Introduction to Computer Vision (CS 484 – CS 555)

Introduction

Dr. Sedat Ozer
Department of Computer Engineering
Bilkent University
sedat@cs.bilkent.edu.tr

Computer Vision:

Various definitions. Some common definitions are:

- Study of visual data.
  - Examples for visual data: Images, Videos
- A scientific field that extracts meaning (information) from images or videos. Examples are:
  - Visual recognition,
  - Detecting and locating cars in an image,
  - Look for a particular and a local activity occurring in videos,
  - Track the objects over time in videos, etc.
- A field that builds algorithms to understand & extract the content of images (or videos) and to use it for other applications and for other fields. Examples are:
  - Robotics,
  - Medical field,
  - Surveillance,
  - Entertainment, etc.

Check the course website

Web site for the course is available at:

http://cs.bilkent.edu.tr/~sedat/CS484_555/index.html

Course **syllabus** is available on the website.

Logistics:

Exams:

- See the Syllabus for details
  - Available on the course website
- 1 Midterm, quizzes,

Hands-on Experience:

- Final Project (No Final!)
- Programming assignments (PA)

Presentation Experience:

- Research paper presentation
- Final project presentation

Course hours:

- Wednesdays: 15:30 - 17:20,
- Mondays: 08:30 - 10:20
- Office: EA 524

Notes:

- Follow the slides & course website

Special thanks to:

- Dr. Selim Aksoy

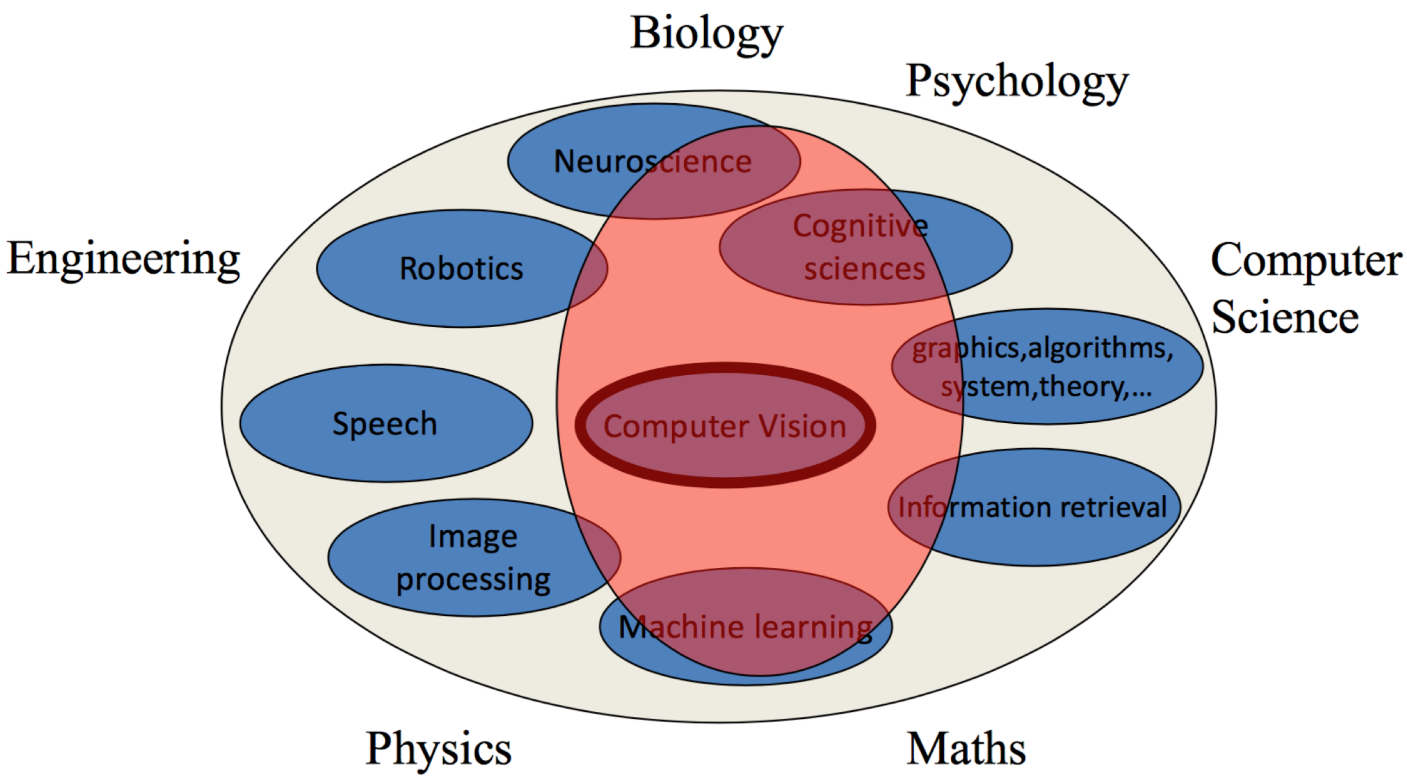
Teaching Assistants:

- Google
- Bing
- TBA

- Attend the class! (mandatory), follow the lecture notes and in-class material (and read the chapters given for that week on the website).
- Python and Matlab are the required languages for PA. (Some PAs may require you to use Matlab, however main programming language will be Python. You must use the specified language in each PA).

Computer Vision

What is it related to?



Computer vision: introduction

Tasks: Image Classification

Problem:
Are there any human in this image?

Answer: Yes / No

⋮

Problem:
Are there any cars in this image?

Answer: Yes / **No**

Tasks: Detection (localization)

Problem:
Tell me if there is any person in this image and if there is, where in the image?

Answer: is in the form of bounding boxes on the image.

Tasks: Object Categorization

Problem:
Detect all the objects in this image and give me a list of their type.

