CS 4476 / 6476: Computer Vision

Instructor: James Hays

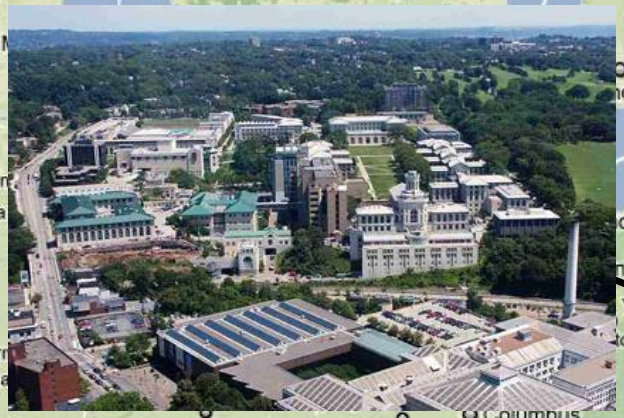
Tas: Varun Agrawal, Samarth Brahmbhatt, Cusuh Ham, Eunji Chong, Wenqi Xian, Wengling Chen, Albert Shaw, Stefan Stojanov

Image by kirkh.deviantart.com

Today's Class

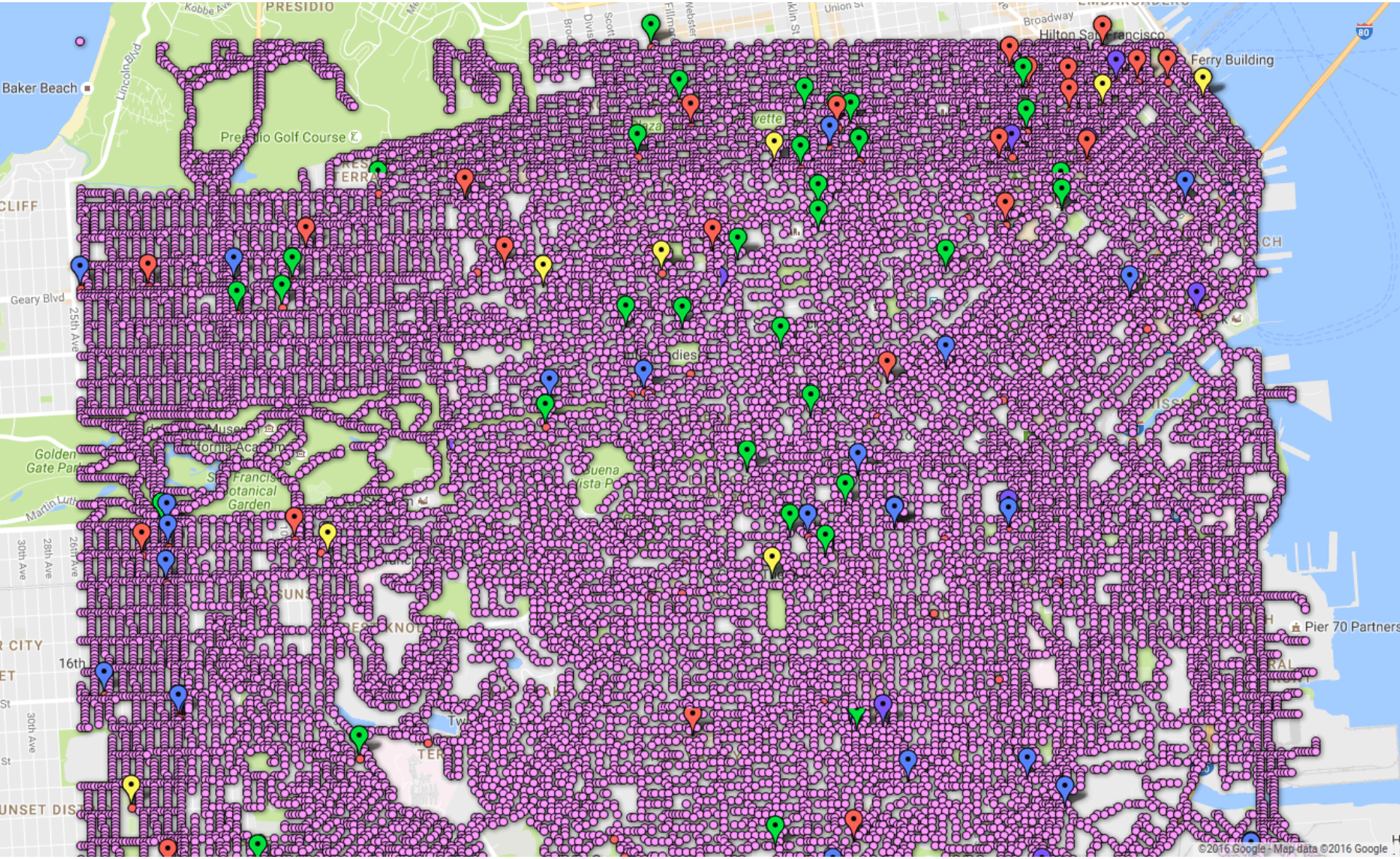
- Course enrollment
- Who am I?
- What is Computer Vision?
- Specifics of this course
- Geometry of Image Formation
- Questions

A bit about me

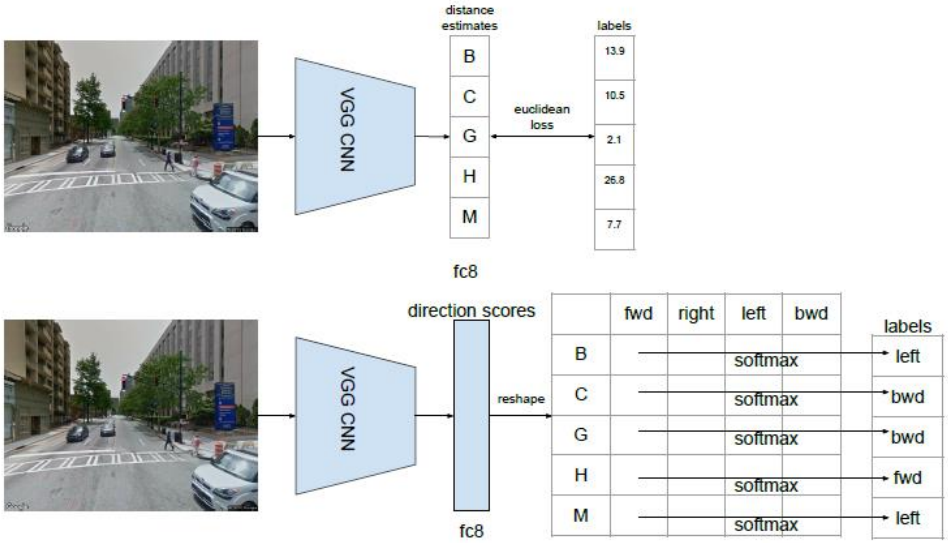


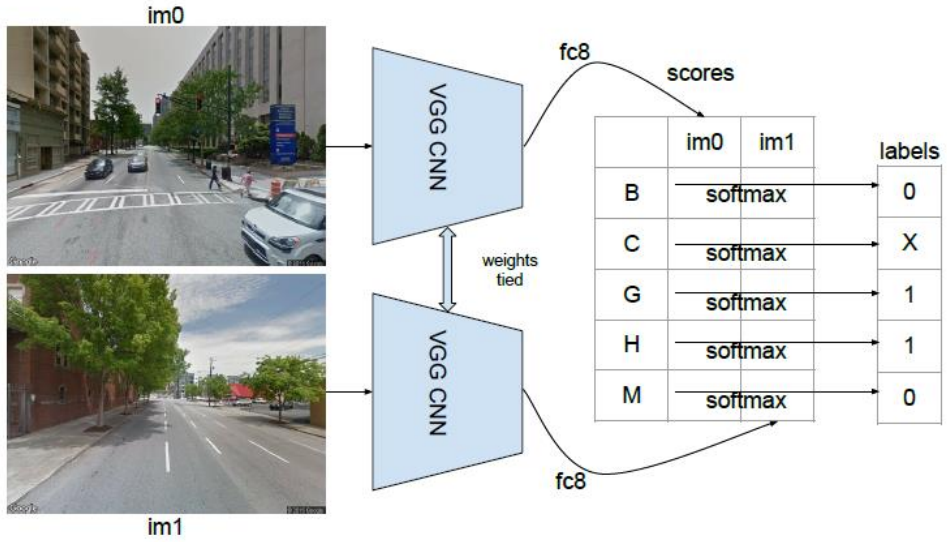
DeepNav

DeepNav: Learning to Navigate Large Cities
Samarth Brahmbhatt and James Hays. CVPR 2017

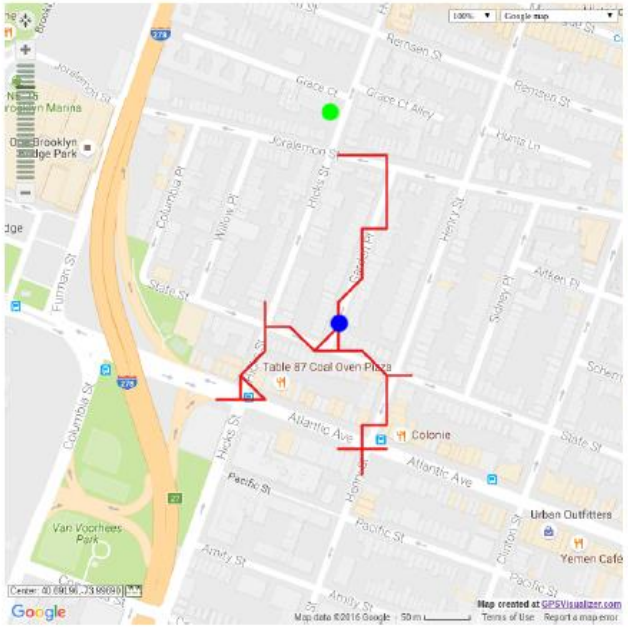

PRESIDIO
Baker Beach
Lincoln Blvd
Presidio Golf Course
Geary Blvd
25th Ave
Golden Gate Park
Martin Luther
30th Ave
28th Ave
26th Ave
16th
30th Ave
Union St
Broadway
Hilton San Francisco
Ferry Building
Pier 70 Partners
©2016 Google - Map data ©2016 Google

Network Architectures

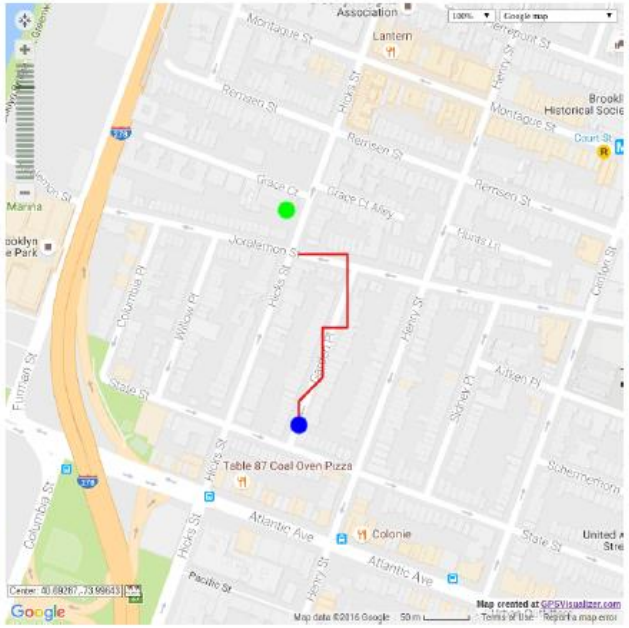




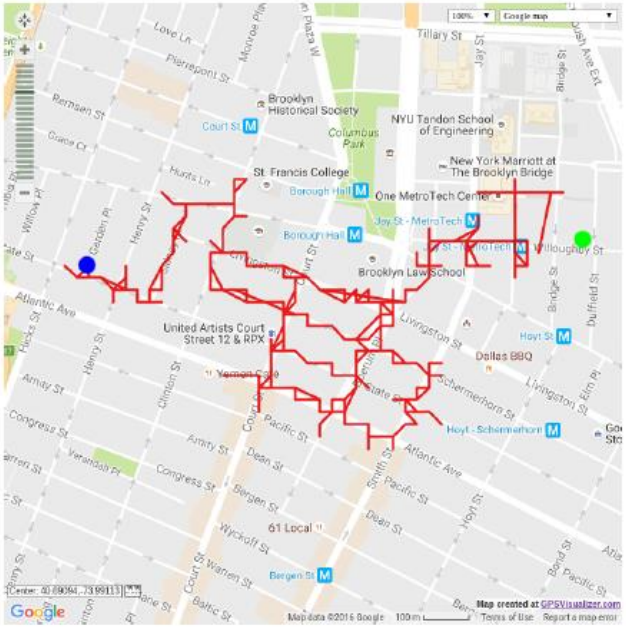
Qualitative Results



Found, length = 60



Found, length = 10



Found, length = 520

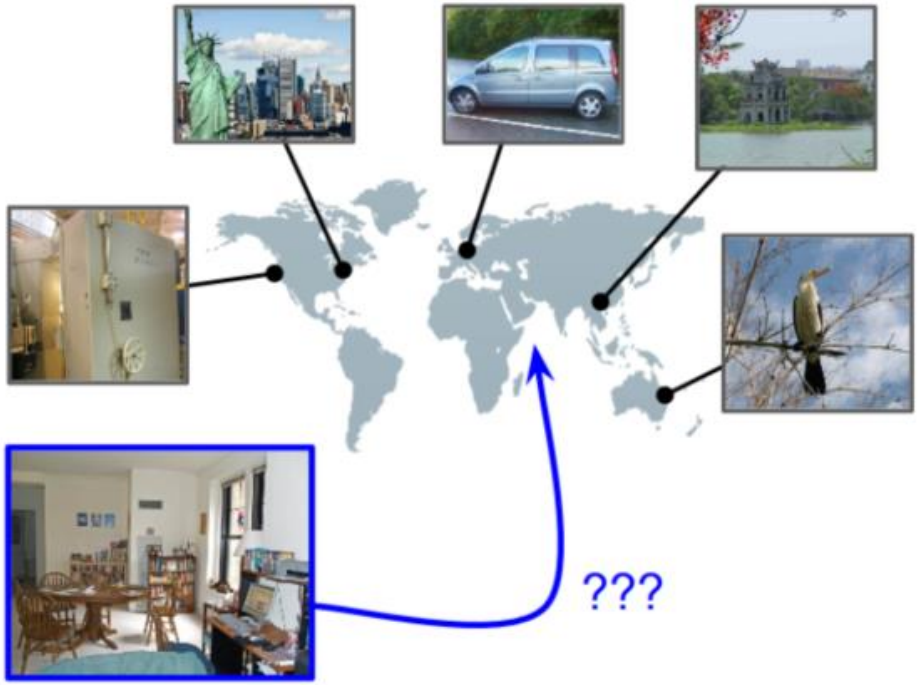
Revisiting IM2GPS in the Deep Learning Era
Nam Vo, Nathan Jacobs, James Hays. ICCV 2017

The problem set up:

- Give a large set of GPS-tagged images.
- Learn to infer GPS coordinate of query images with unknown location.

