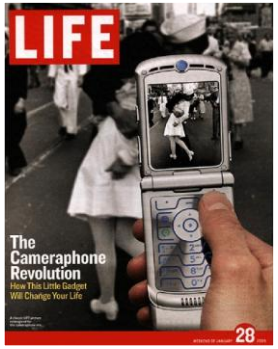
Introduction

CS 554 – Computer Vision

Pinar Duygulu

Bilkent University

Why study computer vision?



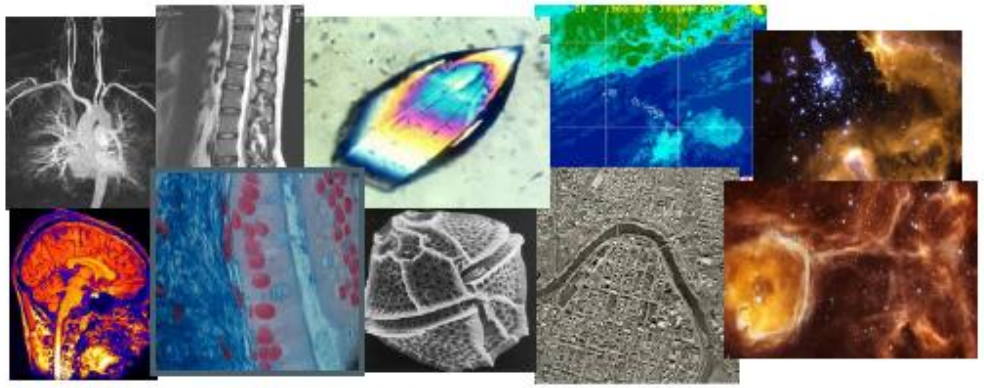
Source: Svetlana Lazebnik, UIUC

Why study computer vision?

An image is worth 1000 words

Images and movies are everywhere

Surveillance and security



Medical and scientific images

Source: Fei Fei Li, Stanford University

The goal of computer vision

- To extract “meaning” from pixels

What we see

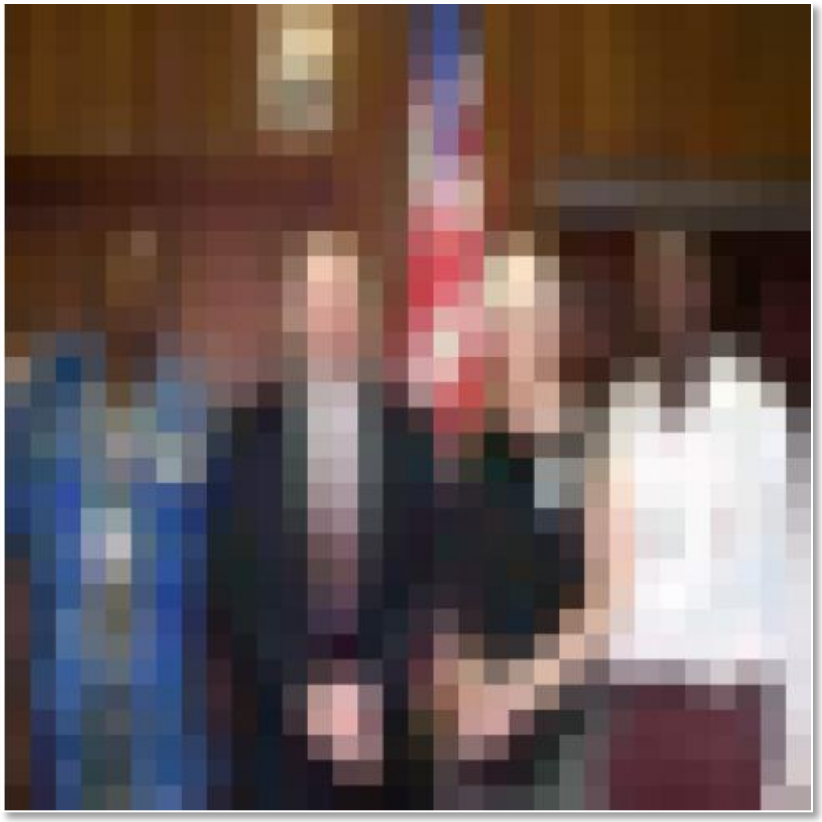
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

The goal of computer vision

- To extract “meaning” from pixels



Humans are remarkably good at this…

Source: “80 million tiny images” by Torralba et al.

What do you see in the picture?

Source: Martial Hebert, CMU

What do you see in the picture?

Black background
Two objects
  One teapot
  One toy
There is a light coming from right
One object is shiny the other is not

Toy:
Consists of 5 layers, in different colors
There is a text : Fisher Price
The layers are in donut shape
Layers are plastic
Bottom is wood

Teapot:
Consists of body and handle
Body is metal
Handle is ceramic
Handle: Dark blue on white
Body : golden
Reflection of toy on the body

Source: Martial Hebert, CMU

Challenge – What do you see in the picture?

Source: Octavia Camps, Penn State

Challenge – What do you see in the picture?

A hand holding a man

A hand holding a shiny sphere

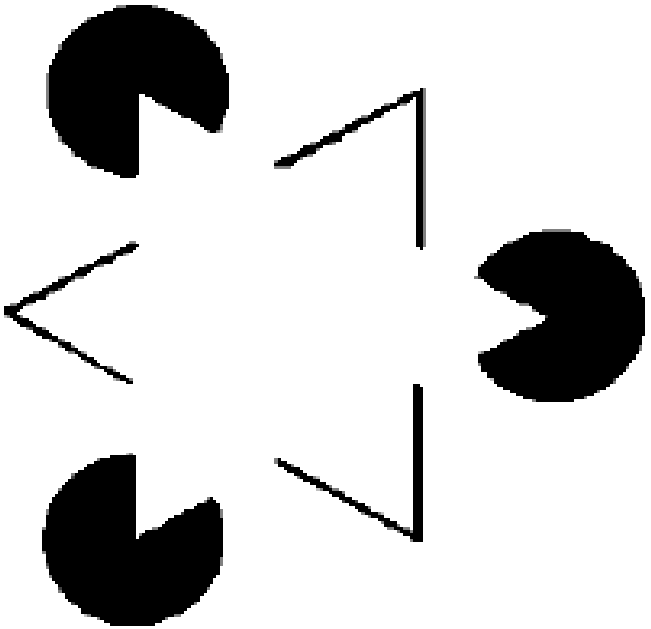
An Escher drawing

Source: Octavia Camps, Penn State

Perception and grouping

Subjective contours

Subjective contours



Kaniza triangle

Source: Michael Black, Brown University

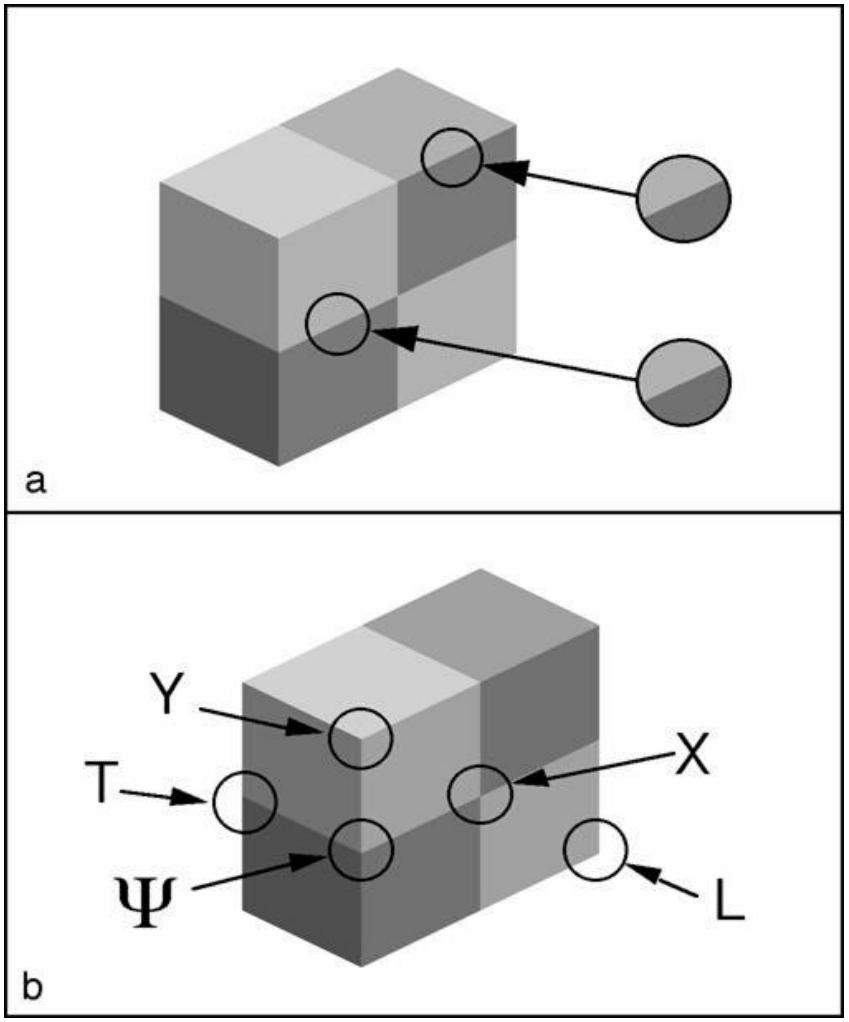
Occlusion

* types of "junctions" give cues about surfaces, occlusion, and light.

I. Rock, The Logic of Perception, 1983.

Source: Michael Black, Brown University

Junctions



•The shape of junctions constrains the possible interpretations of the scene.

• Ambiguous : Paint and surface boundaries can be confused.