Answer: Building, Tree, Sky, People, Human, Grass

Semantic Segmentation: What we want

Image source: Antonio Torralba

Segmentation

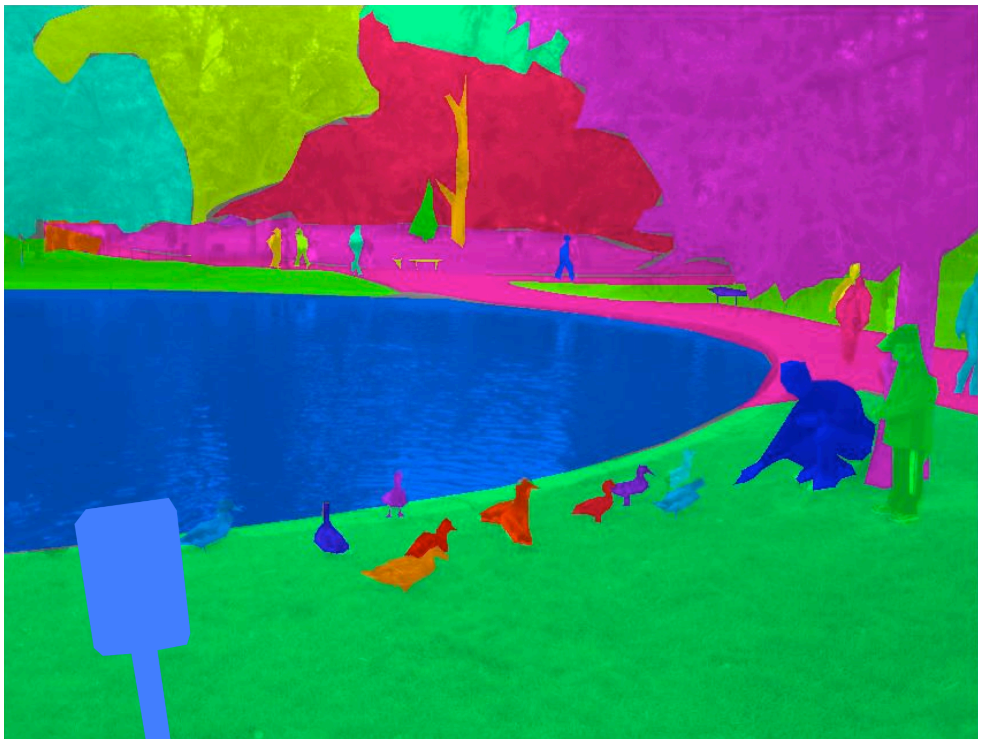


Image source: Antonio Torralba

Segmentation

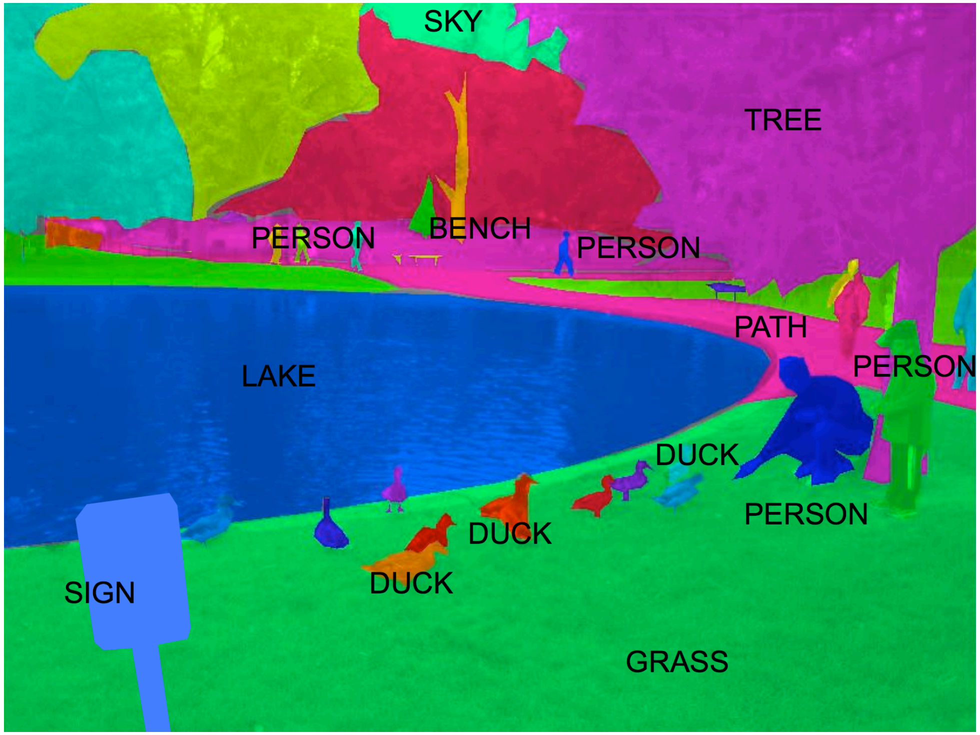


Image source: Antonio Torralba

Segmentation

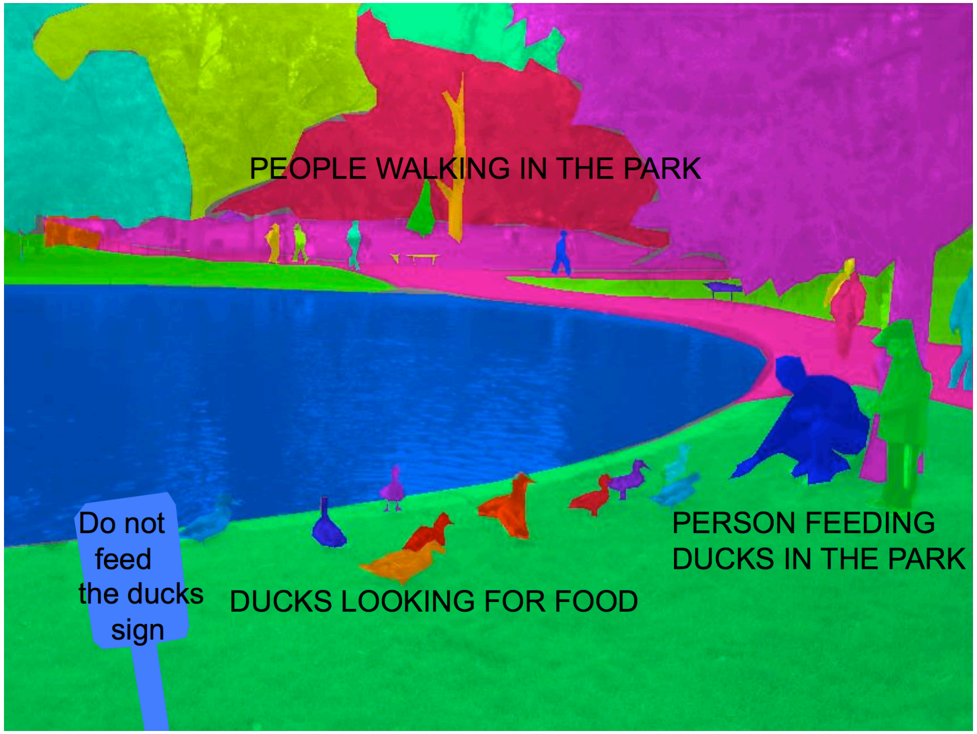
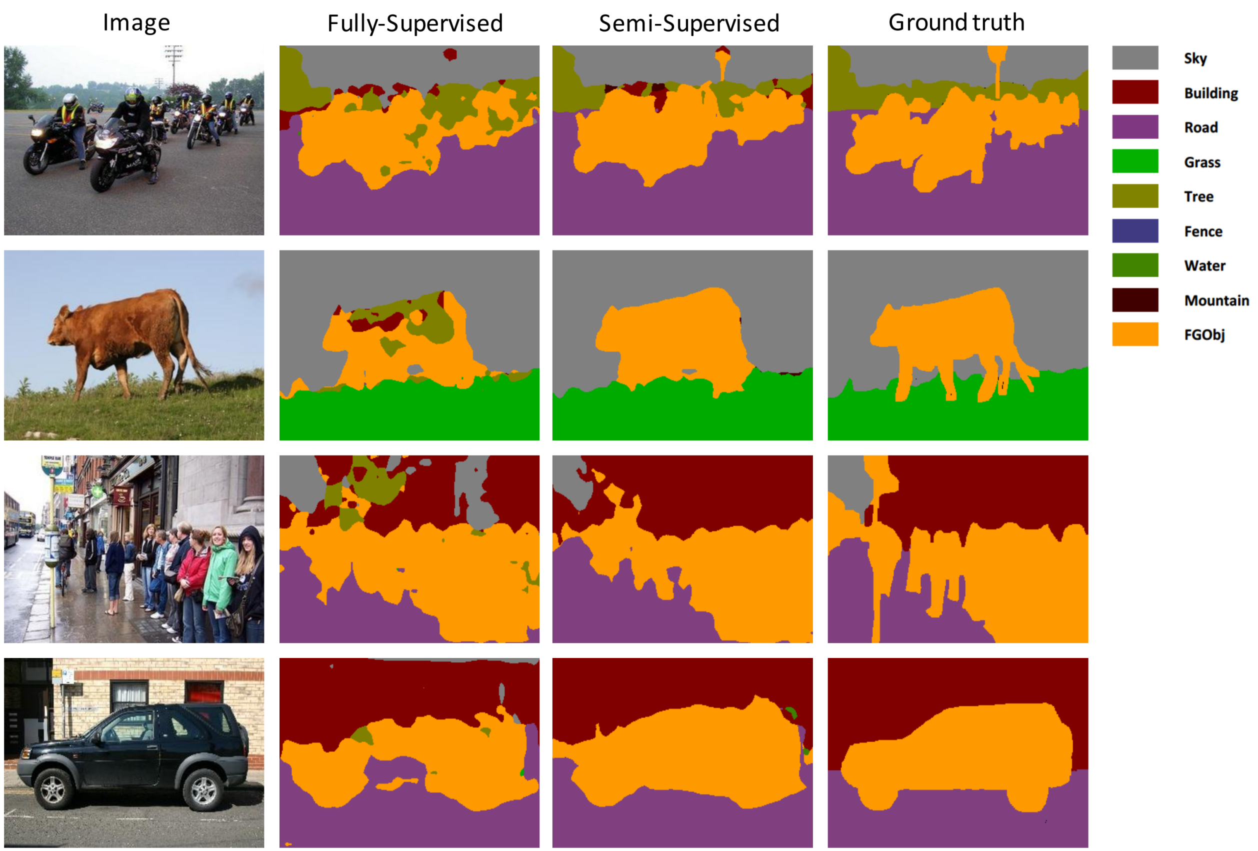


Image source: Antonio Torralba

Semantic Segmentation: Early Results



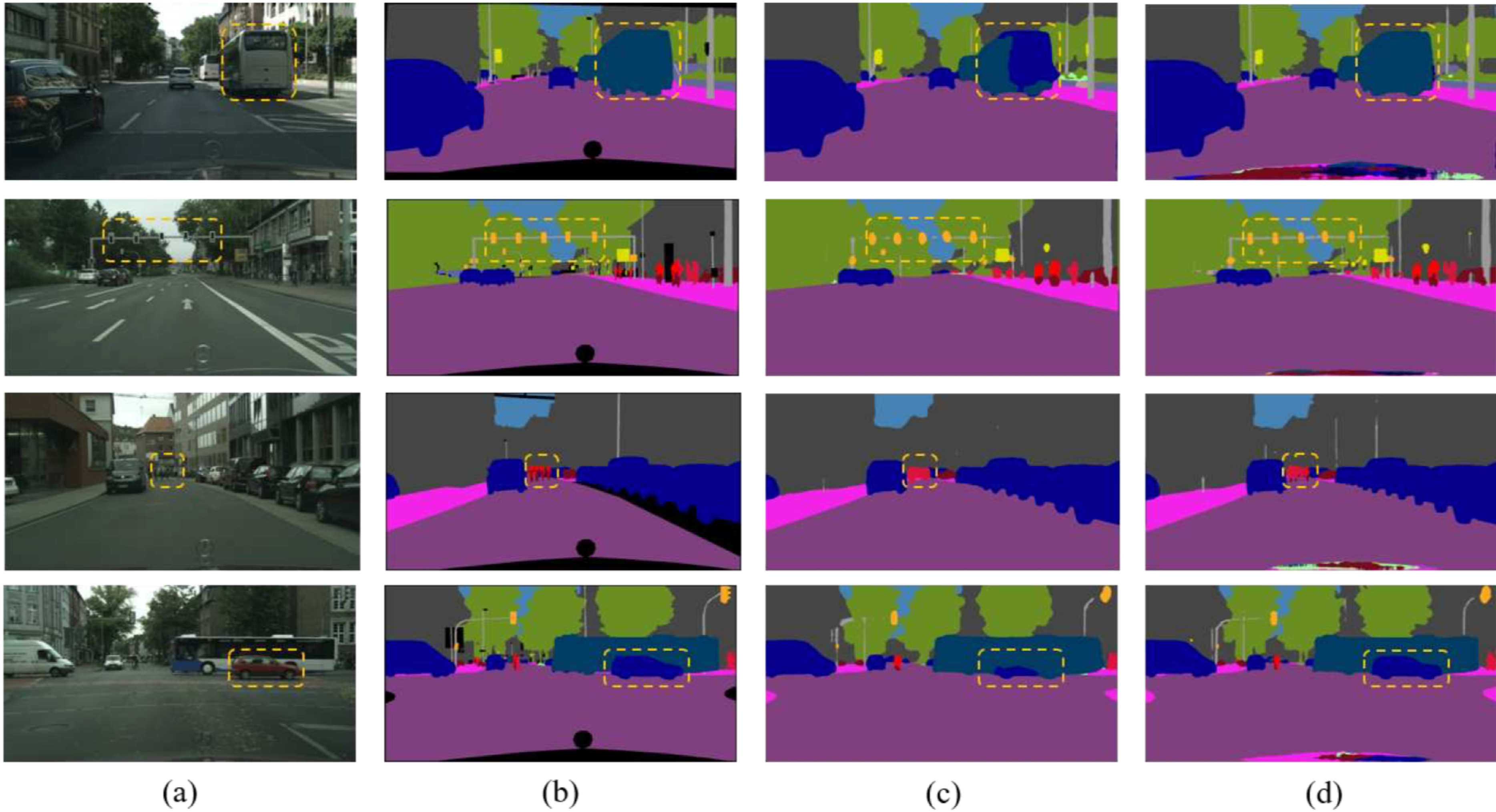
Semantic Segmentation Using GAN, Nasim, Concetto, and Mubarak, 2017.