Approaches:

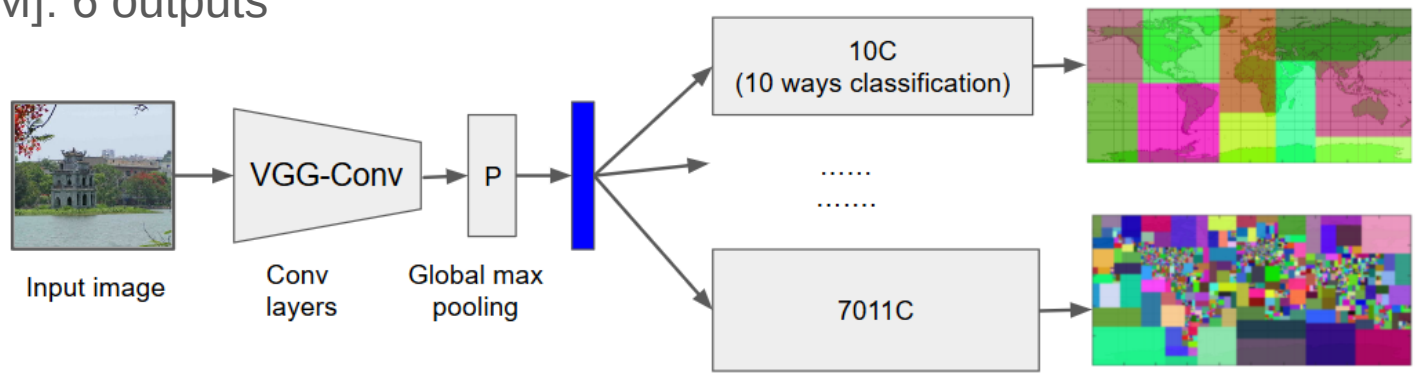
- Image retrieval
- Image classification



Geolocalization at planet scale

Caffe library, Vgg-16 imagenet initialization, training data: Im2GPS (~6m images)

Model [M]: 6 outputs



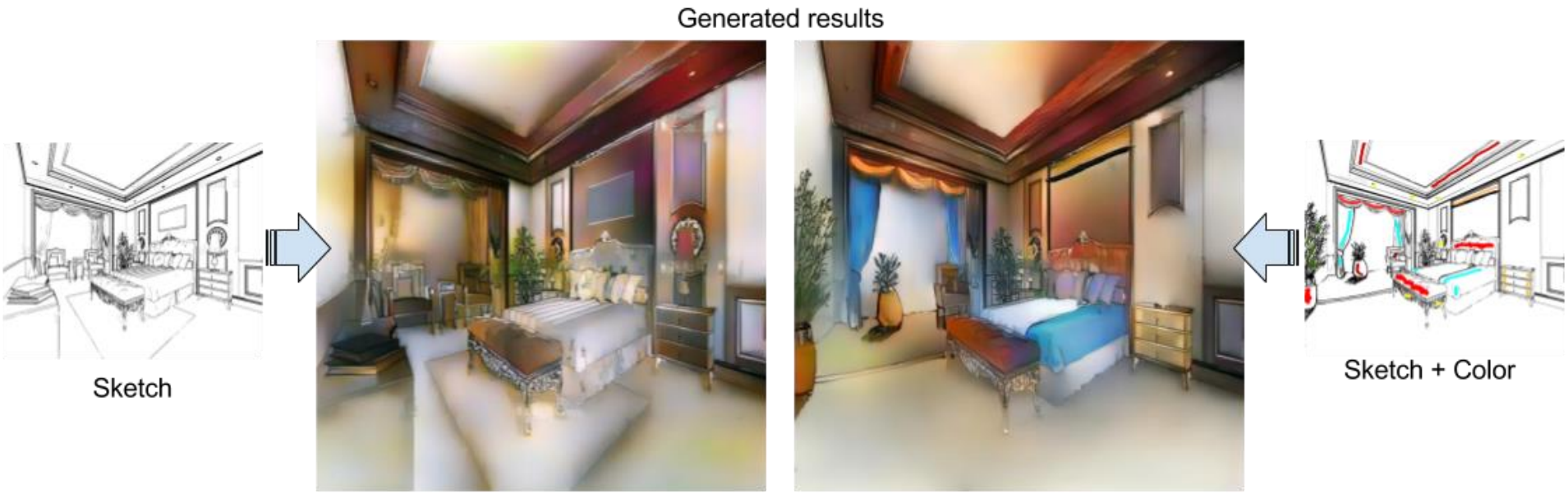
Model [L]: 7011C only
Model [L2]: 359C only
Model [R]: finetuned from [M] with ranking loss

Geolocalization at planet scale, Quantitative result

Table 1. Retrieval performance on Im2GPS test set. (Human* performance is average from 30 mturk workers over 940 trials, so it might not be directly comparable)

	Street	City	Region	Country	Cont.
Threshold (km)	1	25	200	750	2500
Human*			3.8	13.9	39.3
Im2GPS [9]		12.0	15.0	23.0	47.0
Im2GPS [10]	02.5	21.9	32.1	35.4	51.9
PlaNet [35]	08.4	24.5	37.6	53.6	**71.3**
[L] 7011C	06.8	21.9	34.6	49.4	63.7
[L] kNN, σ=4	**12.2**	**33.3**	**44.3**	**57.4**	**71.3**
... 28m database	**14.4**	**33.3**	**47.7**	**61.6**	**73.4**

Scribbler: Controlling Deep Image Synthesis with Sketch and Color



Patsorn Sangkloy, Jingwan Lu, Chen Fang , Fisher Yu, and James Hays. CVPR 2017

Training Data – (Mostly) Synthetic Sketches

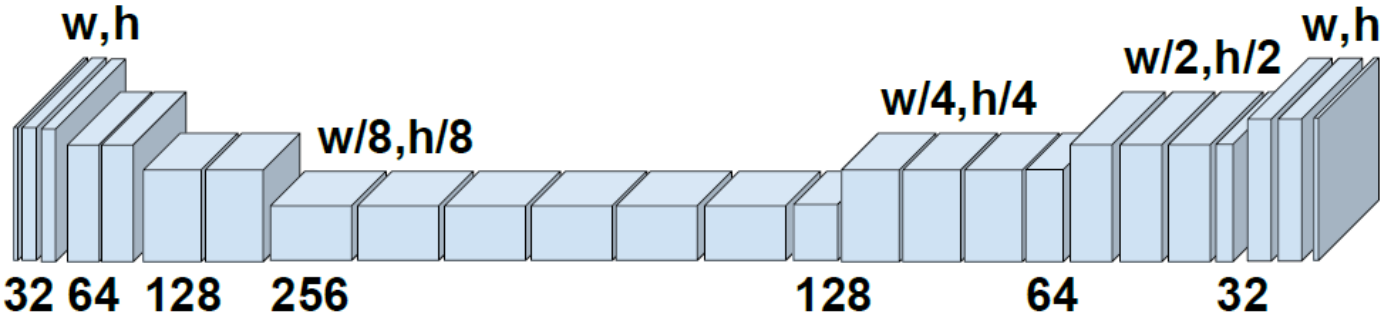
Style Transfer
[Ulyanov et al. 2016]

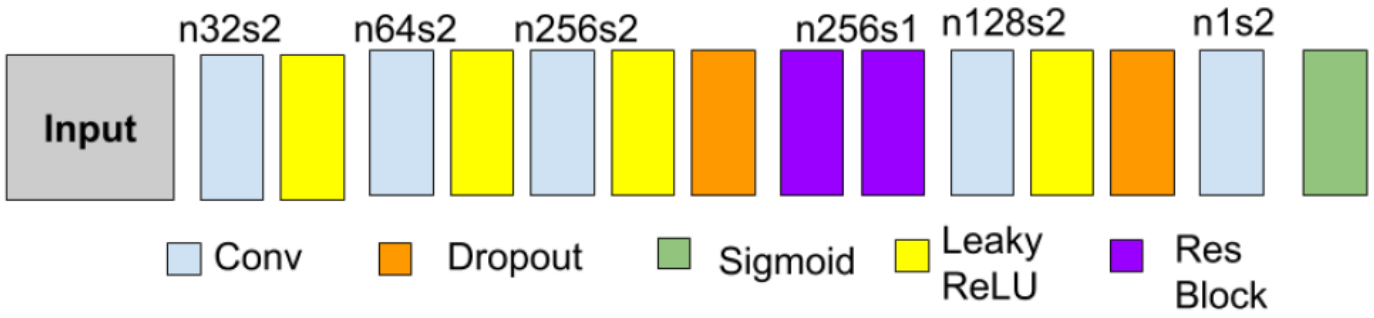
XDoG
[Holger et al. 2012]

Photoshop

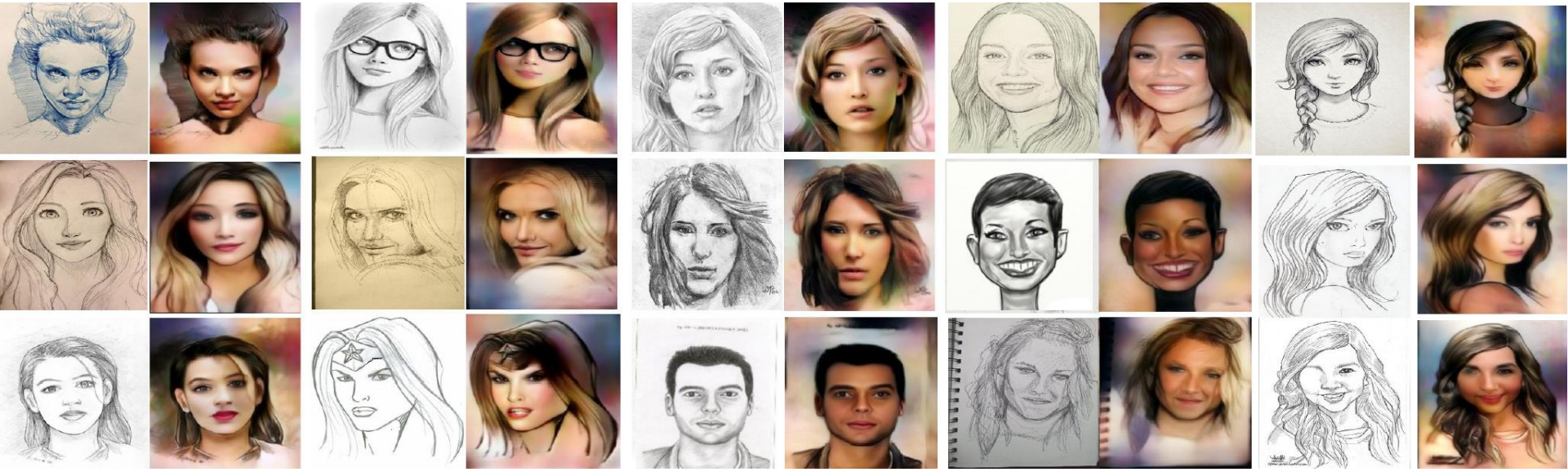
CUHK
[Wang and Tang 2009]

Network Architecture – Adversarial Learning

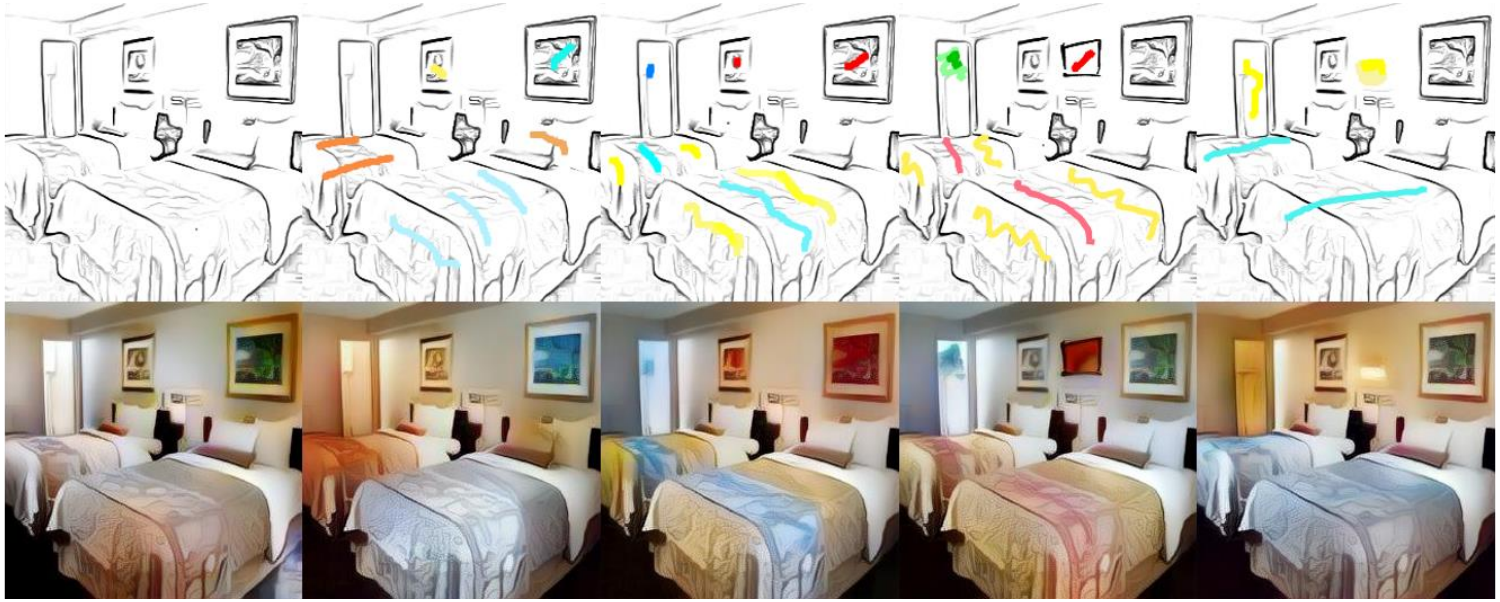




Results on held out sketches



Results on held out sketches



Groundtruth Image

Shoe

Tools

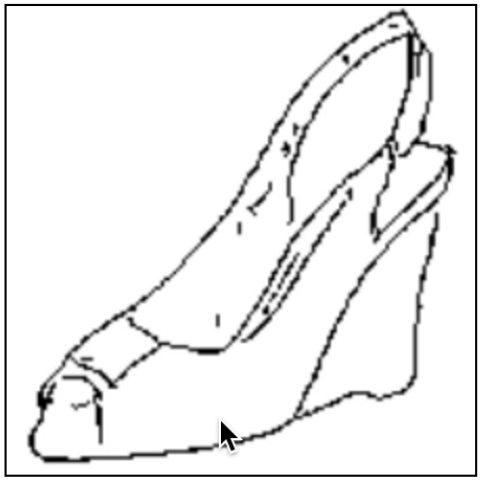
Current Mode: Color Stroke

Color Picker

Brush Size

Enter sketch URL

Input Sketch



Final Result

Given a simple sketch drawn by user, we hallucinate colors and shadings

What is Computer Vision?

Computer Vision and Nearby Fields

- Computer Graphics: Models to Images
- Comp. Photography: Images to Images
- Computer Vision: Images to Models

Derogatory summary of computer vision:

Machine learning applied to visual data

Computer Vision

Make computers understand images and video.

What kind of scene?

Where are the cars?

How far is the building?