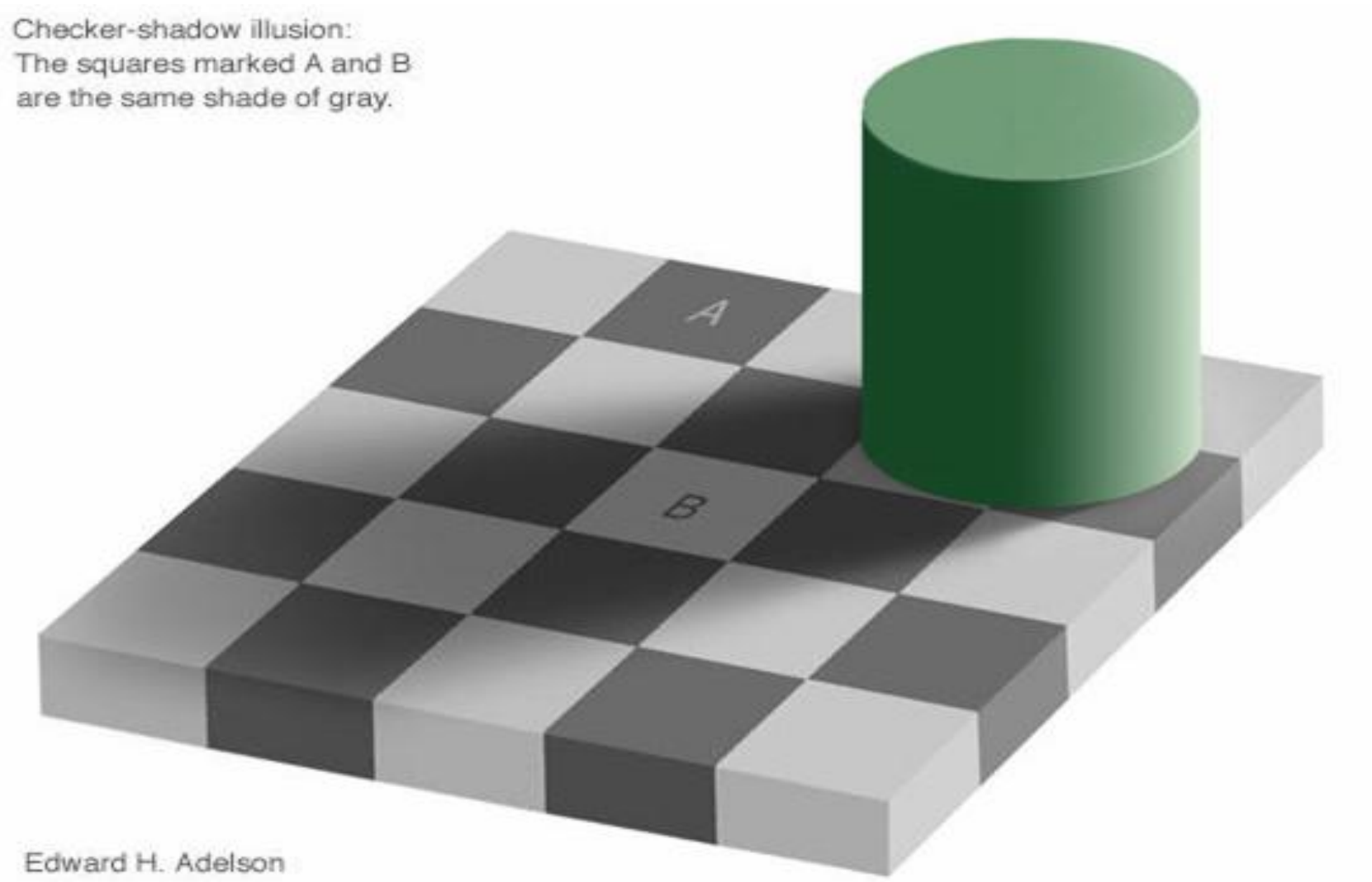
Adelson

Source: Michael Black, Brown University

Texture

Source: Forsyth&Ponce

Shading



Source: Michael Black, Brown University

Shading

Source: Michael Black, Brown University

Stereo

Source: http://www.well.com/user/jimg/index.html

Stereo

Source: http://www.well.com/user/jimg/index.html

Parts and relations

How flexible are the spatial relations of the parts?

Source: Michael Black, Brown University

How good are our models?

Thompson, P. (1980). "Margaret Thatcher: a new illusion." *Perception* **9**:483-484

Source: Michael Black, Brown University

How good are our models?

Thompson, P. (1980). "Margaret Thatcher: a new illusion." *Perception* **9**:483-484

Source: Michael Black, Brown University

Is it only about matching?

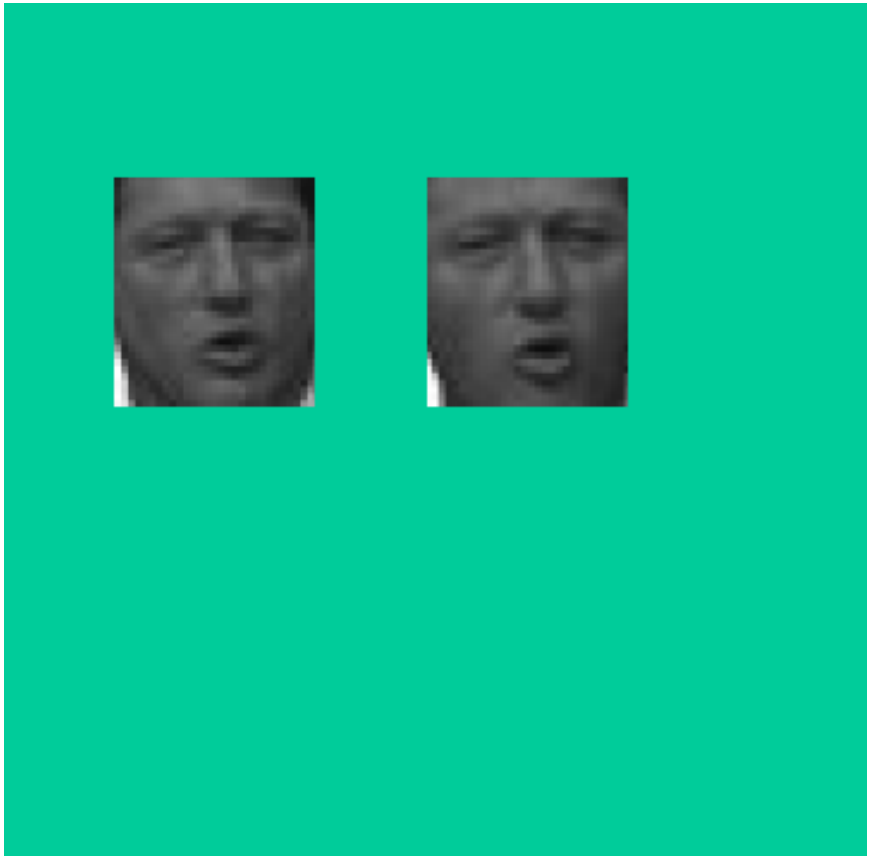
What are our "models"?

How good are they?

Ron Rensick

Source: Michael Black, Brown University

Is it only about matching?



Source: Michael Black, Brown University

Context

Antonio Torralba

Context

a person?

Antonio Torralba

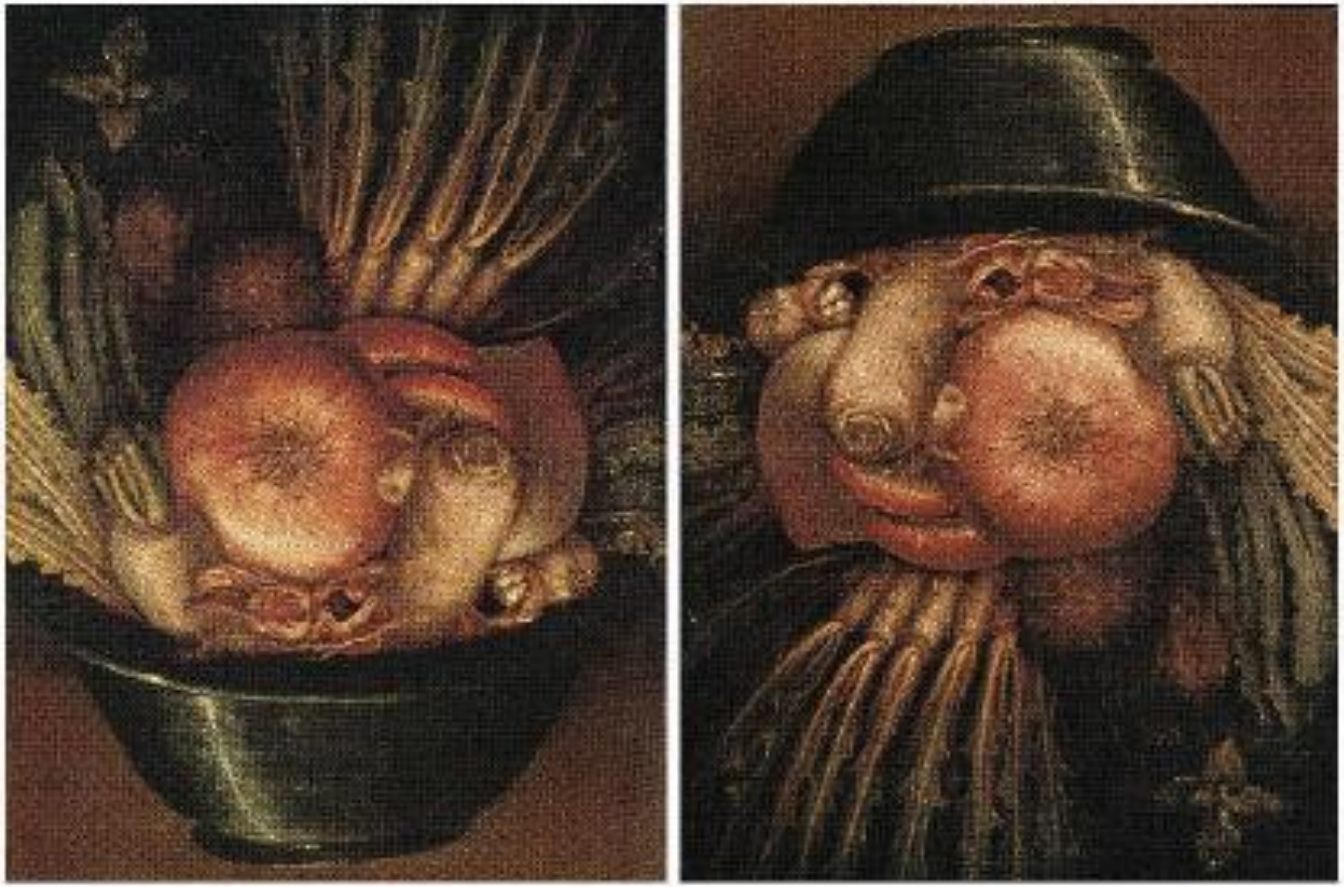
Context

a person?