Semantic Segmentation: Recent Results (a)

(a): original images
(b): ground truth
(c): Deep Lab results

Image source: Wang et al., "Self-supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation", CVPR 2020.

Semantic Segmentation: Recent Results (b)



(a): original images, **(b):** ground truth
(c): Deep Lab v3+ results
(d): Improved DeepLabv3+ results

Image source: Wang et al., "Dual Super-Resolution Learning for Semantic Segmentation ", CVPR 2020.

Instance Segmentation

Images are from: Wang et al. ”CenterMask: single shot instance segmentation with point representation”, CVPR 2020.

Q & A: What will we study in this course?

- How is this different than “Computer Vision” course?
- “I hear that deep learning is the big thing in CV. Will we learn that in this course”?

A Brief History

The early optimism (1960-1970)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

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

50 Years ago

25 Years ago

15 Years ago

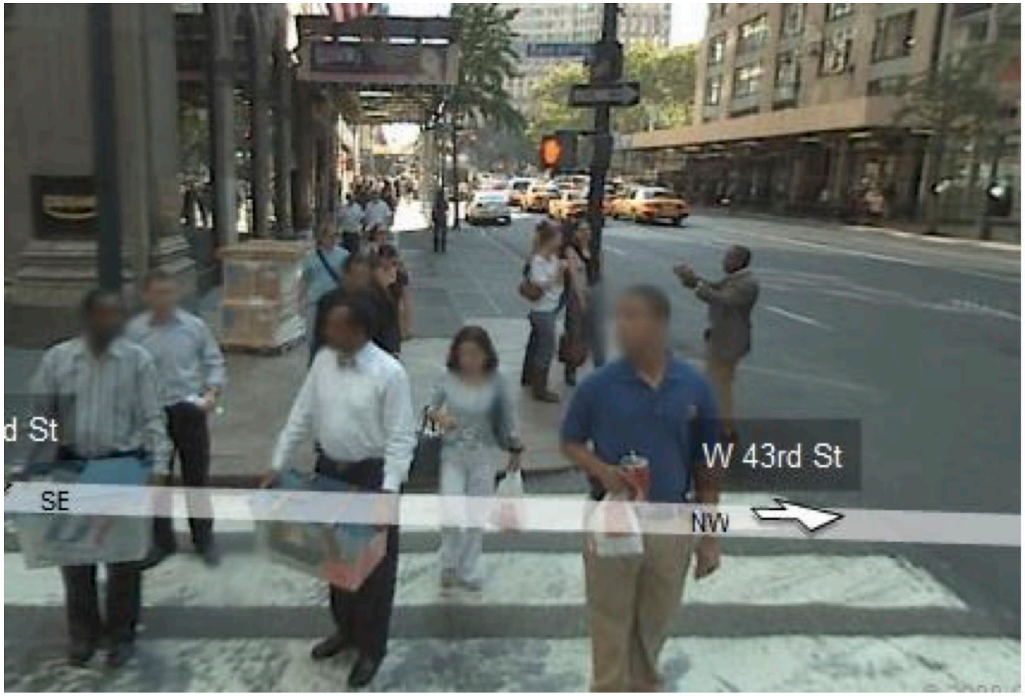
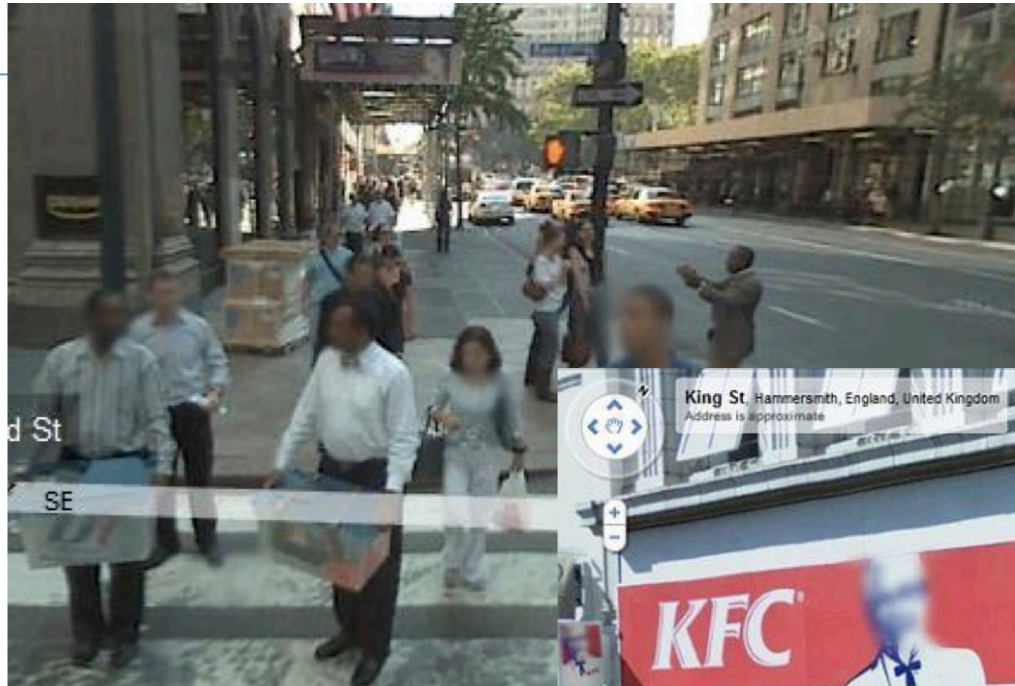


Image source: Antonio Torralba

• The representation and matching of pictorial structures Fischler, Elschlager (1973).
• Face recognition using eigenfaces M. Turk and A. Pentland (1991).
• Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
• Graded Learning for Object Detection - Fleuret, Geman (1999)
• Robust Real-time Object Detection - Viola, Jones (2001)
• Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
•….



- The representation and matching of pictorial structures Fischler, Elschlager (1973).
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time Object Detection - Viola, Jones (2001)
- Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
- ….

