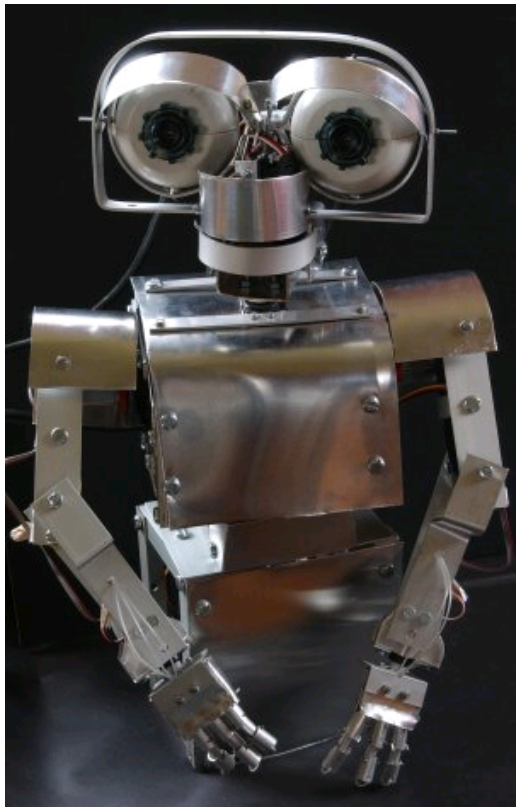
Computer Vision and Machine Learning

Erik Sudderth
Dept. of Computer Science

(slides courtesy Charless Fowlkes)

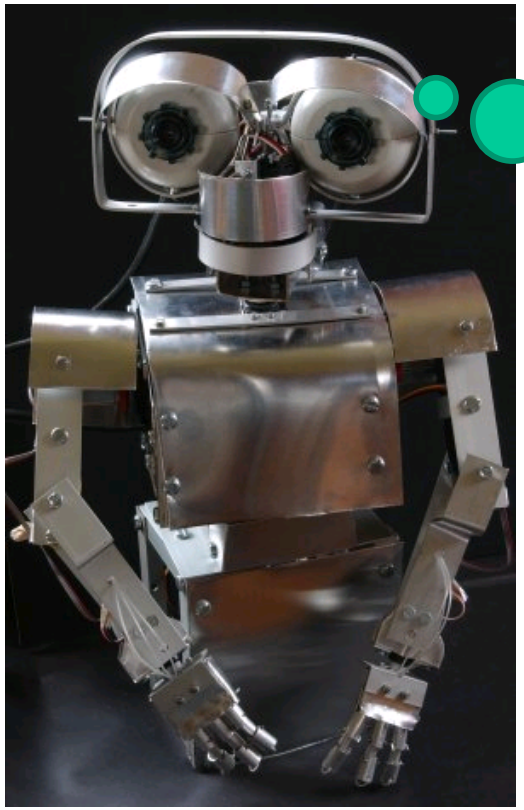
Computer Vision

How can we build artificial systems that can “make sense” of what is in an image?



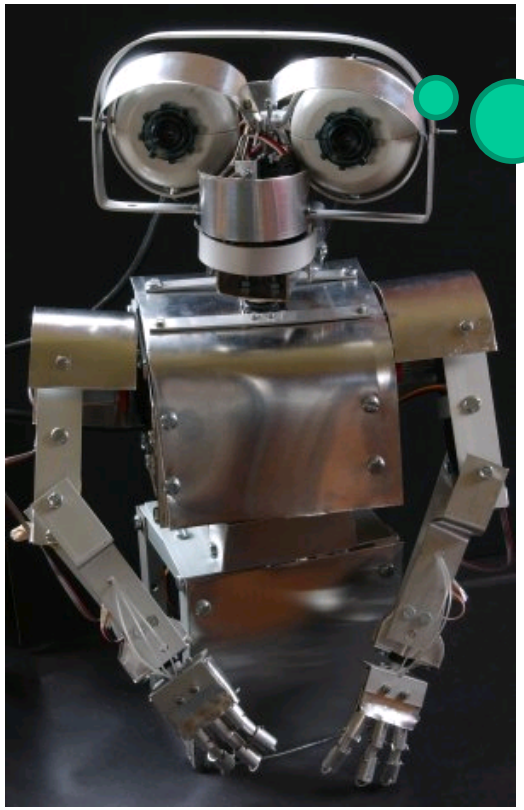
Computer Vision

How can we build artificial systems that can “make sense” of what is in an image?



Computer Vision

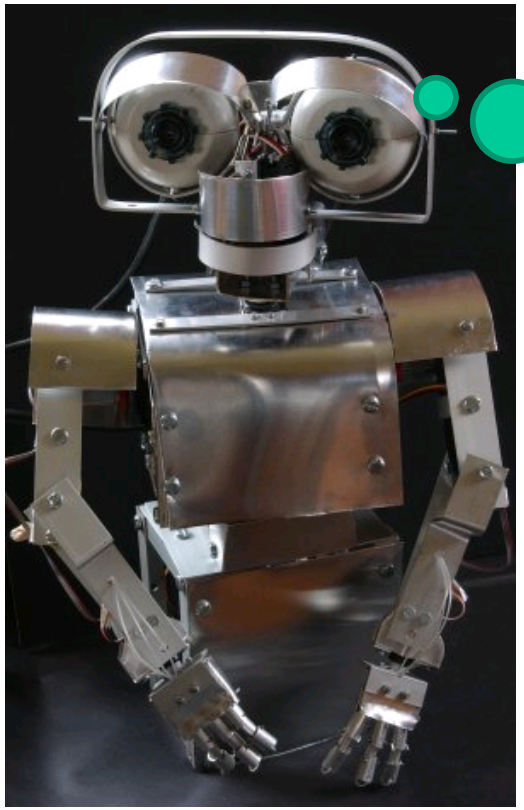
How can we build artificial systems that can “make sense” of what is in an image?



three bunnies

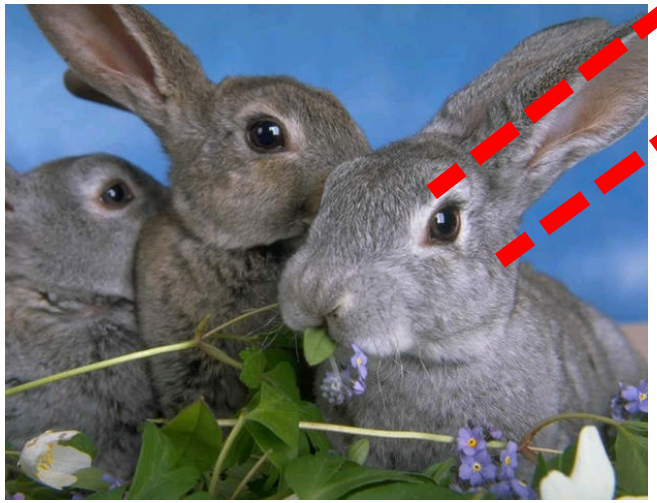
Computer Vision

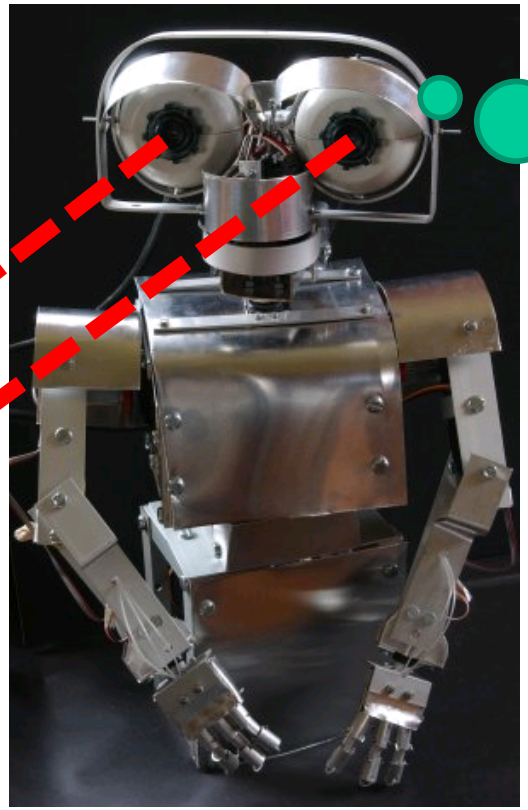
How can we build artificial systems that can “make sense” of what is in an image?



Computer Vision

How can we build artificial systems that can “make sense” of what is in an image?



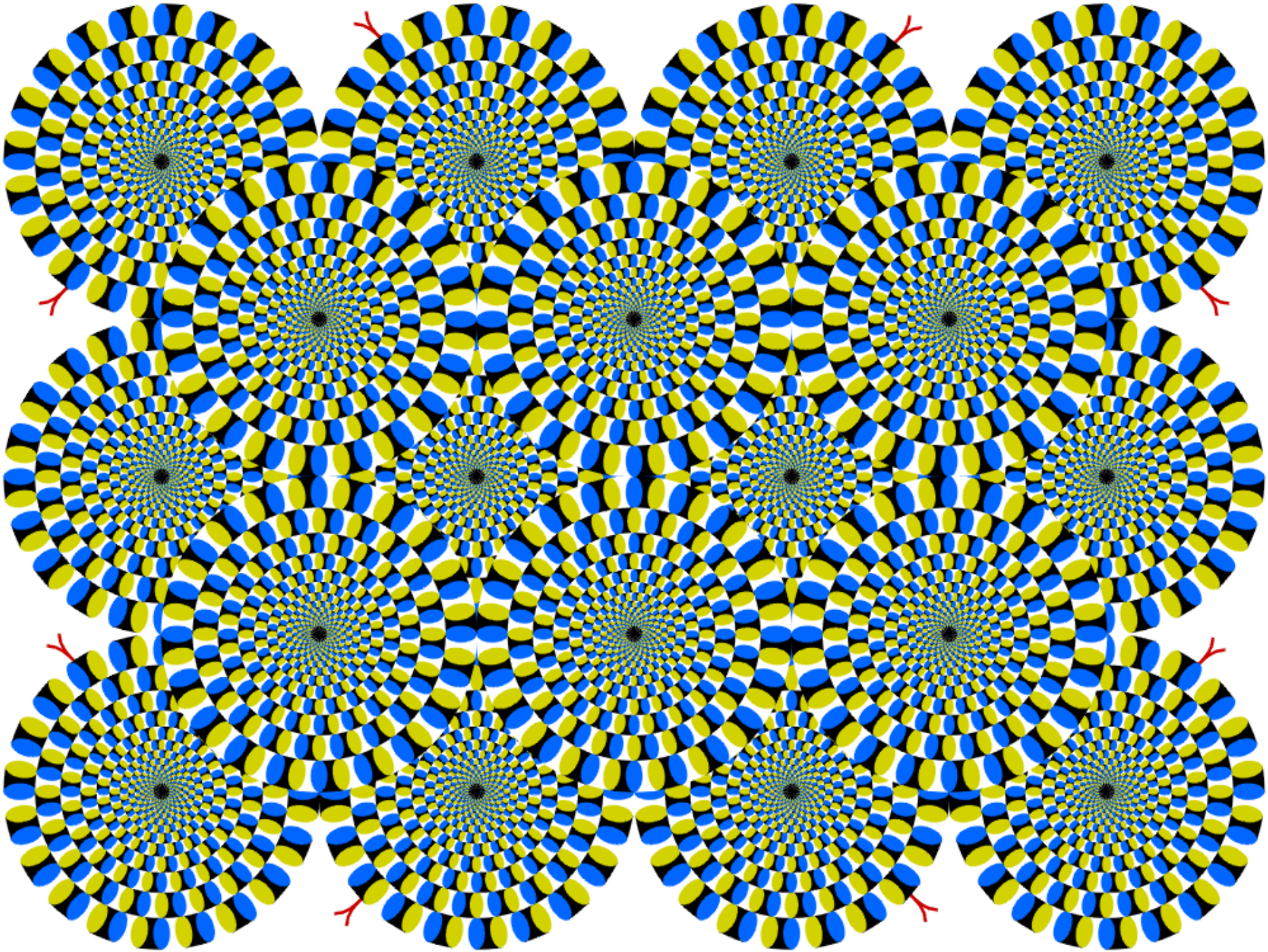


Every picture tells a story

Can computers match (or beat) human vision?

Yes and no (but mostly no!)

- humans are still much better at high-level image understanding



Current state of the art

What can current vision systems do?