Antonio Torralba

the blob is identical to the one on the previous slide after a 90deg rotation

Prior Expectations

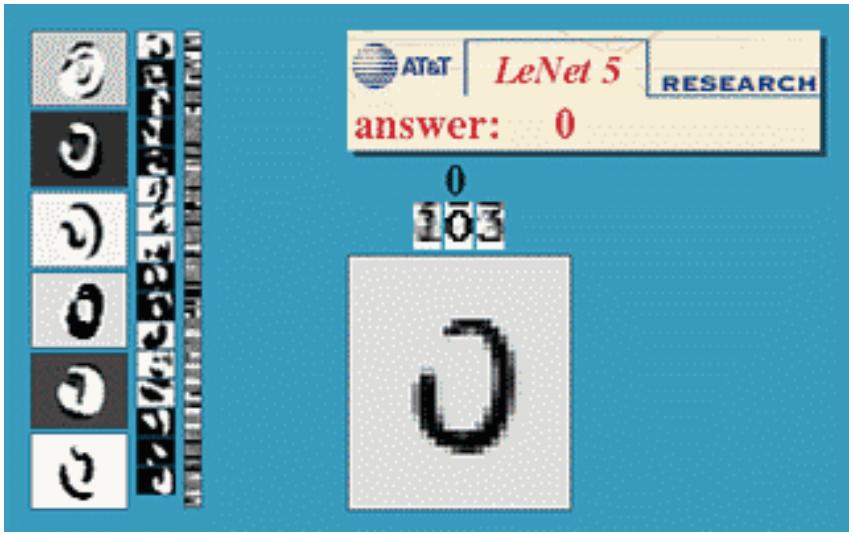


Giuseppe Arcimboldo

Source: Michael Black, Brown University

Applications

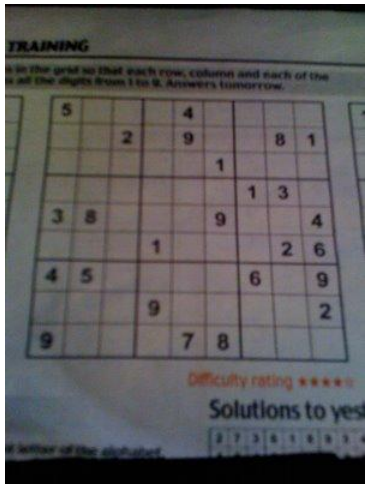
Optical character recognition (OCR)



Digit recognition
yann.lecun.com

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

Automatic check processing



Sudoku grabber
http://sudokugrab.blogspot.com/

Source: S. Seitz, N. Snavely

Biometrics

Fingerprint scanners on many new laptops, other devices

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

Source: S. Seitz

Face detection

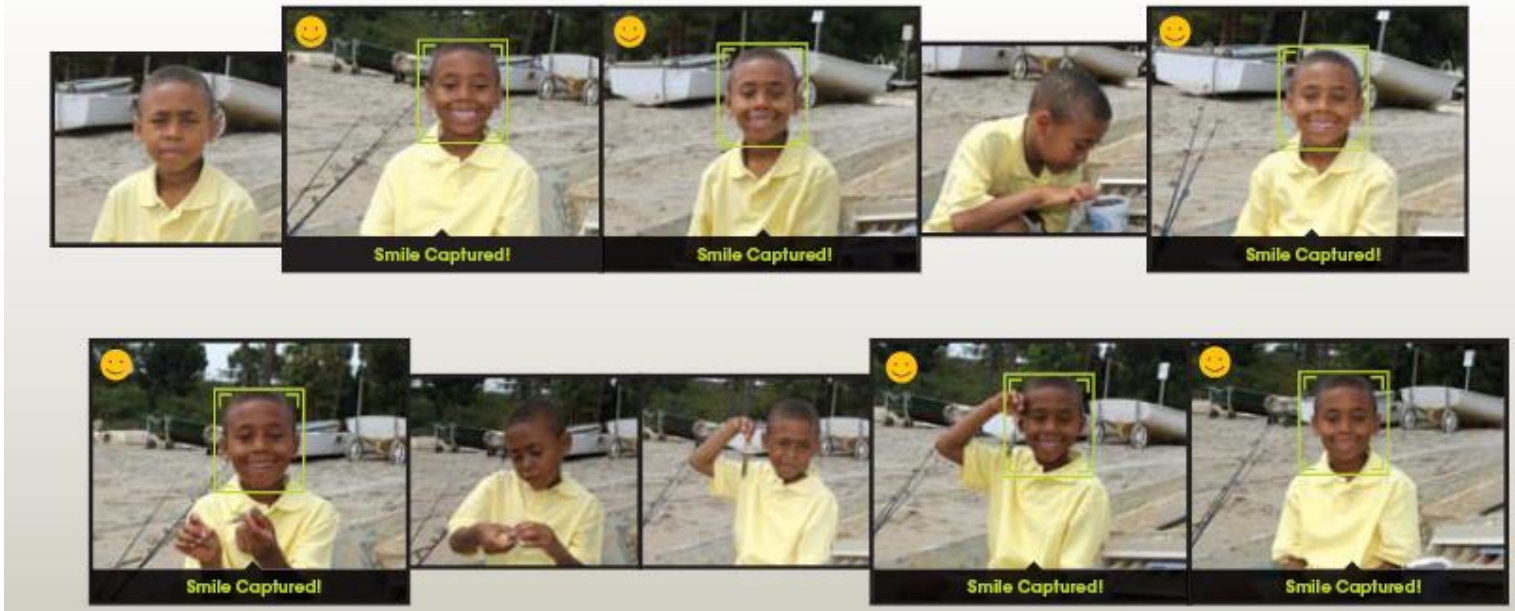
- Many consumer digital cameras now detect faces

Source: S. Seitz

Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



[Sony Cyber-shot® T70 Digital Still Camera]

Source: S. Seitz

Face recognition: Apple iPhoto software

http://www.apple.com/ilife/iphoto/

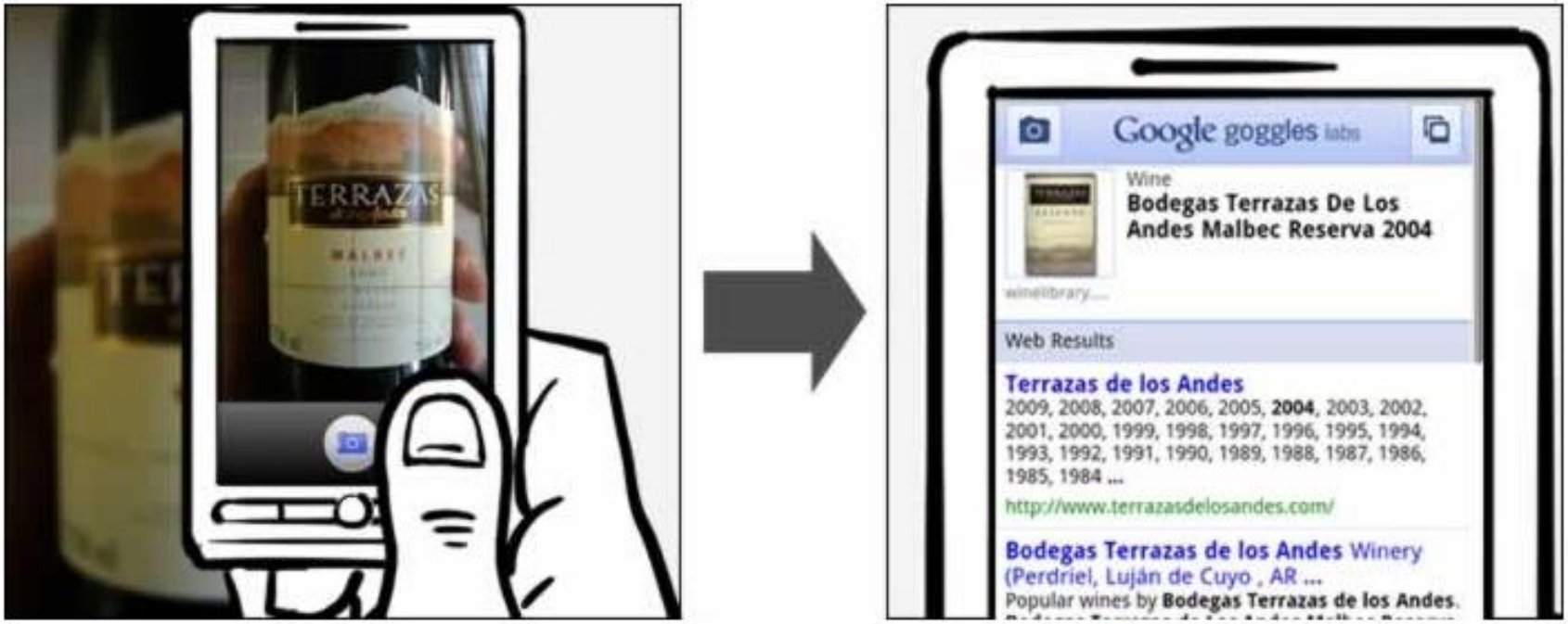
Source: S. Lazebnik, UIUC

Mobile visual search: Google Goggles

Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.