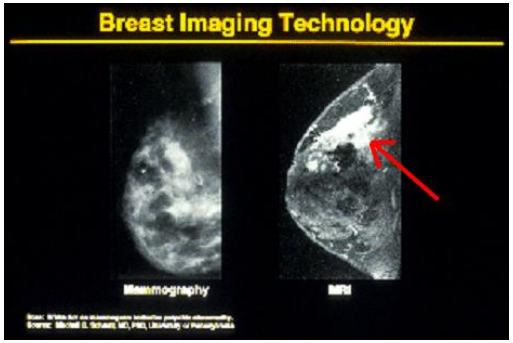
…

Vision is really hard

- Vision is an amazing feat of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - One third of human brain devoted to vision (more than anything else)

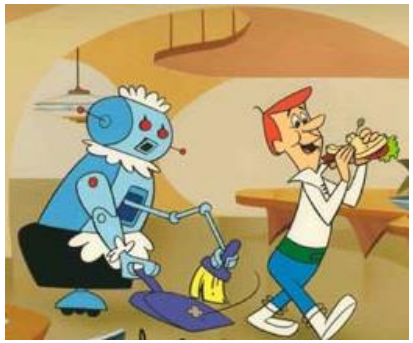
Why computer vision matters

Safety



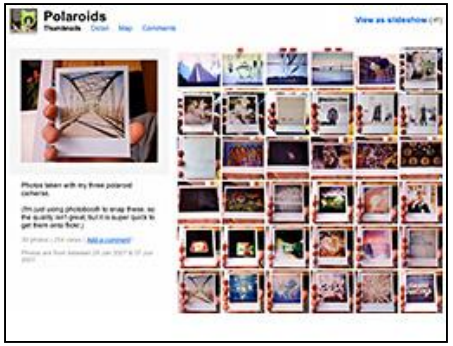
Health

Security



Comfort

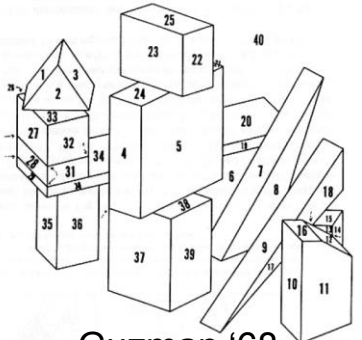
Fun



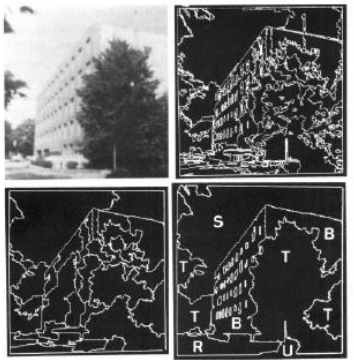
Access

Ridiculously brief history of computer vision

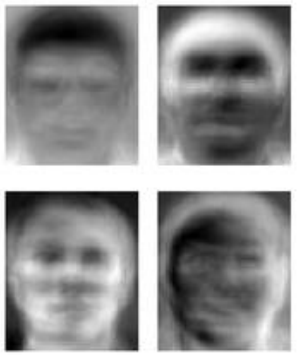
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis 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

How vision is used now

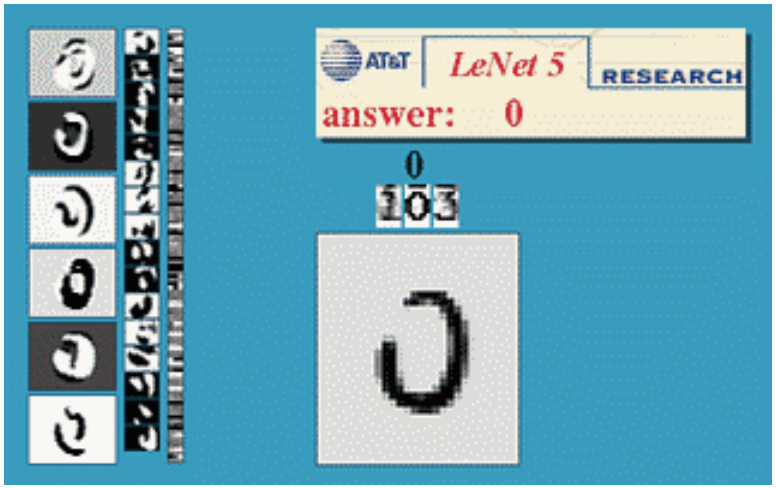
- Examples of real world applications

Some of the following slides by Steve Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



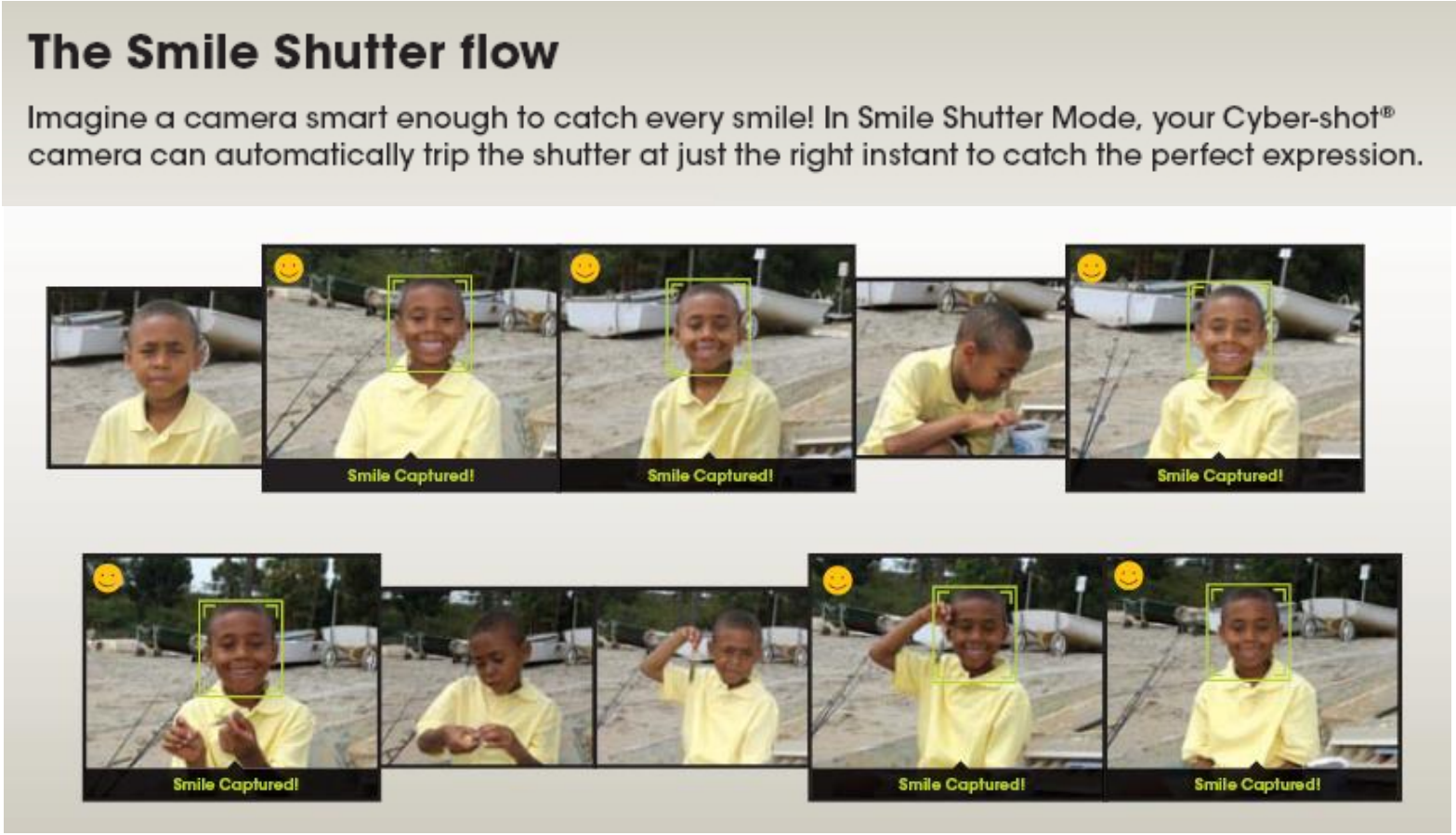
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

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

Face detection

- Digital cameras detect faces

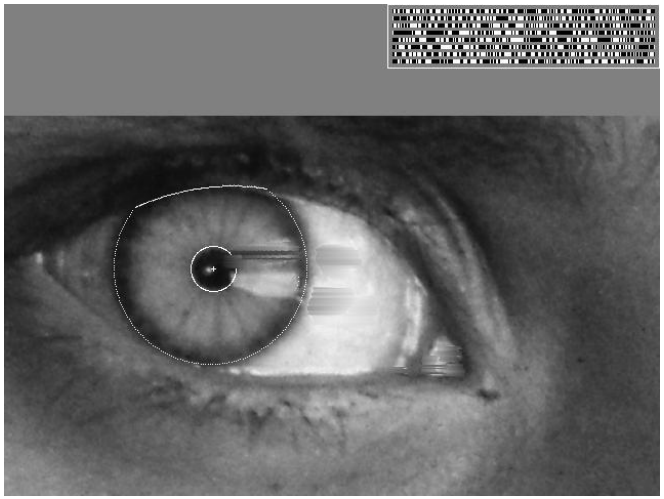
Smile detection

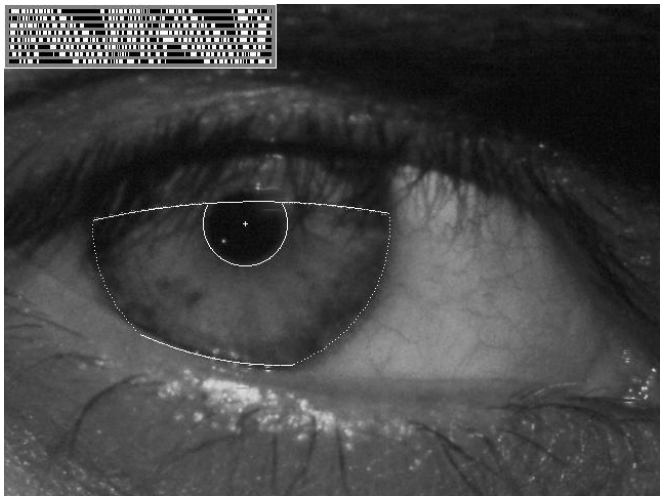


[Sony Cyber-shot® T70 Digital Still Camera]

Vision-based biometrics

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





Login without a password…

Fingerprint scanners on many new laptops, other devices

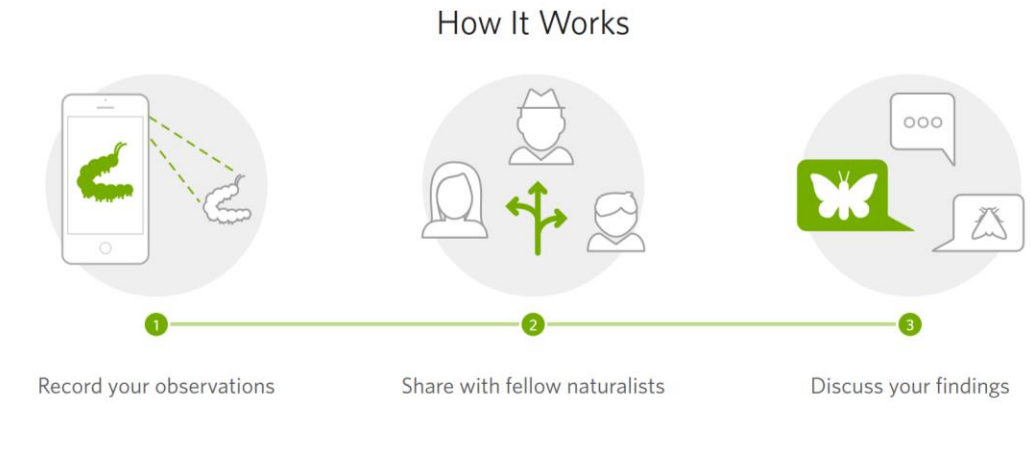
Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

iNaturalist



https://www.inaturalist.org/pages/computer_vision_demo

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

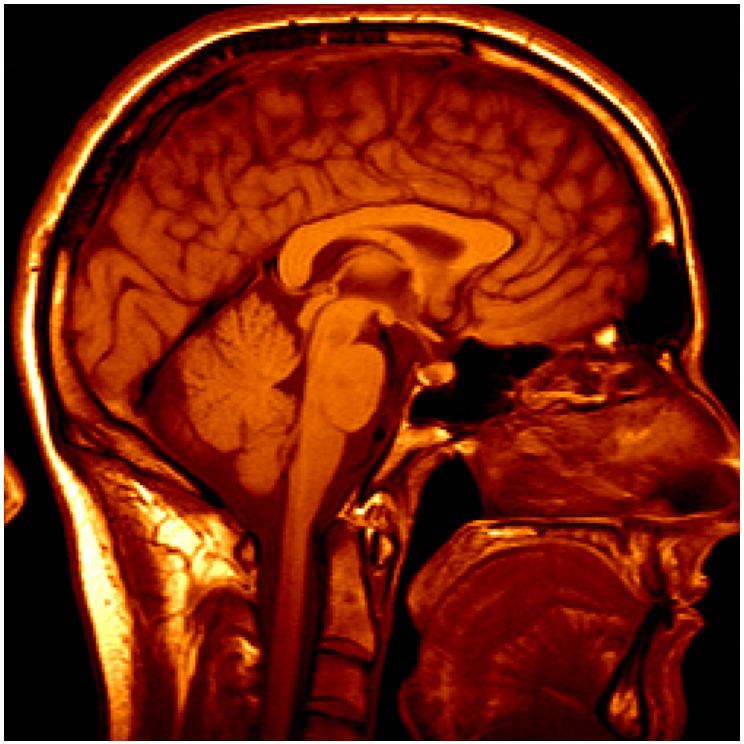
Pirates of the Carribean, Industrial Light and Magic

Sports

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

http://www.sportvision.com/video.html

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Smart cars

Slide content courtesy of Amnon Shashua

- Mobileye
  - Market Capitalization: 11 Billion dollars

Google 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)

Ford acquires SAIPS for self-driving machine learning and computer vision tech

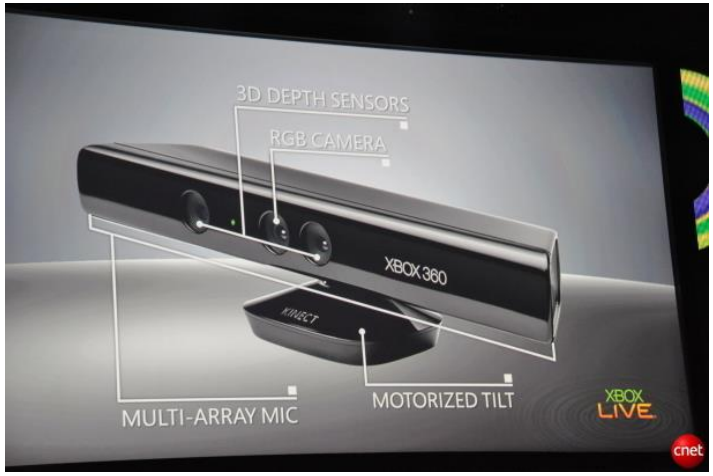
Posted Aug 16, 2016 by **Darrell Etherington** *(@etherington)*

Ford outlined a few of the ways it's aiming to ship driverless cars by 2021, and part of the plan involves acquisitions. CEO Mark Fields revealed at a press event in Palo Alto today that the automaker acquired SAIPS, an Israeli company focusing on machine learning and computer vision. It's also partnering exclusively with Nirenberg Neuroscience, to bring more "humanlike intelligence" to machine learning components of driverless car systems.

SAIPS' technology brings image and video processing algorithms, as well as deep learning tech focused on processing and classifying input signals, all key ingredients in the special sauce that makes up autonomous vehicle tech. This company's expertise should help with on-board interpretation of data captured by sensors on Ford's self-driving cars, and turning that data into usable info for the car's virtual driver system. SAIPS' offerings include detection of anomalies, persistent tracking of objects detected by sensors, and much more. The company's past clients include HP and Trax, but its partner group doesn't appear to have included much in the way of driving-specific applications.