Image source: Antonio Torralba

Era of hand-crafted features

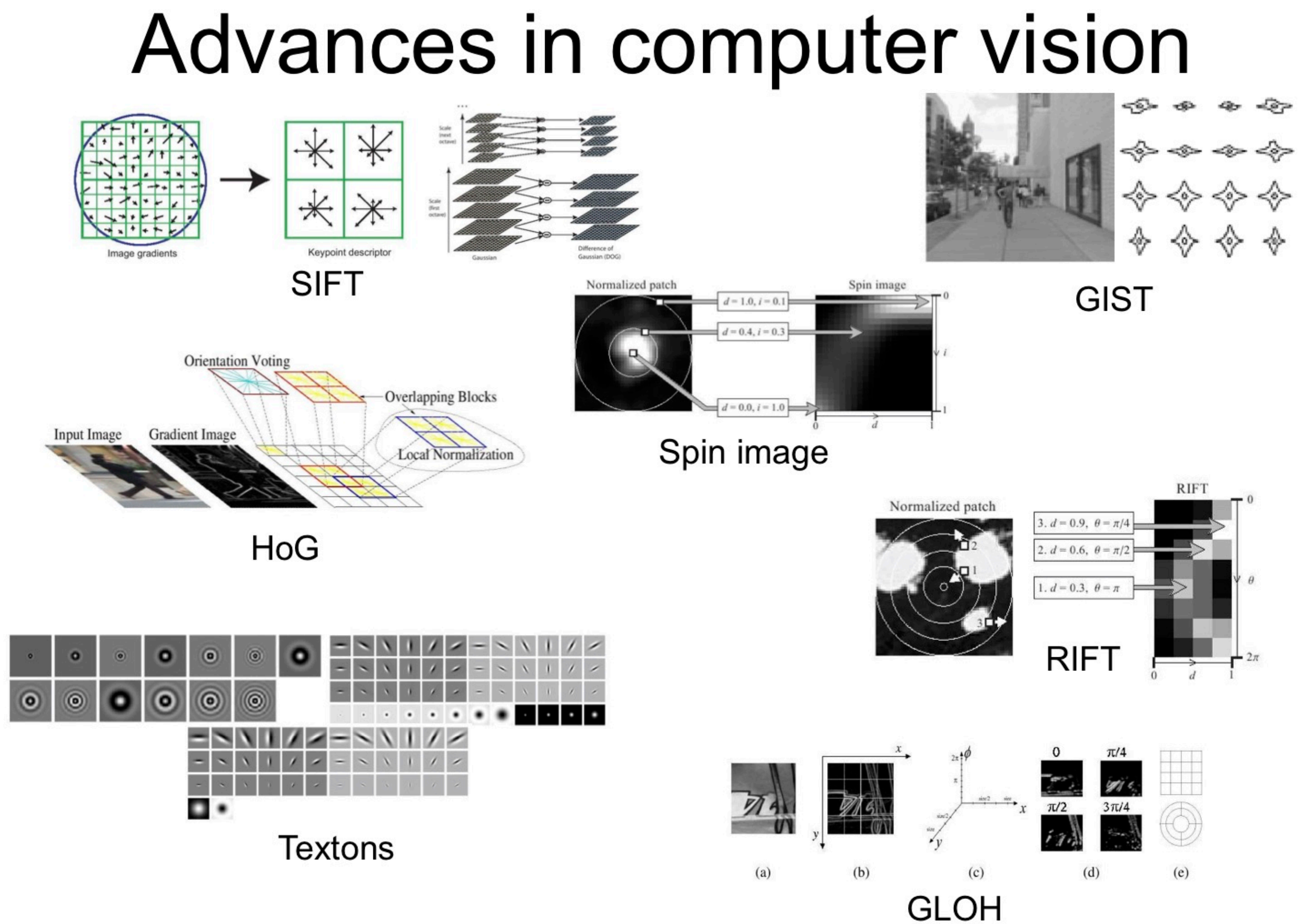


Image source: Antonio Torralba

Big Data

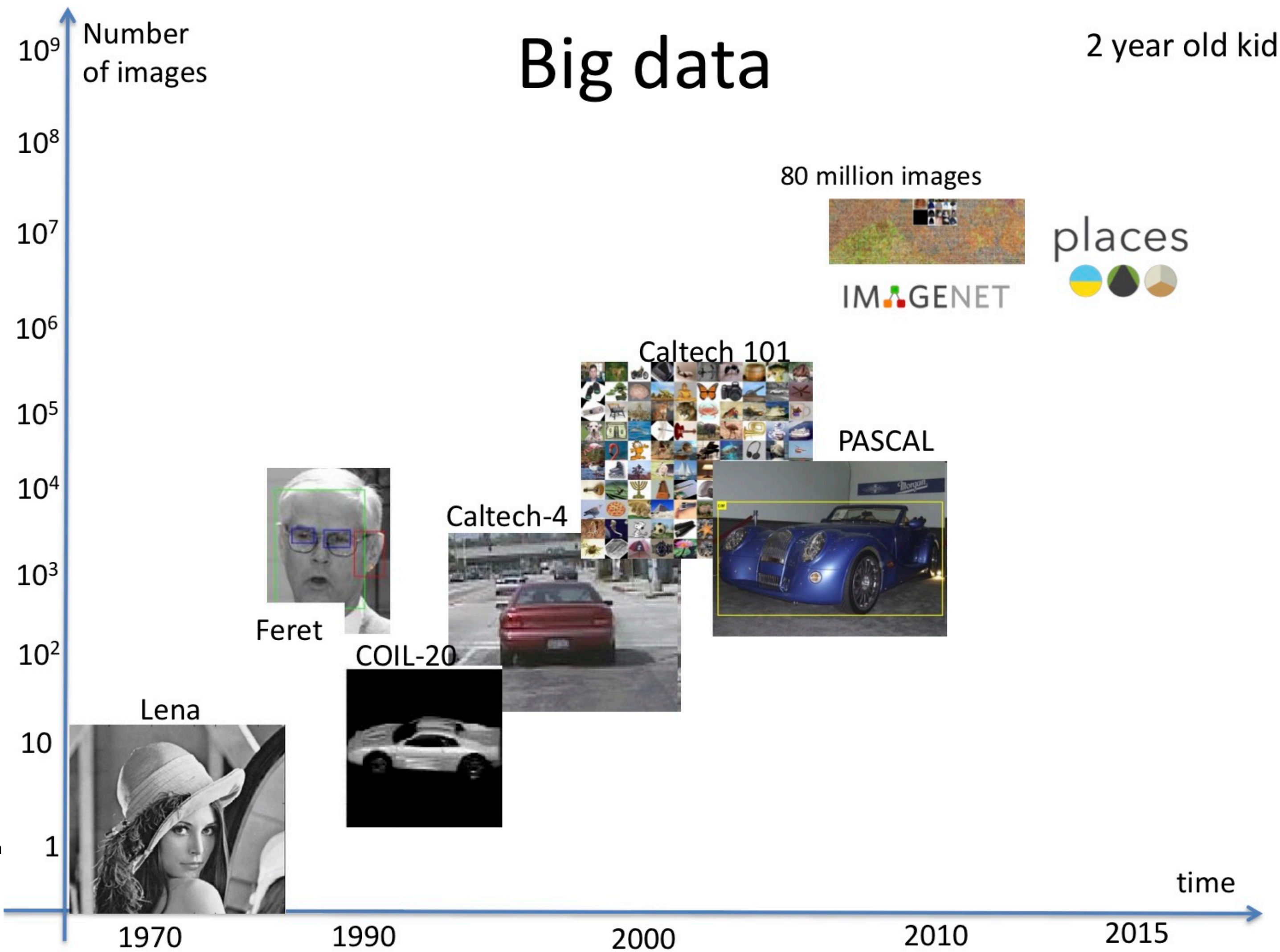


Image source: Antonio Torralba

Big data tools and companies

The time of big data

Image source: Antonio Torralba

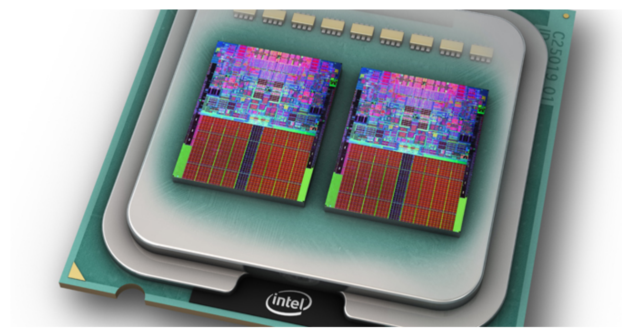
Distributed / GPU / Cloud Computing

amazon web services™

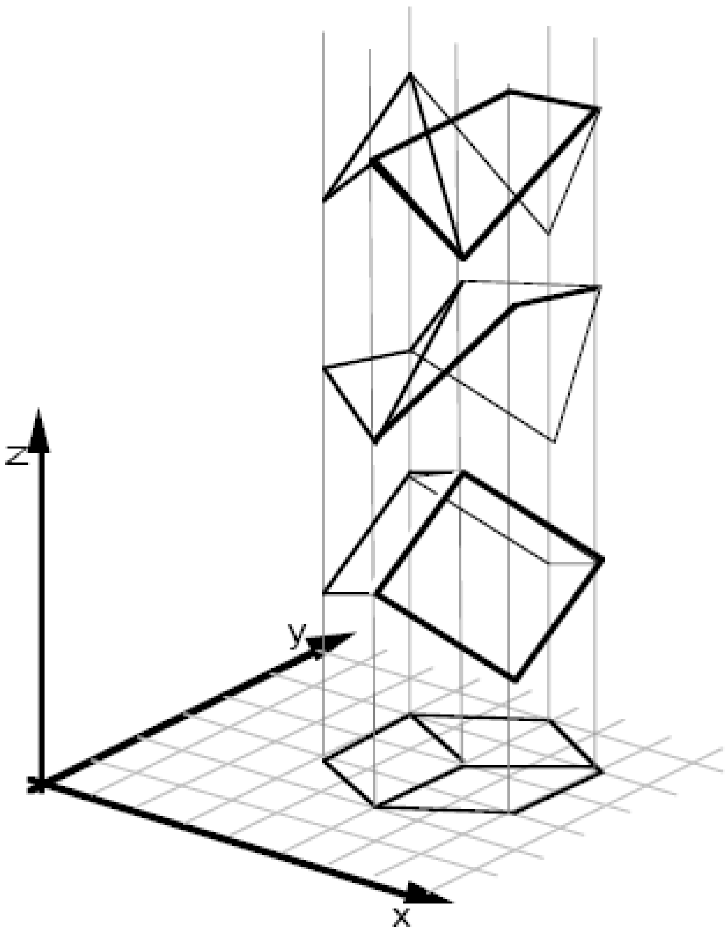
Google Cloud

Microsoft Azure

NVIDIA.
GPU CLOUD



Why vision is so hard?



[Sinha and Adelson 1993]

Ill-posed problem

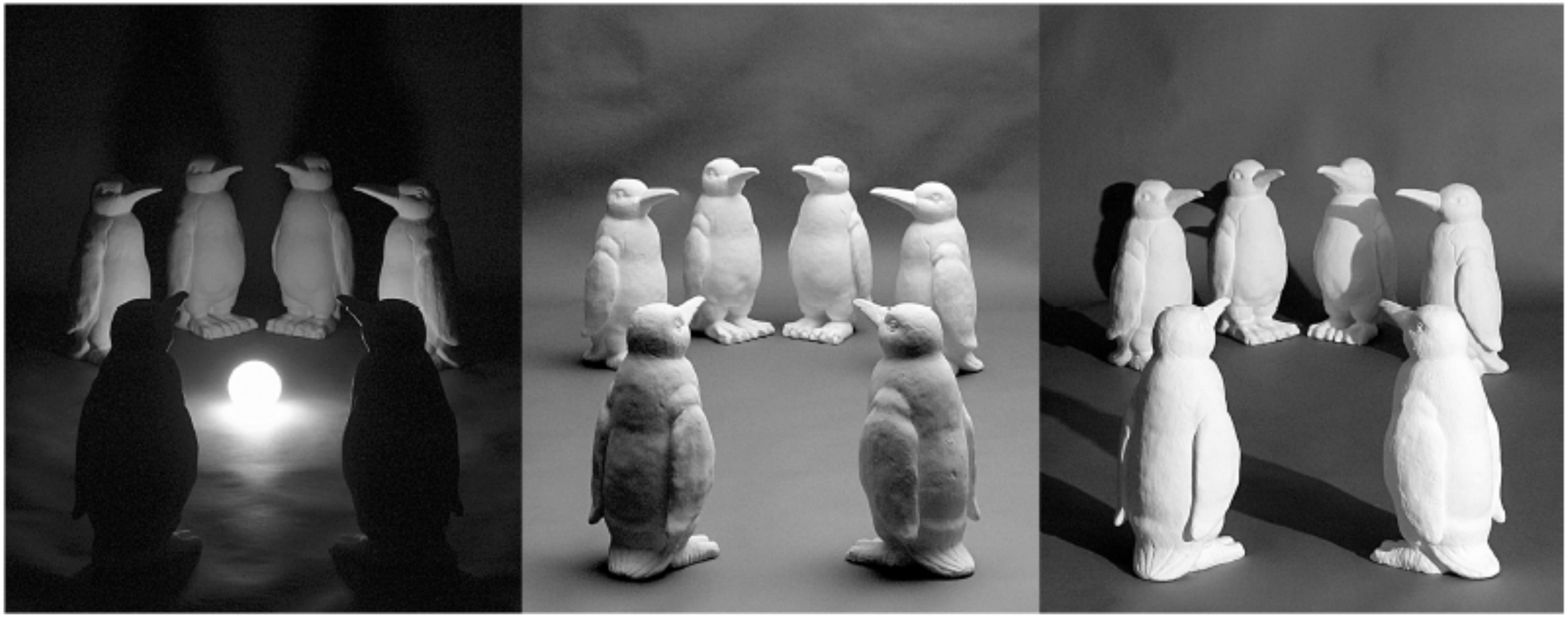
Adapted from Ali Farhadi, U of Washington