Google goggles labs



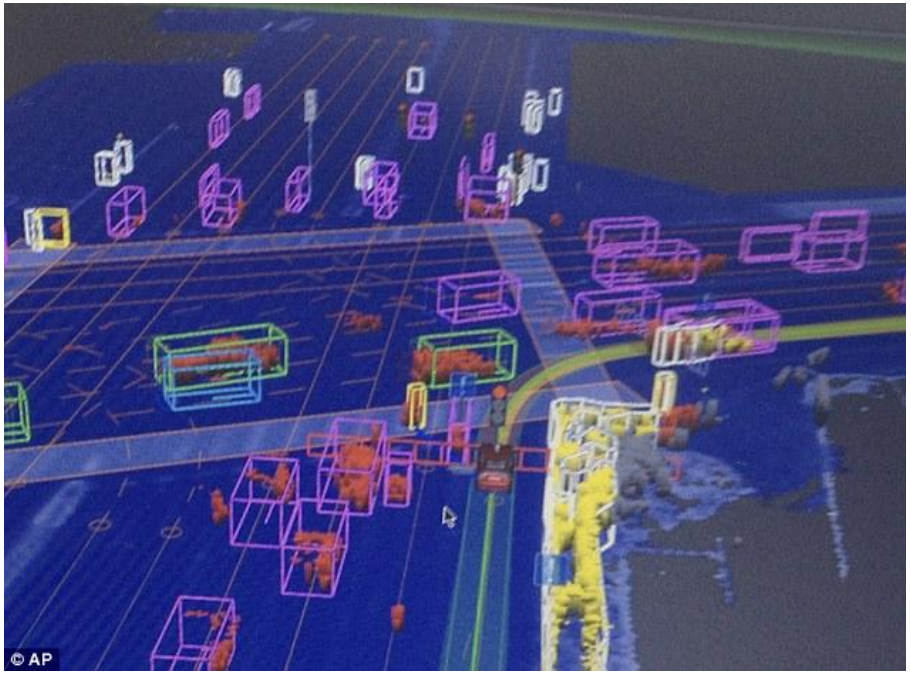
Source: S. Lazebnik, UIUC

Automotive safety

- Mobileye: Vision systems in high-end BMW, GM, Volvo models
  - Pedestrian collision warning
  - Forward collision warning
  - Lane departure warning
  - Headway monitoring and warning