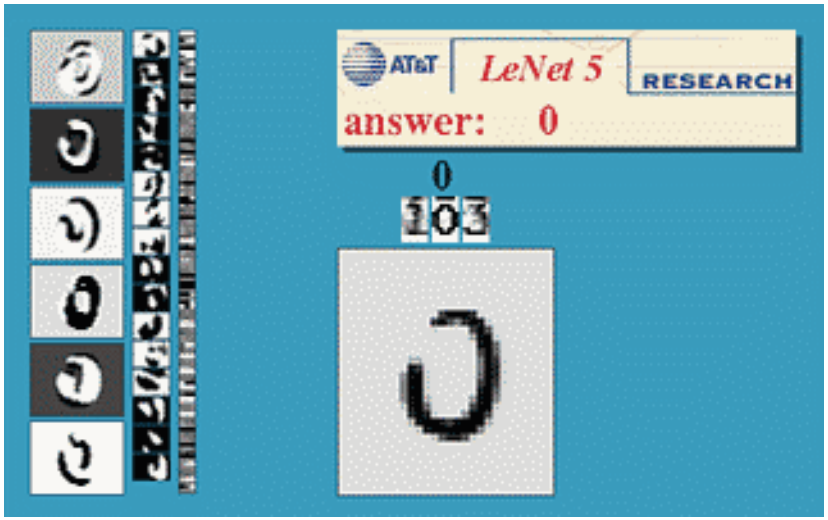
- recognize and detect (some) categories of objects
- recover (some) aspects of 3D scene geometry from multiple images

Some examples of working computer vision systems…

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

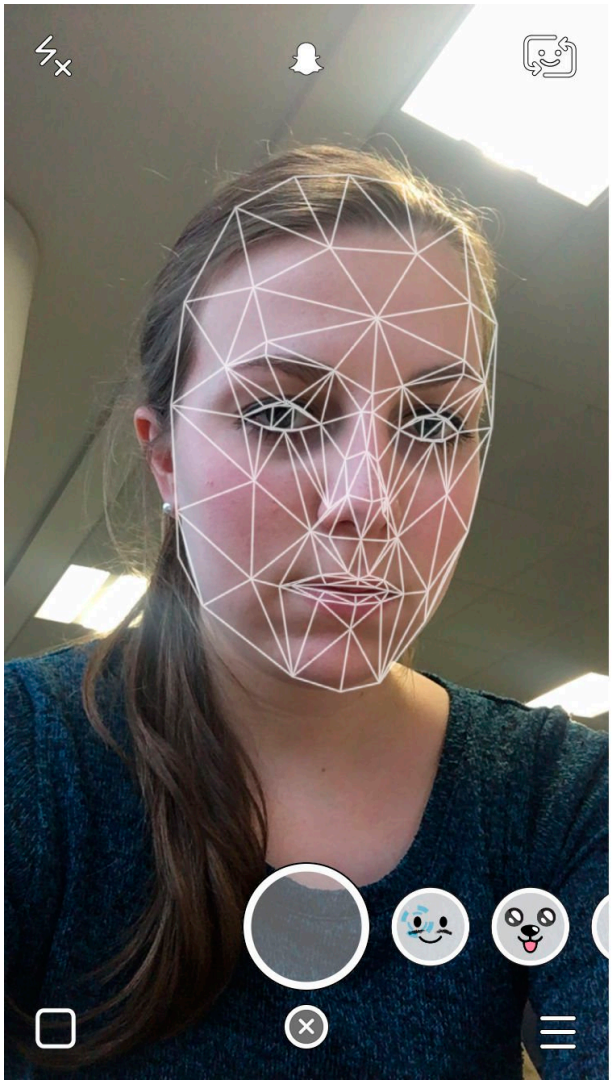
Face detection

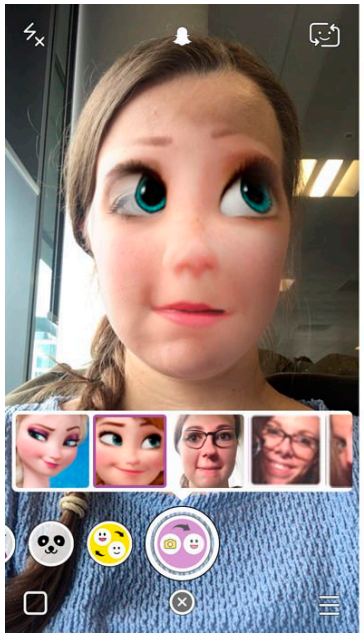
Many digital cameras now detect faces

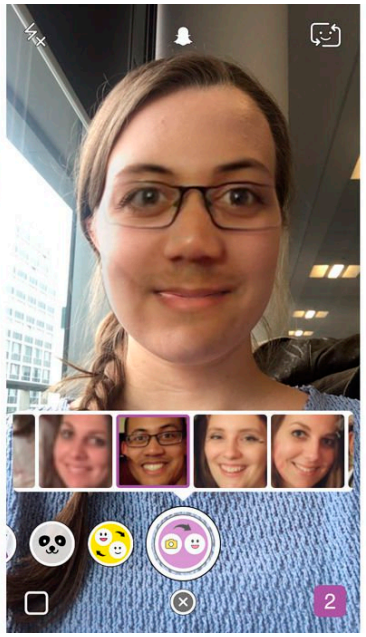
- Canon, Sony, Fuji, …

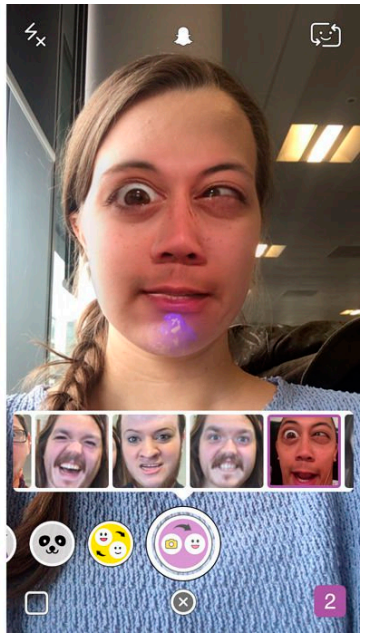
Online photo organization (Google, Facebook, etc.)

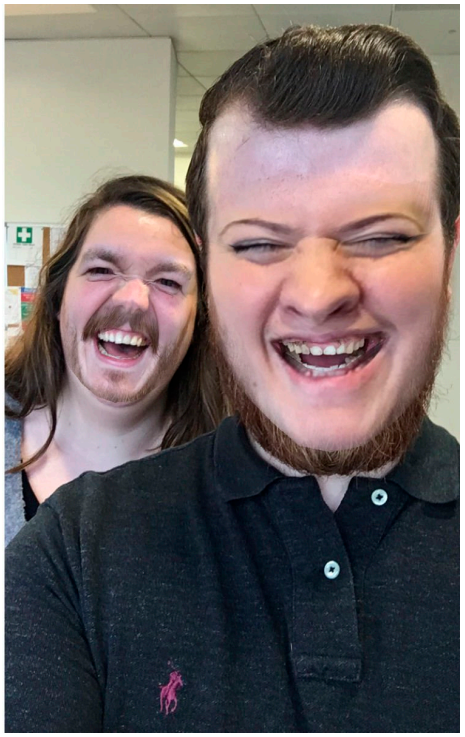
SnapChat Face Swap





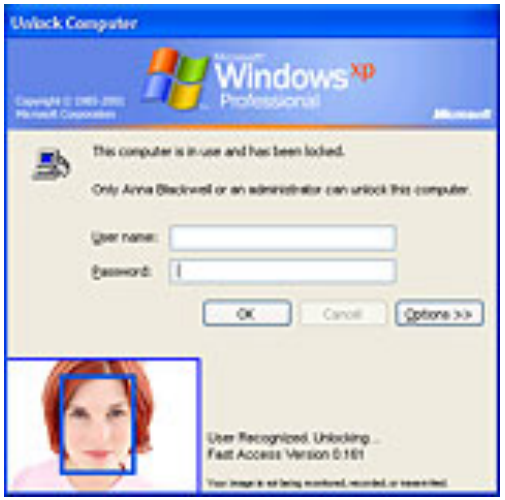






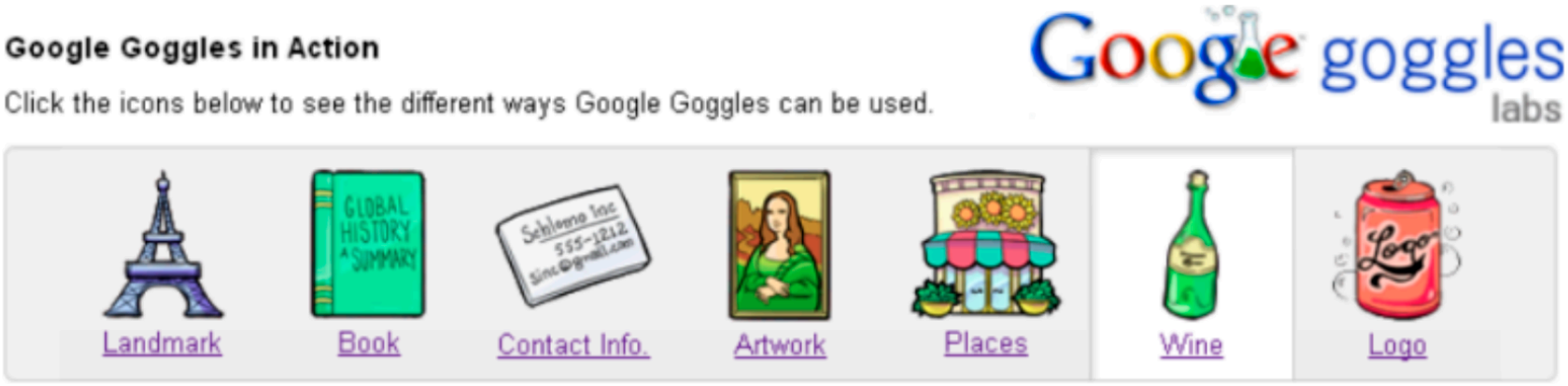
Login without a password…

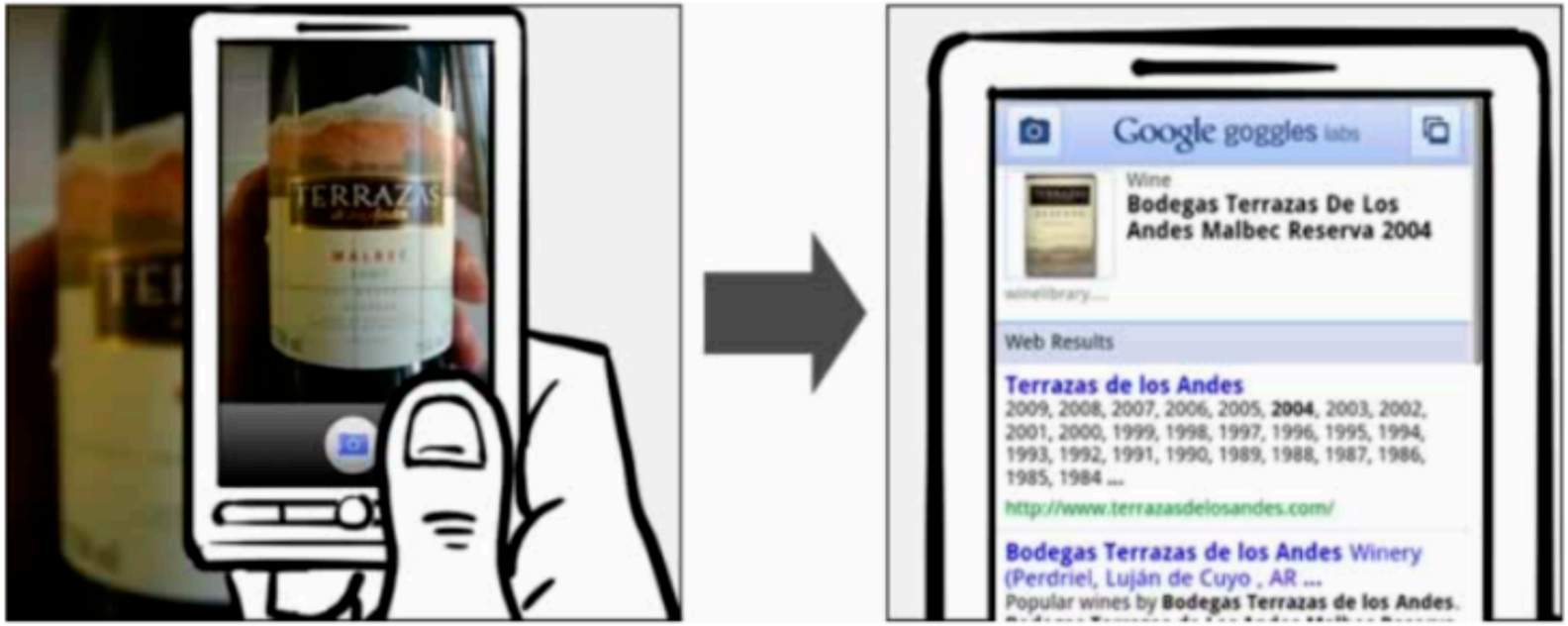
Fingerprint scanners on many new laptops, other devices



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

Object recognition (in mobile phones)



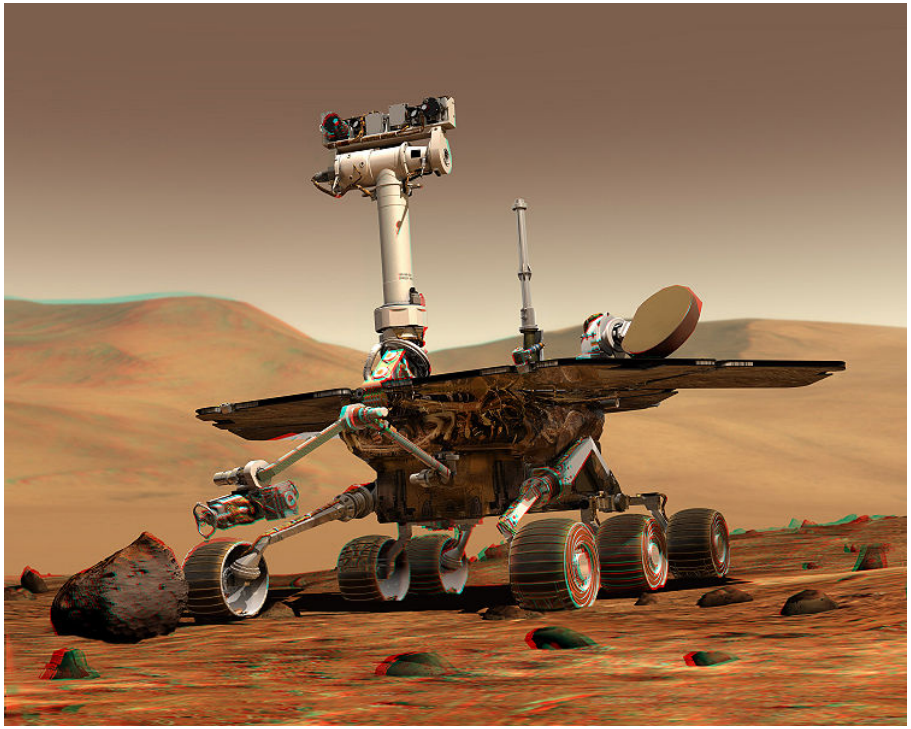


Smart cars

[Mobileye]

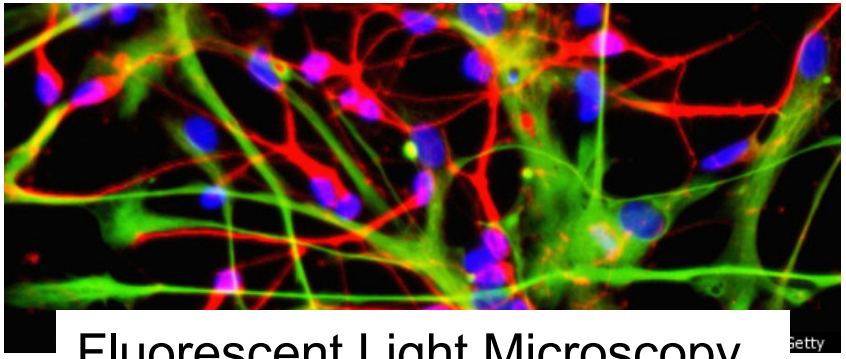
- Vision systems currently in high-end BMW, GM, Volvo models
- Early warning of collisions, pedestrians, lane changes etc.