Challenges 1: view point variation

Michelangelo 1475-1564

Adapted from L. Fei-Fei,
R. Fergus, A. Torralba

Challenges 2: illumination



Adapted from Fei-Fei Li

Challenges 3: occlusion

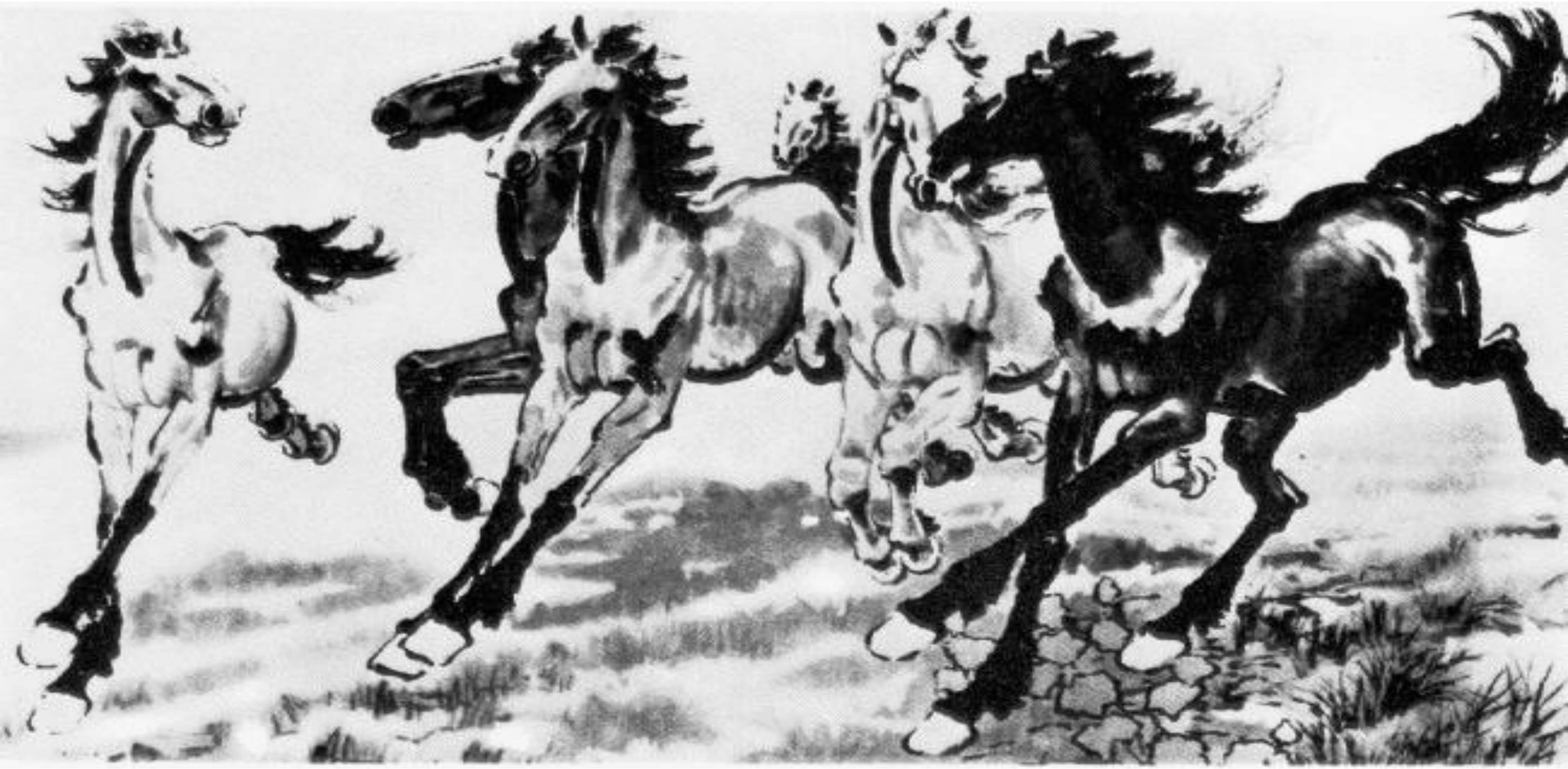
Magritte, 1957

Adapted from L. Fei-Fei,
R. Fergus, A. Torralba

Challenges 4: scale

Adapted from L. Fei-Fei, R. Fergus, A. Torralba

Challenges 5: deformation



Xu, Beihong 1943

Adapted from L. Fei-Fei, R. Fergus, A. Torralba

Challenges 6: background clutter

Adapted from Fei-Fei Li

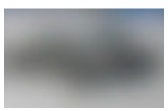
Challenges 7: intra-class variation






Adapted from L. Fei-Fei, R. Fergus, A. Torralba

Challenges 8: local ambiguity



Adapted from Antonio Torralba, MIT

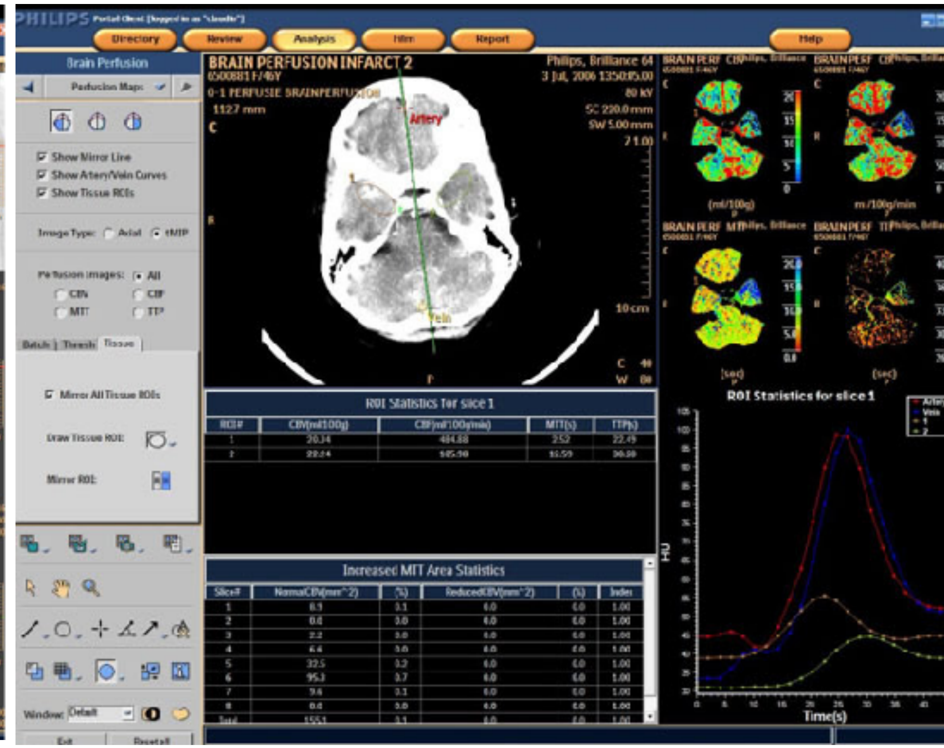
What works today?

- Driver assistance and autonomous cars
- Medical image analysis
- Security
  - Biometrics
  - Surveillance
  - Tracking
  - Target recognition
- Remote sensing
- Robotics
- Industrial inspection, quality control
- Document analysis
- Multimedia
- Assisted living
- Human-computer interfaces
- …

Test: what do you see in this image?

