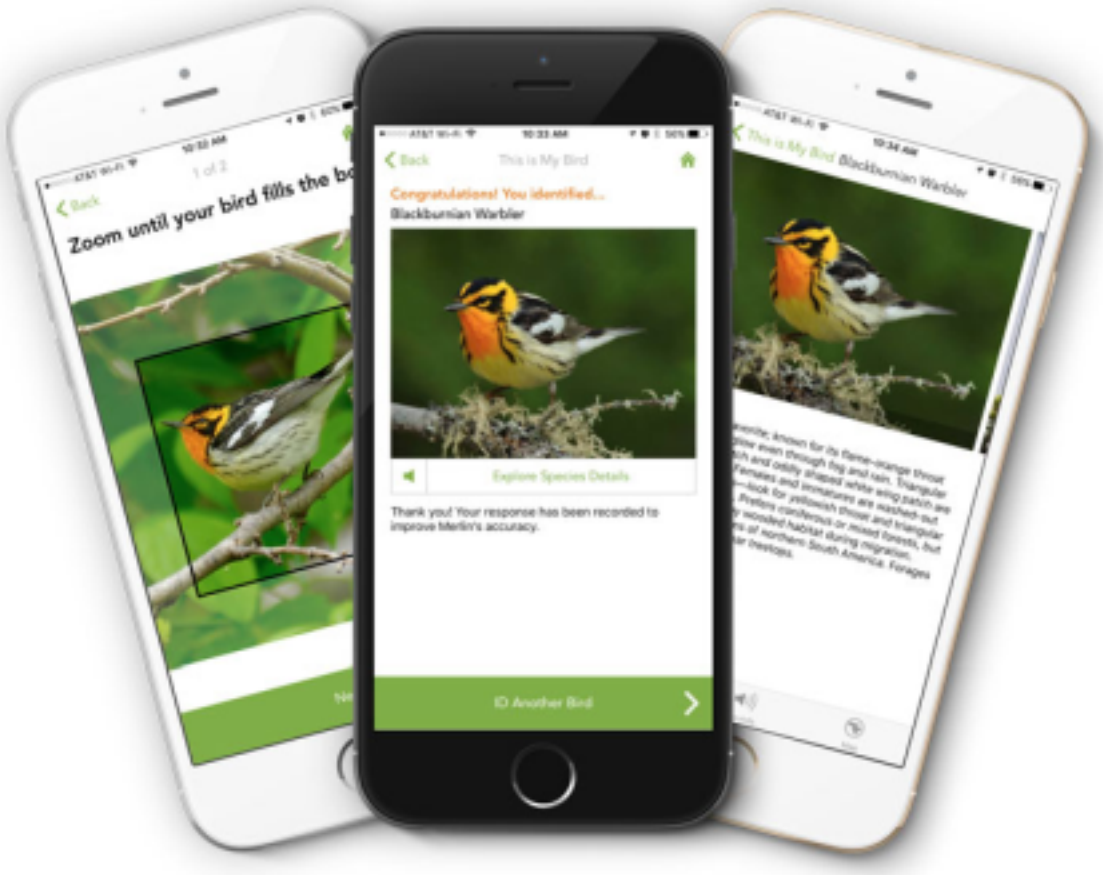
This website shows the tree species included in Leafsnap, the collections of its users, and the team of research volunteers working to produce it.

Free for iPhone: Available on the App Store **and iPad:** Available on the iPad App Store

You Tube

guardian.co.uk

Bird Identification



Merlin Bird ID (based on Cornell Tech technology!)

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
- For more, read "Computer Vision On Mars" by Matthies et al.

Slide from James Hayes

Automatic Image and Video Editing

Data driven approach of image editing

Given a single image at day

Input image at "blue hour" (just after sunset)