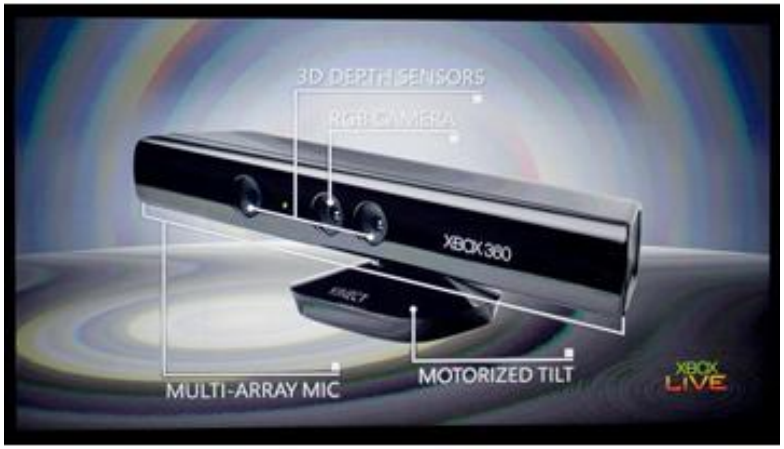
Source: A. Shashua, S. Seitz

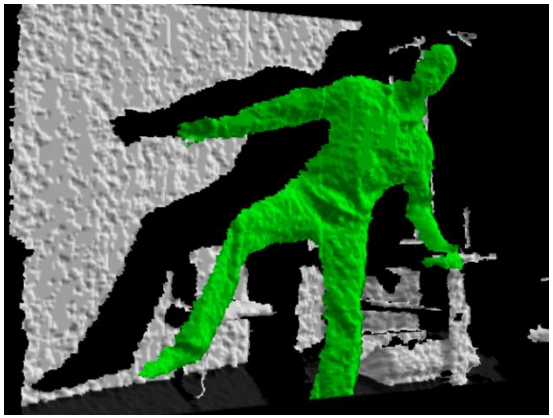
Self-driving cars

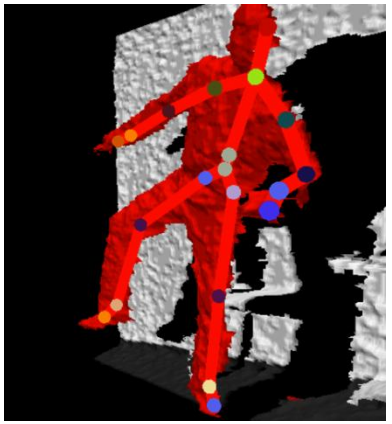


Source: S. Lazebnik, UIUC

Vision-based interaction: Xbox Kinect







Source: S. Lazebnik, UIUC

3D Reconstruction: Kinect Fusion

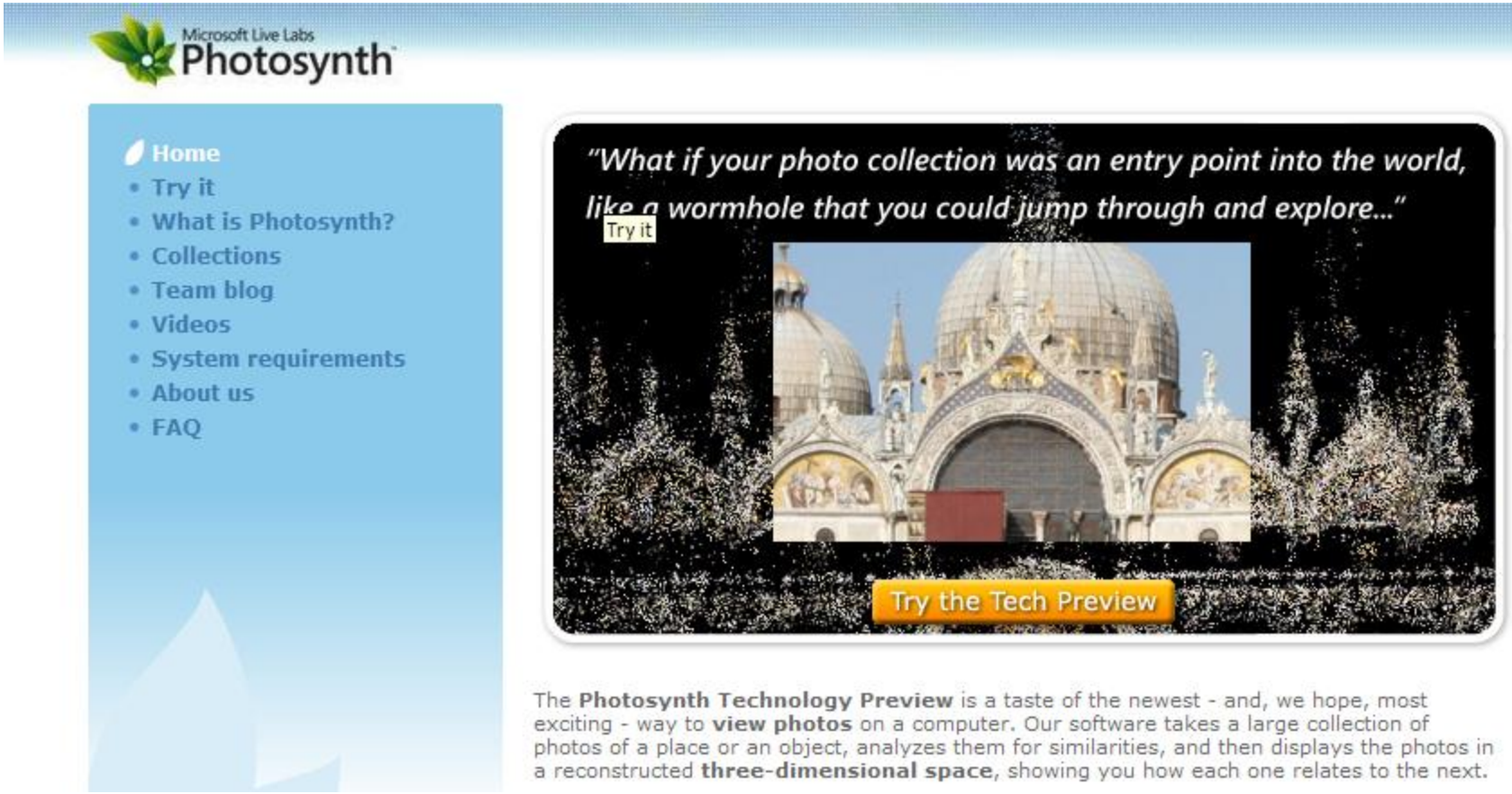
YouTube Video

Source: S. Lazebnik, UIUC

3D Reconstruction: Multi-View Stereo

YouTube Video

Source: S. Lazebnik, UIUC

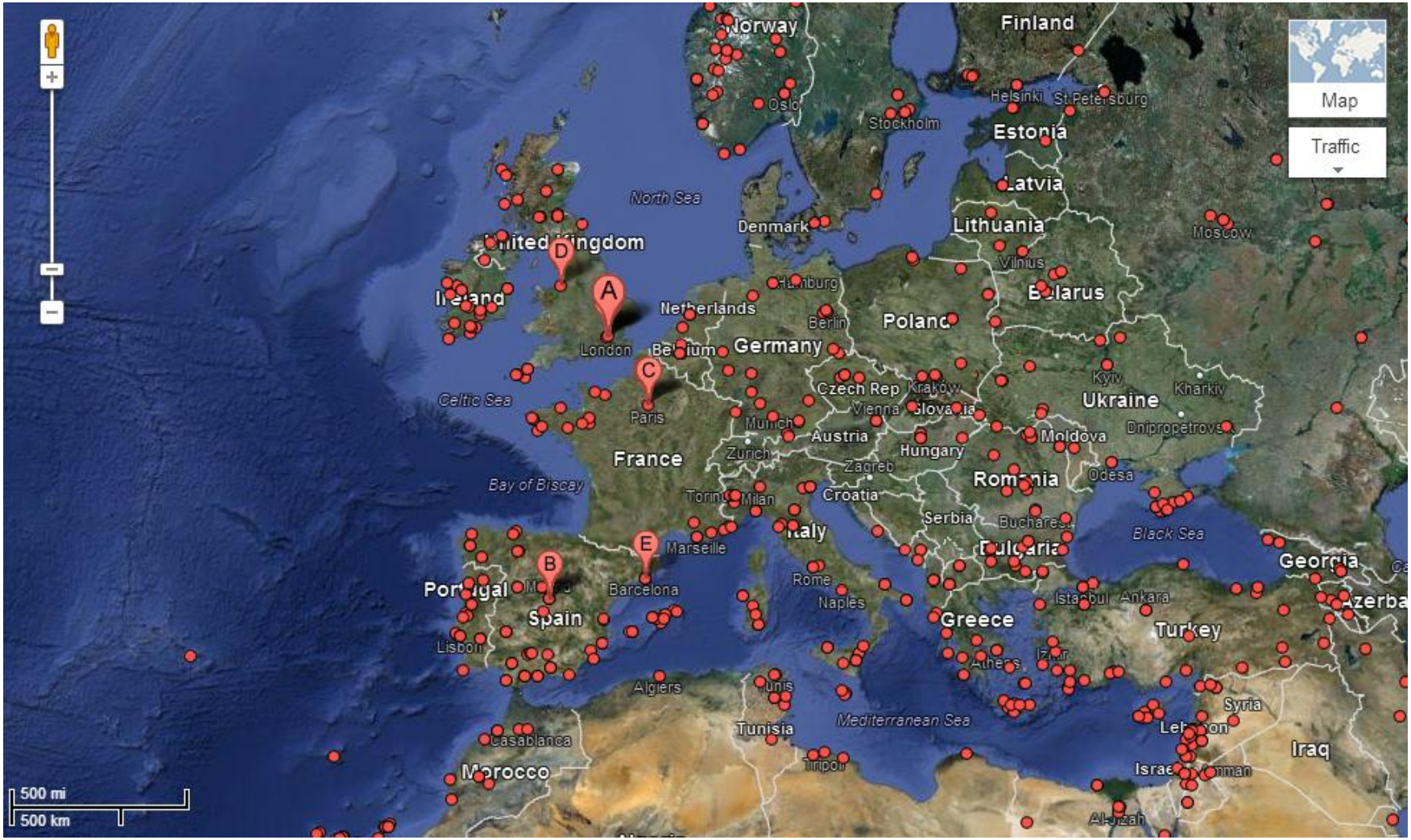


http://labs.live.com/photosynth/

Based on Photo Tourism technology developed by Noah Snavely, Steve Seitz, and Rick Szeliski

Source: Szeliski, Seitz. Chen

Google Maps Photo Tours



http://maps.google.com/phototours

Earth viewers (3D modeling)

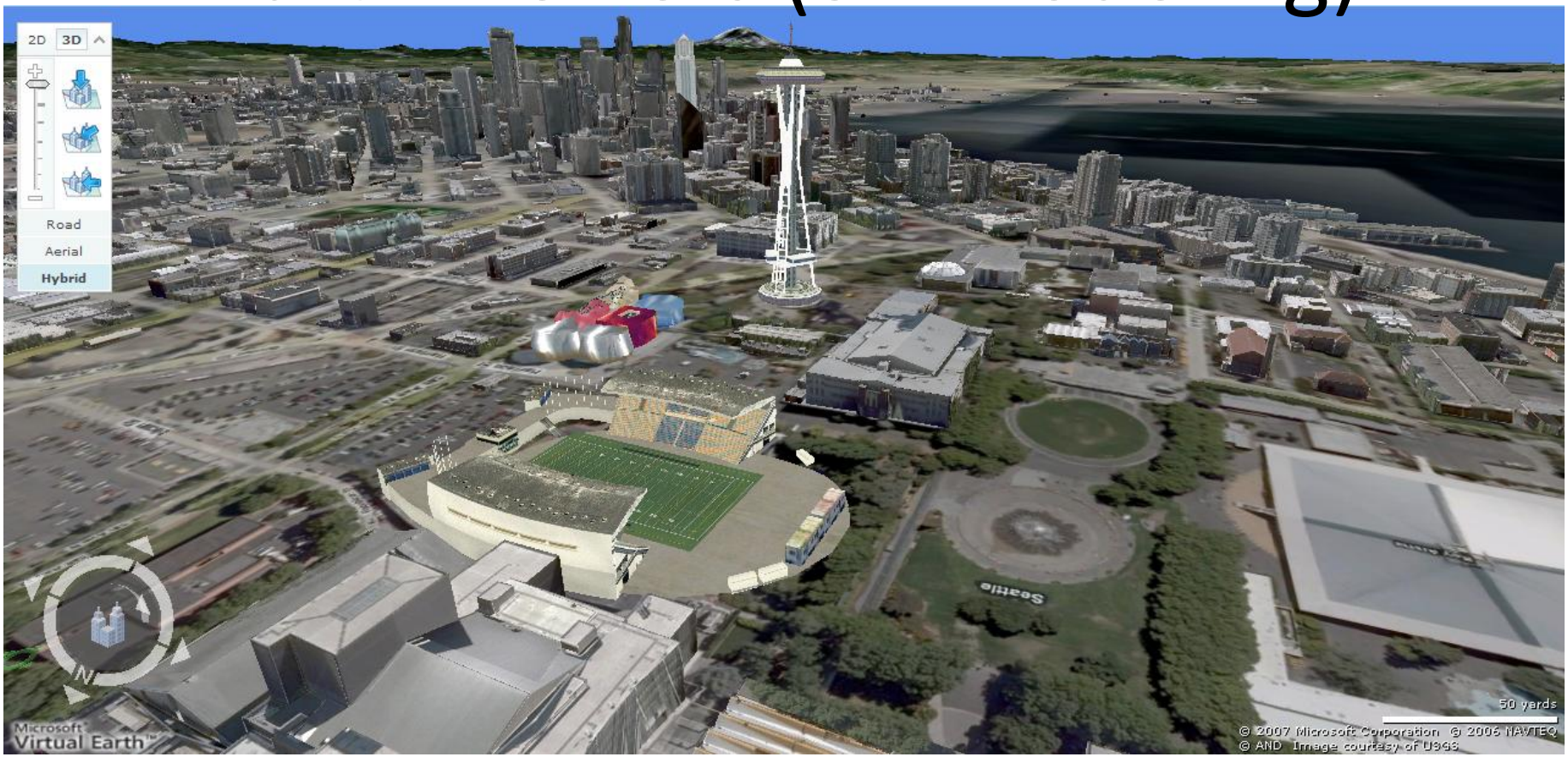


Image from Microsoft's Virtual Earth
(see also: Google Earth)

Source: Szeliski, Seitz. Chen

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… "

Source: Szeliski, Seitz. Chen

Object recognition (in mobile phones)

- This is becoming real:
  - Lincoln Microsoft Research
  - Point & Find, Nokia

Source: Szeliski, Seitz. Chen

Special effects: shape and motion capture

Source: Szeliski, Seitz. Chen

Vision in space

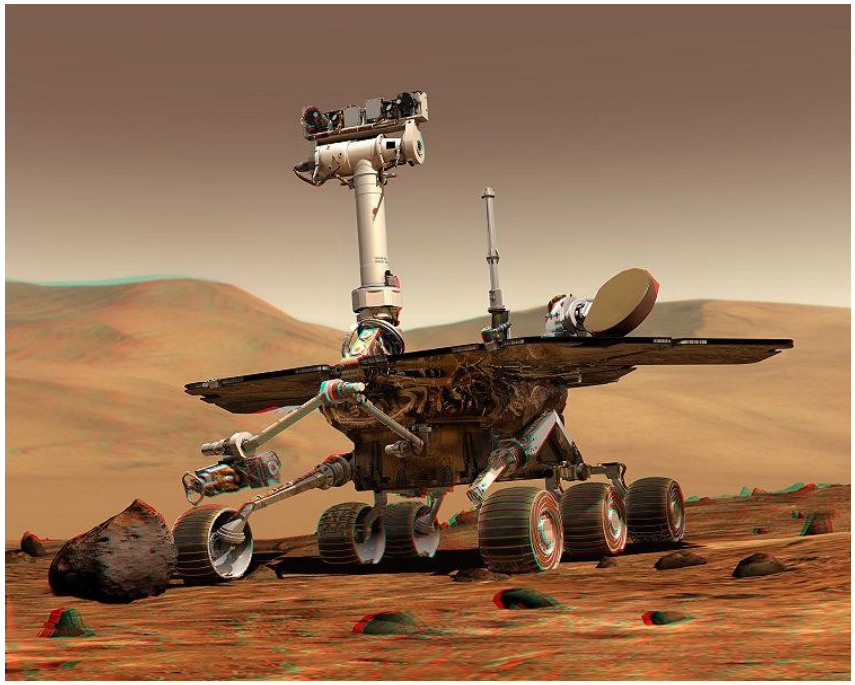
NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

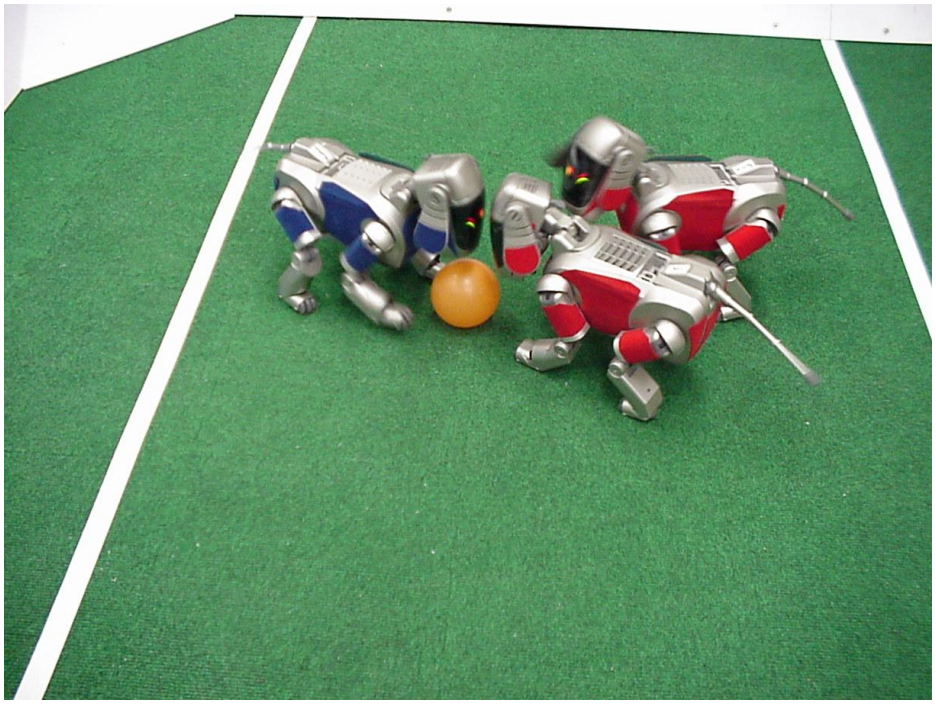
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Source: Szeliski, Seitz. Chen