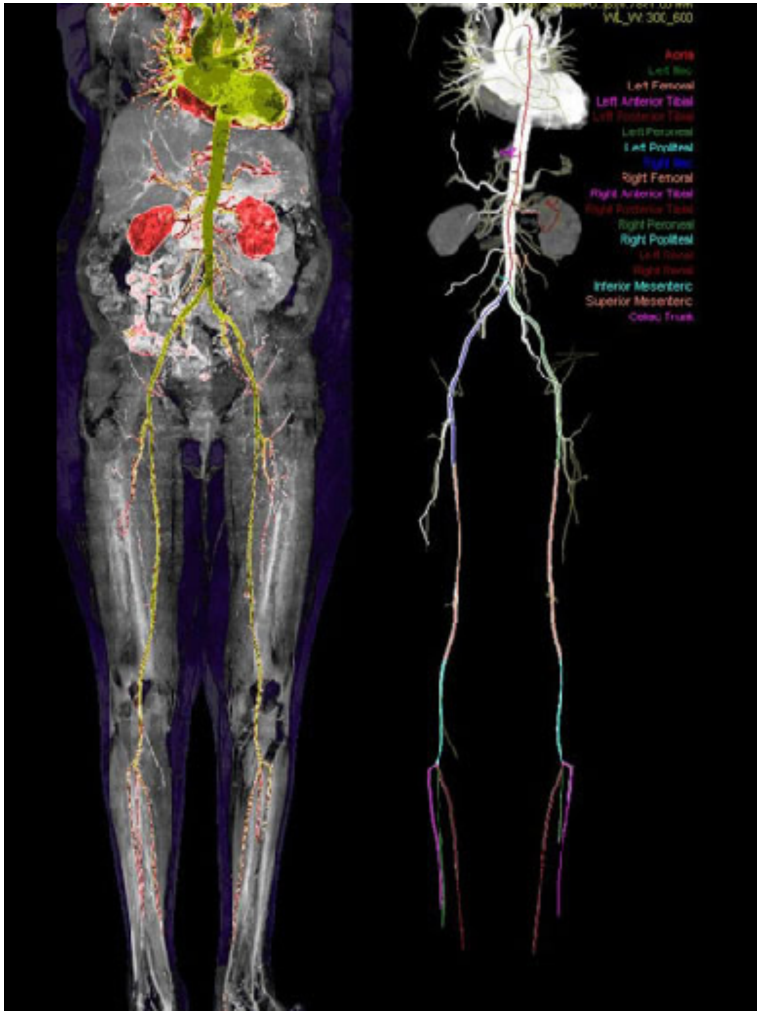
Image source: Antonio Torralba

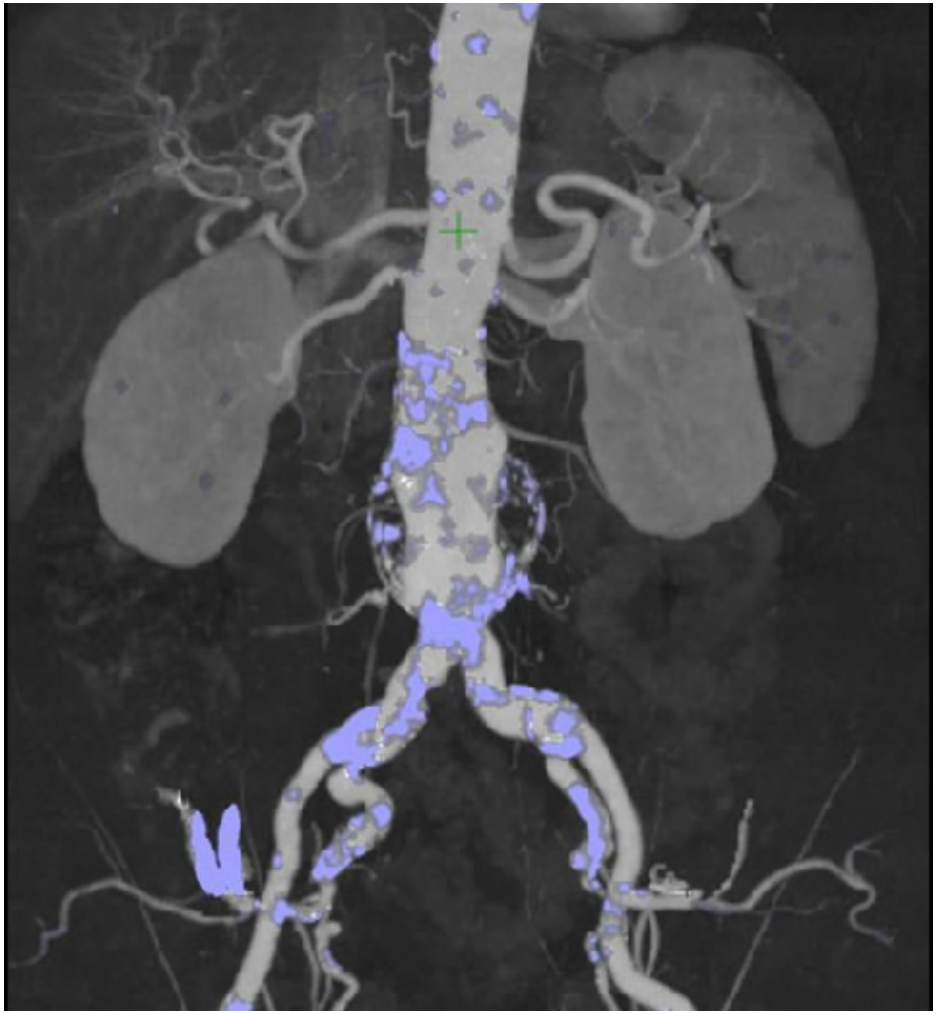
Medical image analysis



http://www.clarontech.com

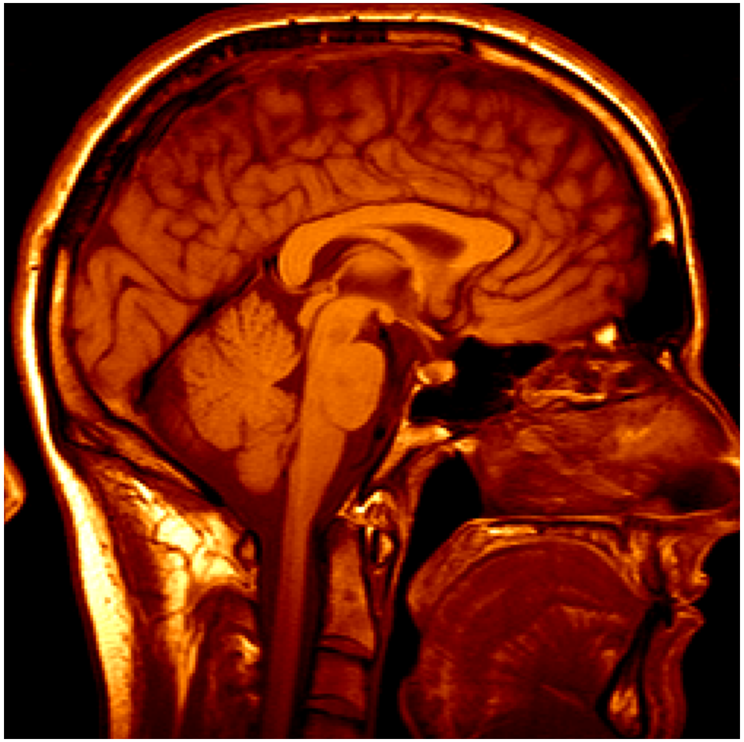
Medical image analysis





http://www.clarontech.com

Medical image analysis

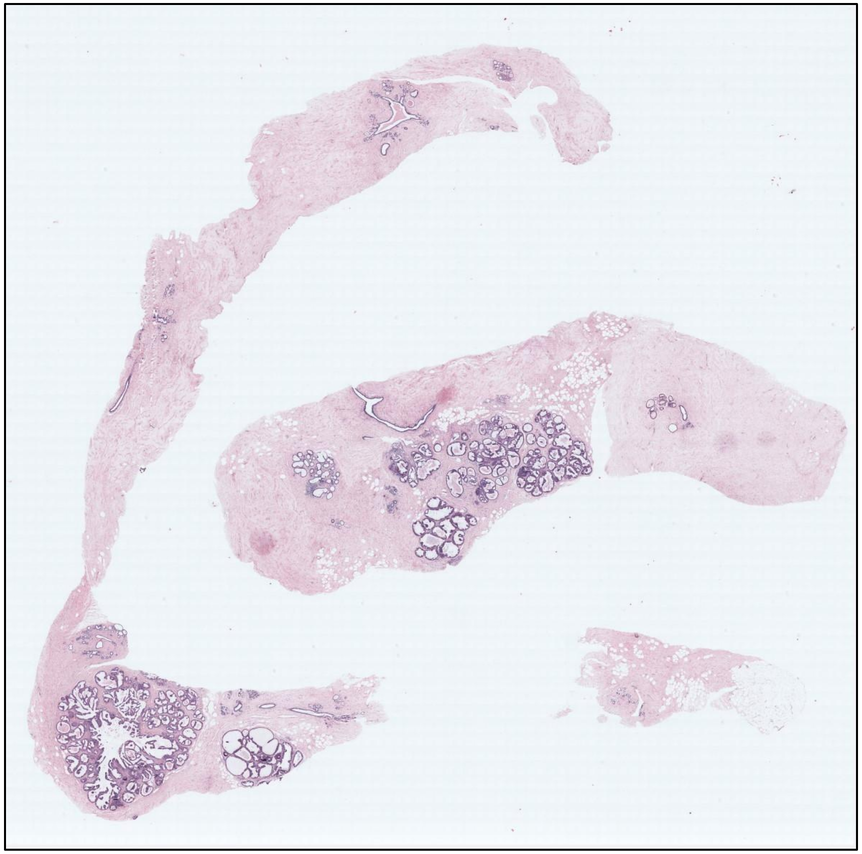


3D imaging: MRI, CT

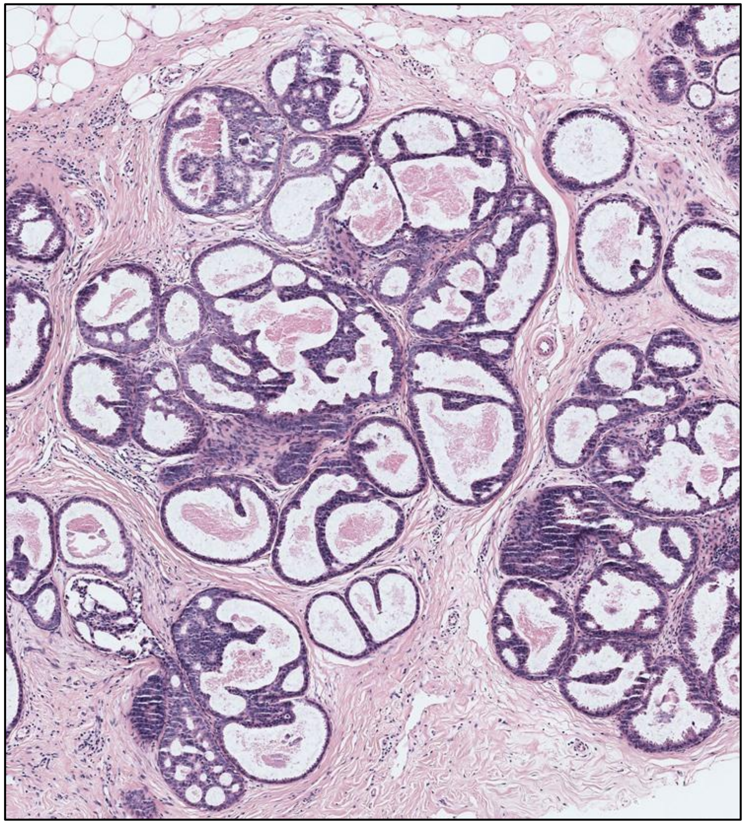
Image guided surgery
Grimson et al., MIT