Robotics

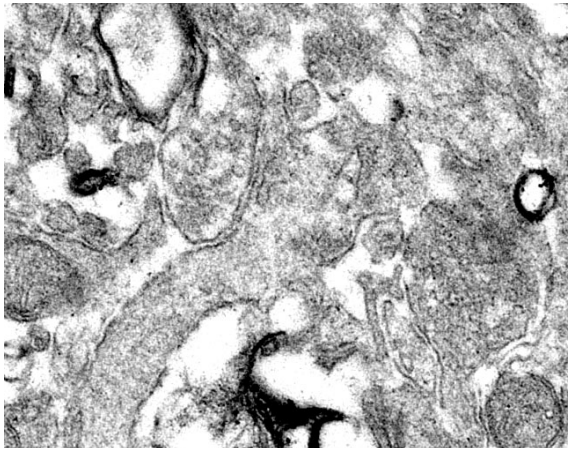


NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

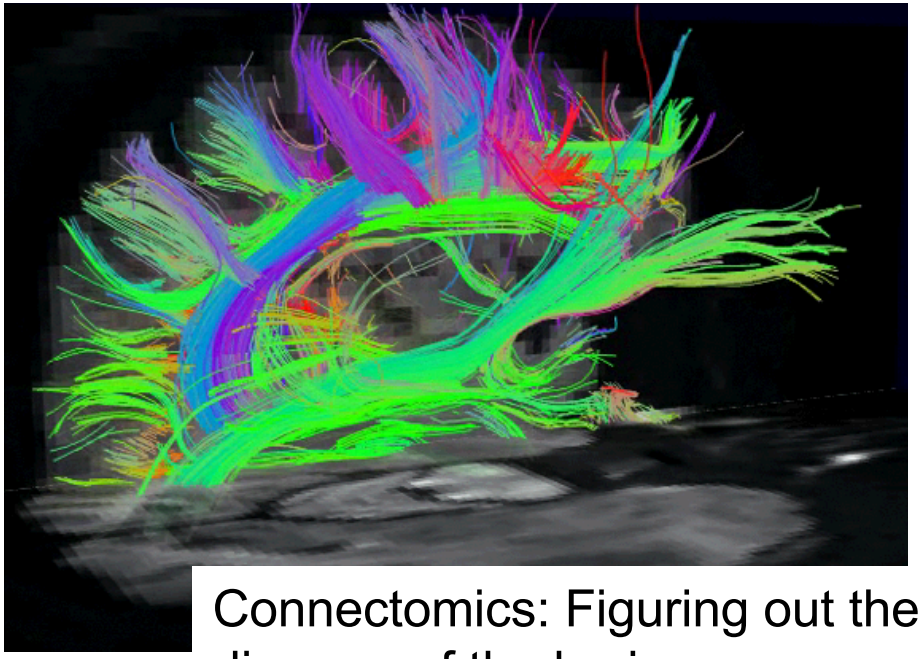
Biological Image Analysis



Fluorescent Light Microscopy



Scanning Electron Microscopy



Connectomics: Figuring out the wiring diagram of the brain

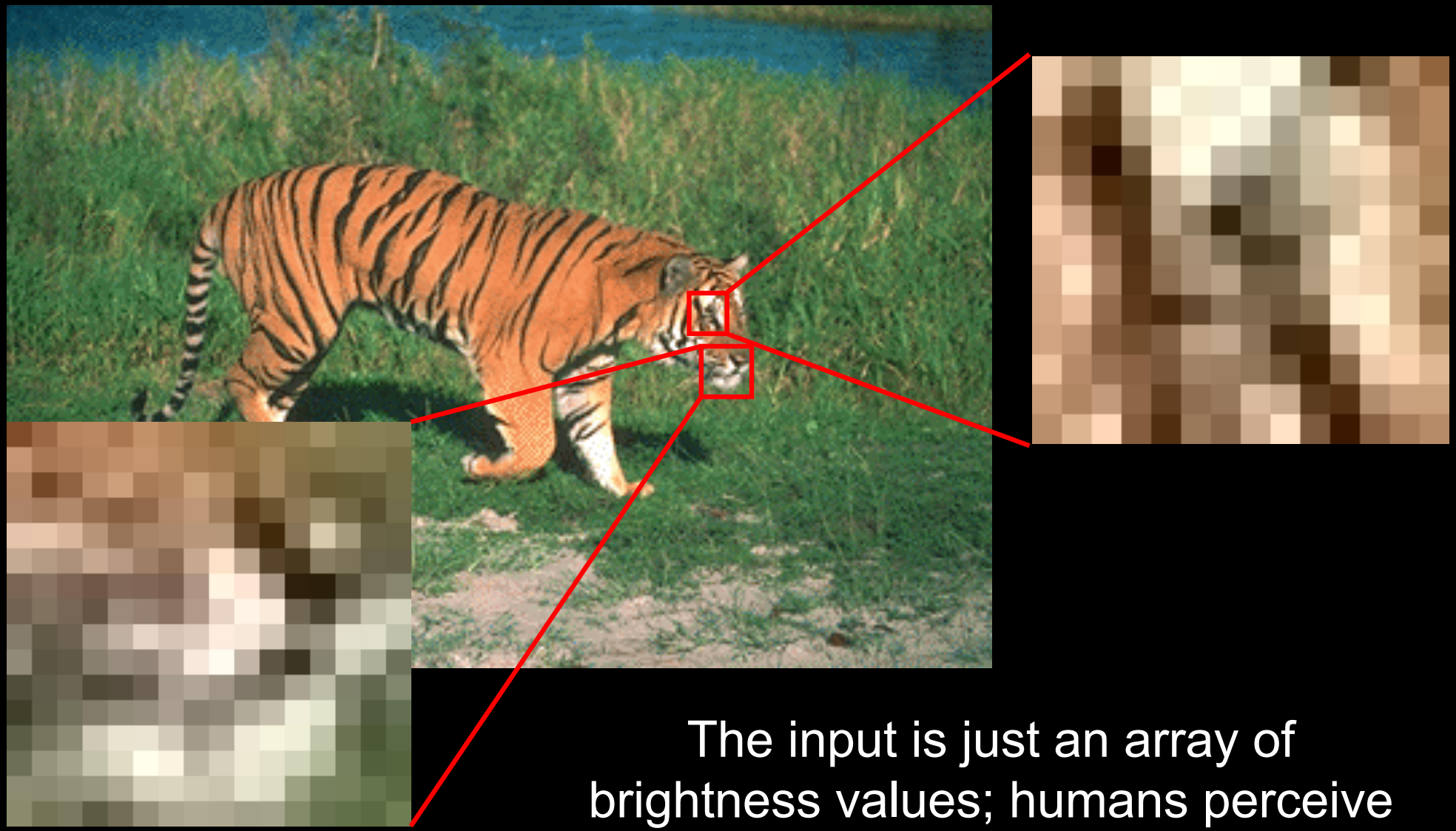
Overview

Introduction

Understanding what is in an image

Learning to see by modeling data

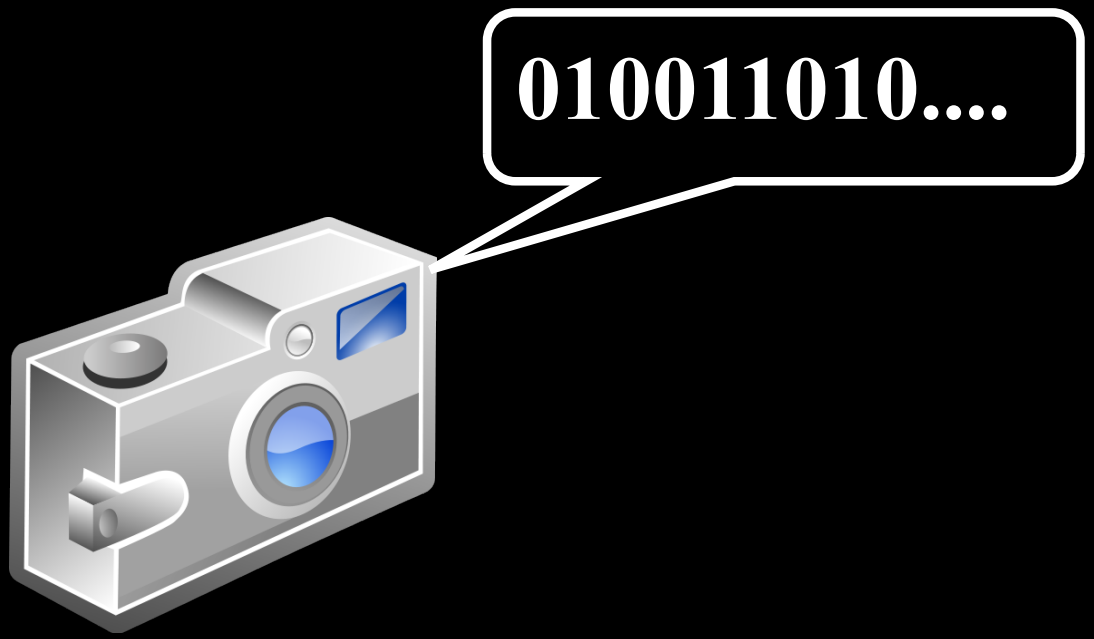
What is in an image?

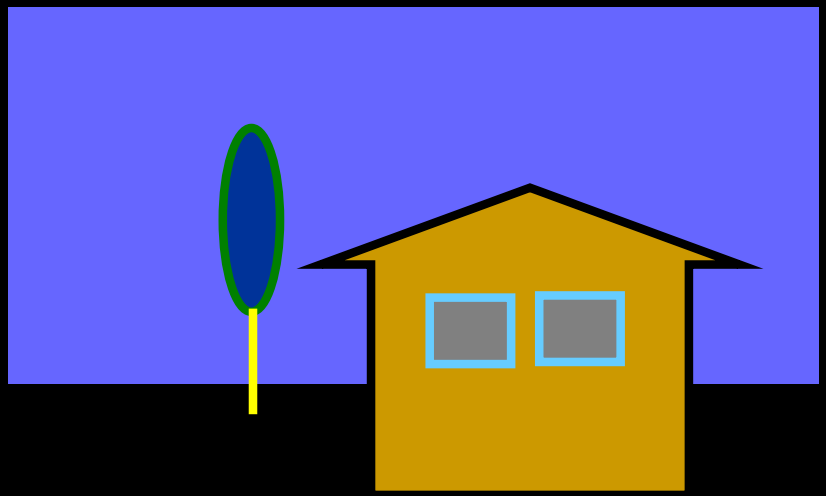


The input is just an array of brightness values; humans perceive structure in it.

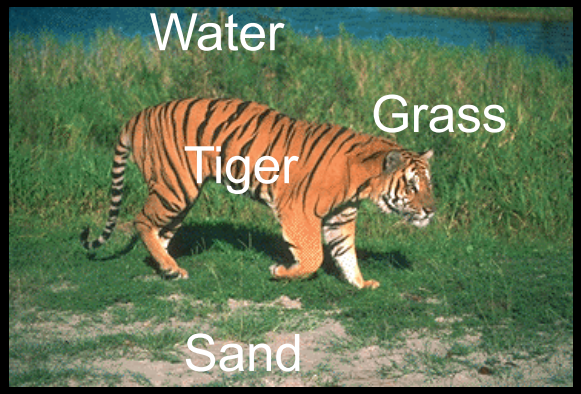
“ I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of color. Do I *have* 327? No. I have sky, house, and trees.”

Laws of Organization in Perceptual Forms
Max Wertheimer (1923)