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Industrial robots

Vision-guided robots position nut runners on wheels

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.

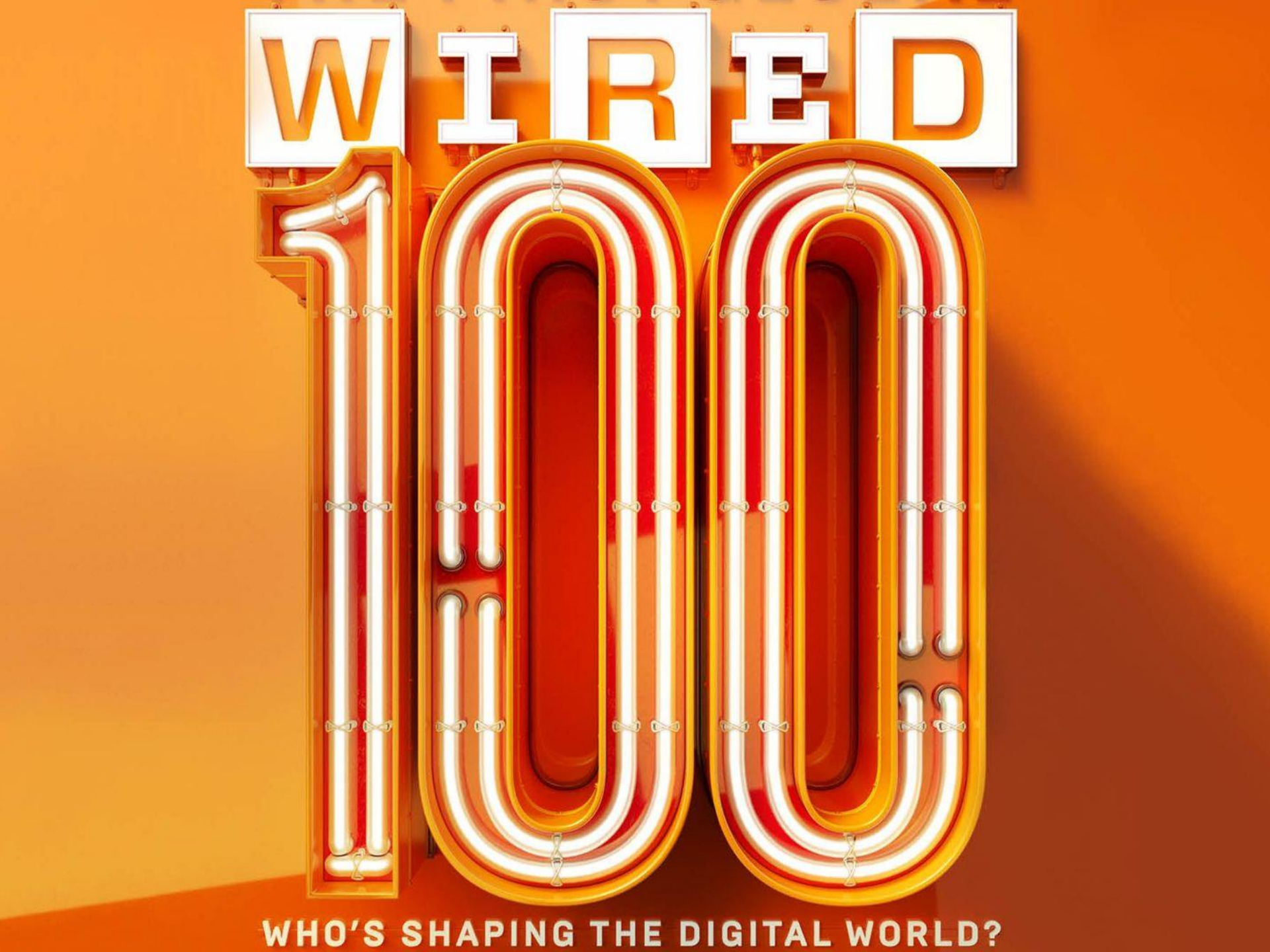
Augmented Reality and Virtual Reality

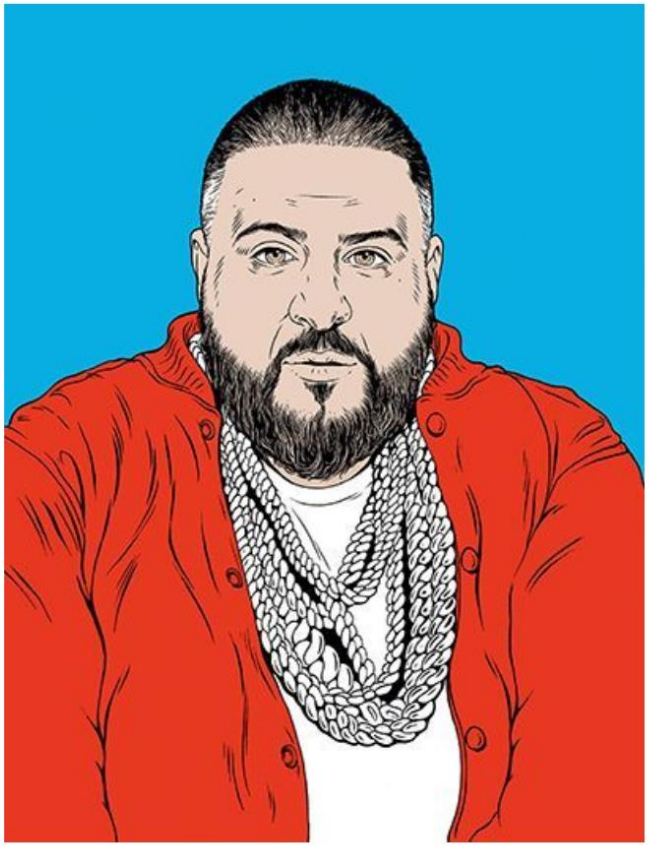
Magic Leap, Oculus, Hololens, etc.

State of the art today?

With enough training data, computer vision nearly matches human vision at most recognition tasks

Deep learning has been an enormous disruption to the field. More and more techniques are being “deepified”.


WIRED
100
WHO'S SHAPING THE DIGITAL WORLD?



DJ Khaled

Credit **Louise Zergaeng Pomeroy**

73. DJ Khaled

Snapchat icon; DJ and producer

Louisiana-born Khaled Mohamed Khaled, aka DJ Khaled, cut his musical chops in the early 00s as a host for Miami urban music radio WEDR. He proceeded to build a solid if not dazzling career as a mixtape DJ and music producer (he founded his label We The Best Music Group in 2008, and was appointed president of Def Jam South in 2009).

69. Geoffrey Hinton

Psychologist, computer scientist; researcher, Google Toronto

British-born Hinton has been dubbed the "godfather of deep learning". The Cambridge-educated cognitive psychologist and computer scientist started being an ardent believer in the potential of neural networks and deep learning in the 80s, when those technologies enjoyed little support in the wider AI community.

But he soldiered on: in 2004, with support from the Canadian Institute for Advanced Research, he launched a University of Toronto programme in neural computation and adaptive perception, where, with a group of researchers, he carried on investigating how to create computers that could behave like brains.

Hinton's work – in particular his algorithms that train multilayered neural networks – caught the attention of tech giants in Silicon Valley, which realised how deep learning could be applied to voice recognition, predictive search and machine vision.

The spike in interest prompted him to launch a free course on neural networks on e-learning platform Coursera in 2012. Today, 68-year-old Hinton is chair of machine learning at the University of Toronto and moonlights at Google, where he has been using deep learning to help build internet tools since 2013.

63. Yann Lecun

Director of AI research, Facebook, Menlo Park

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

62. Richard Branson

Founder, Virgin Group, London

Branson saw his personal fortune grow £550 million when Alaska Air bought Virgin America for $2.6 billion in April. He is pressing on with civilian space travel with Virgin Galactic.

61. Taylor Swift

Entertainer, Los Angeles

Credit **Google DeepMind**

Google-backed startup DeepMind Technologies has built an artificial intelligence agent that can learn to successfully play 49 classic Atari games by itself, with minimal input.

The story of AlphaGo so far

AlphaGo is the first computer program to defeat a professional human Go player, the first program to defeat a Go world champion, and arguably the strongest Go player in history.

AlphaGo's first formal match was against the reigning 3-times European Champion, Mr Fan Hui, in October 2015. Its 5-0 win was the first ever against a Go professional, and the results were published in full technical detail in the international journal, **Nature**. AlphaGo then went on to compete against legendary player Mr Lee Sedol, winner of 18 world titles and widely considered to be the greatest player of the past decade.

AlphaGo's 4-1 victory in Seoul, South Korea, in March 2016 was watched by over 200 million people worldwide. It was a landmark achievement that experts agreed was a decade ahead of its time, and earned AlphaGo a 9 dan professional ranking (the highest certification) - the first time a computer Go player had ever received the accolade.

During the games, AlphaGo played a handful of **highly inventive winning moves,** several of which - including move 37 in game two - were so surprising they overturned hundreds of years of received wisdom, and have since been examined extensively by players of all levels. In the course of winning, AlphaGo somehow taught the world completely new knowledge about perhaps the most studied and contemplated game in history.

Since then, AlphaGo has continued to surprise and amaze. In January 2017, an improved AlphaGo version was revealed as the online player "Master" which achieved **60 straight wins in online fast time-control games** against top international Go players.

In May 2017, Alpha Go took part in The Future of Go Summit in the birthplace of Go, China, to delve deeper into the mysteries of Go in a spirit of mutual collaboration with the country's top players. You can read more about the five day summit **here**.

› More on The Future of Go Summit in this video

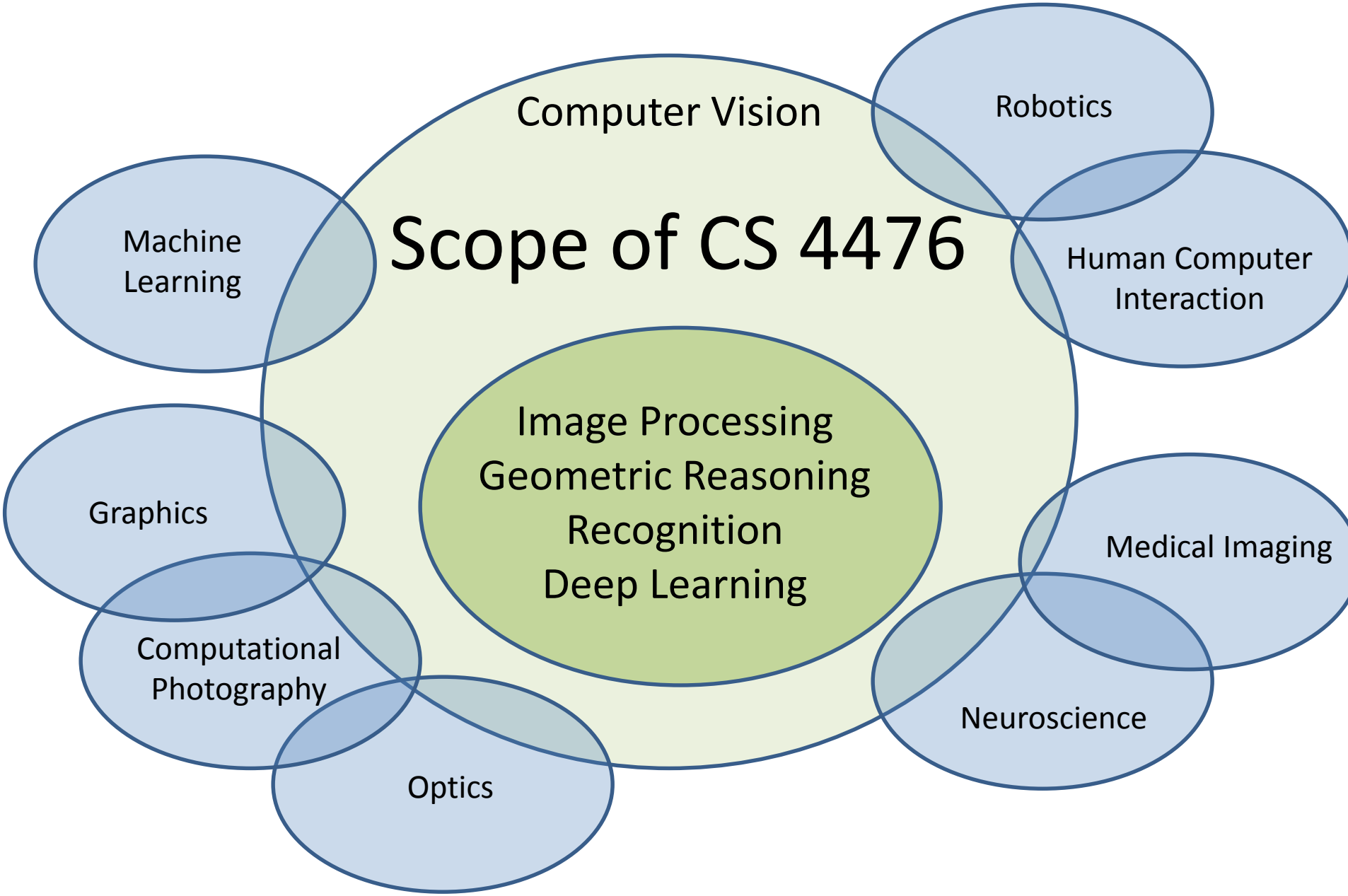
› Watch the video here

Course Syllabus (tentative)

http://www.cc.gatech.edu/~hays/compvision

Grading

- 80% programming projects (6 total)
- 20% quizzes (2 total)


Computer Vision
Robotics
Scope of CS 4476
Machine Learning
Human Computer Interaction
Image Processing
Geometric Reasoning
Recognition
Deep Learning
Graphics
Medical Imaging
Computational Photography
Neuroscience
Optics

Course Topics

- Interpreting Intensities
  - What determines the brightness and color of a pixel?
  - How can we use image filters to extract meaningful information from the image?
- Correspondence and Alignment
  - How can we find corresponding points in objects or scenes?
  - How can we estimate the transformation between them?
- Grouping and Segmentation
  - How can we group pixels into meaningful regions?
- Categorization and Object Recognition
  - How can we represent images and categorize them?
  - How can we recognize categories of objects?
- Advanced Topics
  - Action recognition, 3D scenes and context, human-in-the-loop vision…

Textbook

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research

http://szeliski.org/Book/

Prerequisites

- **Linear algebra**, basic calculus, and probability
- Experience with image processing or Matlab will help but is not necessary