Adapted from CSE 455, U of Washington

Medical image analysis



75568 x 74896 pixel
whole slide image



7440 x 8260 pixel
region of interest

Cancer detection and grading

Medical image analysis

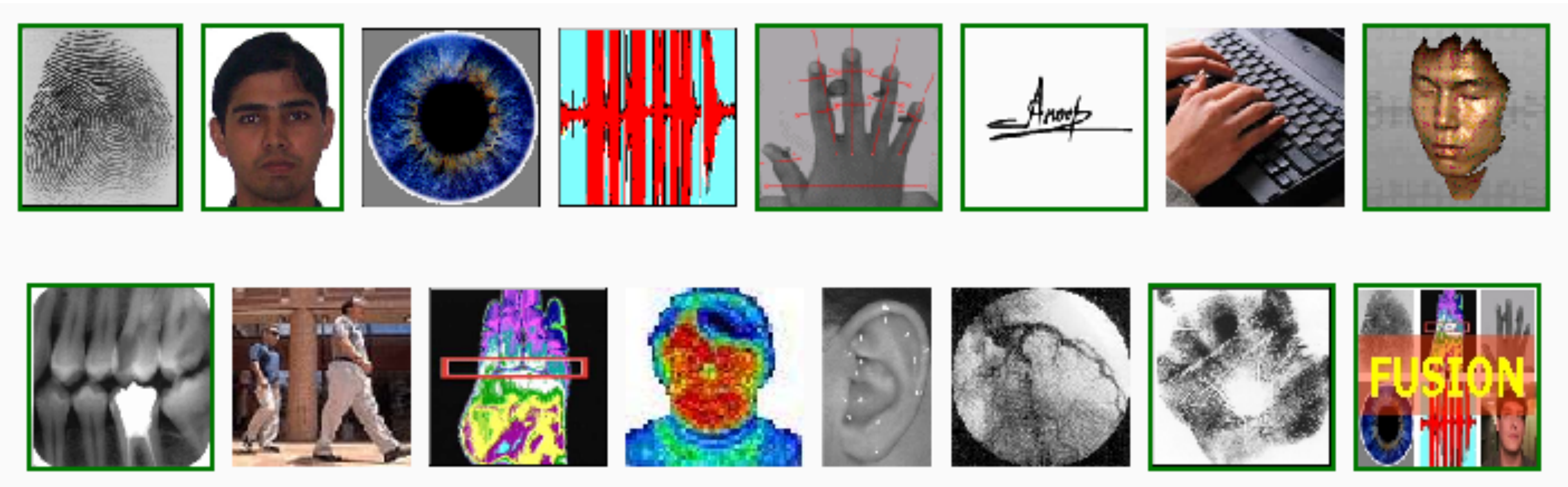
Slice of lung

Adapted from Linda Shapiro, U of Washington

Medical image analysis

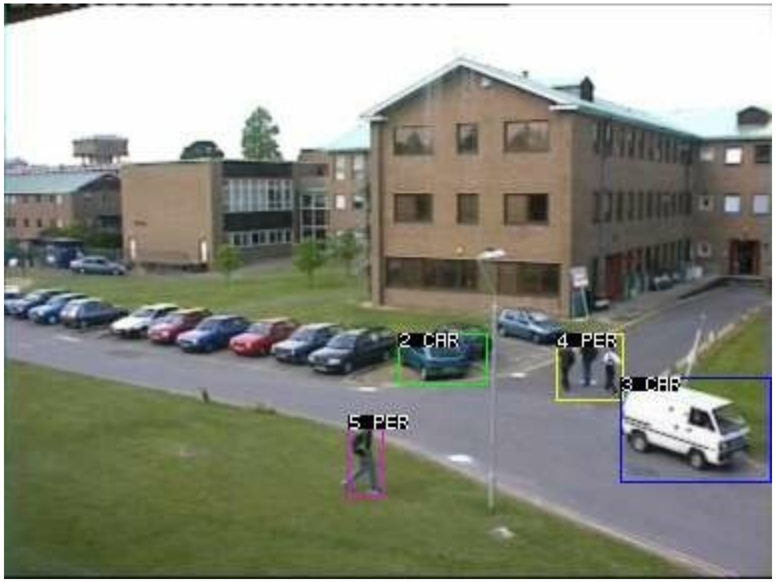
"Your x-ray showed a broken rib, but we fixed it with Photoshop."

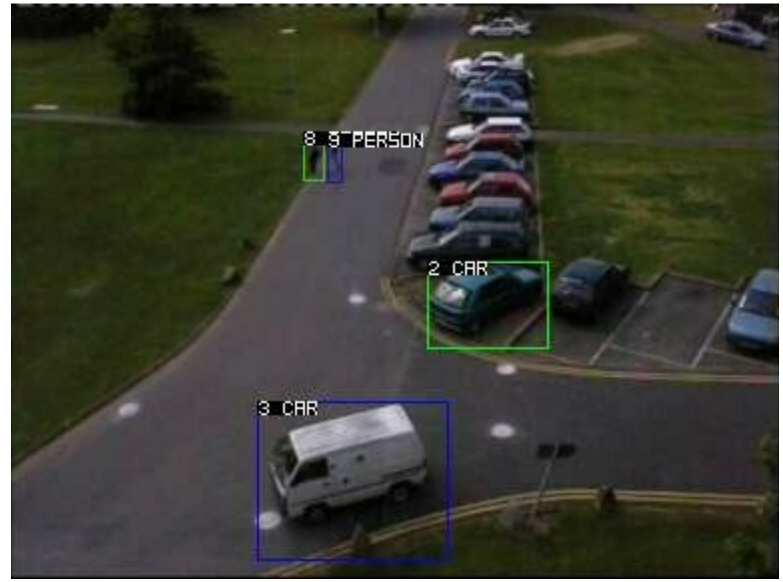
Biometrics

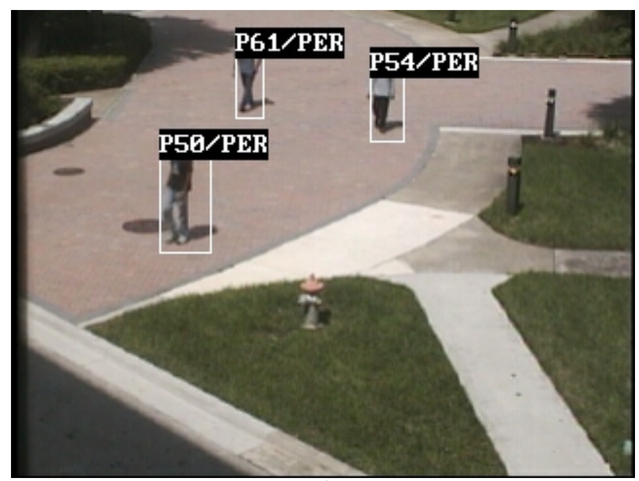


Adapted from Anil Jain, Michigan State

Surveillance and tracking



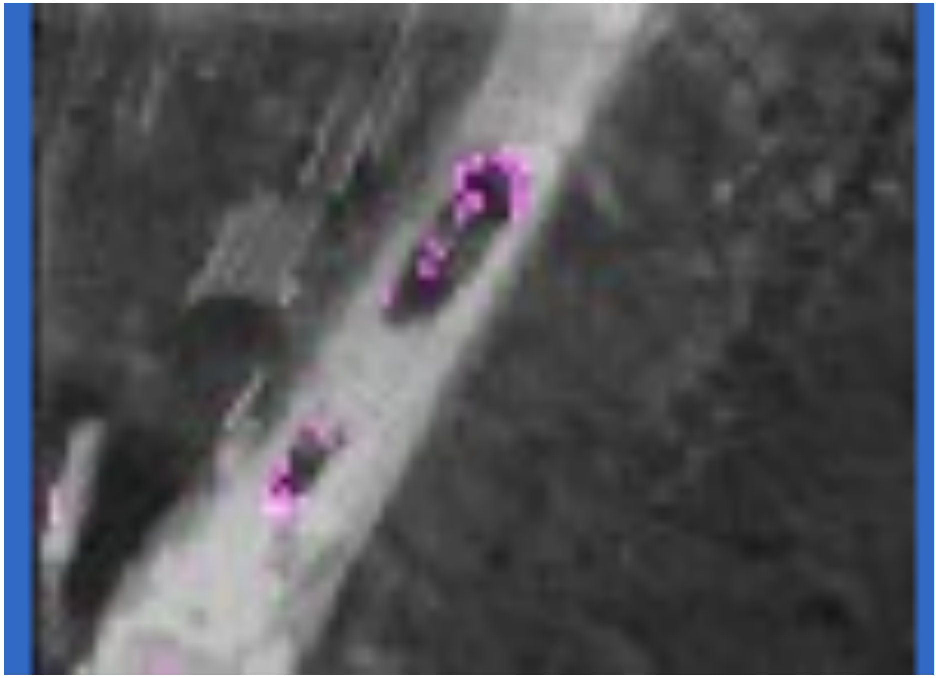




Surveillance and tracking

Adapted from Martial Hebert, CMU

Surveillance and tracking



Tracking in UAV videos

Adapted from Martial Hebert, CMU, and
Masaharu Kobashi, U of Washington

Vehicle and pedestrian protection

Lane departure warning, collision warning, traffic sign recognition, pedestrian recognition, blind spot warning