Database of time-lapse videos

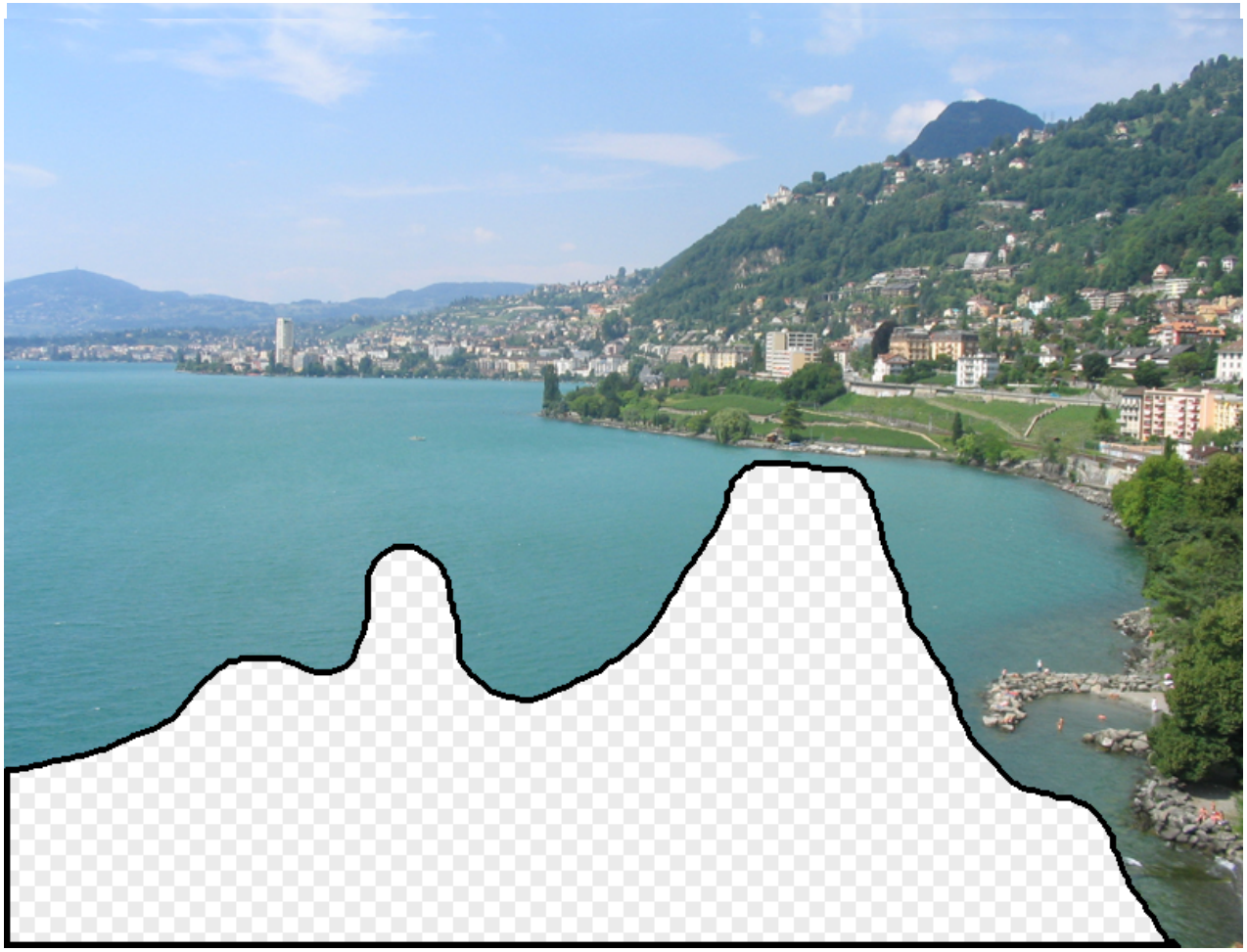
A database of time-lapse videos

A hallucination of the same scene at night

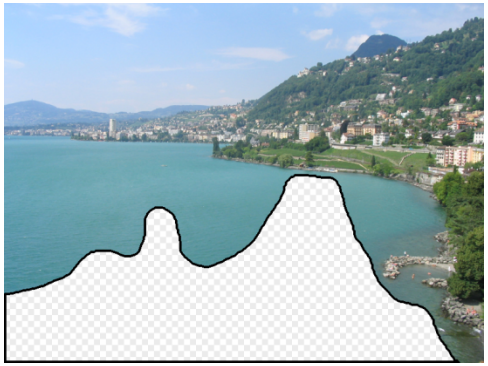
Hallucinate at night

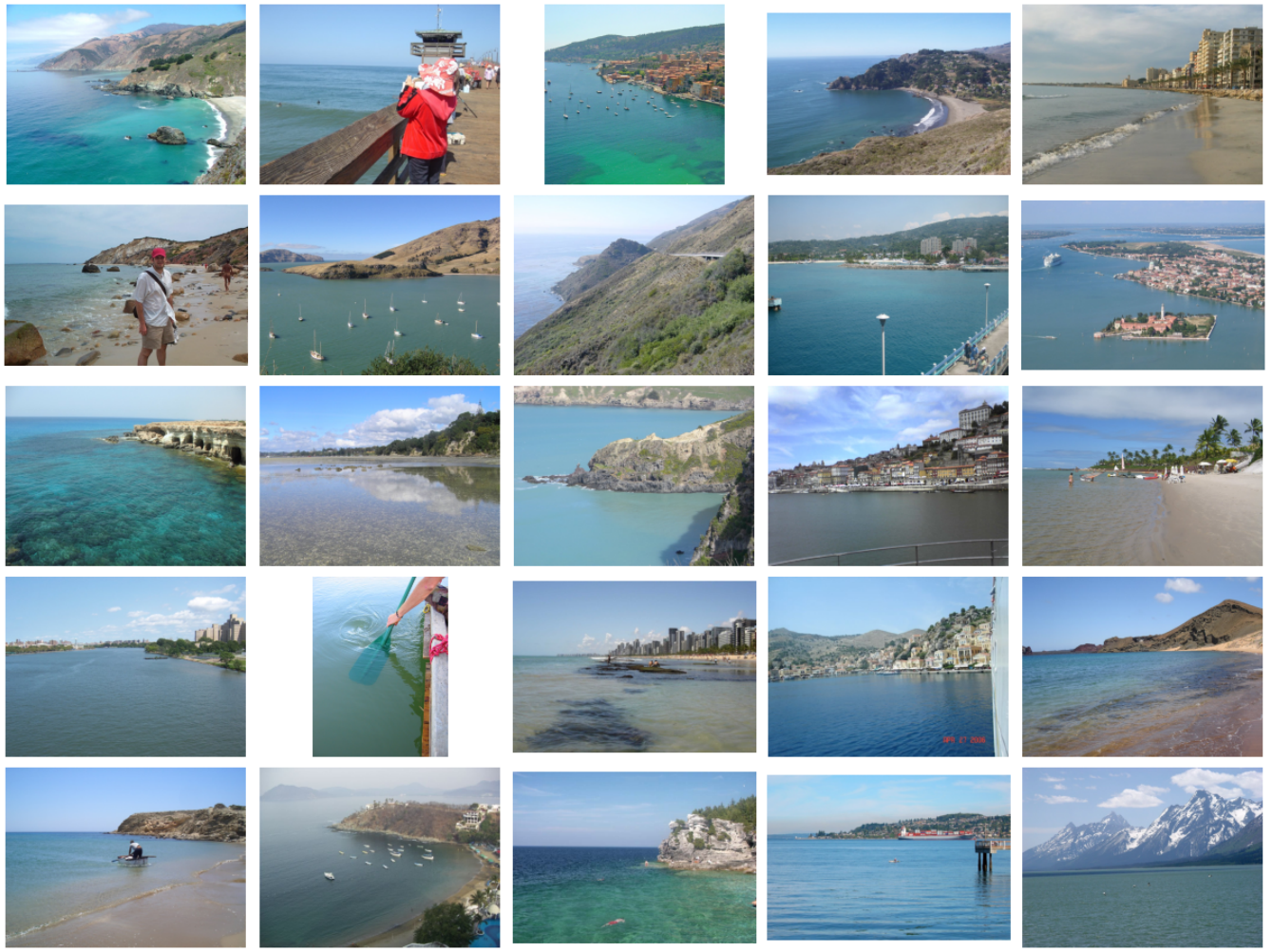
YiChang Shi, Sylvain Pairs, Fredo Durand, and William T Freeman, SIGGRAPH ASIA 2013