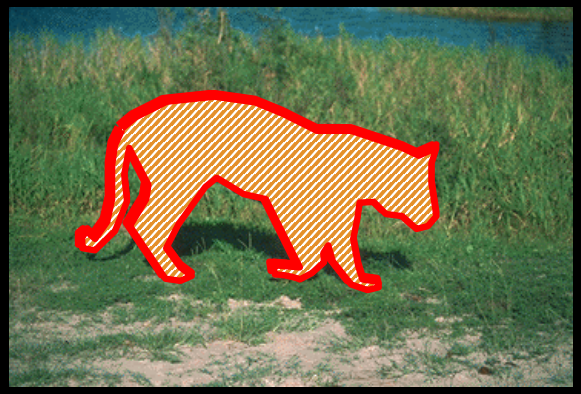


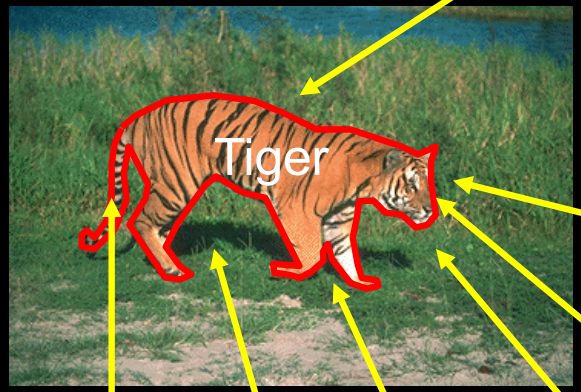


From Pixels to Perception



outdoor
wildlife





back
head
eye
mouse
legs
tail
shadow

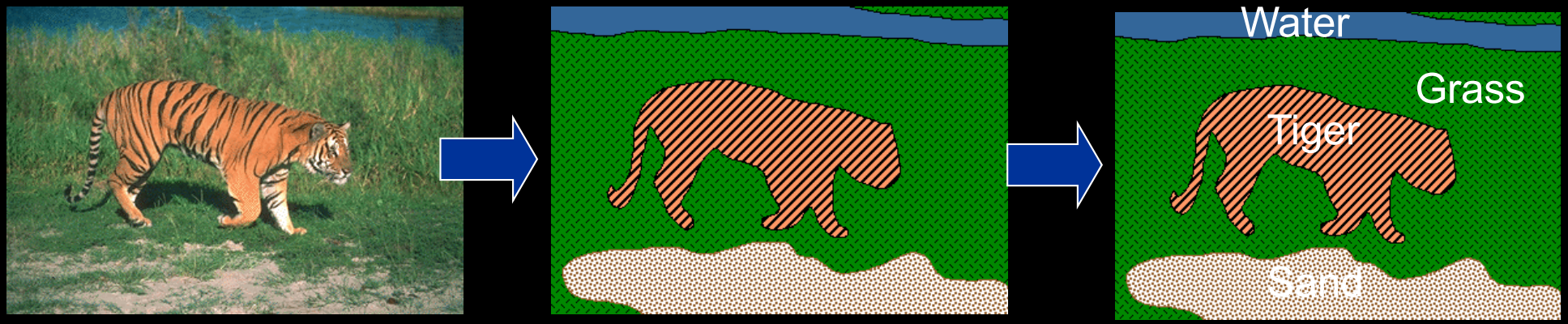
Slide from J. Malik, UC Berkeley

Feed-forward processing

Pixels
Local Neighborhoods

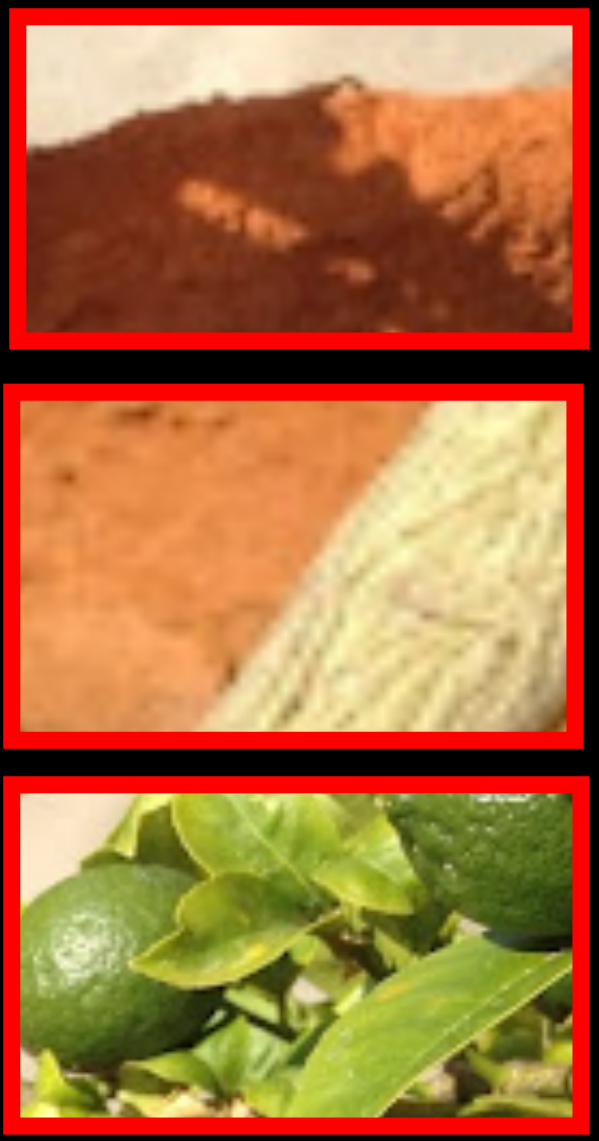
Contours
Surfaces

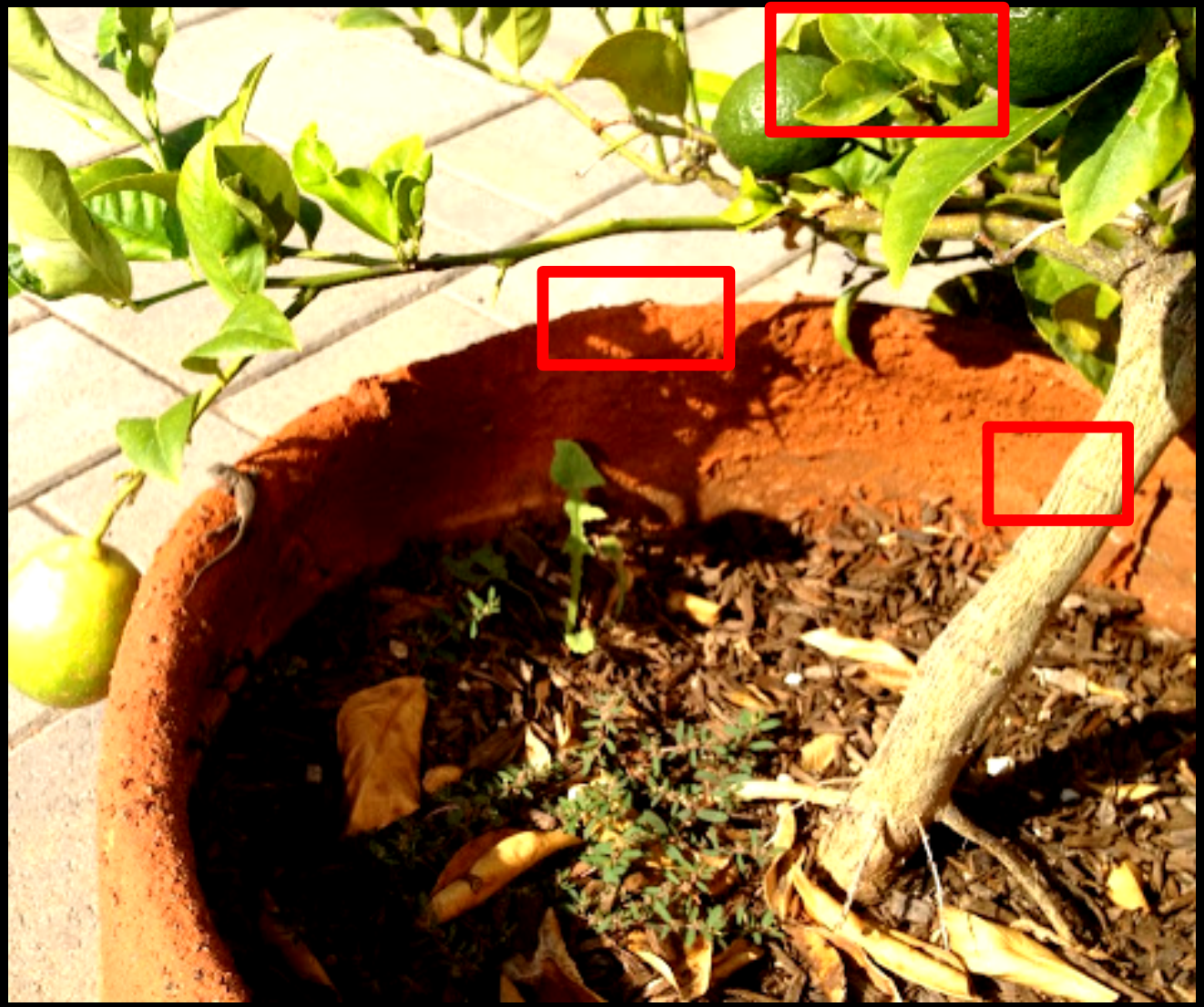
Objects
Scenes



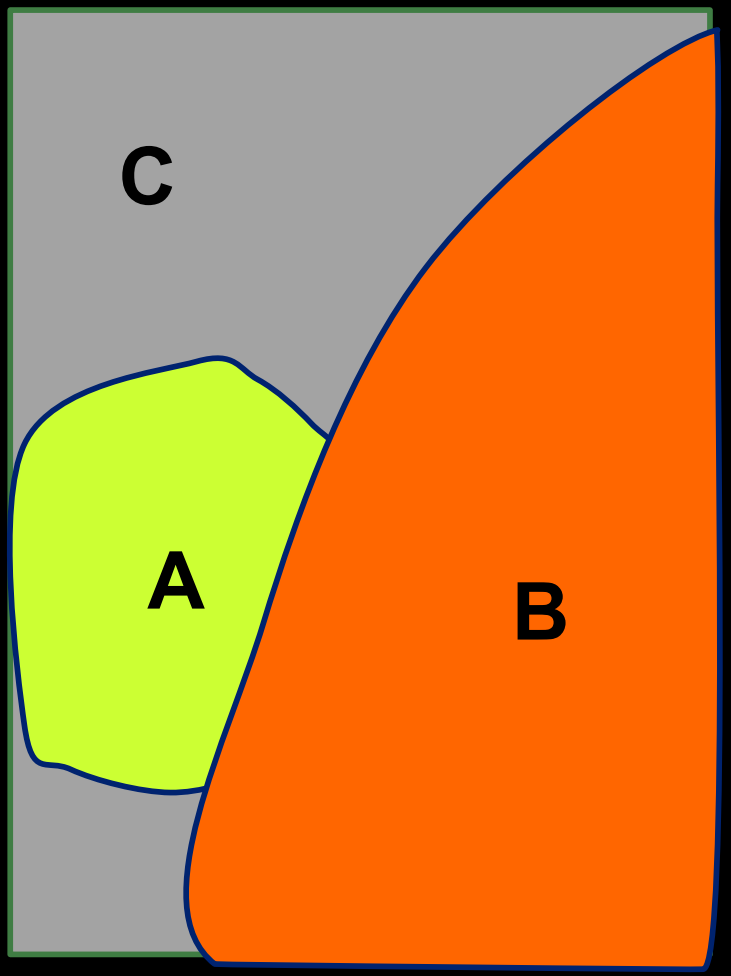
Low-level → Mid-level → High-level

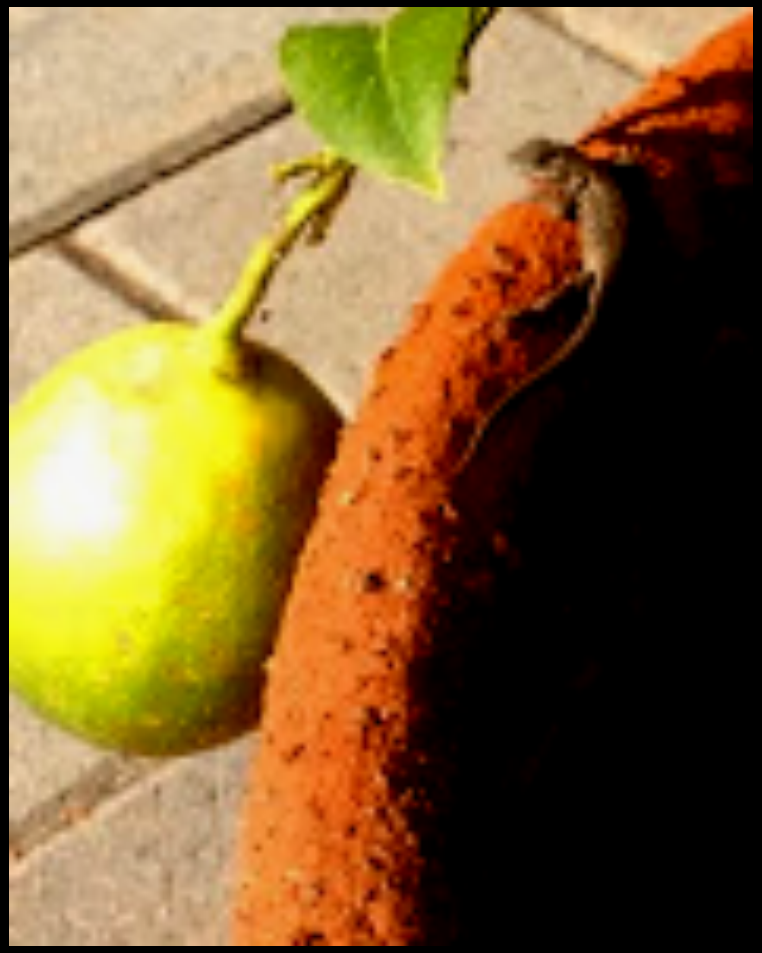
Boundaries are often defined by discontinuities in image brightness, color and/or texture





Humans assign a depth ordering to surfaces across a contour



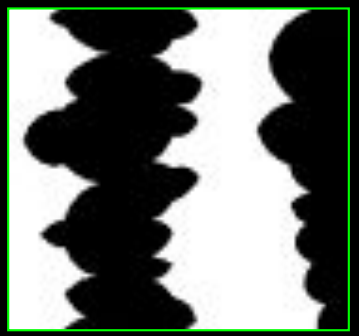


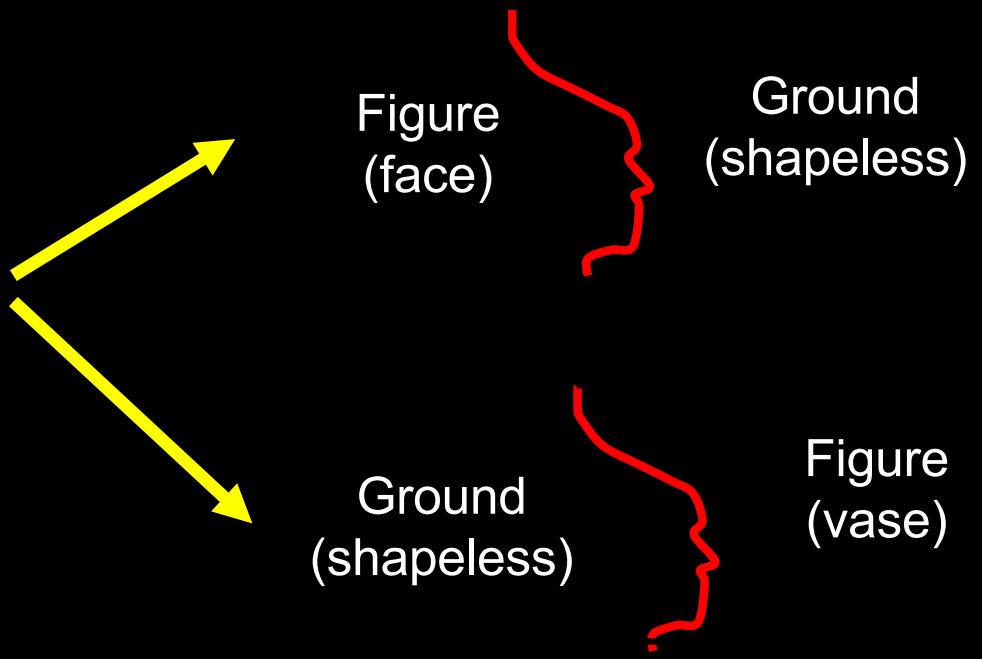
B is in front of A. A is in front of C. B is in front of C