Projects

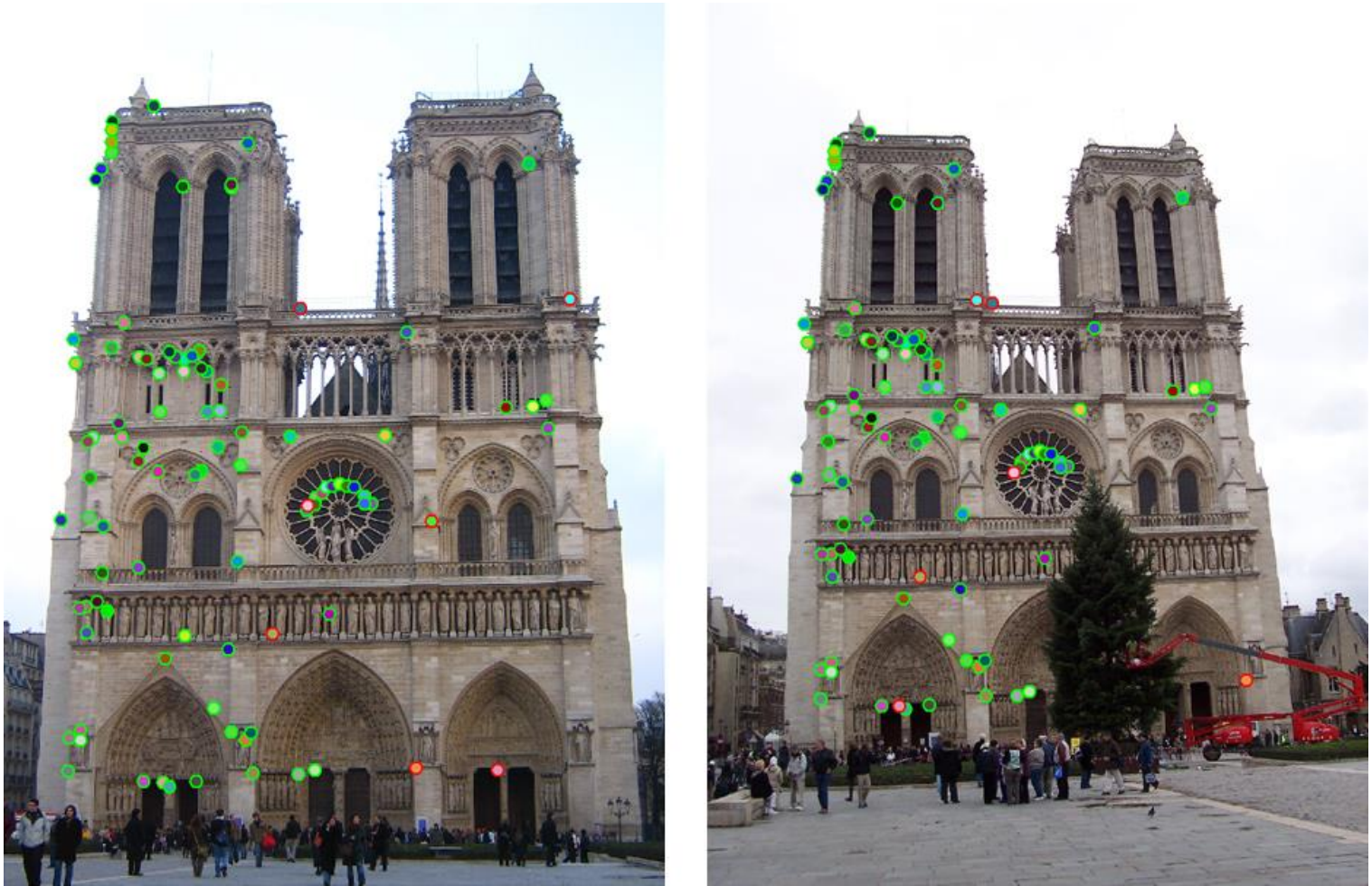
- Image Filtering and Hybrid Images
- Local Feature Matching
- Camera Calibration and Fundamental Matrix Estimation with RANSAC
- Scene Recognition with Bag of Words
- Object Detection with a Sliding Window
- Recognition with Deep Learning

Proj1: Image Filtering and Hybrid Images

- Implement image filtering to separate high and low frequencies
- Combine high frequencies and low frequencies from different images to create an image with scale-dependent interpretation

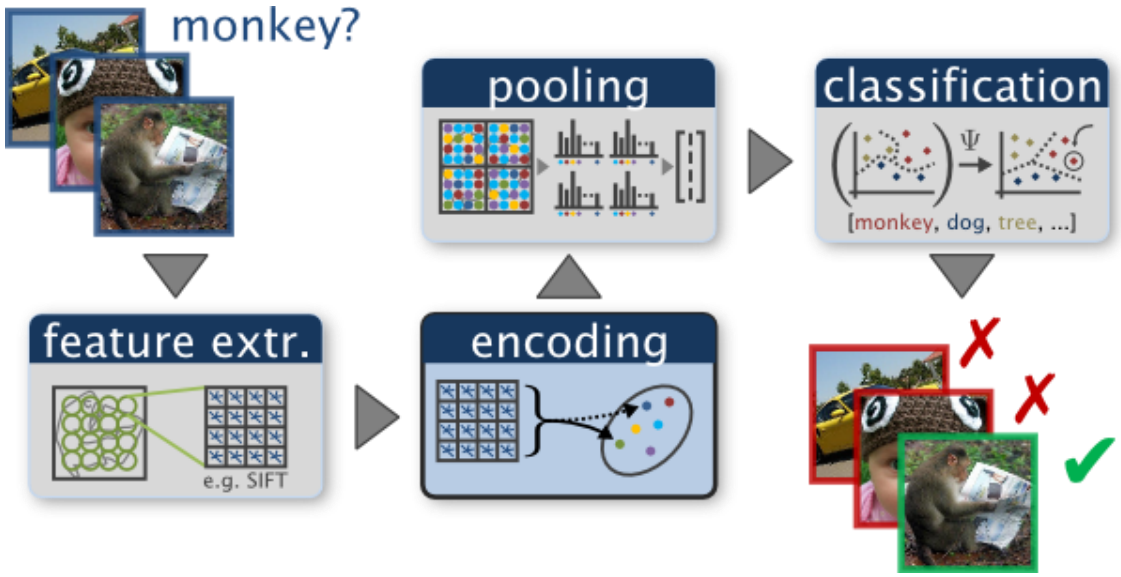
Proj2: Local Feature Matching

- Implement interest point detector, SIFT-like local feature descriptor, and simple matching algorithm.



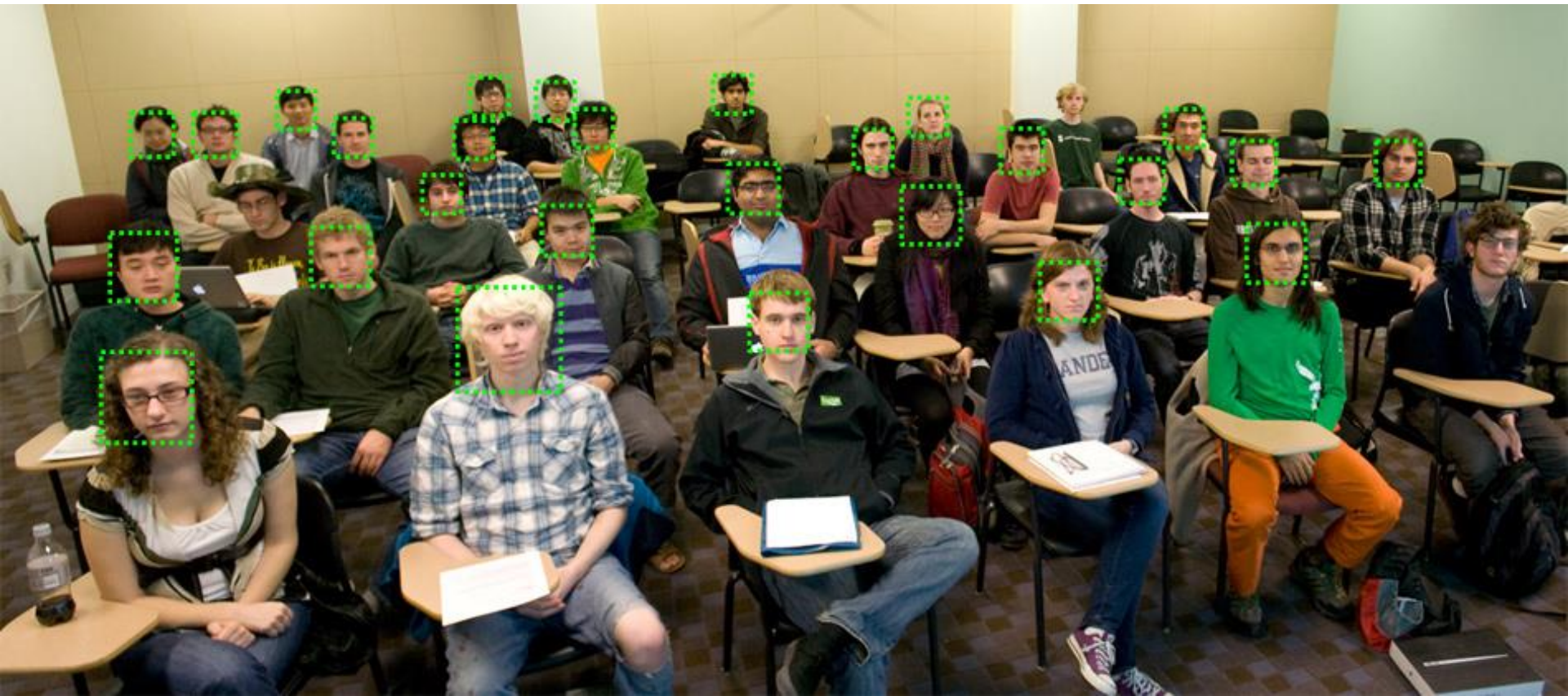
Proj4: Scene Recognition with Bag of Words

- Quantize local features into a “vocabulary”, describe images as histograms of “visual words”, train classifiers to recognize scenes based on these histograms.



Proj5: Object Detection with a Sliding Window

- Train a face detector based on positive examples and “mined” hard negatives, detect faces at multiple scales and suppress duplicate detections.



Course Syllabus (tentative)

http://www.cc.gatech.edu/~hays/compvision

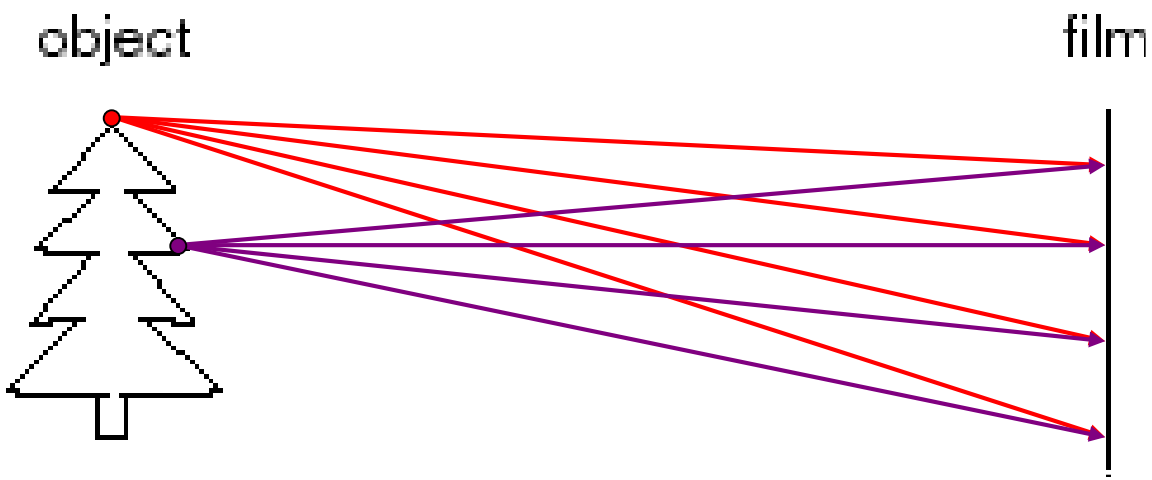
The Geometry of Image Formation

Mapping between image and world coordinates

- Pinhole camera model
- Projective geometry
  - Vanishing points and lines
- Projection matrix

What do you need to make a camera from scratch?

Image formation

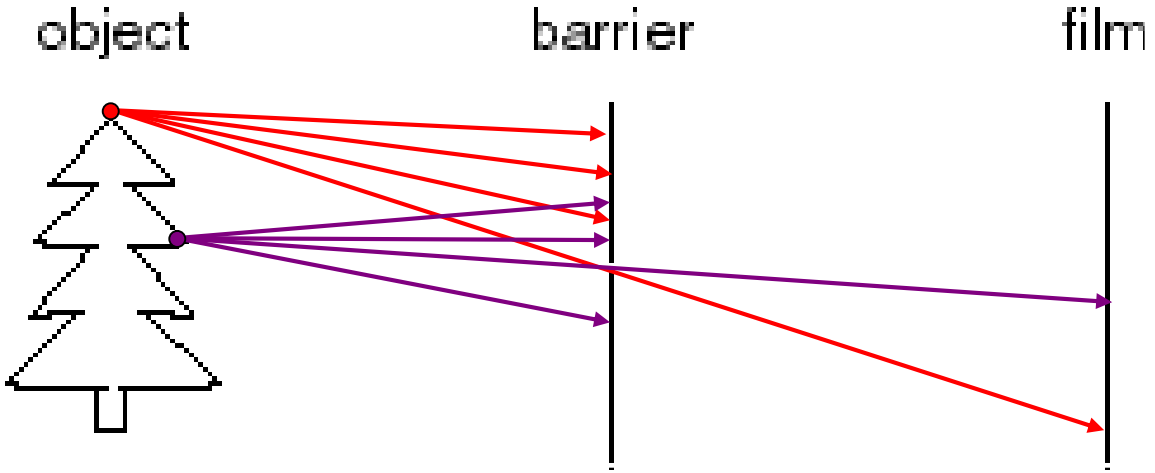


Let's design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

Slide source: Seitz

Pinhole camera

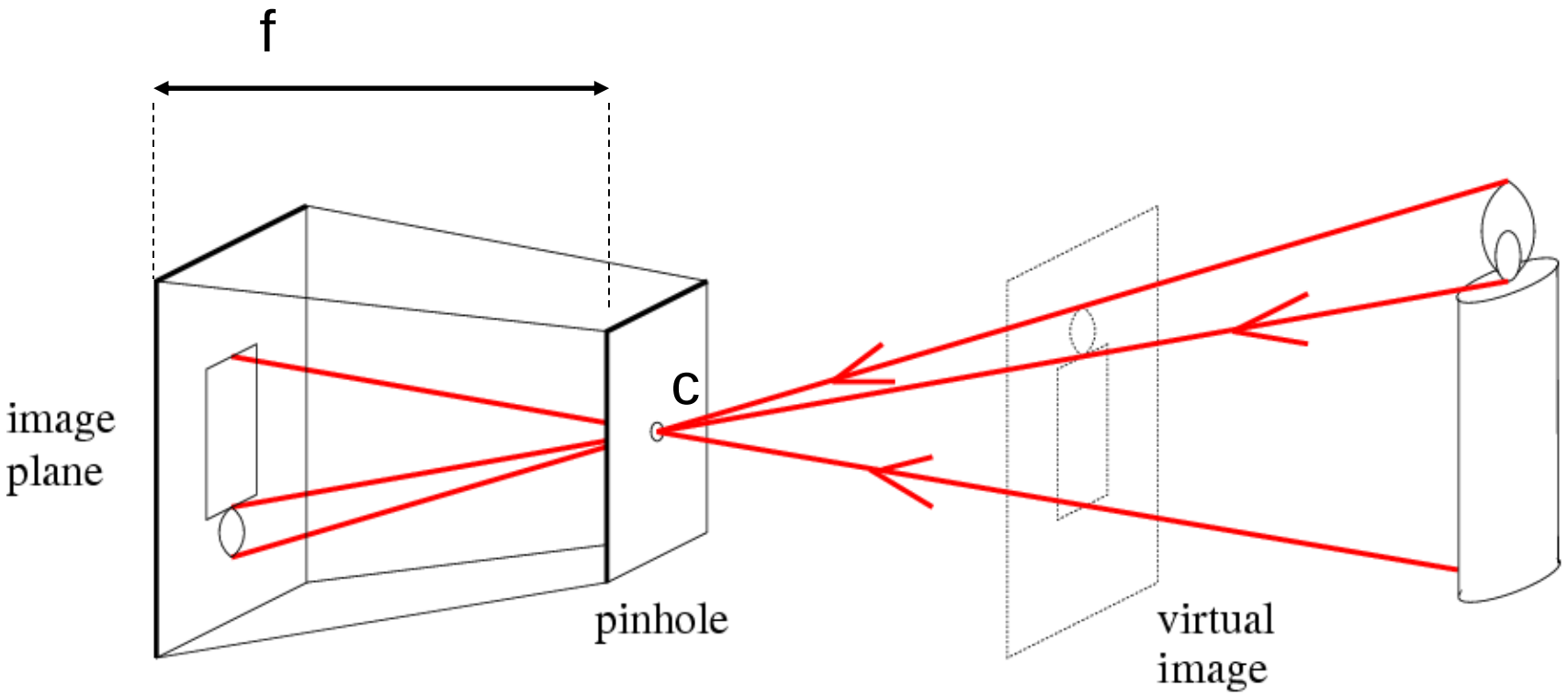


Idea 2: add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**

Slide source: Seitz

Pinhole camera



f = focal length
c = center of the camera

Figure from Forsyth

Camera obscura: the pre-camera

- Known during classical period in China and Greece (e.g. Mo-Ti, China, 470BC to 390BC)