Scene Completion



[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]





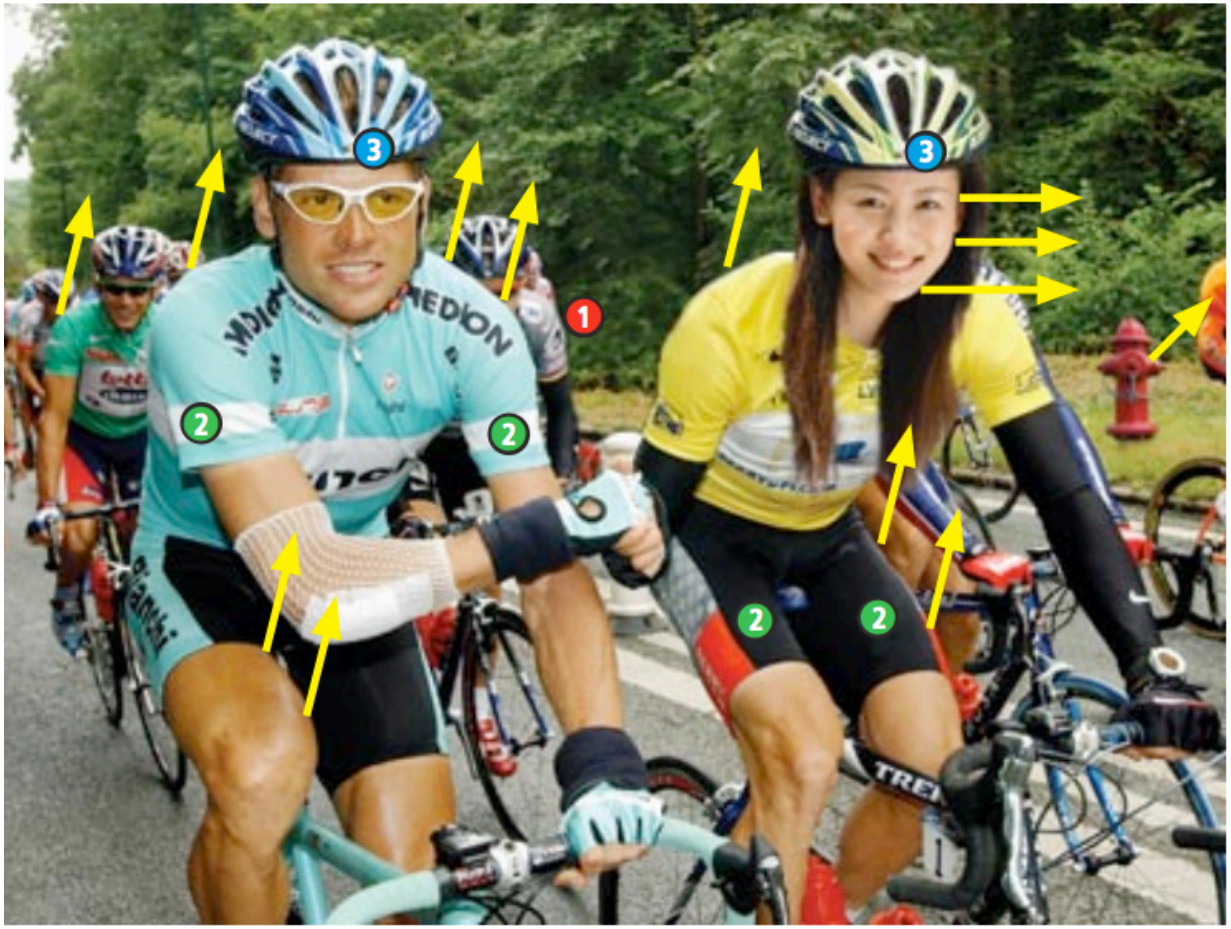
Nearest neighbor scenes from database of 2.3 million photos