Face-Vase Illusion

Figure/Ground Organization

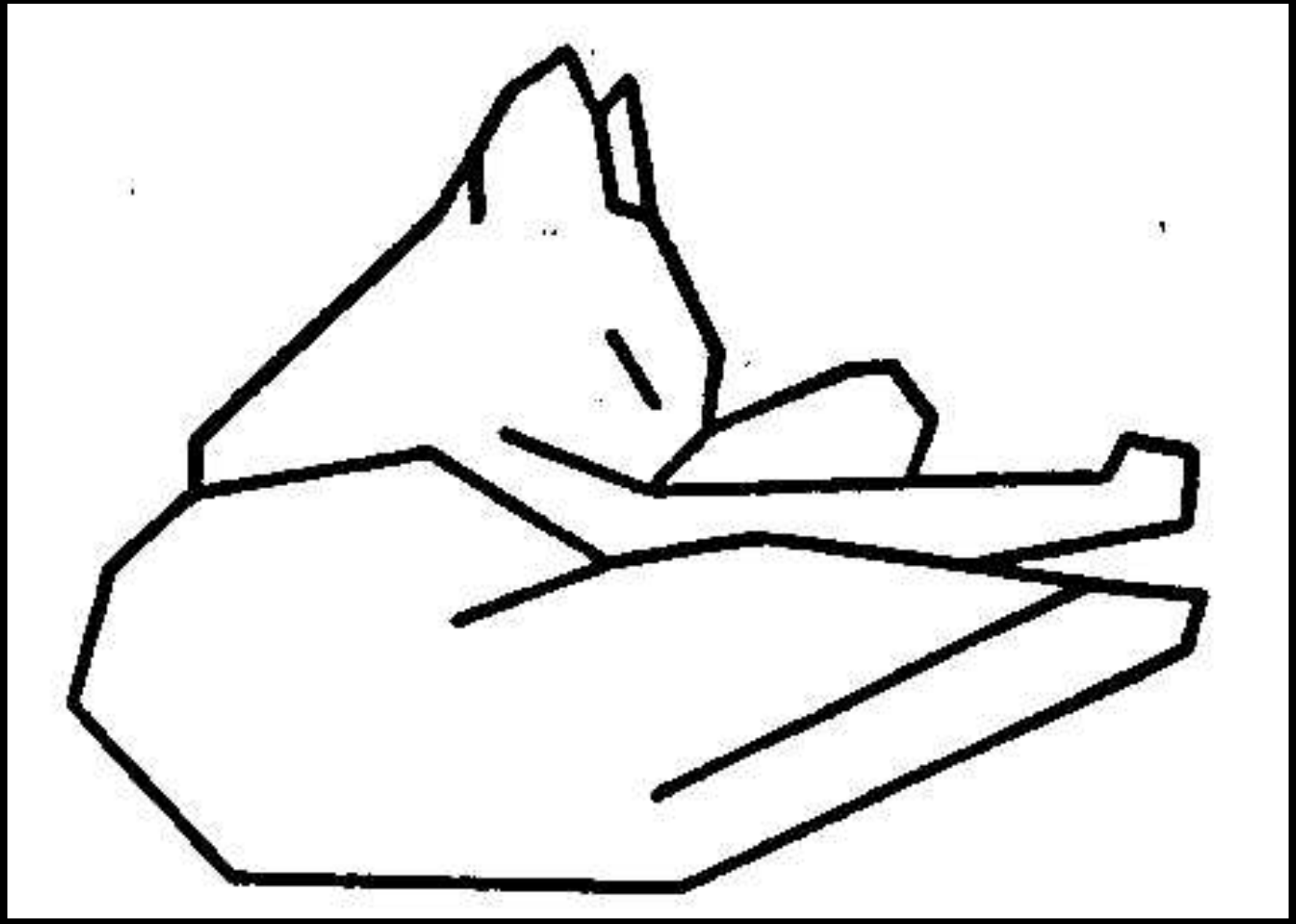
- A contour belongs to one of the two (but not both) abutting regions.





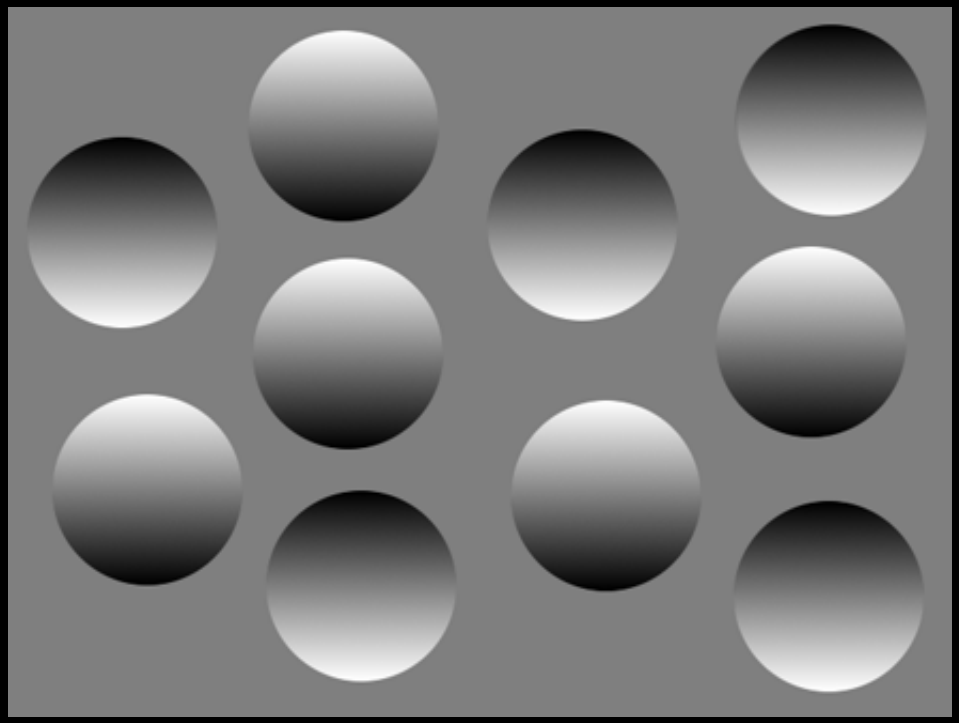
Important for the perception of shape

Line drawings convey lots of information about shape, identity, etc.



Attneave's Cat (1954)

Shading

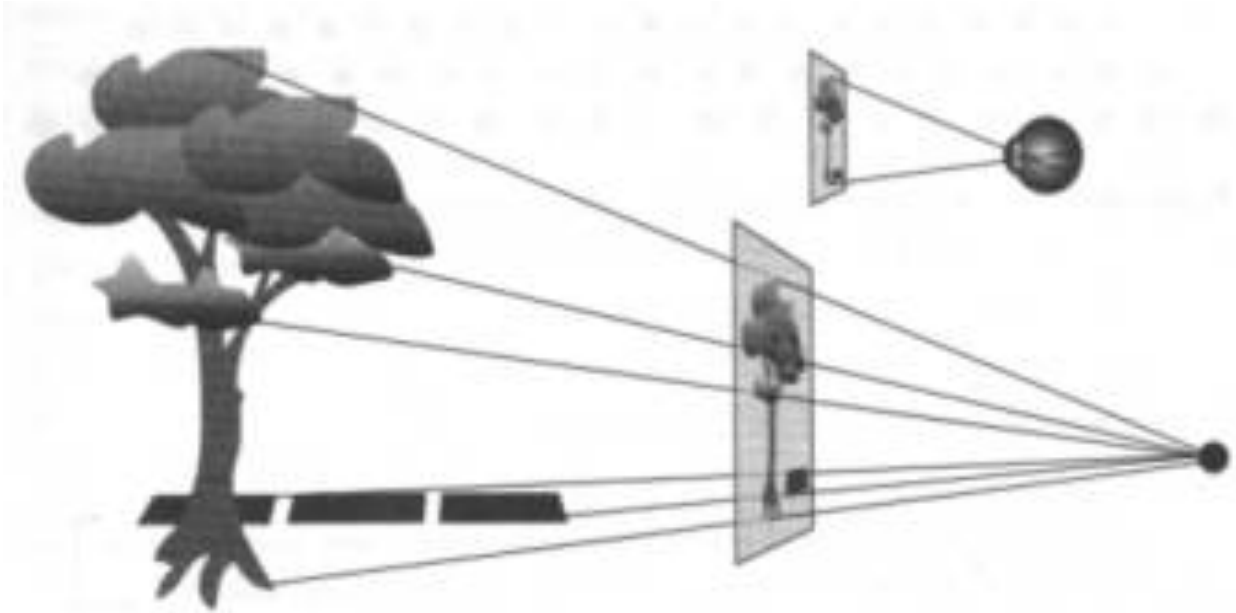


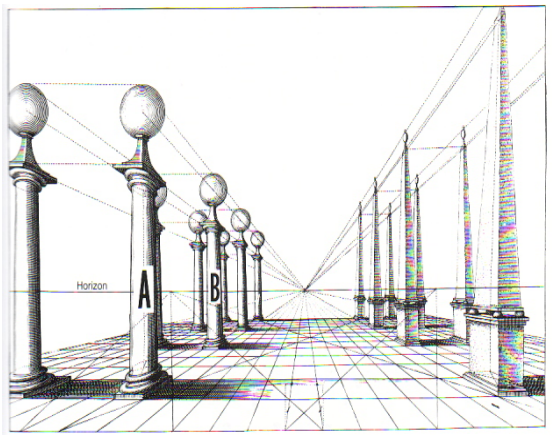
What do we see here?

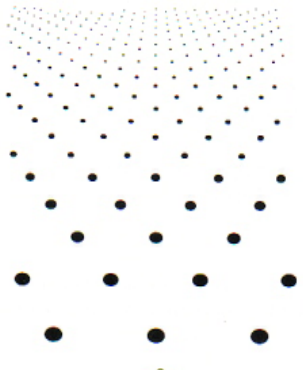
And here?

Perspective Cues

- Parallel lines converge
- Distant objects appear smaller
- Textured Elements become smaller with distance





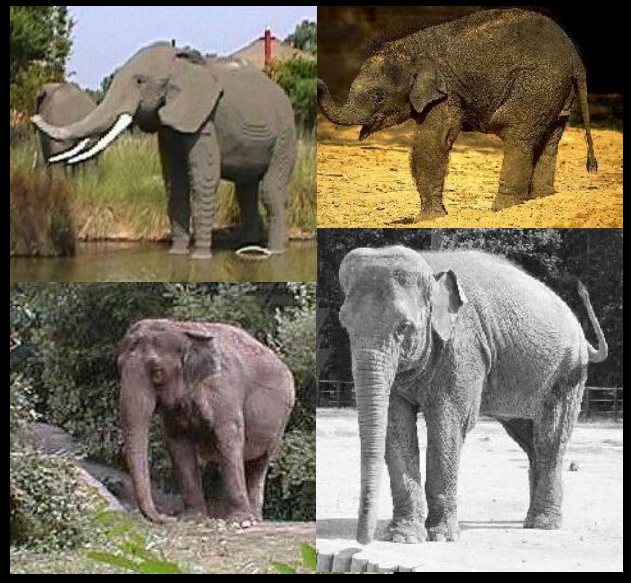


Size Constancy

Size Constancy

Why doesn’t it bother you than one person is 3x smaller than another?

Object Category Recognition



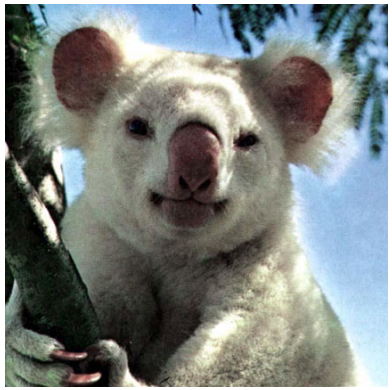
Challenges in object recognition

Illumination

Object pose

Clutter

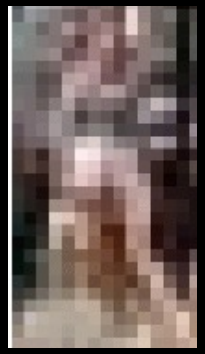
Occlusions



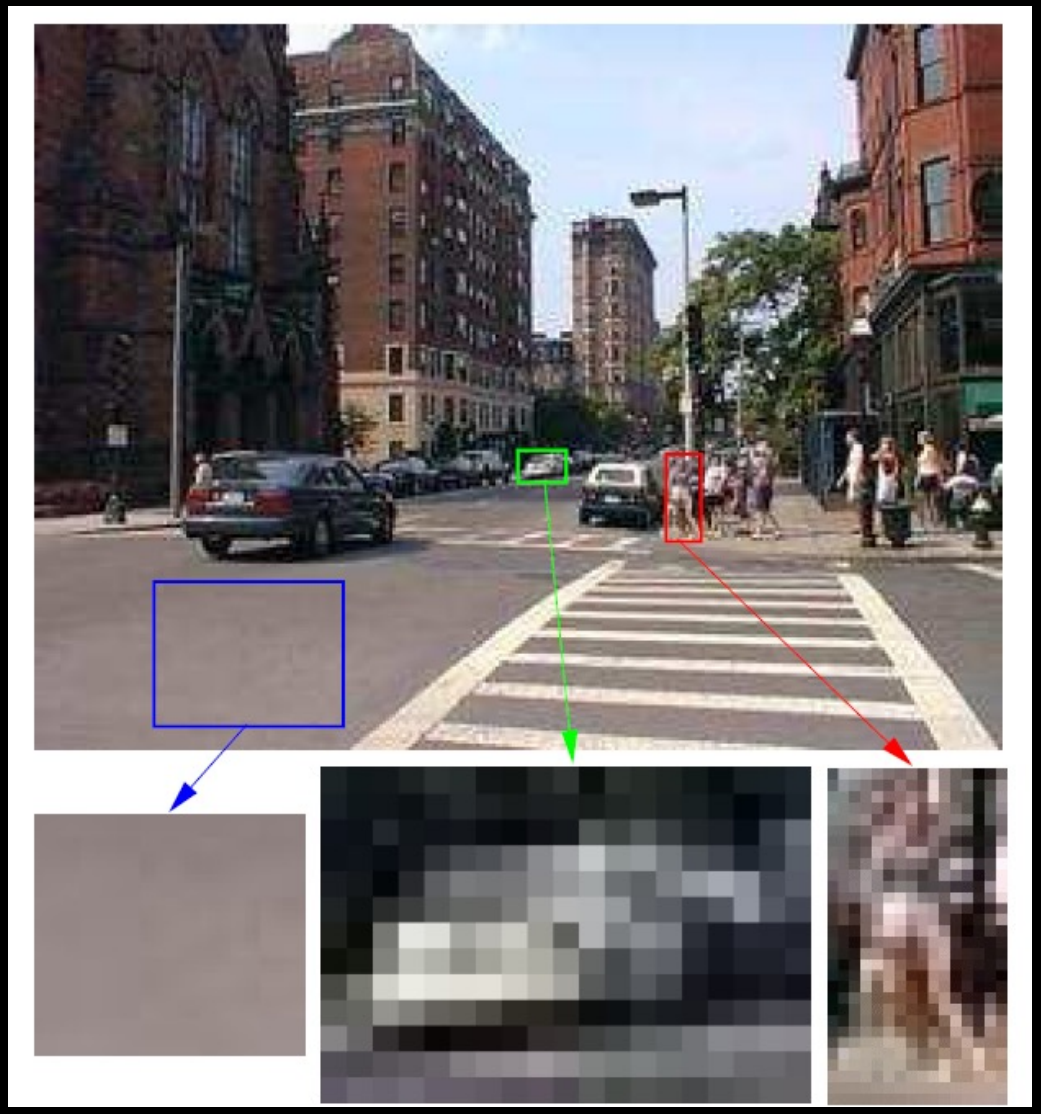
Intra-class appearance

Viewpoint

Slide from K. Grauman, B. Leibe



Objects are in Scenes
(context is powerful!)