Robotics



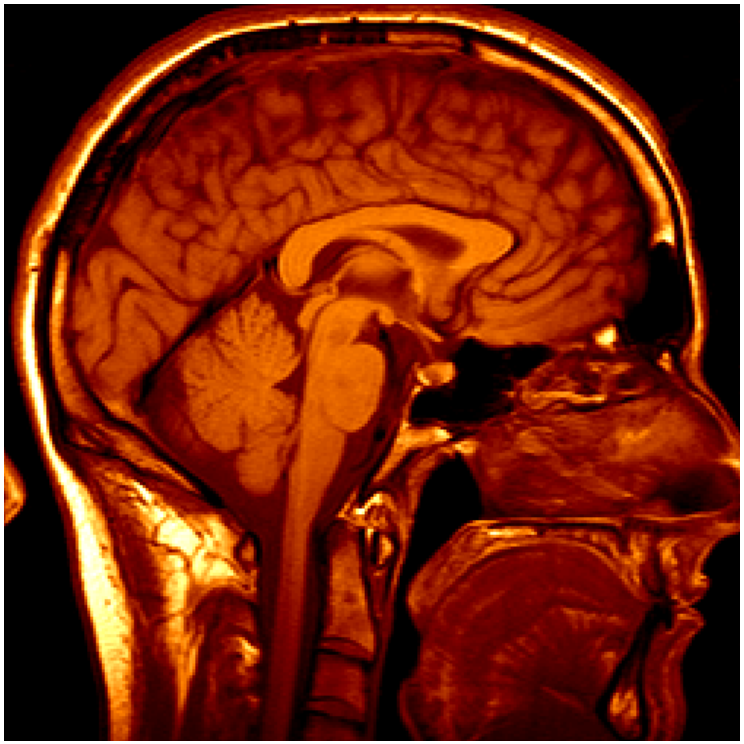
NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



http://www.robocup.org/

Source: Szeliski, Seitz. Chen

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Source: Szeliski, Seitz. Chen

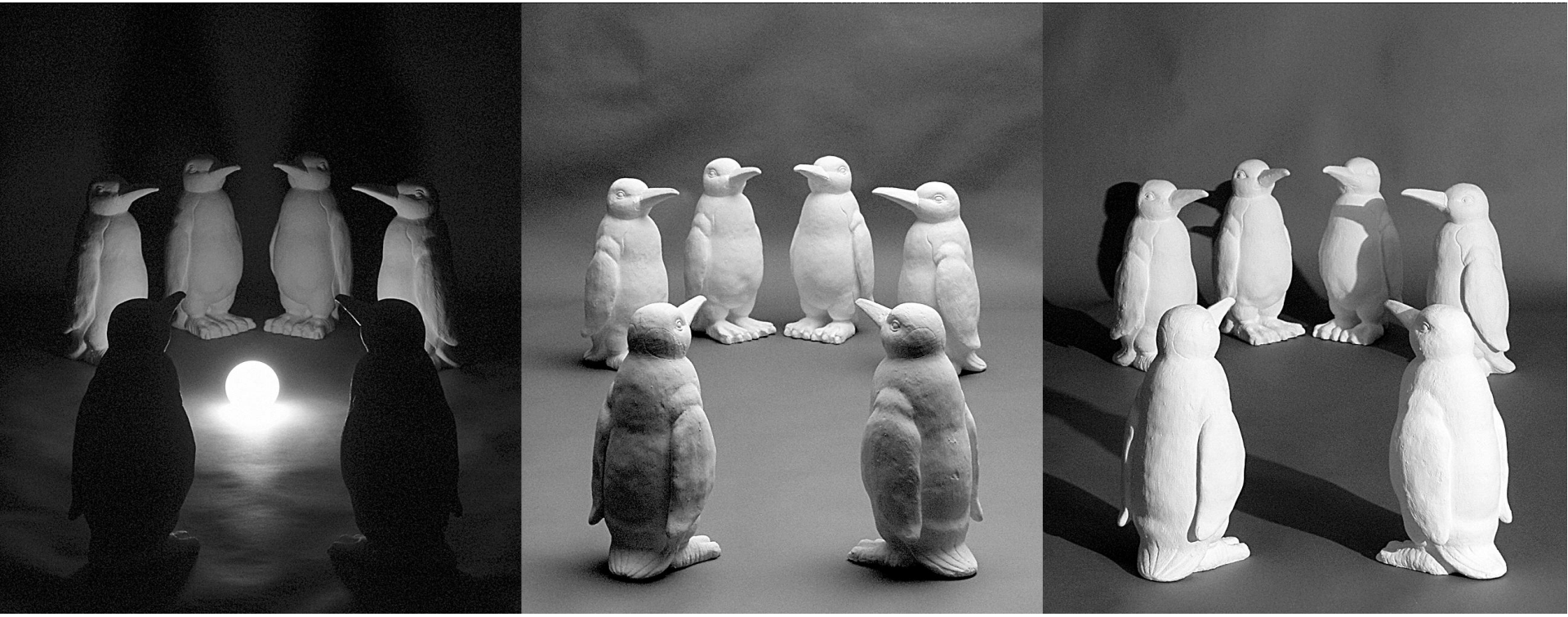
Why is computer vision difficult?