Graph cut + Poisson blending

Image Forensics

From Hany Fraid, Digital Image Frensics
http://www.cs.dartmouth.edu/farid/downloads/publications/scjam08.pdf

Image Forensics



Arrows indicates light source direction

1. They are not shot Together.
2. The girl's helmet is the same as the man's but recolored.

From Hany Fraid, Digital Image Frensics
http://www.cs.dartmouth.edu/farid/downloads/publications/sciam08.pdf

Applications of Computer Vision: Detecting cloning

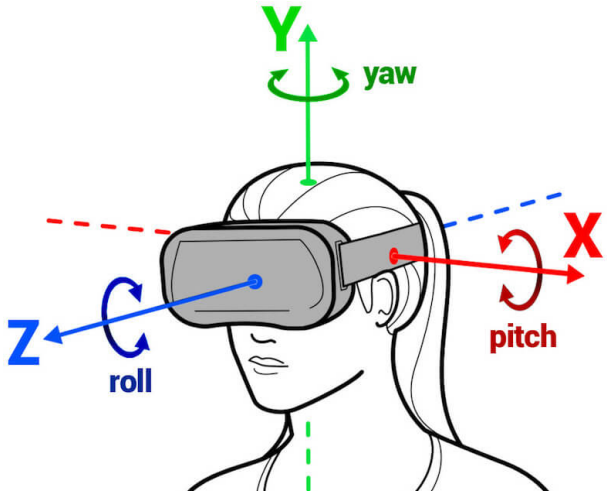
From Hany Fraid, Digital Image Forensics
http://www.cs.dartmouth.edu/farid/downloads/publications/sciam08.pdf

INVESTING 3/25/2014 @ 5:43PM | 70,399 views

Facebook Buys Oculus, Virtual Reality Gaming Startup, For $2 Billion

+ Comment Now + Follow Comments

Virtual & Augmented Reality



6DoF head tracking

Hand & body tracking

3D scene understanding

3D-360 video capture

Smart phone and augmented reality

Can the computer match human perception?

- What does human vision have:
  - Amazing eyes (sensor) and big brain
  - Experiences in the real-world (Prior knowledge)
  - Good at solving hard problems with ambiguous information
- What does computer vision have?
  - Lots of CPUs, fast speed
  - Mathematical techniques
  - Data, data, data ((could we learn from constraints from data)
  - Fast at solving simple and tedious problems
  - But tremendous progress has been made of solving hard problem such as human activity understanding

Slide courtesy from Noah Snavely

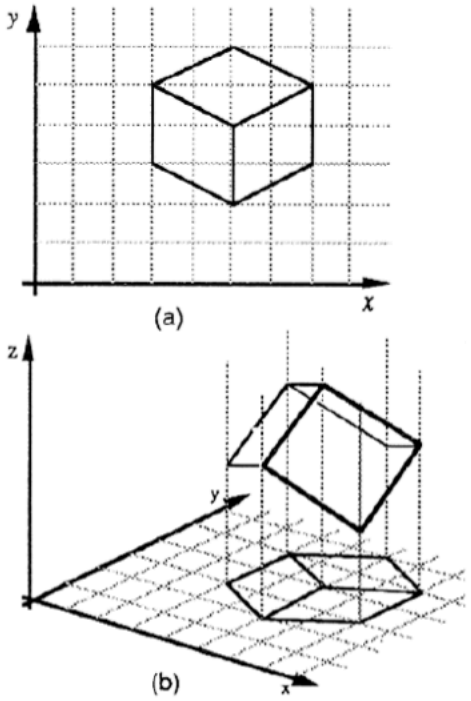
Why vision is hard?

- We must estimate 3D properties (texture, color, surface properties) of a visual scene from 2D images?
- It is an inverse problem under-constrained.

Slide courtesy from Noah Snavely

Vision is an inverse problem

Why is this hard?



Construct 3D from 2D images

Sinha & Adelson 93

Vision is an inverse problem

Why is this hard?