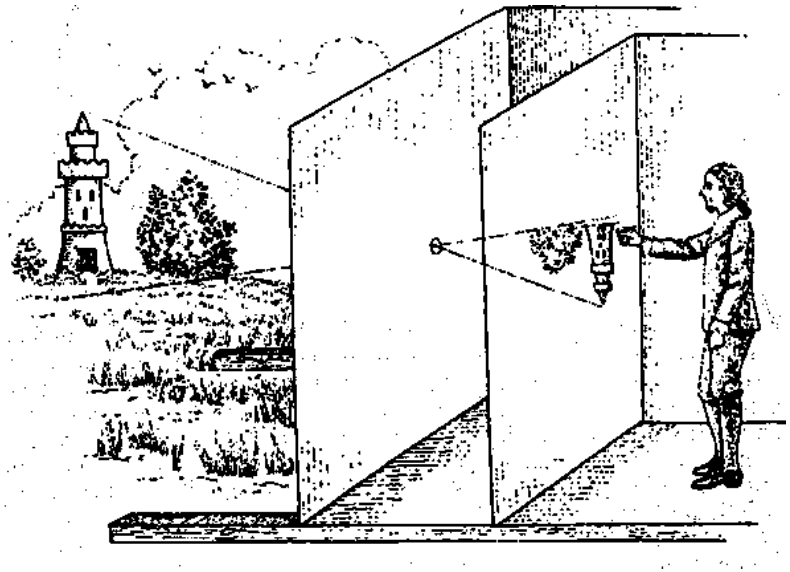
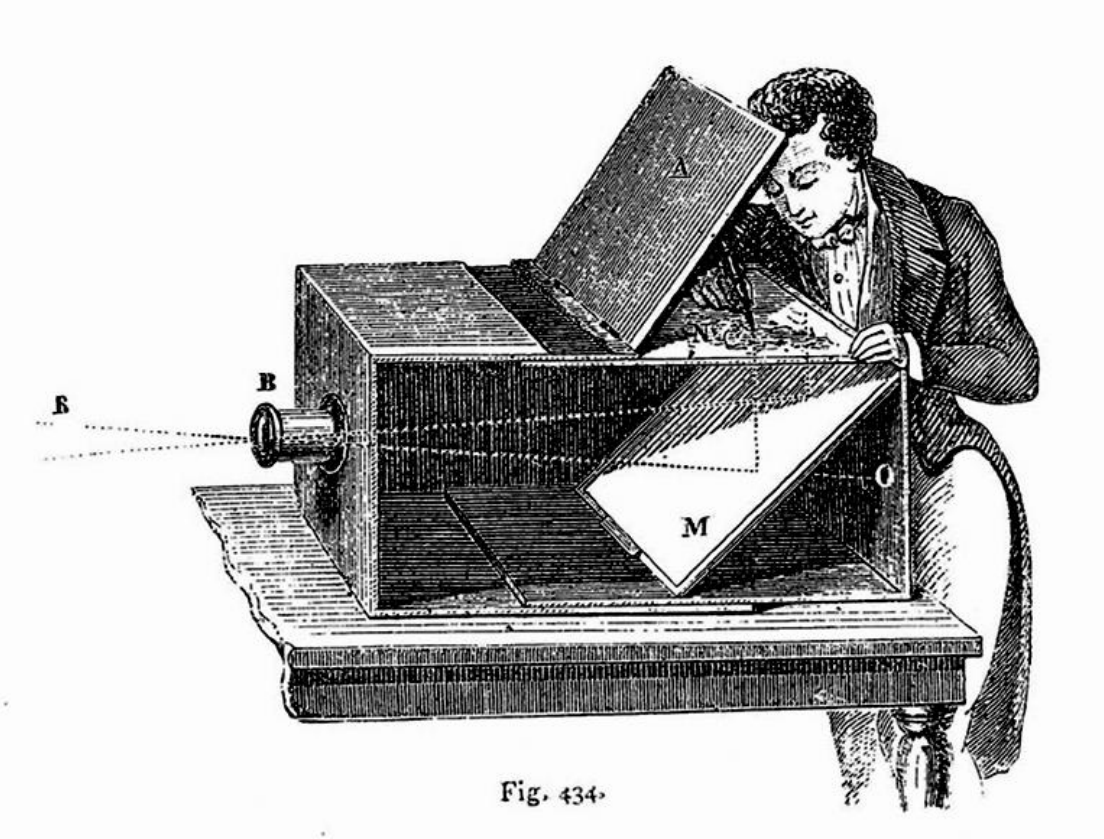


Illustration of Camera Obscura

Freestanding camera obscura at UNC Chapel Hill

Photo by Seth Ilys

Camera Obscura used for Tracing

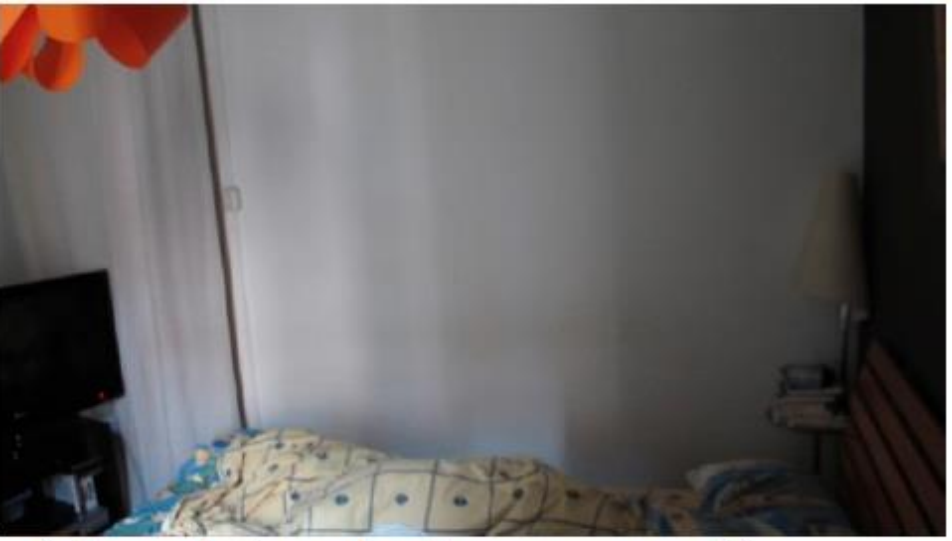


Lens Based Camera Obscura, 1568

Accidental Cameras

Accidental Pinhole and Pinspeck Cameras
Revealing the scene outside the picture.
Antonio Torralba, William T. Freeman

Accidental Cameras



a) Input (occluder present)

b) Reference (occluder absent)

c) Difference image (b-a)

d) Crop upside down

e) True view

First Photograph

Oldest surviving photograph

- Took 8 hours on pewter plate

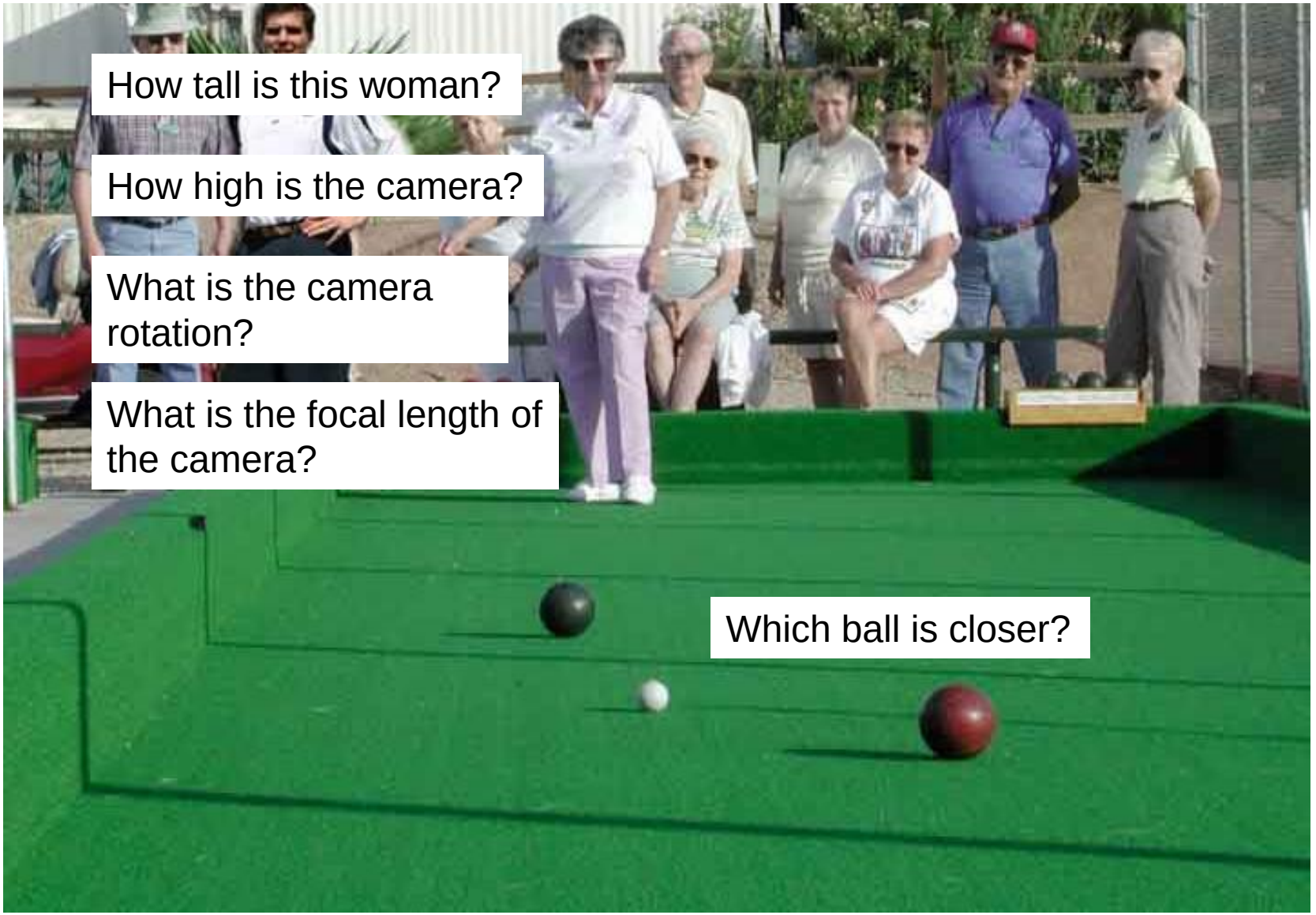
Joseph Niepce, 1826

Photograph of the first photograph

Stored at UT Austin

Niepce later teamed up with Daguerre, who eventually created Daguerrotypes

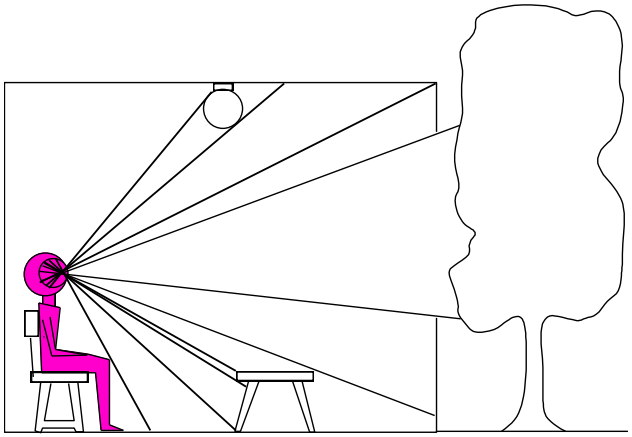
Today's class: Camera and World Geometry



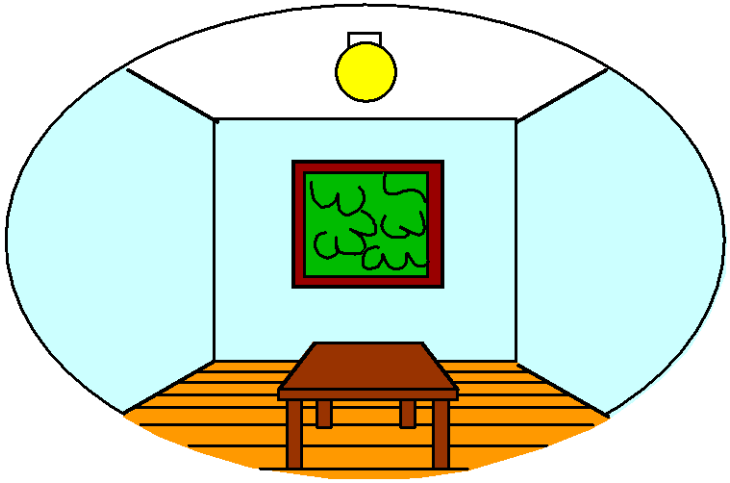
Dimensionality Reduction Machine (3D to 2D)

3D world

2D image



Point of observation



Slide source: Seitz

Projection can be tricky…

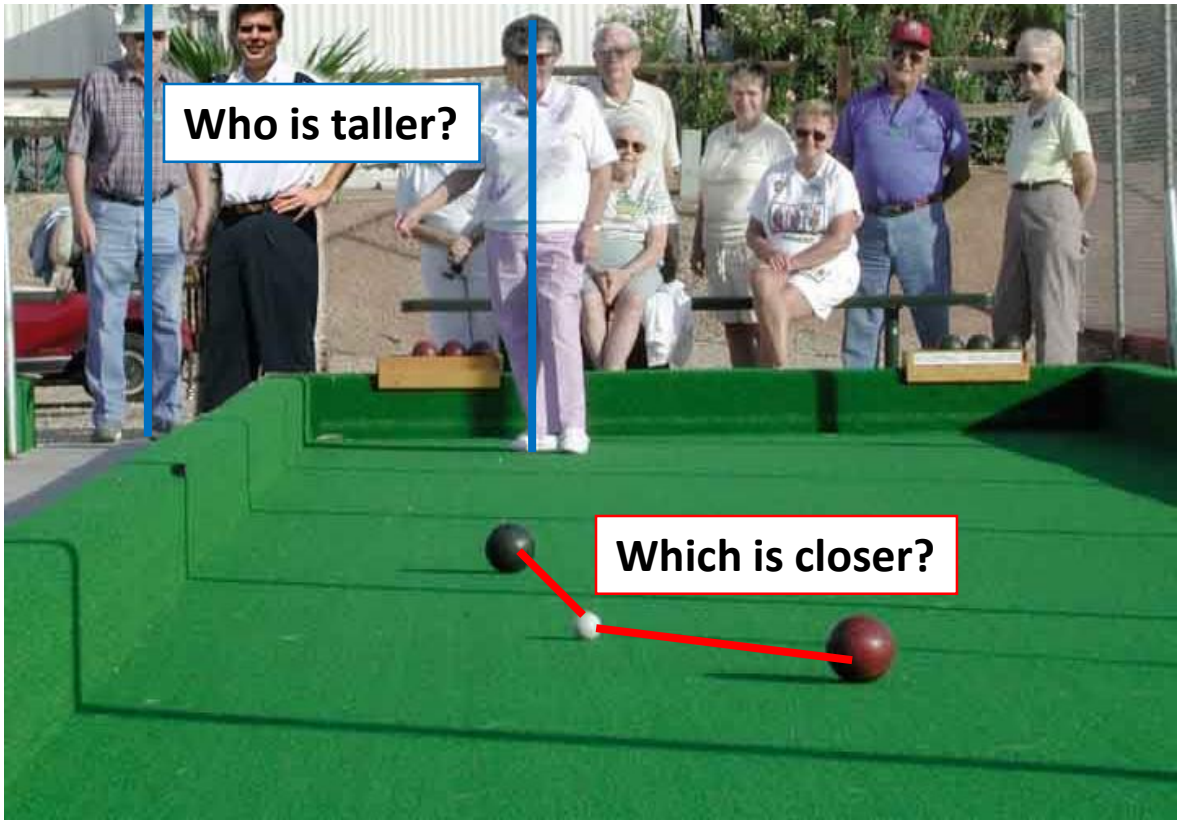
Slide source: Seitz

Projection can be tricky…

Projective Geometry

What is lost?