Challenges: viewpoint variation

Michelangelo 1475-1564

Source: Fei-Fei, Fergus & Torralba

Challenges: illumination

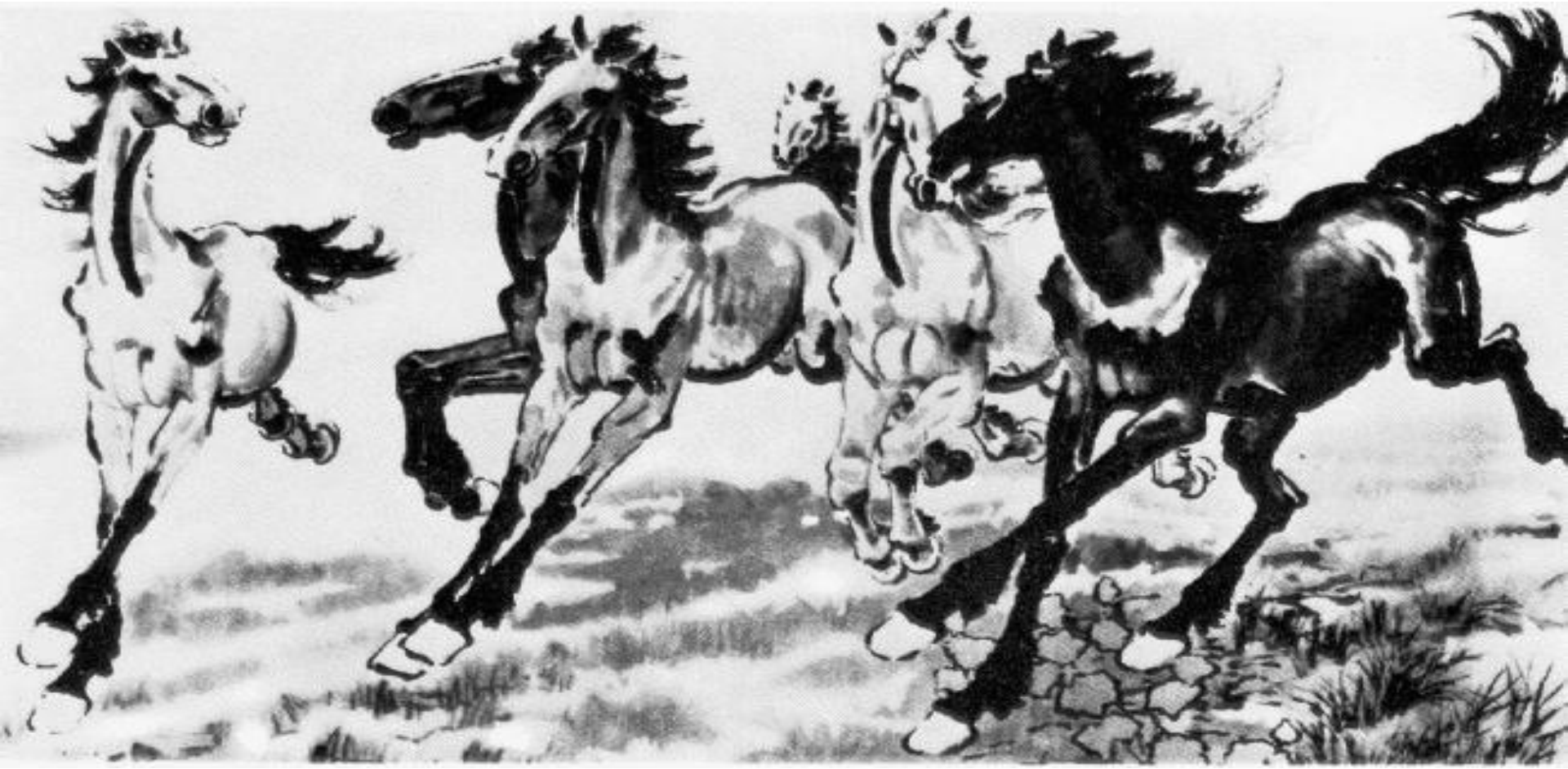


Source: J. Koenderink

Challenges: scale

Source: Fei-Fei, Fergus & Torralba

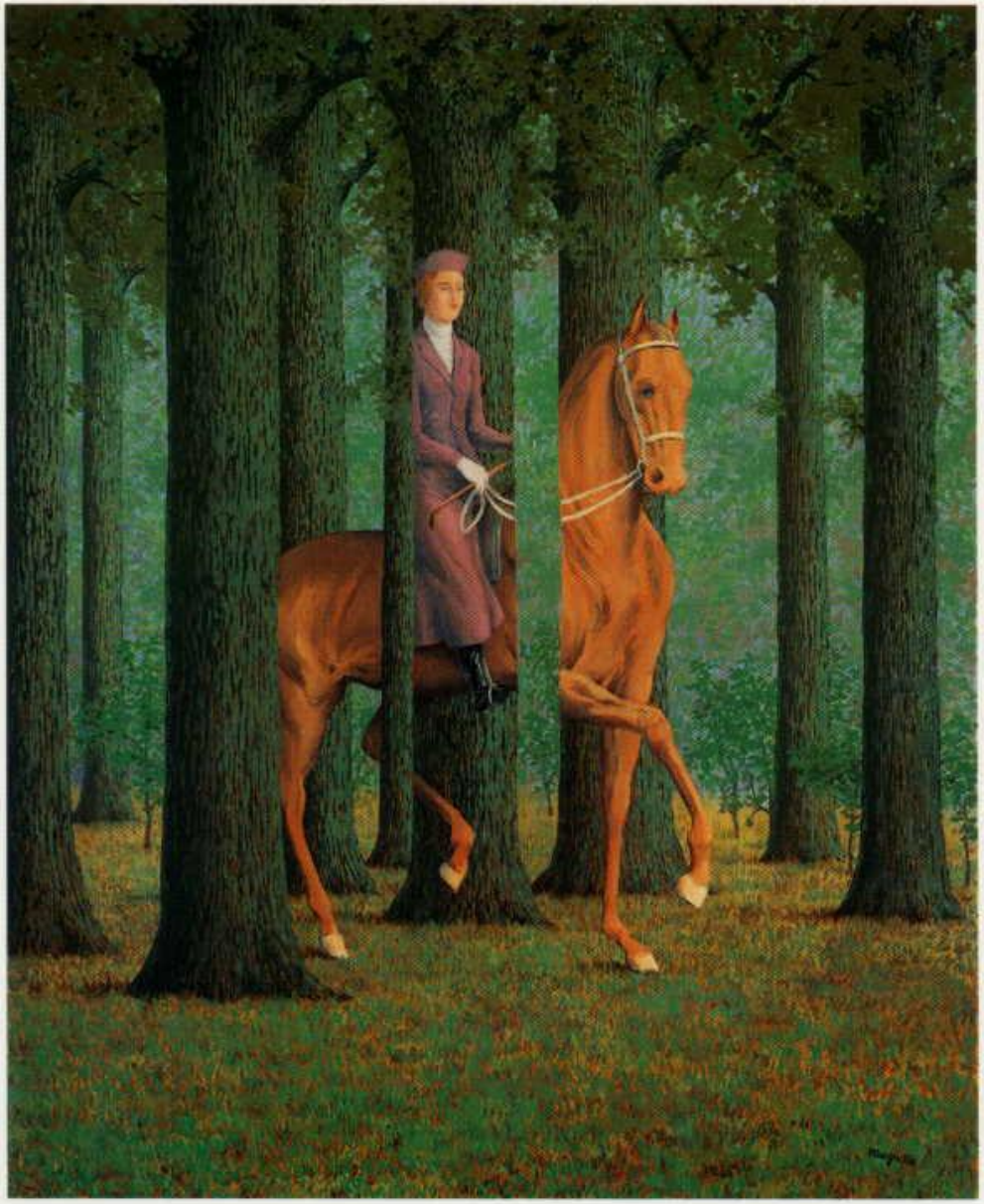
Challenges: deformation



Xu, Beihong 1943

Source: Fei-Fei, Fergus & Torralba

Challenges: occlusion



Source: Fei Fei, Fergus, Torralba

Challenges: Background Clutter

Source: Svetlana Lazebnik, UIUC

Challenges: Motion

Source: Svetlana Lazebnik, UIUC

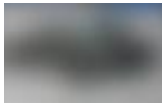
Challenges: object intra-class variation

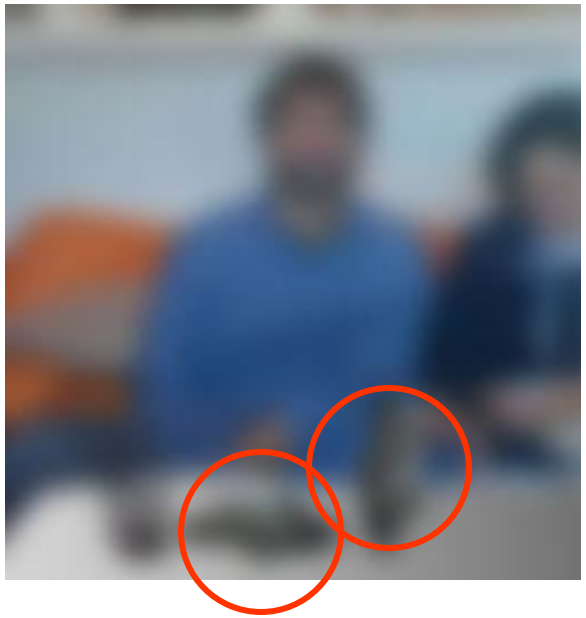


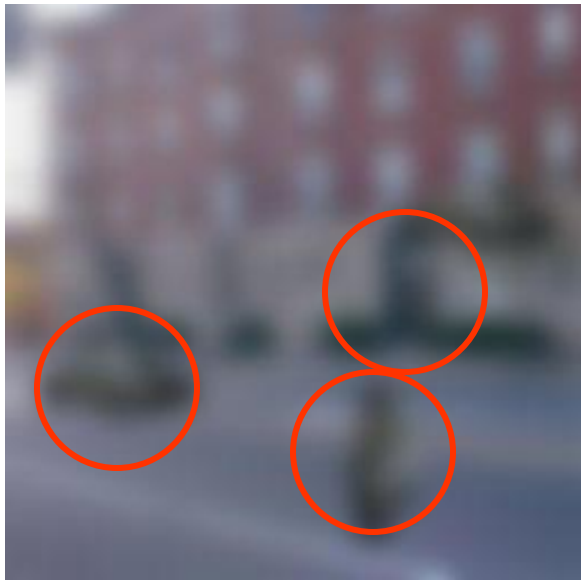



Source: Fei-Fei, Fergus & Torralba

Challenges: local ambiguity







Source: Fei-Fei, Fergus & Torralba

Challenges: local ambiguity

Source: Rob Fergus and Antonic

Challenges: local ambiguity

Source: Rob Fergus and Antonic

Challenges: Inherent ambiguity

- Many different 3D scenes could have given rise to a particular 2D picture

Image source: Svetlana Lazebnik, UIUC

Challenges or opportunities?

- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Image source: J. Koenderink