From work of Torralba and Oliva, Hoeim, Efros and Hebert

Overview

Introduction

Understanding what is in an image

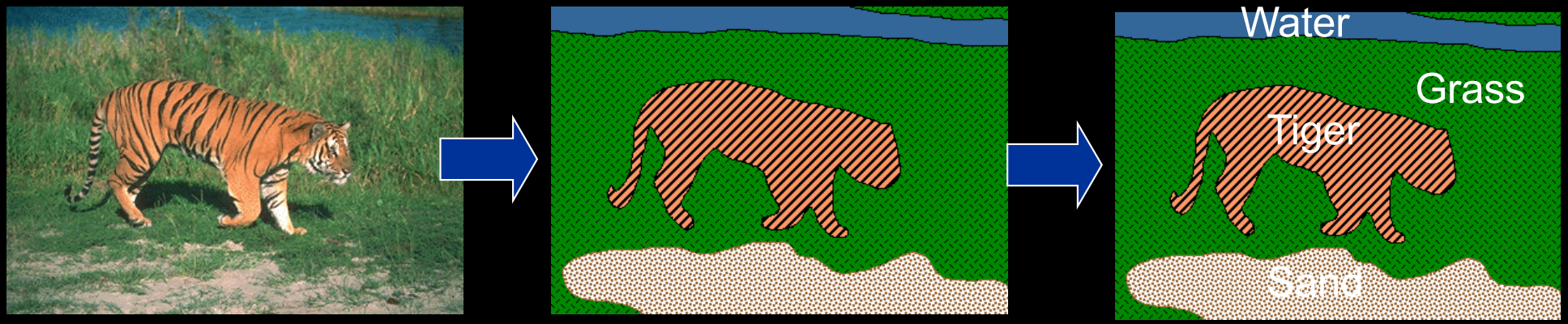
Learning to see by modeling data

Feed-forward processing

Pixels
Local Neighborhoods

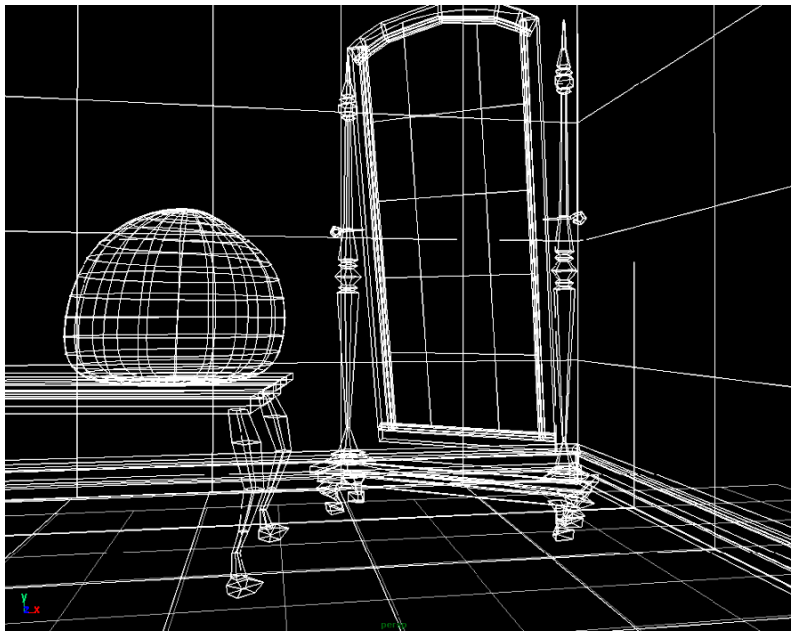
Contours
Surfaces

Objects
Scenes

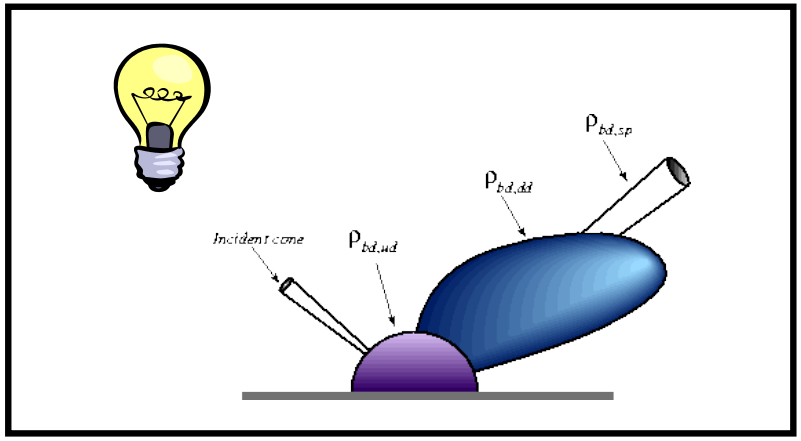


Low-level → Mid-level → High-level

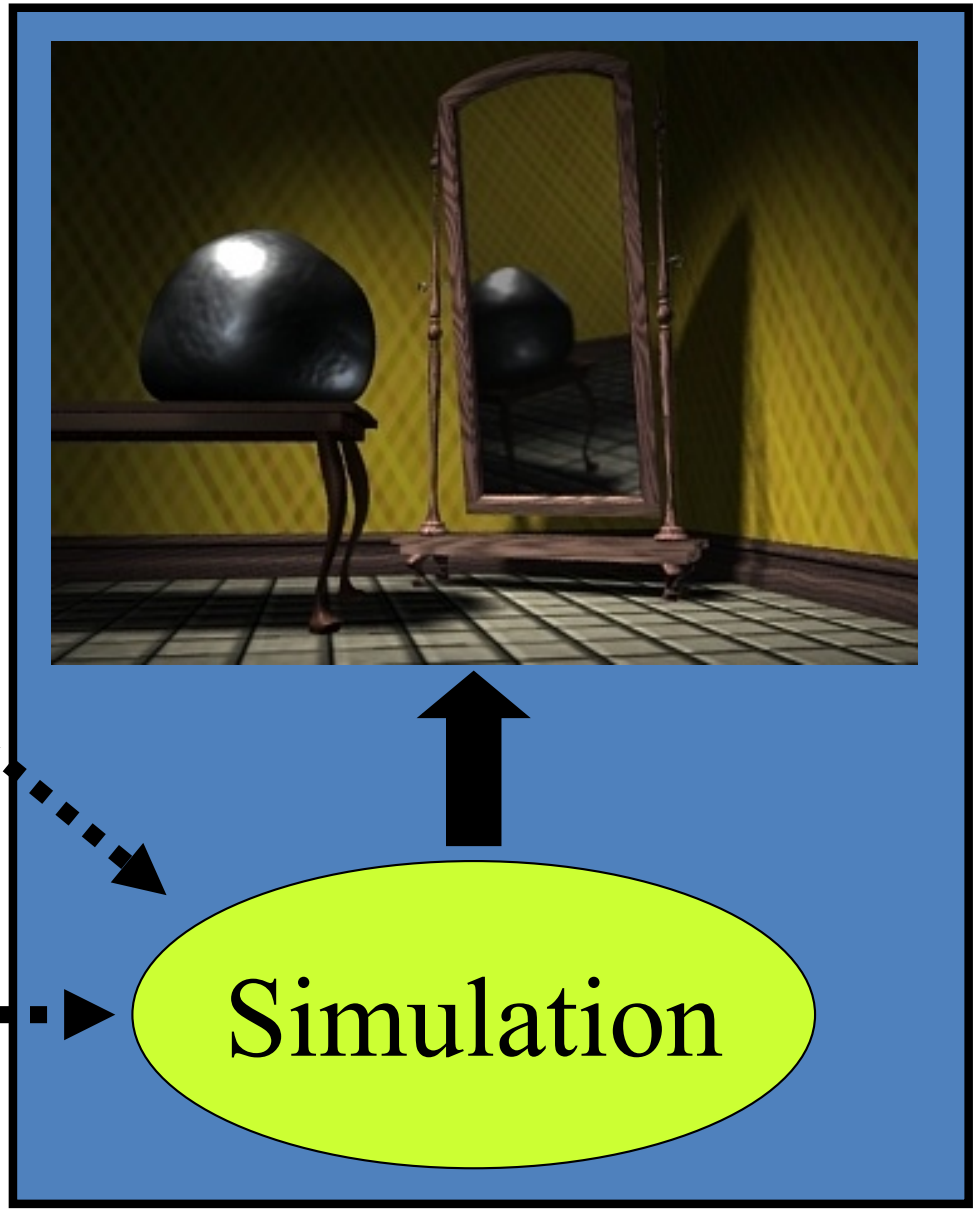
Feed-forward processing, a naïve proposal?



3D geometry + materials

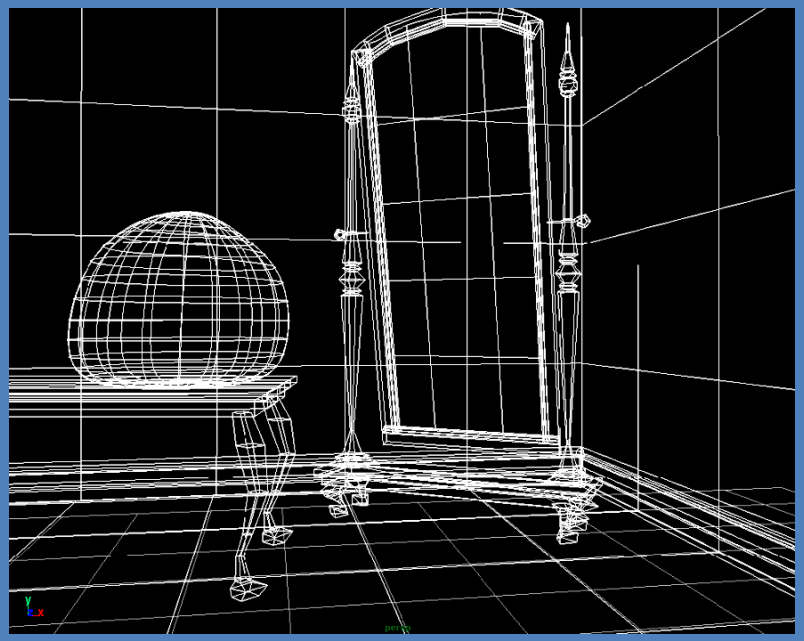


physics

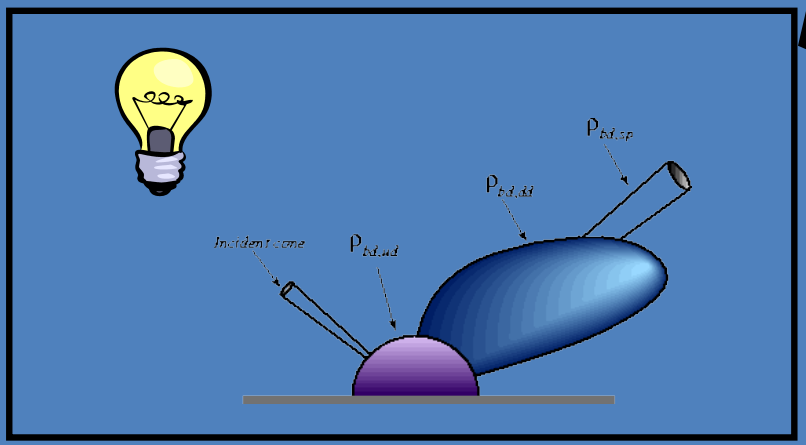


GRAPHICS

Computer vision as inverse computer graphics



3D geometry + materials



physics

Estimation

VISION!

The inverse problem is under-constrained!