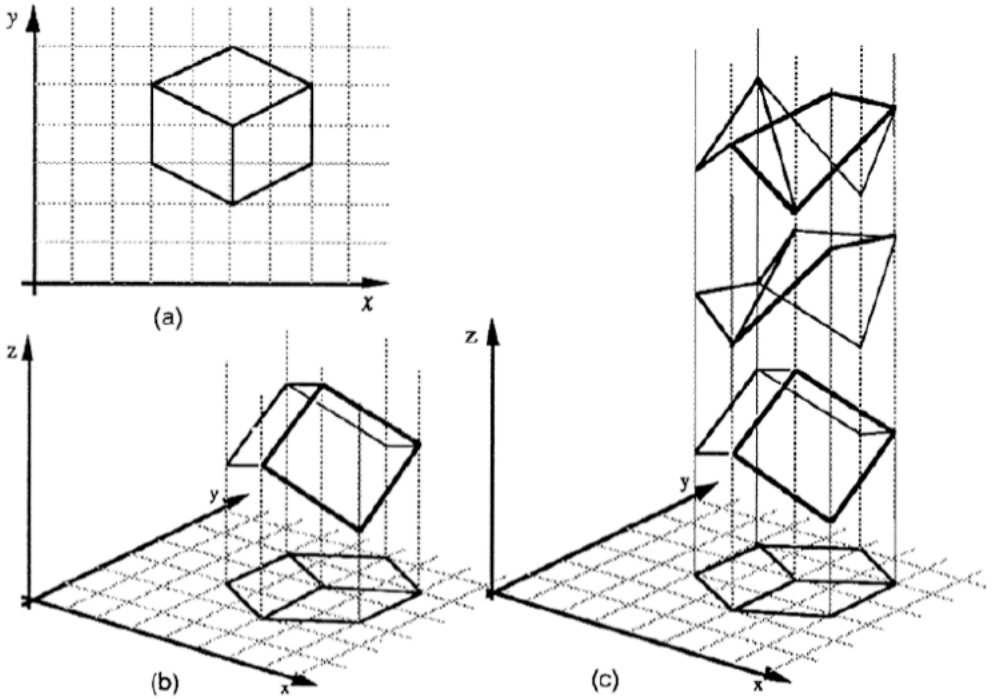


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

Sinha & Adelson 93

Vision is hard

Some things have strong variations in appearance






Slide courtesy from Antonio Torralba

Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Slide courtesy from Noah Snavely

Why is computer vision difficult?

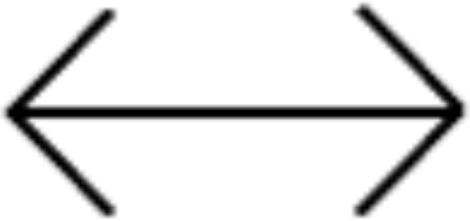
Intra-class variation

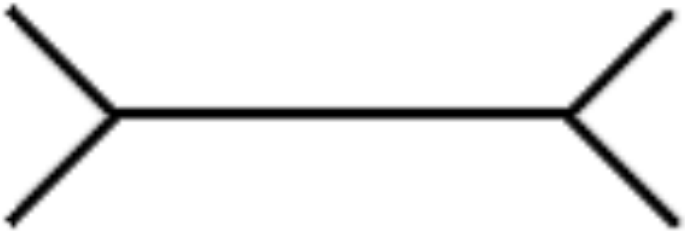
Motion (Source: S. Lazebnik)

Background clutter

Occlusion

Vision is hard





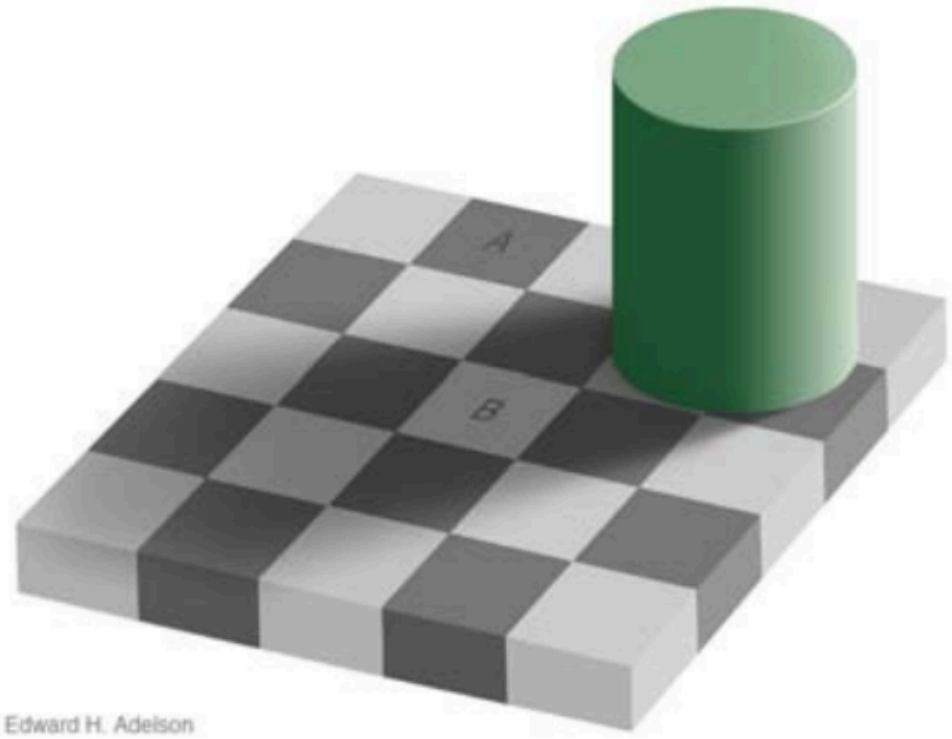
Muller-Lyer illusion

Vision is hard

What is the most likely scenario?