Depth cues: Linear perspective

Image source: Svetlana Lazebnik, UIUC

Depth cues: Aerial perspective

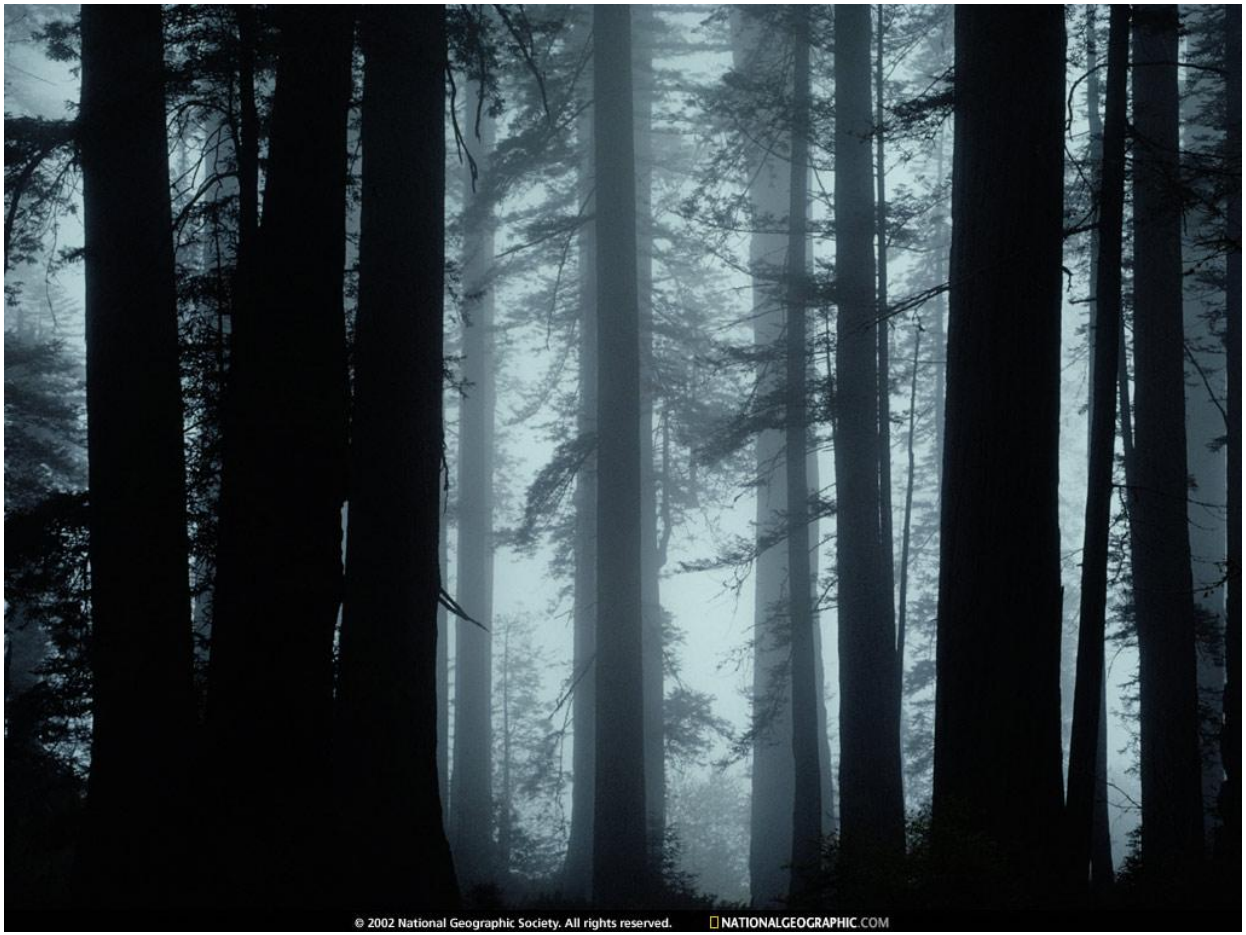


Image source: Svetlana Lazebnik, UIUC

Depth ordering cues: Occlusion

Source: J. Koenderink

Shape cues: Texture gradient

Image source: Svetlana Lazebnik, UIUC

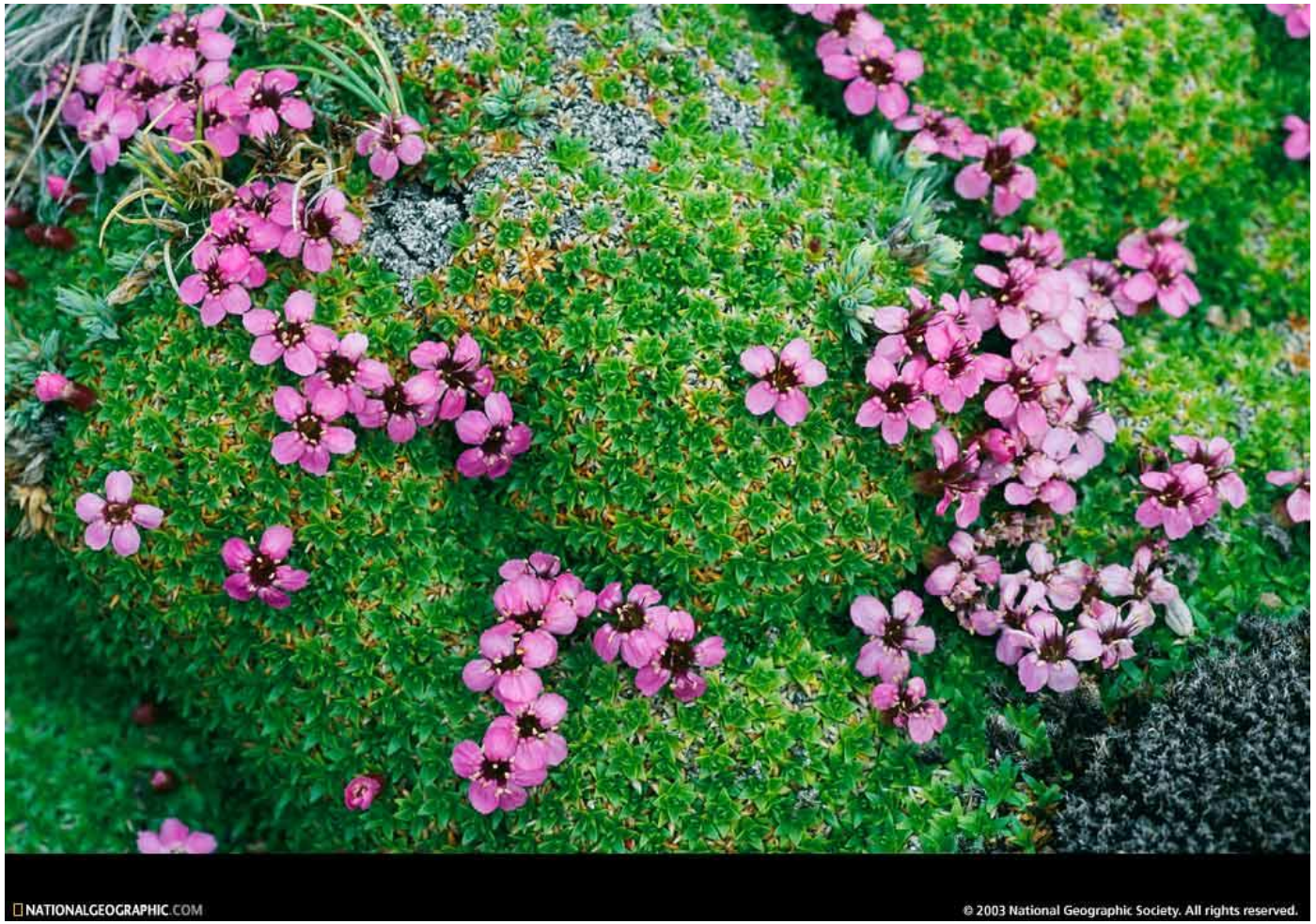
Shape and lighting cues: Shading

Image source: Svetlana Lazebnik, UIUC

Position and lighting cues: Cast shadows

Source: J. Koenderink

Grouping cues: Similarity (color, texture, proximity)

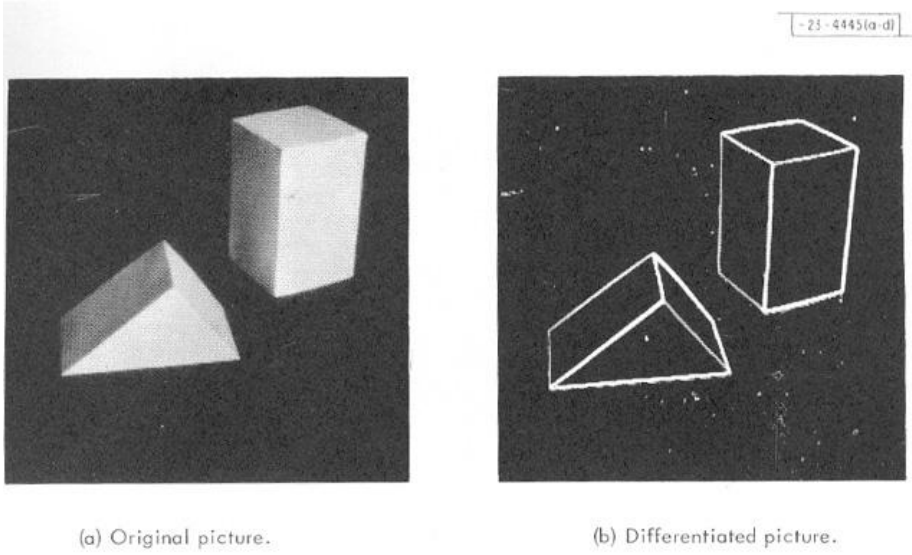


Source: Svetlana Lazebnik, UIUC

Grouping cues: “Common fate”

Source: Fei Fei Li, Stanford University

Origins of computer vision



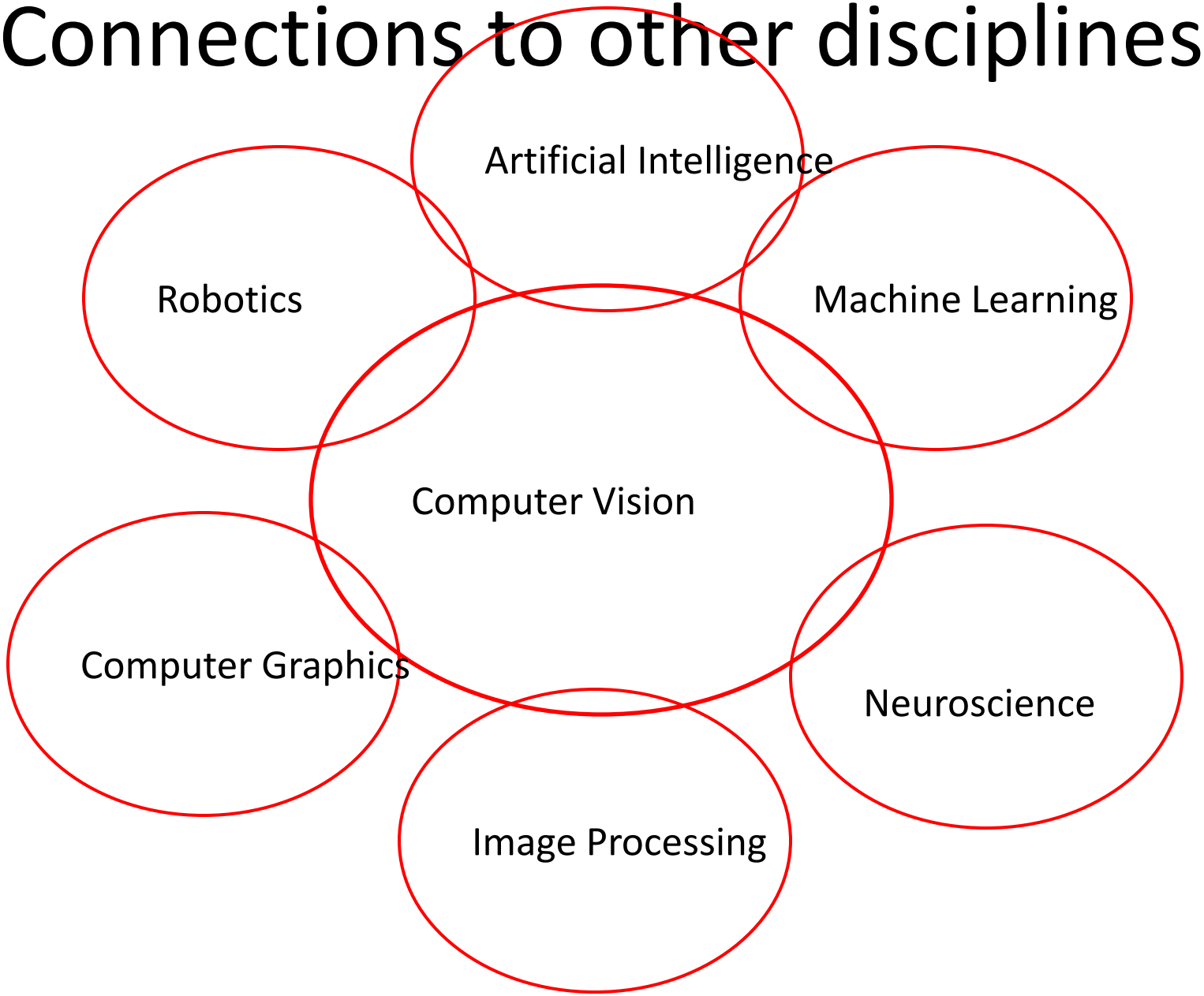
(a) Original picture. (b) Differentiated picture.

(c) Line drawing. (d) Rotated view.

L. G. Roberts *Machine Perception of Three Dimensional Solids*

Source: Svetlana Lazebnik, UIUC

Connections to other disciplines



Source: Svetlana Lazebnik, UIUC