- Length



Length and area are not preserved

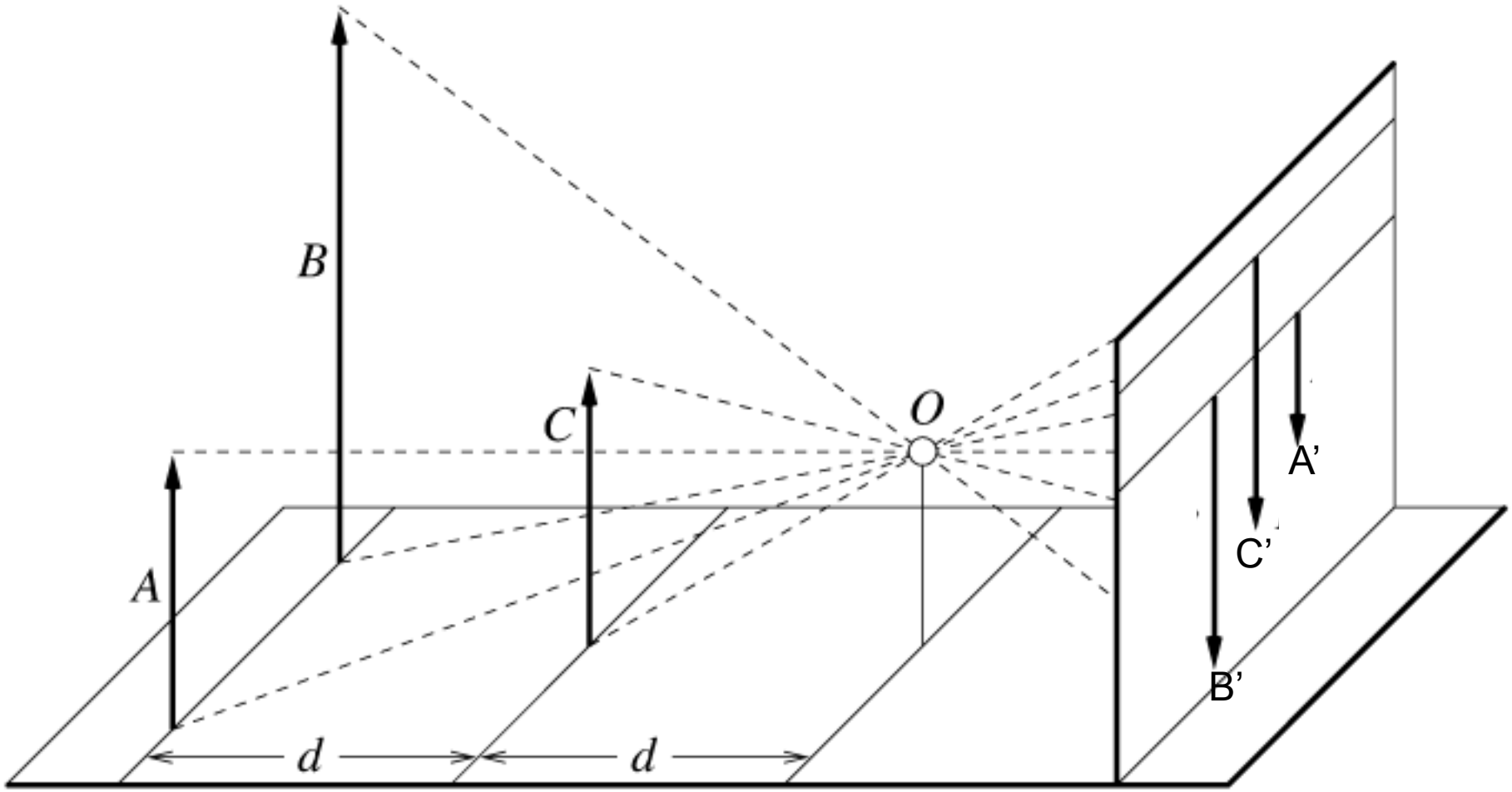
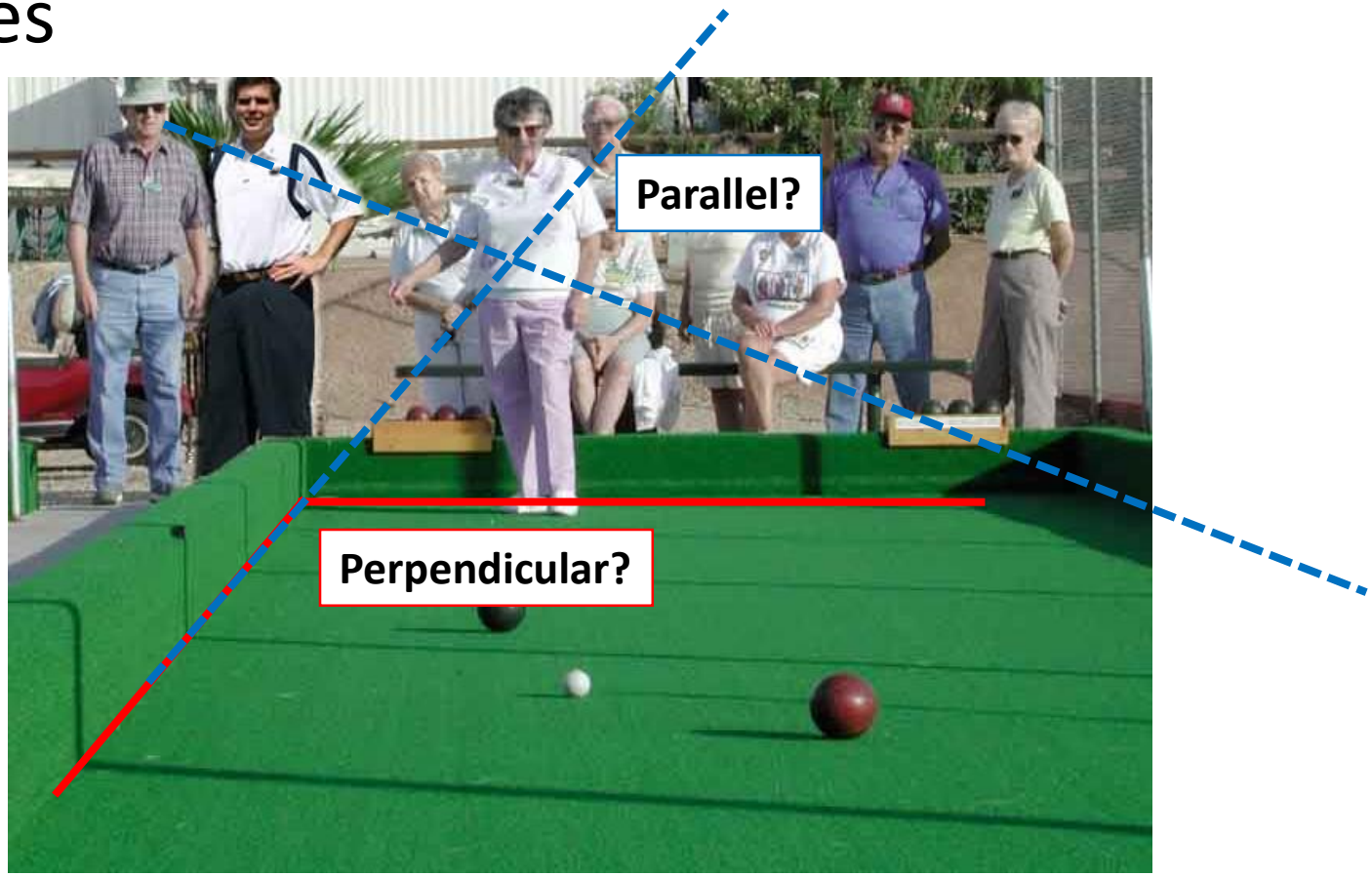


Figure by David Forsyth

Projective Geometry

What is lost?

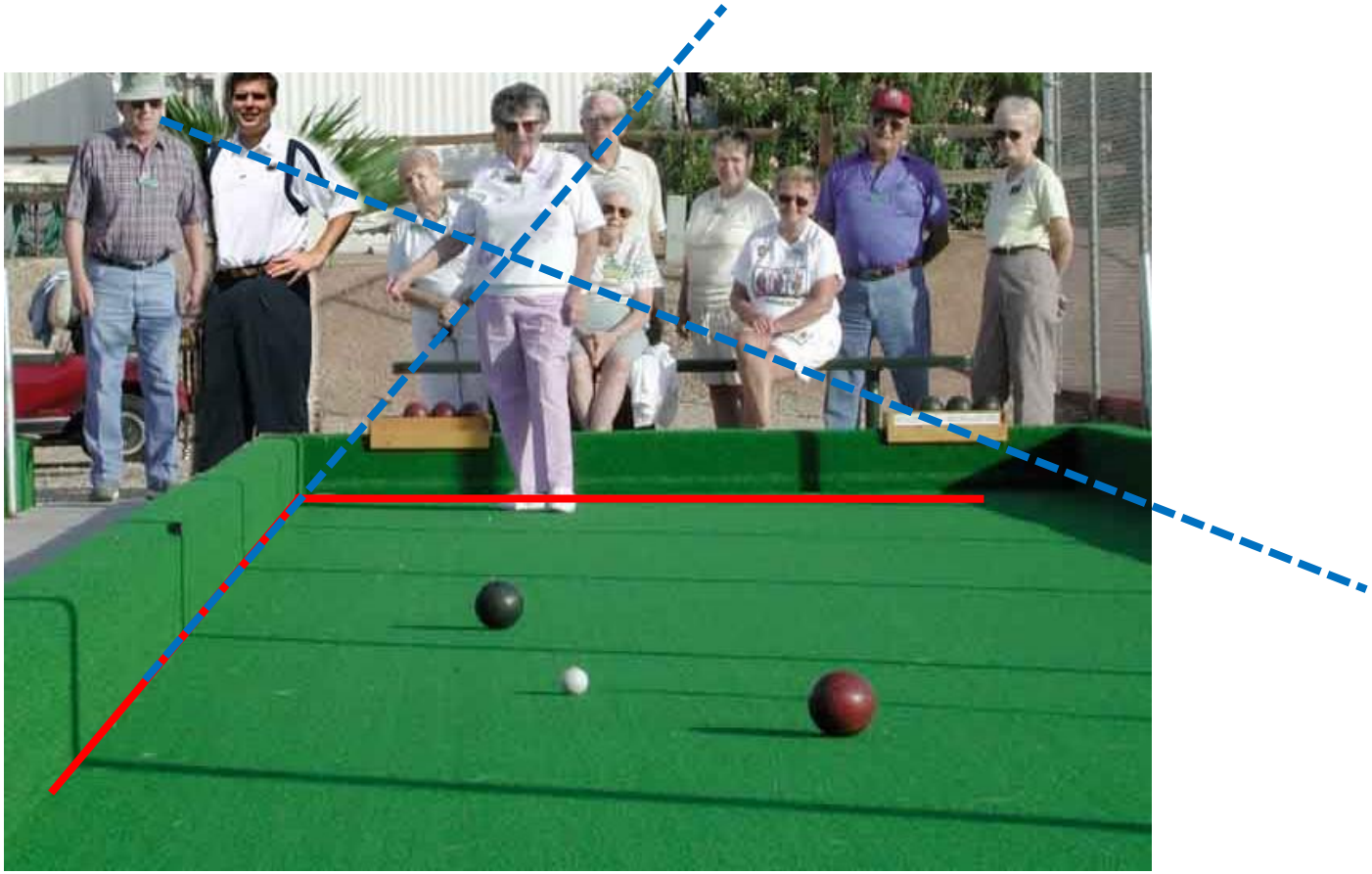
- Length
- Angles



Projective Geometry

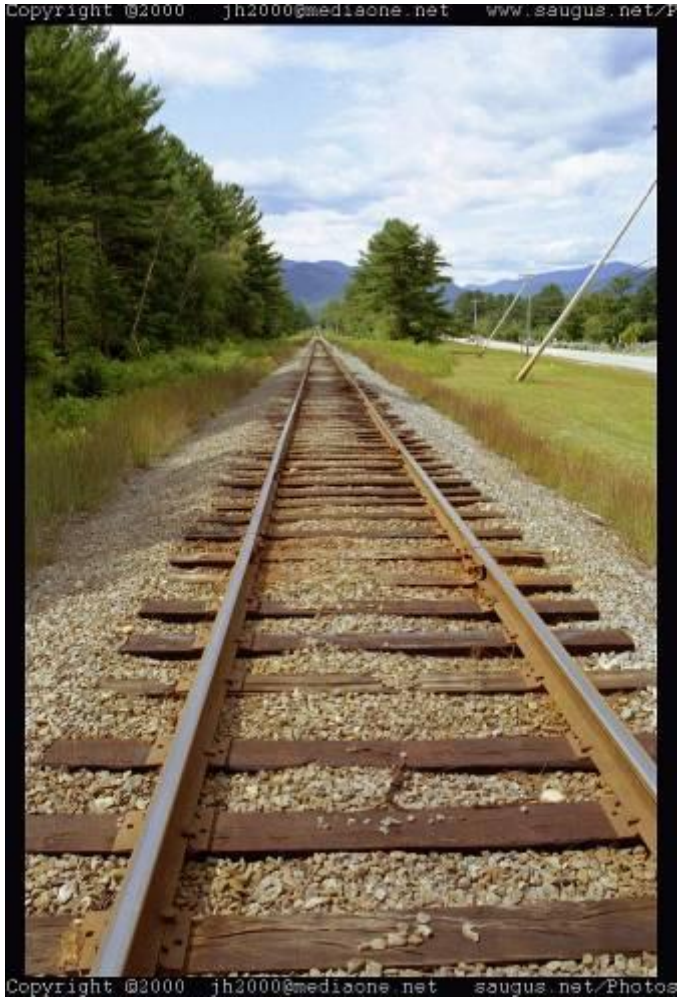
What is preserved?