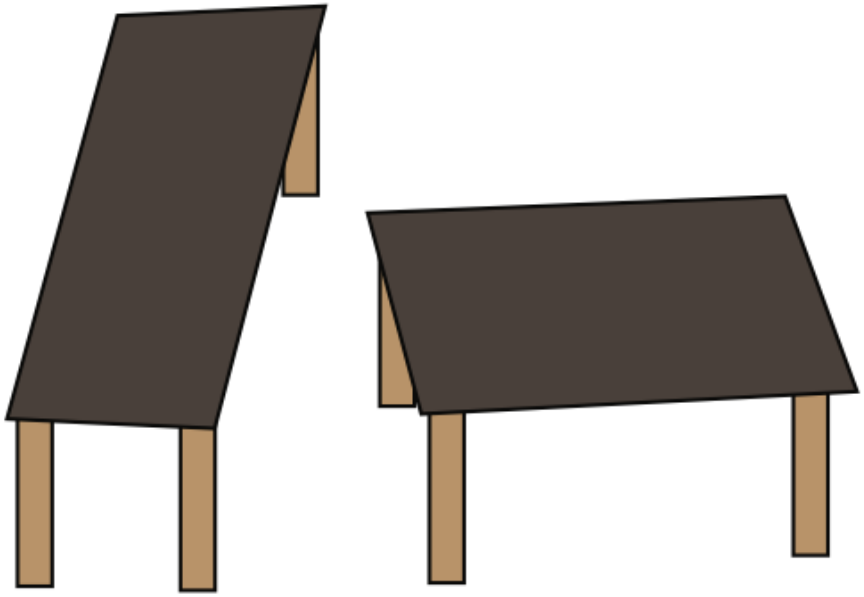
Prior knowledge: Occlusion is more common than L shaped object

Vision is hard



Edward Adelson

Vision is hard



Roger Shepard

http://www.opticalillusion.net/optical-illusions/shepards-tables-whats-up/

Bottom line

- Perception is an inherently ambiguous problem
  - Many different 3D scenes could have given rise to a particular 2D picture

  - We often need to use prior knowledge about the structure of the world

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

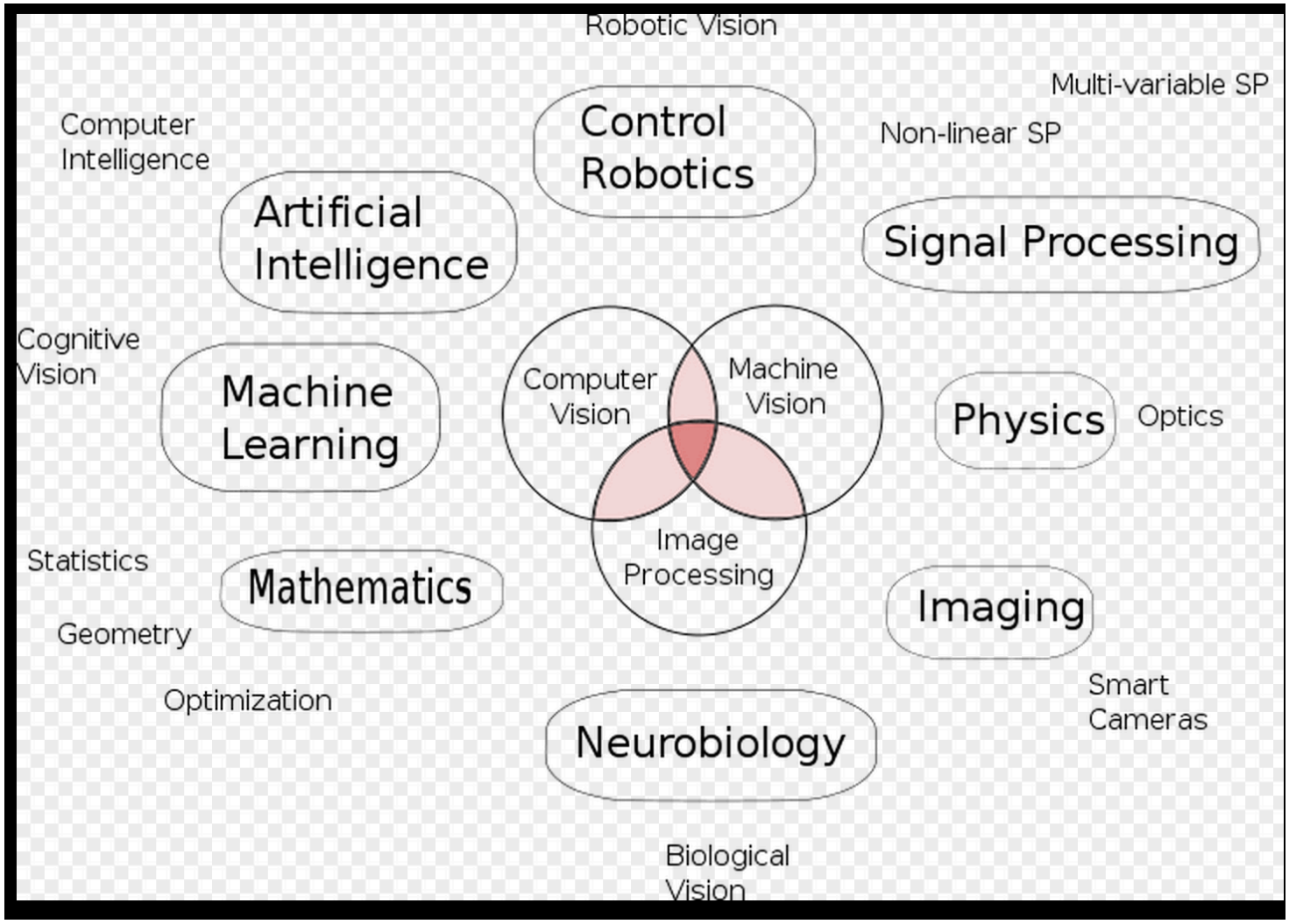
But what do we have?