Perception as Inference

Herman von Helmholtz
(1821-1894)

Thomas Bayes
(1701-1761)

"Unconscious Inference"

Measurements delivered by the eye are incomplete. They must be combined with some prior beliefs and assumptions to yield useful percepts

Bayes' Rule

Herman von Helmholtz
(1821-1894)

Thomas Bayes
(1701-1761)

W = state of the world, I = image observed

$$P(W \mid I) \propto P(I \mid W) P(W)$$

Bayes' Rule

Herman von Helmholtz
(1821-1894)

Thomas Bayes
(1701-1761)

W = state of the world, I = image observed

$$P(W \mid I) \propto P(I \mid W) P(W)$$

image formation

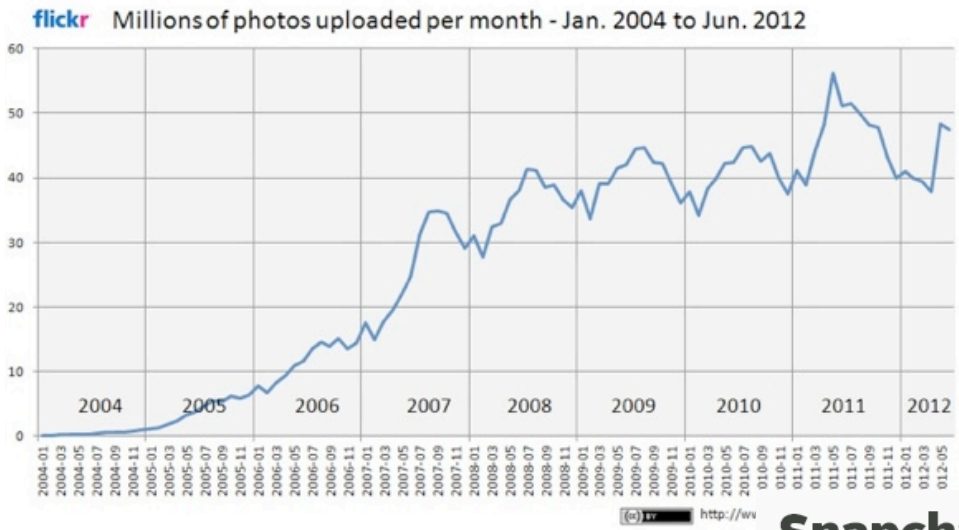
prior belief

Where do we learn these statistics about images and scenes?

Ecological statistics

Internalizing the statistics of natural scenes conveys survival advantage

Internet statistics



100M photos updated *daily to Flickr in 2011.* 3B public photos in 2011

People Upload an Average of 250 Million Photos Per Day to Facebook

That's 7.5 billion photos a month

BY STAN HORACZEK ON FEBRUARY 1, 2012 0

RELATED TAGS: NEWS, FACEBOOK, INTERNET, PHOTO SHARING

Like 85 Tweet 79 +1 15 Submit submit

Snapchat Sees More Daily Photos Than Facebook

Posted Nov 19, 2013 by Jordan Crook (@jordanrcrook)

66 Like 1k Tweet Share 175

Snapchat now shares 400 million snaps each day, according to CEO Evan Spiegel, who confirmed the number to TechCrunch today.

In September, Evan Spiegel announced on the Disrupt stage that the disappearing photo-sharing service was seeing 350 million snaps sent per day, up from 200 million in June. Clearly, growth on the hot new service is staggering.

Even more staggering, perhaps, is the fact that the 400 million/day figure surpasses the photo-sharing activity on both Instagram *and* Facebook.

Facebook reportedly sees 350 million photos uploaded each day from its worldwide audience of more than 1 billion users, and Instagram's 150 million users upload around 50 million photos each day.

CrunchBase

Snapchat

FOUNDED May 2011 | TOTAL FU $123M

OVERVIEW

Snapchat is the fastest way to share a r friends. You control how long your frie your message – simply set the timer up and send. They'll have that long to view and then it disappears forever. We'll let they take a screenshot! Build relationsh points, and view your best friends. Sna instantly fun and insanely ...

WEBSITE http://www.snapchat.com

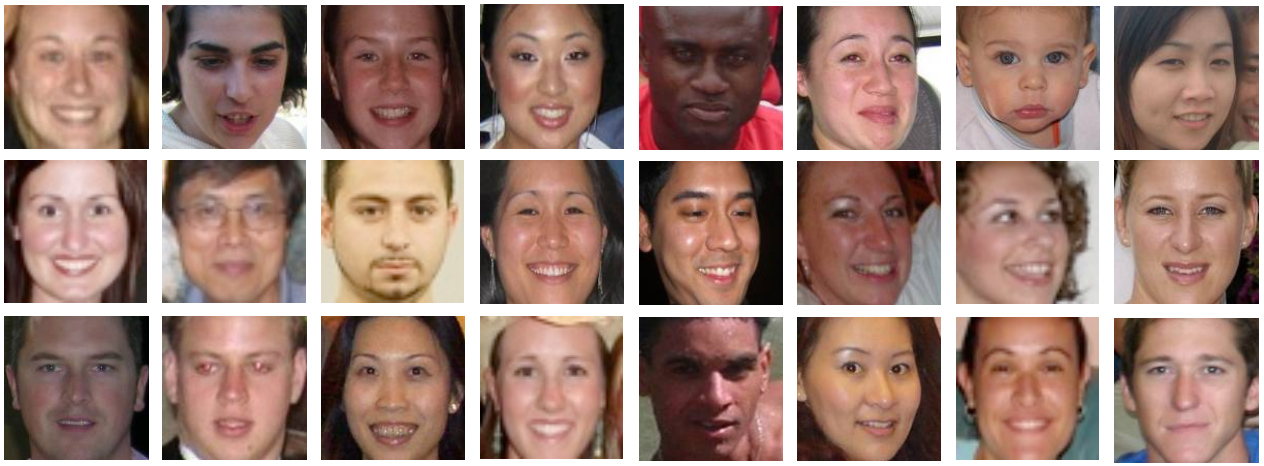
Full profile for Snapchat

CrunchDaily

Category-level Examples

street scenes

Food plates



faces

pedestrians

Instance-specific Examples

Portraits of Bill Clinton

Image Categorization

Training

Training Images

Training Labels

Image Features

Classifier Training

Trained Classifier

Testing

Test Image

Image Features

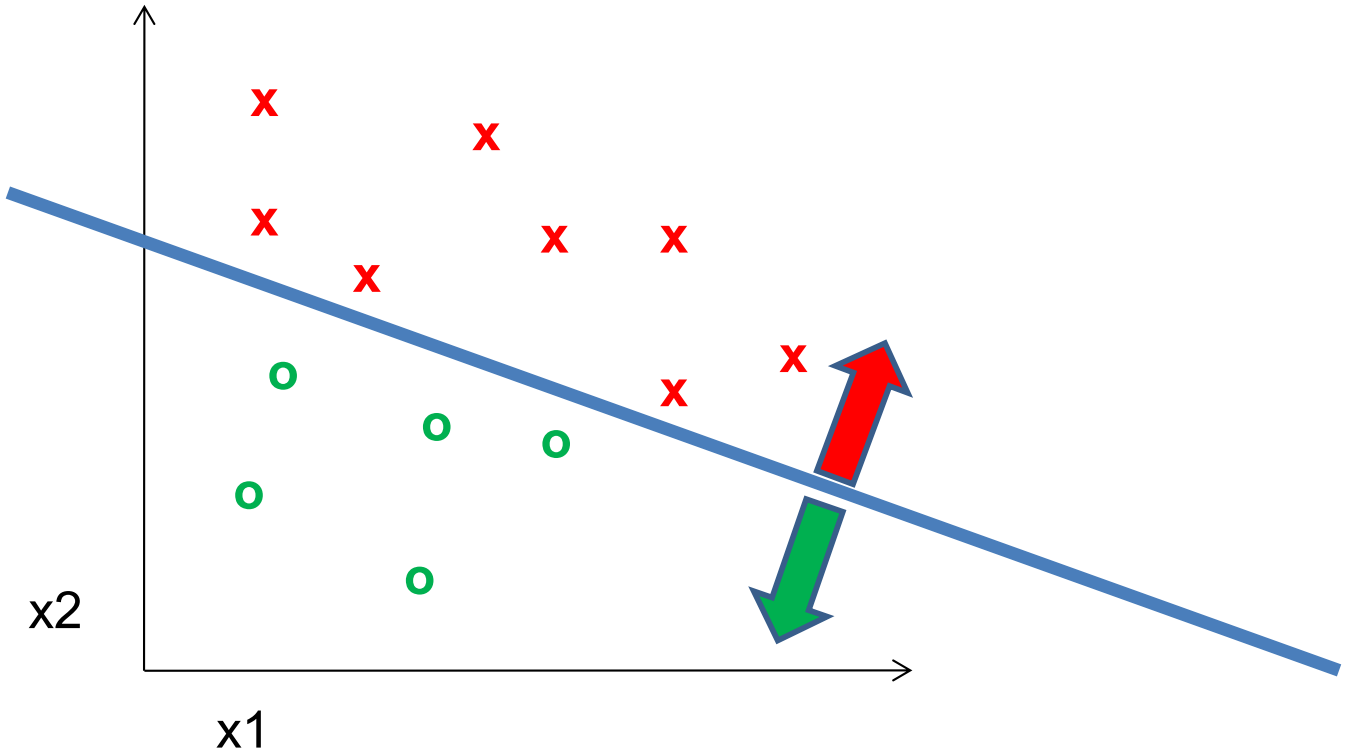
Trained Classifier

Prediction

Outdoor

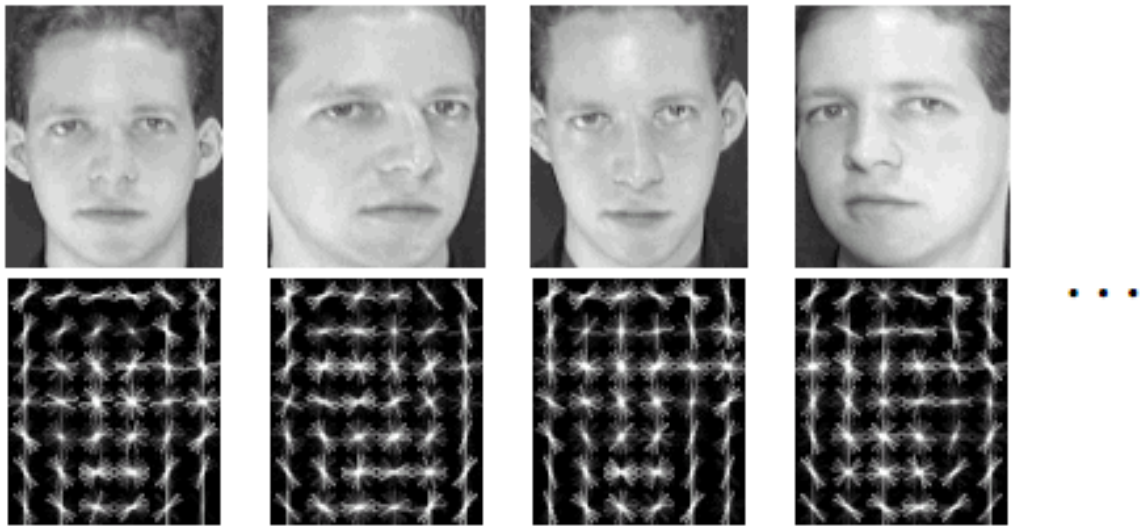
Learning a classifier

Given some set of features with corresponding labels, learn a function to predict the labels from the features



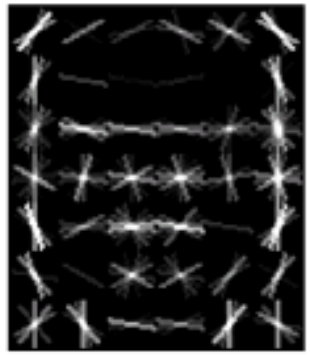
Learning the appearance of faces

- Positive examples:



- Negative examples: random patches from images without faces

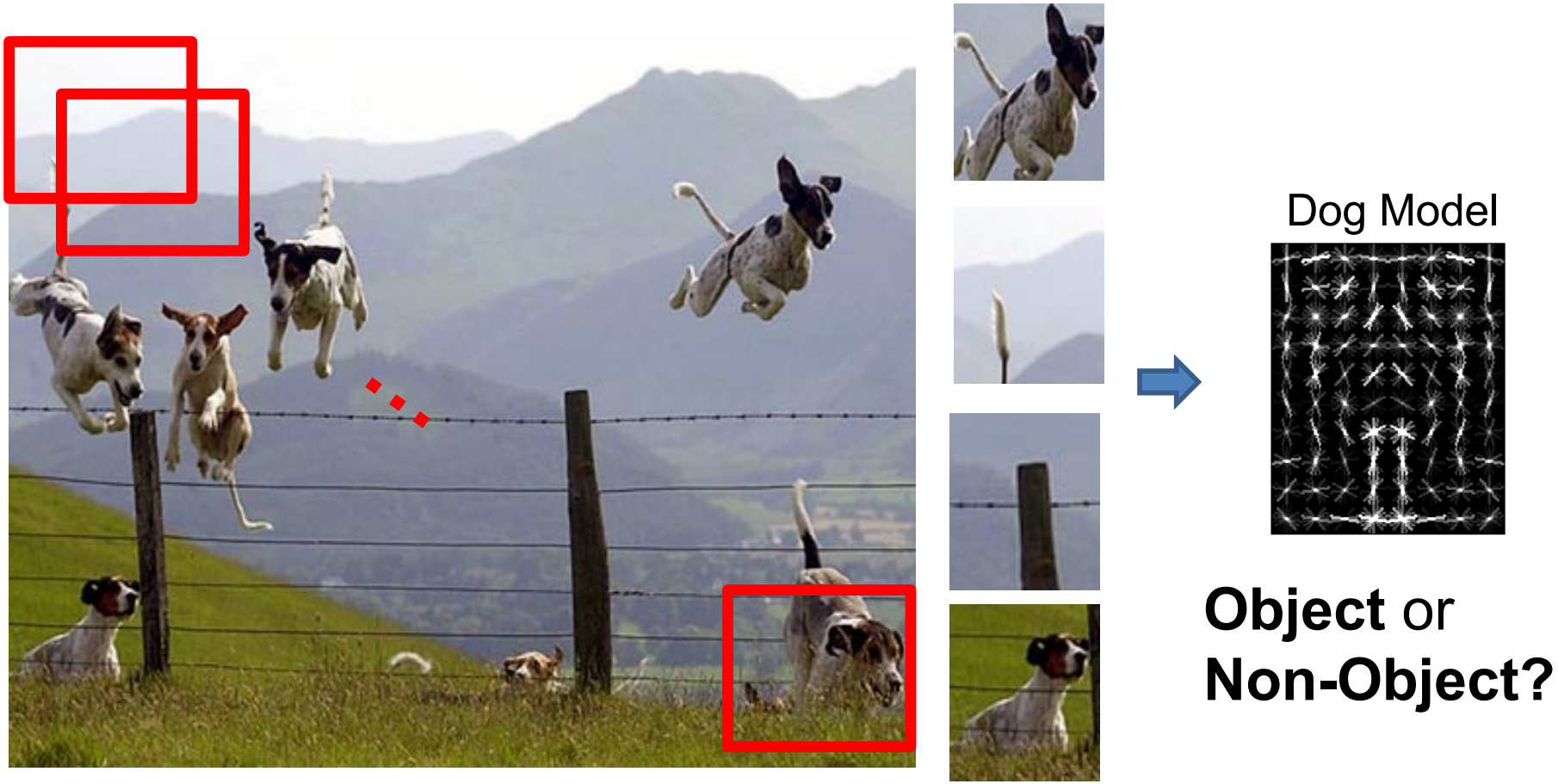
Learn a good discriminative
template from training examples



template

Object Category Detection

- Focus on object search: “Where is it?”
- Build templates that quickly differentiate object patch from background patch