- Straight lines are still straight

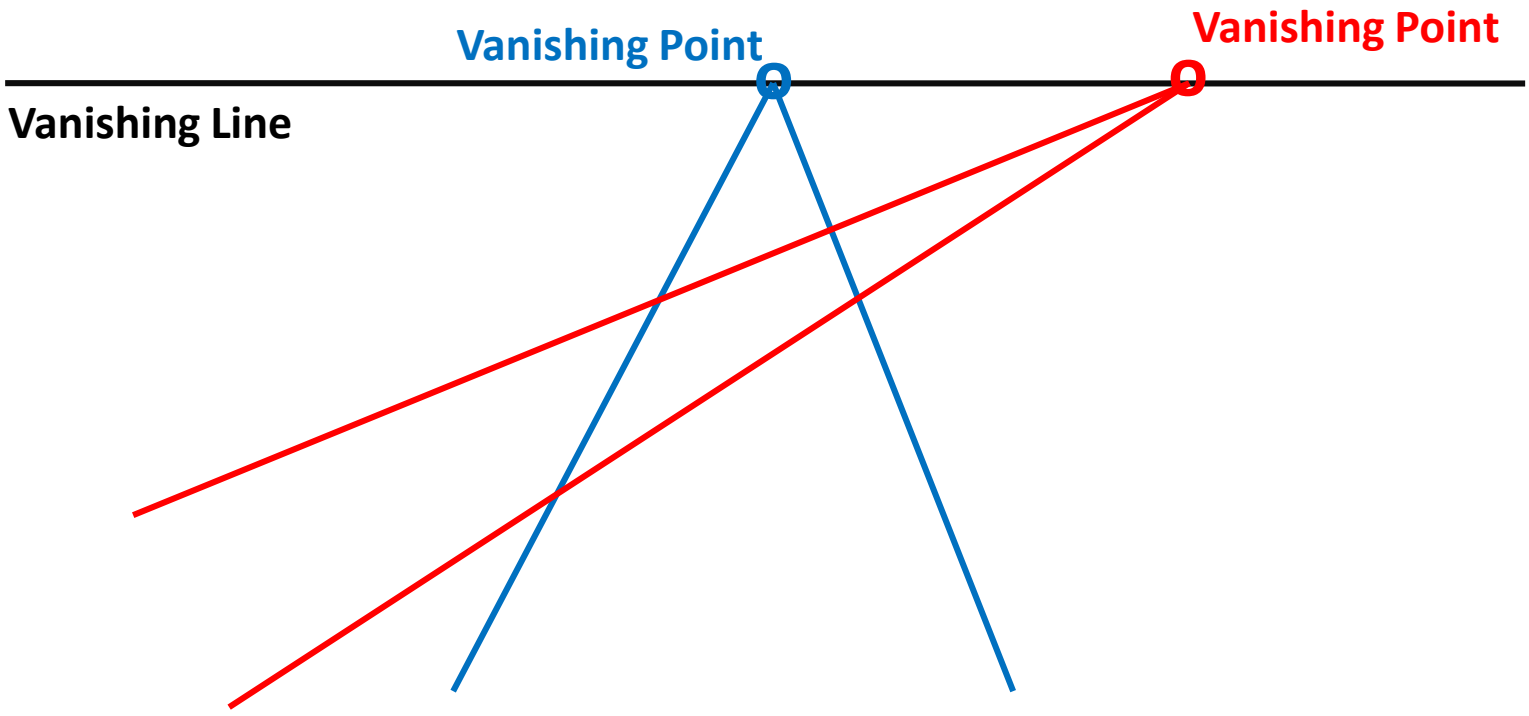


Vanishing points and lines

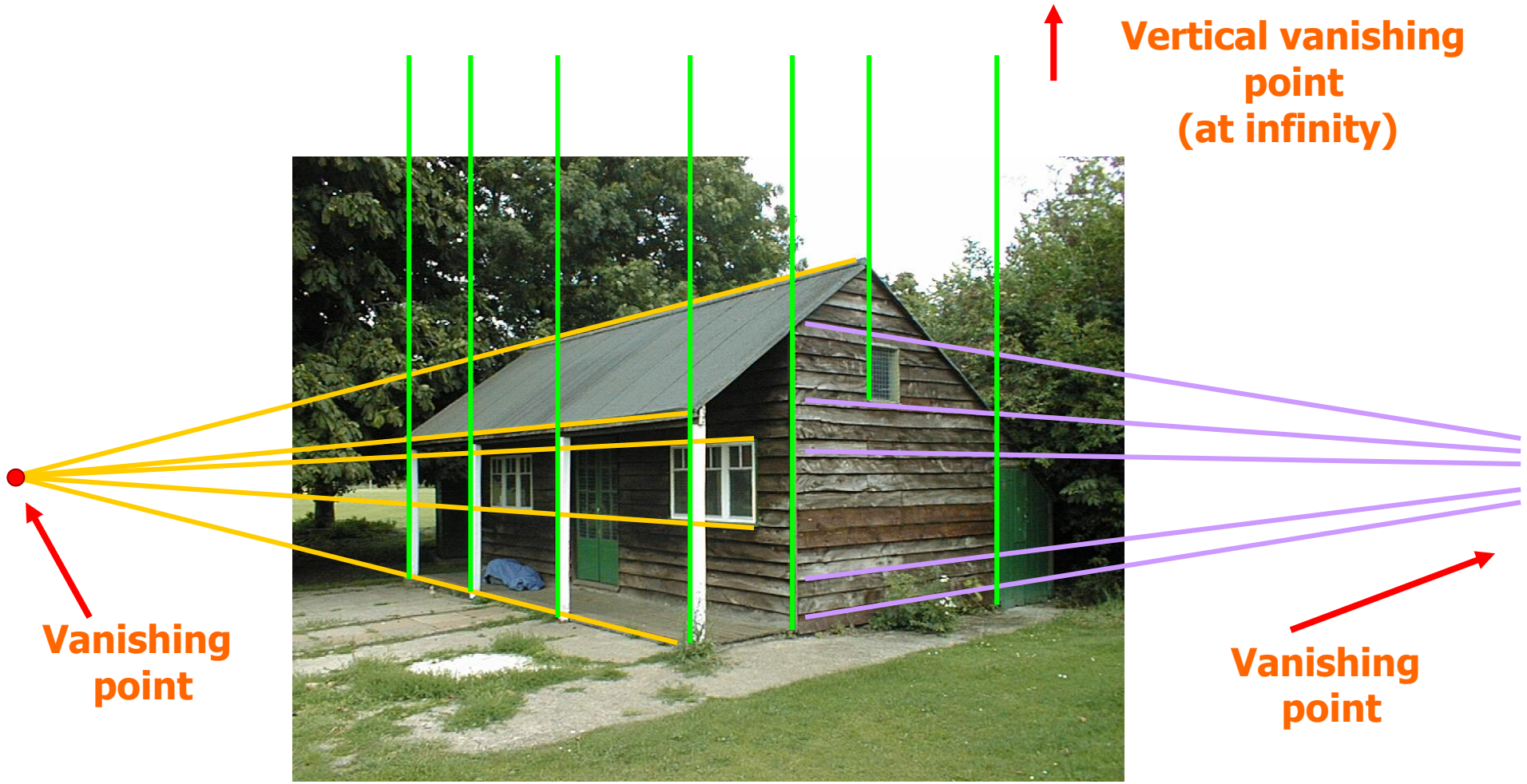
Parallel lines in the world intersect in the image at a “vanishing point”



Vanishing points and lines

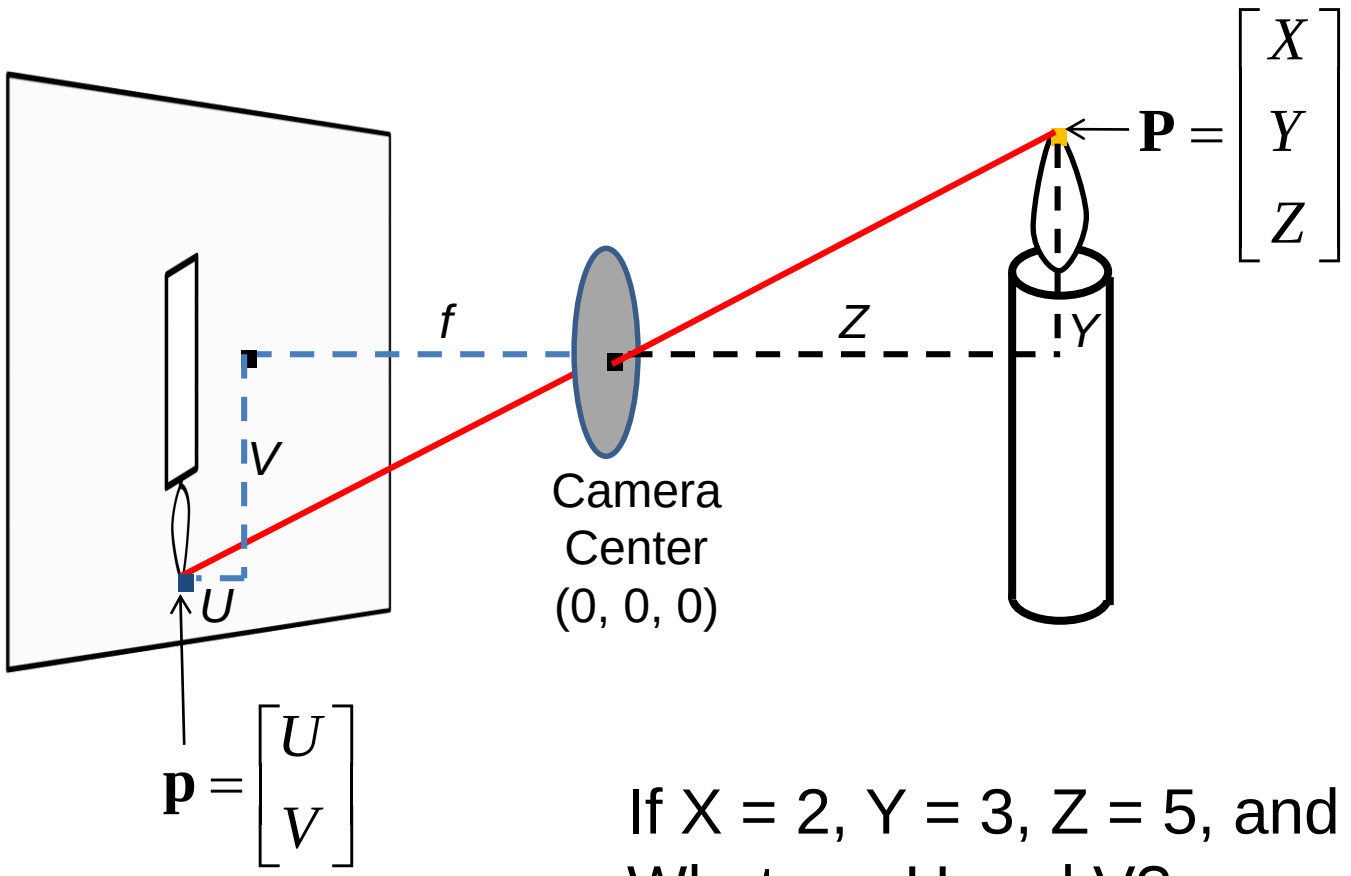


Vanishing points and lines



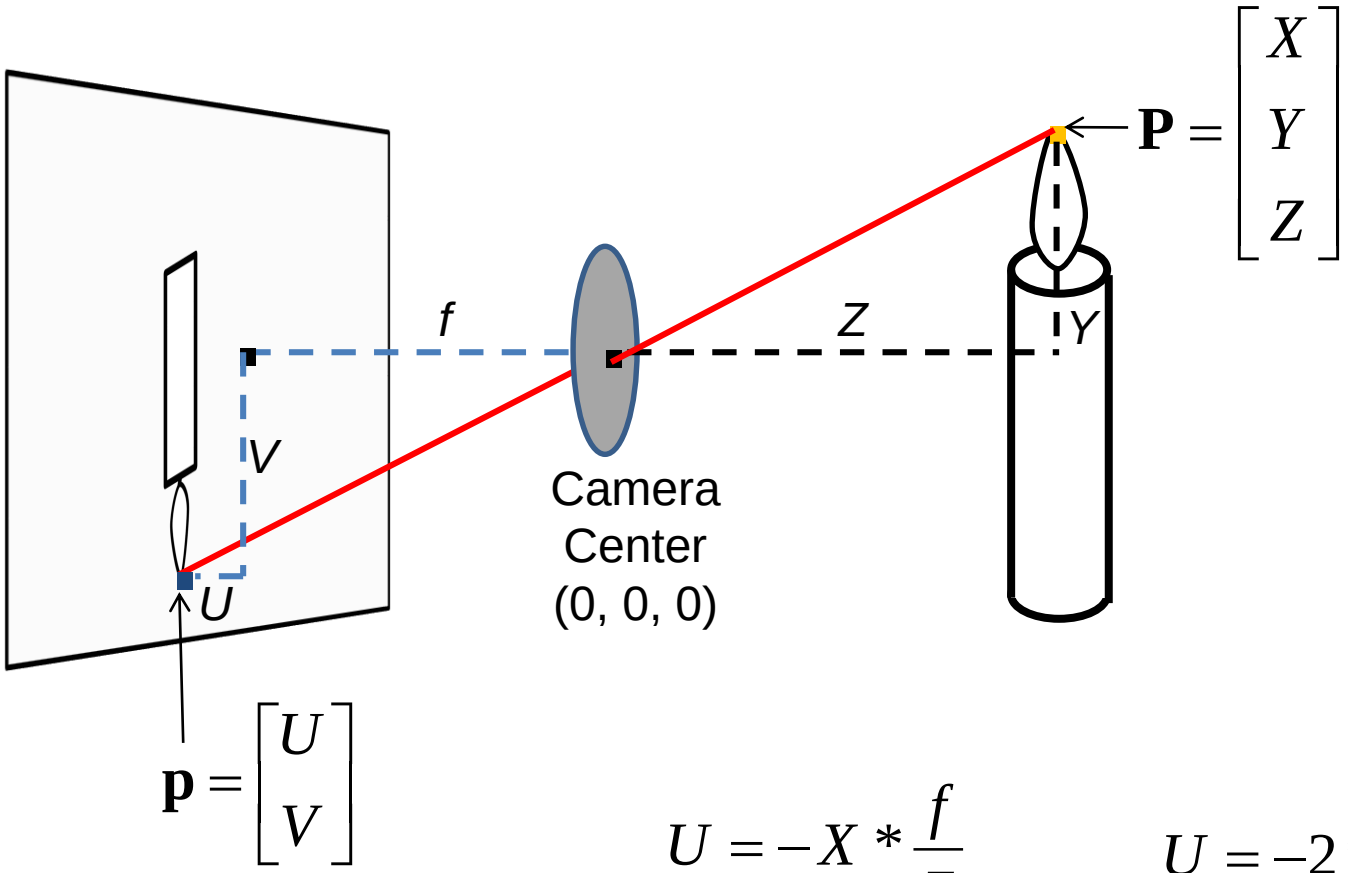
Slide from Efros, Photo from Criminisi

Projection: world coordinates→image coordinates



If X = 2, Y = 3, Z = 5, and f = 2
What are U and V?

Projection: world coordinates→image coordinates



$$U = -X * \frac{f}{Z} \qquad U = -2 * \frac{2}{5}$$

$$V = -Y * \frac{f}{Z} \qquad V = -3 * \frac{2}{5}$$

Sanity check, what if f and Z are equal?

Projection: world coordinates→image coordinates