- Lots and lots and lots of data.
- We can learn from humans.
- We have Prior knowledge, which can provide constrains of the problem.

Computer Vision and Nearby Fields

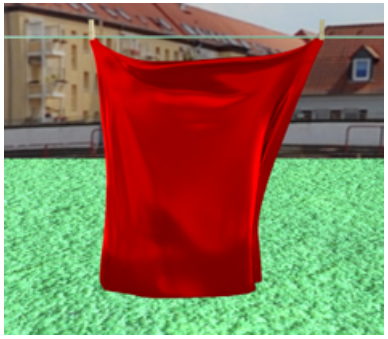
- Computer Graphics: Models to Images
- Computational Photography: Images to Images
- Computer Vision: Images to Models
- Human perception: understand how humans perceive the world
- Machine learning (tools that are common in all above areas): algorithm of how to learn from data

Computer Vision and Nearby Fields

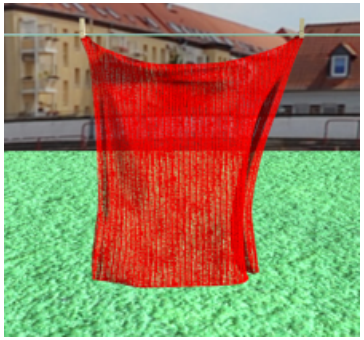


My research: automatic material recognition from images and videos

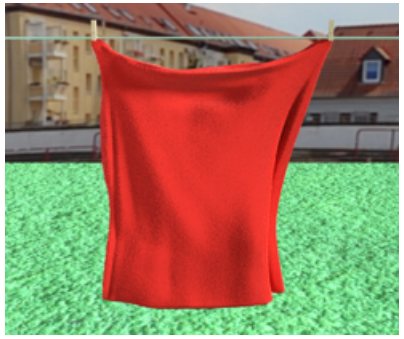
satin



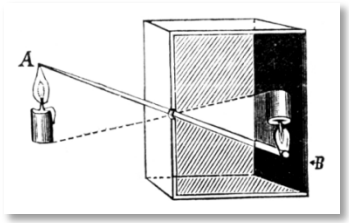
knit



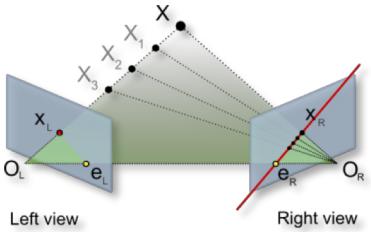
velvet



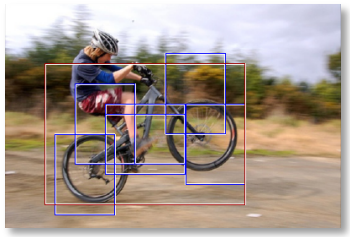
What we are going to learn?