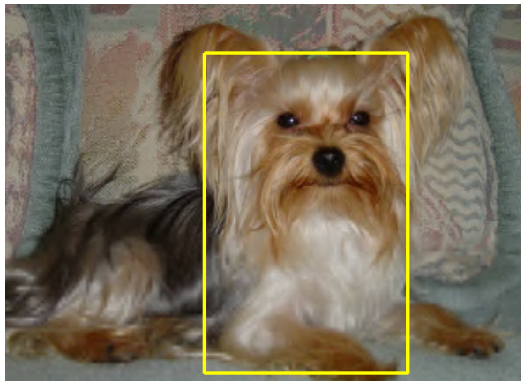
Detection: perform classification on each possible window of an image

What the Detector Sees

Challenges in modeling the non-object class

True Detections

Bad Localization



Confused with Similar Object

Misc. Background

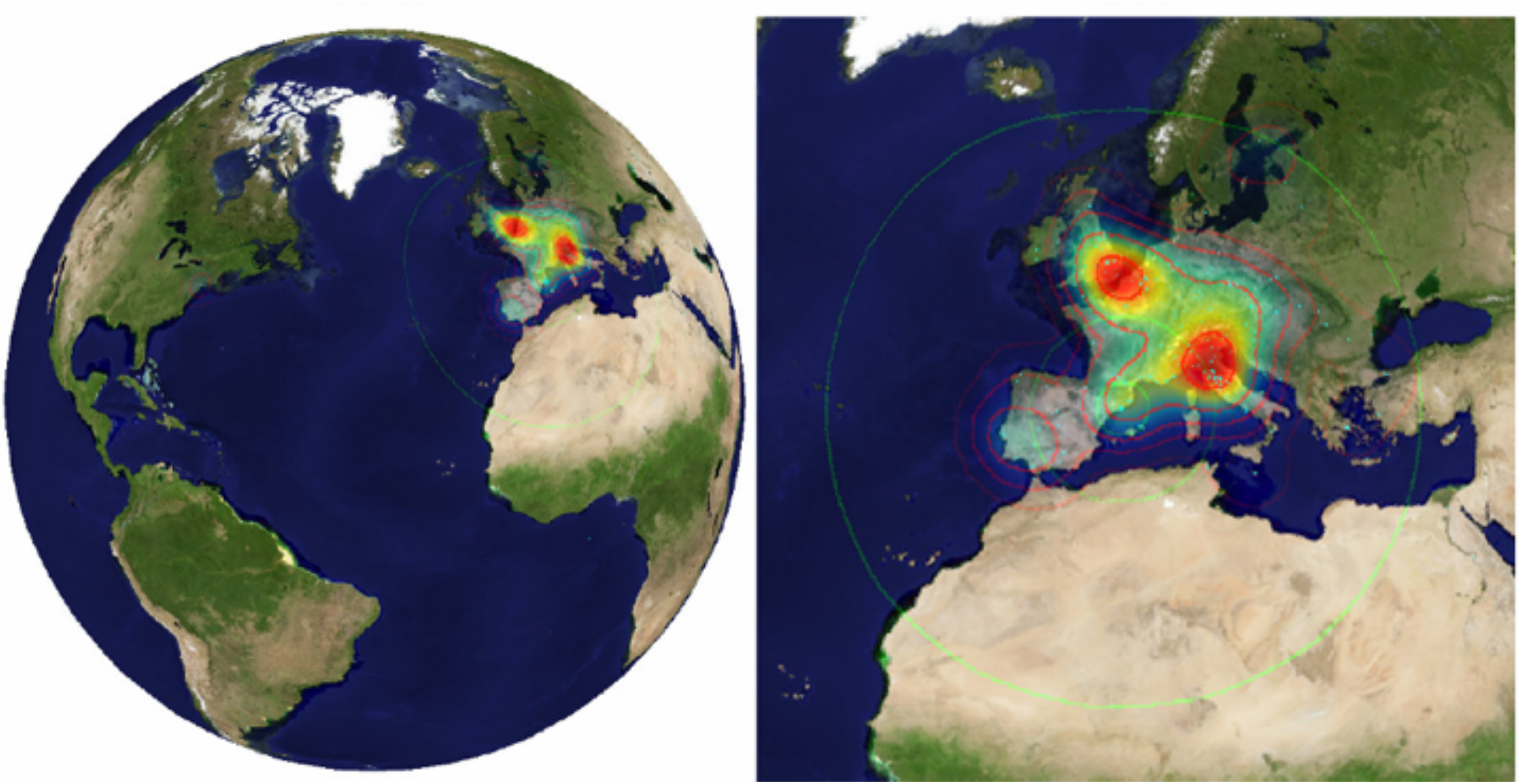
Confused with Dissimilar Objects

Where is this scene?

im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

Example Scene Matches




USA


Utah


Arizona


Utah


Utah


Utah


Tunisia


Kenya


Utah


LosAngeles


Burundi


NewMexico


Utah


Utah


Utah


Mendoza


USA
Utah
Arizona
Utah
Utah
Utah
Tunisia
Kenya
Utah
LosAngeles
Burundi
NewMexico
Utah
Utah
Utah
Mendoza

California

Oklahoma

SouthAfrica

Zambia

Kenya

Hyderabad

Mongolia

SouthAfrica

Kenya

Kenya

Zambia

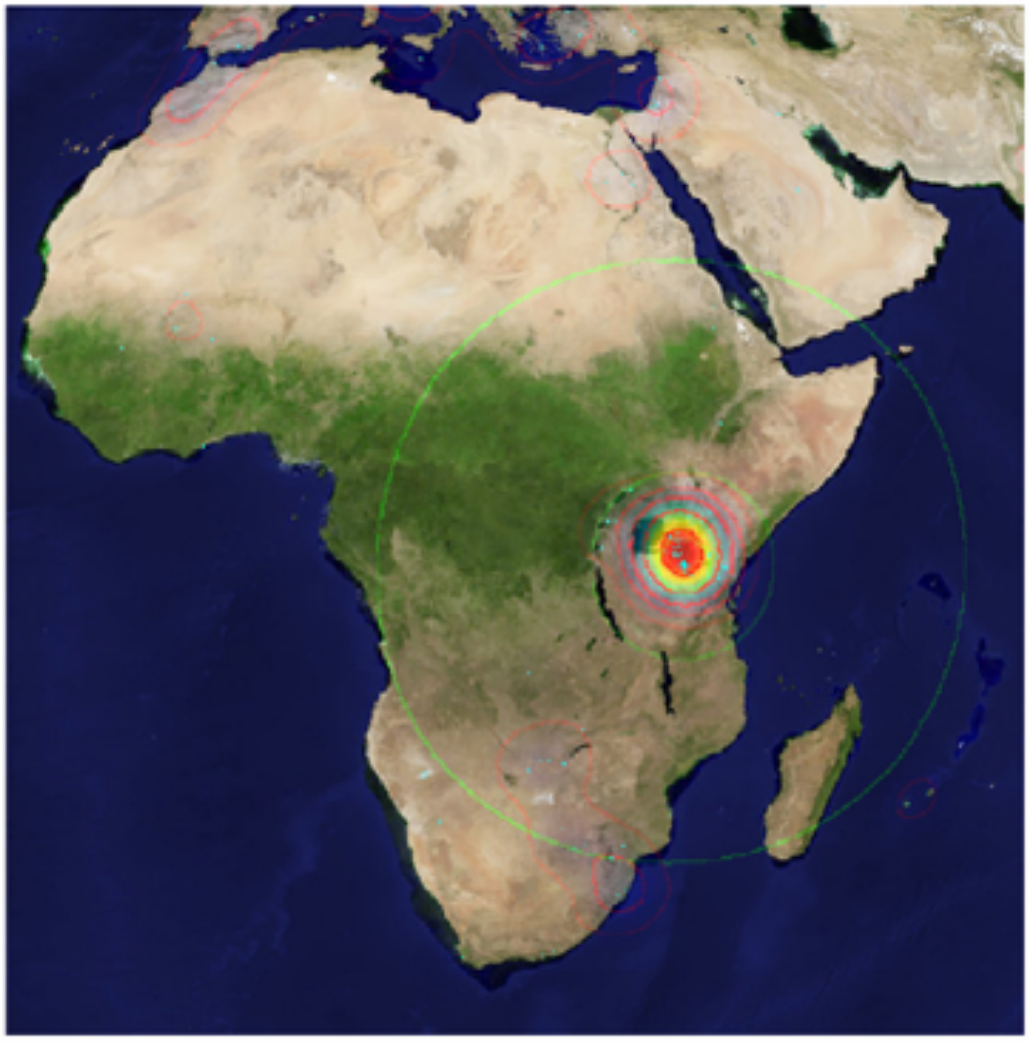
Ethiopia

Nevada

africa

Morocco

Tennessee



Toronto

Florida

NewYork

Boston

Boston

Oregon

Oregon

Oregon

NewYork

Barcelona

Oregon

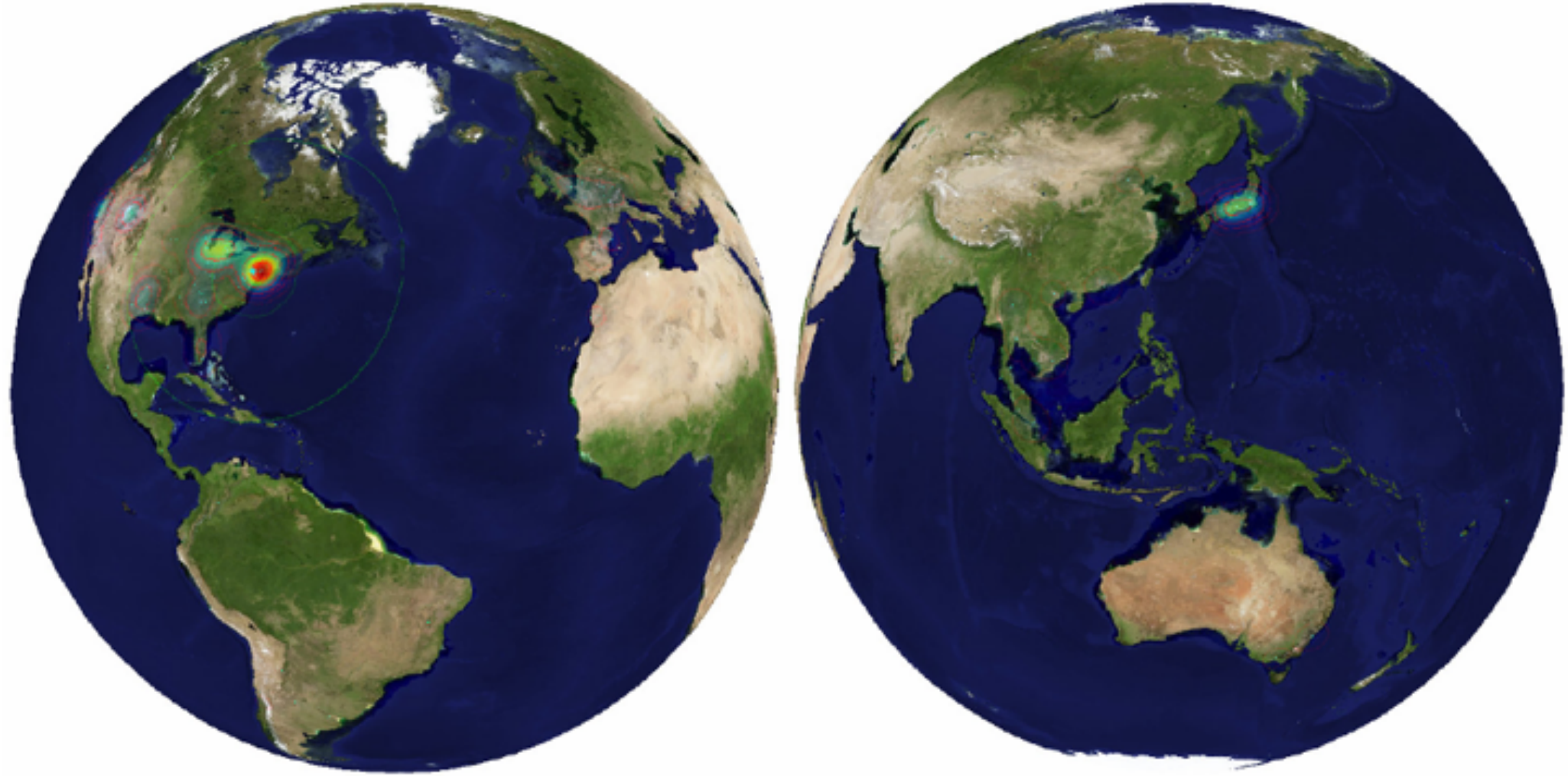
Chicago

Ohio

Philadelphia

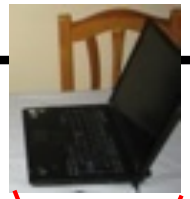
NewYorkCity

Boston



Have all the photos been taken?

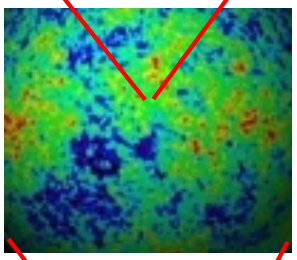
Number of images on my hard drive: 10^5



Number of images seen during my first 10 years: 10^8
(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: 10^{20}
106,456,367,669 humans[1] * 60 years * 3 images/second * 60 * 60 * 16 * 365 =
1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: 10^{88}



Number of all 8-bits 32x32 images: 10^{7373}
$256^{32*32*3} \sim 10^{7373}$

Images are unique

But not all images are that original…

But not all images are that original…