1. Low-level vision
   - image processing, edge detection, feature detection, cameras, image formation



2. Geometry and algorithms
   - projective geometry, stereo, structure from motion, Markov random fields

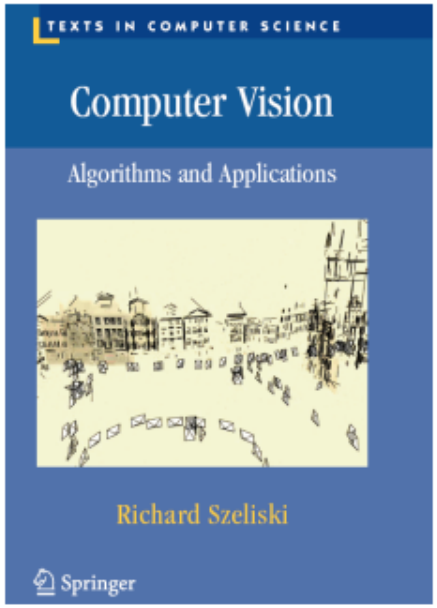


3. Recognition
   - face detection / recognition, category recognition, segmentation

4. Light, color, and reflectance

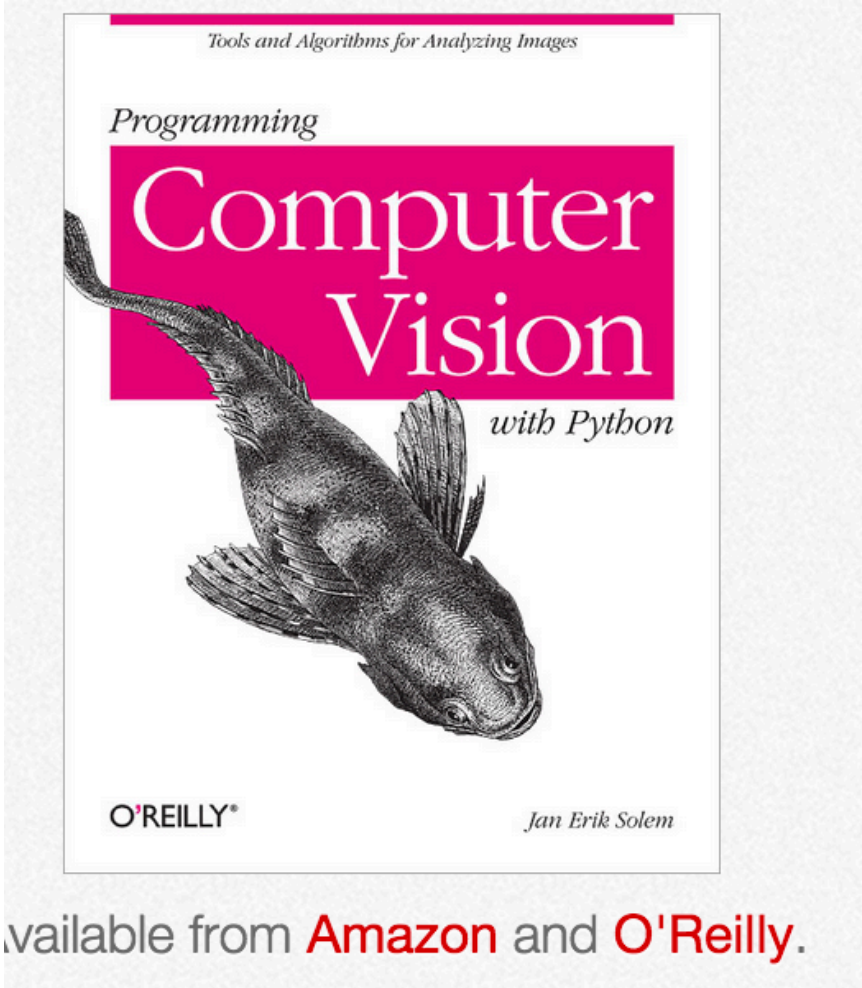
Textbook

Computer Vision: Algorithms and Applications



http://szeliski.org/Book/

Textbook



http://programmingcomputervision.com/

Grading

60 % Homework assignments. 5 projects + 1 homework.

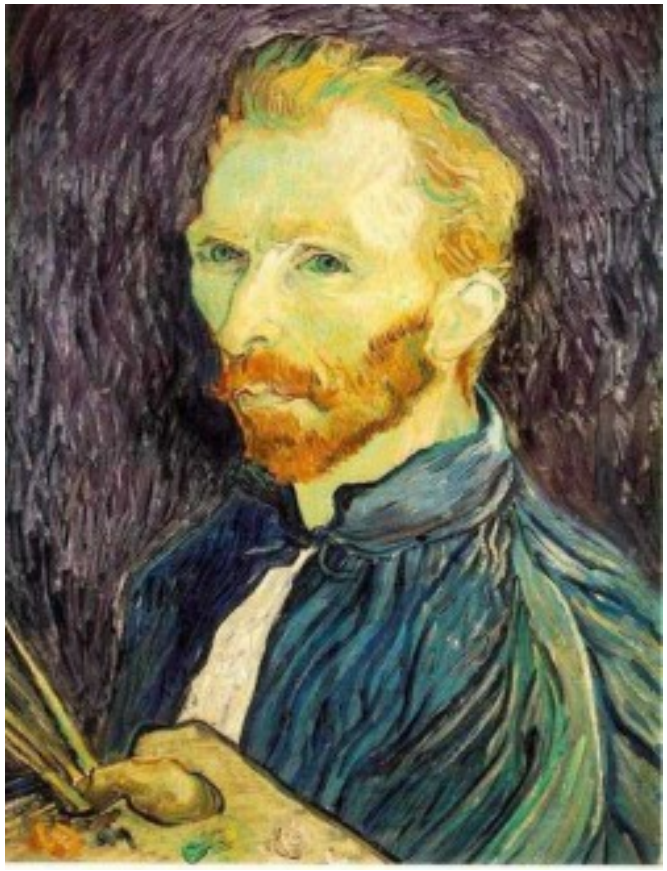
10 % Mid-term exam (written)

15% Final project

5 % Attendance

10% In-class quiz

Project: Hybrid images from image pyramids



Prerequisites

- **CSC 280 and 281 (preferred)**
- Linear algebra, basic calculus, and probability recommended.
- No previous experience with image processing is required.

Course requirements

- Prerequisites—*these are essential*!
  - Data structures
  - A good working knowledge of Python programming
  - Linear algebra
  - Vector calculus

- Course does ***not*** assume prior imaging experience
  - computer vision, image processing, graphics, etc.

Academic Integrity

You must not copy lines of code from other people unless teamwork is allowed. Final project, for example, allows teamwork. But individual homework assignment does not. You can discuss your homework with other people but you must declare with whom you discussed with.

You must not copy lines of code from internet including online forums.

Office hours

Wed: 3:30-4:30pm

Friday: 2:30-3:30pm

Location: Myers Technology building *this building, Room 204.

Connect to course GITHUB site

Take-home Reading

Szelisky : Chapter 1, What is computer vision?

Finish the warm up exercises.