http://www.mobileye-vision.com

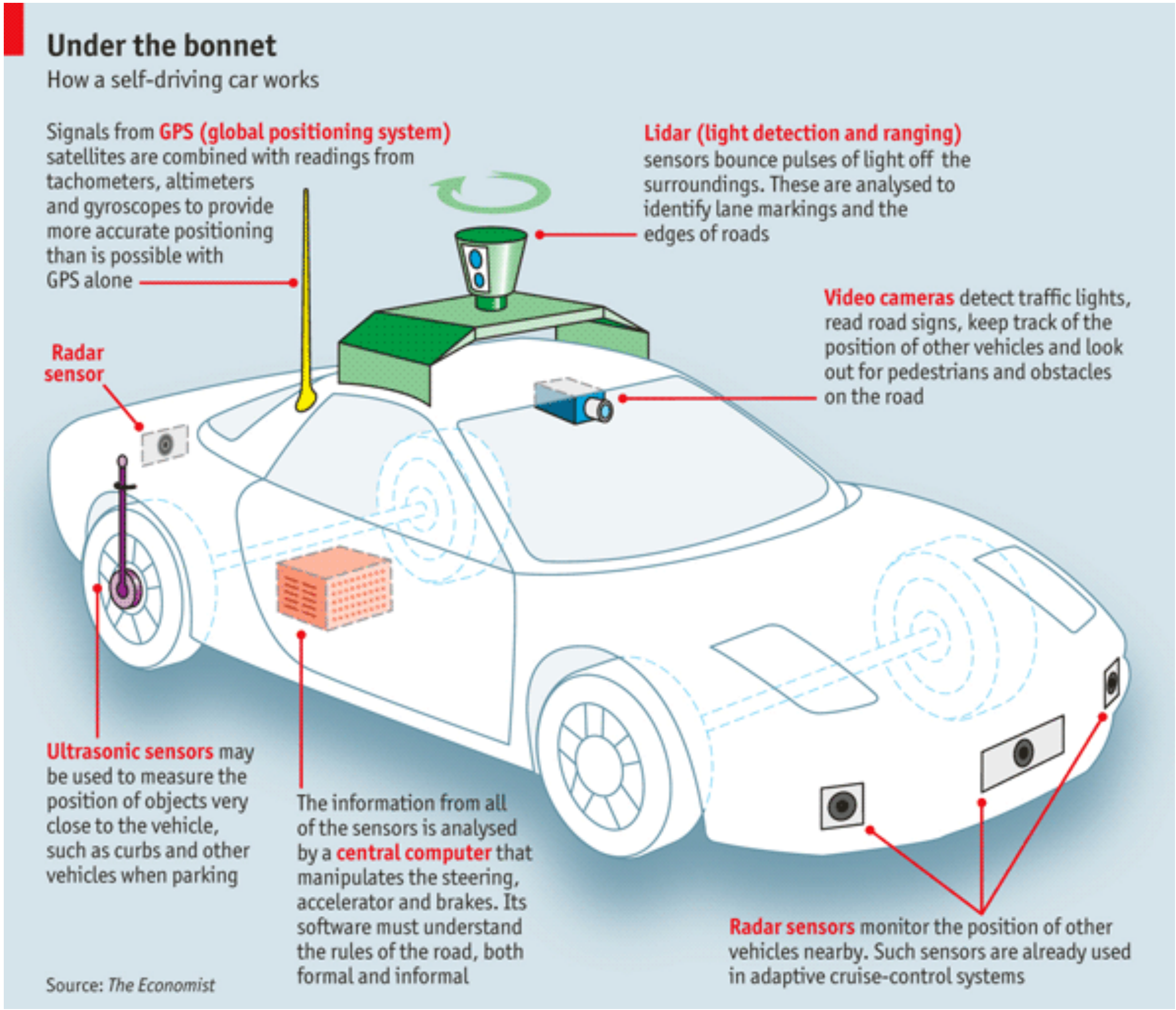
Smart cars

Adapted from CSE 455, U of Washington

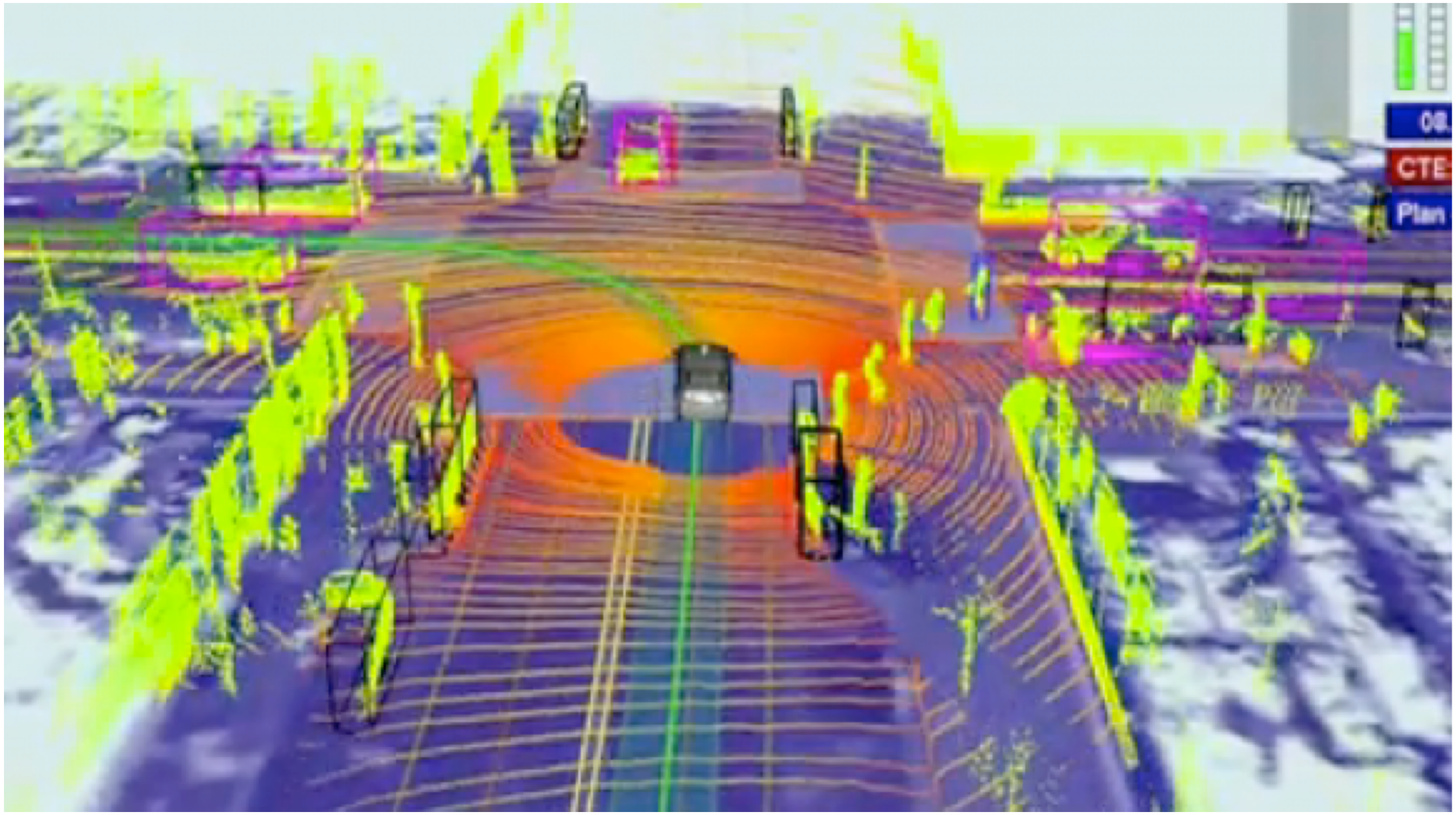
Self-driving cars

http://www.darpa.mil/grandchallenge/index.asp
http://en.wikipedia.org/wiki/DARPA_Grand_Challenge

Self-driving cars



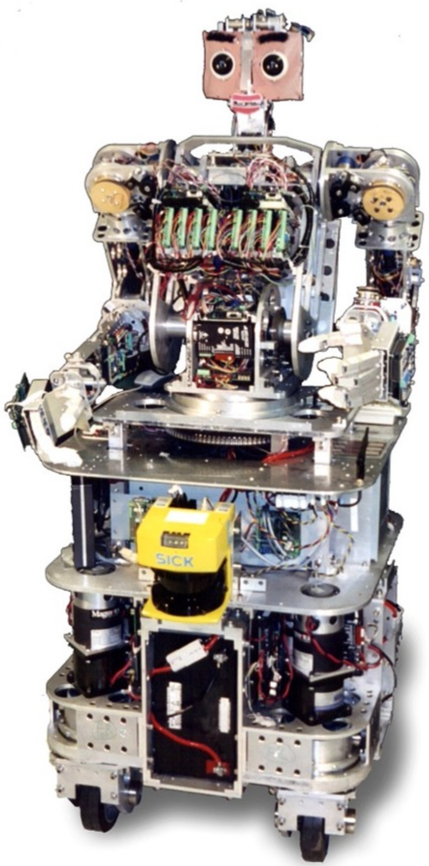
Self-driving cars



"Our self-driving cars have now traveled nearly 200,000 miles on public highways in California and Nevada, 100 percent safely. They have driven from San Francisco to Los Angeles and around Lake Tahoe, and have even descended crooked Lombard Street in San Francisco. They drive anywhere a car can legally drive."

- Sebastian Thrun, Google

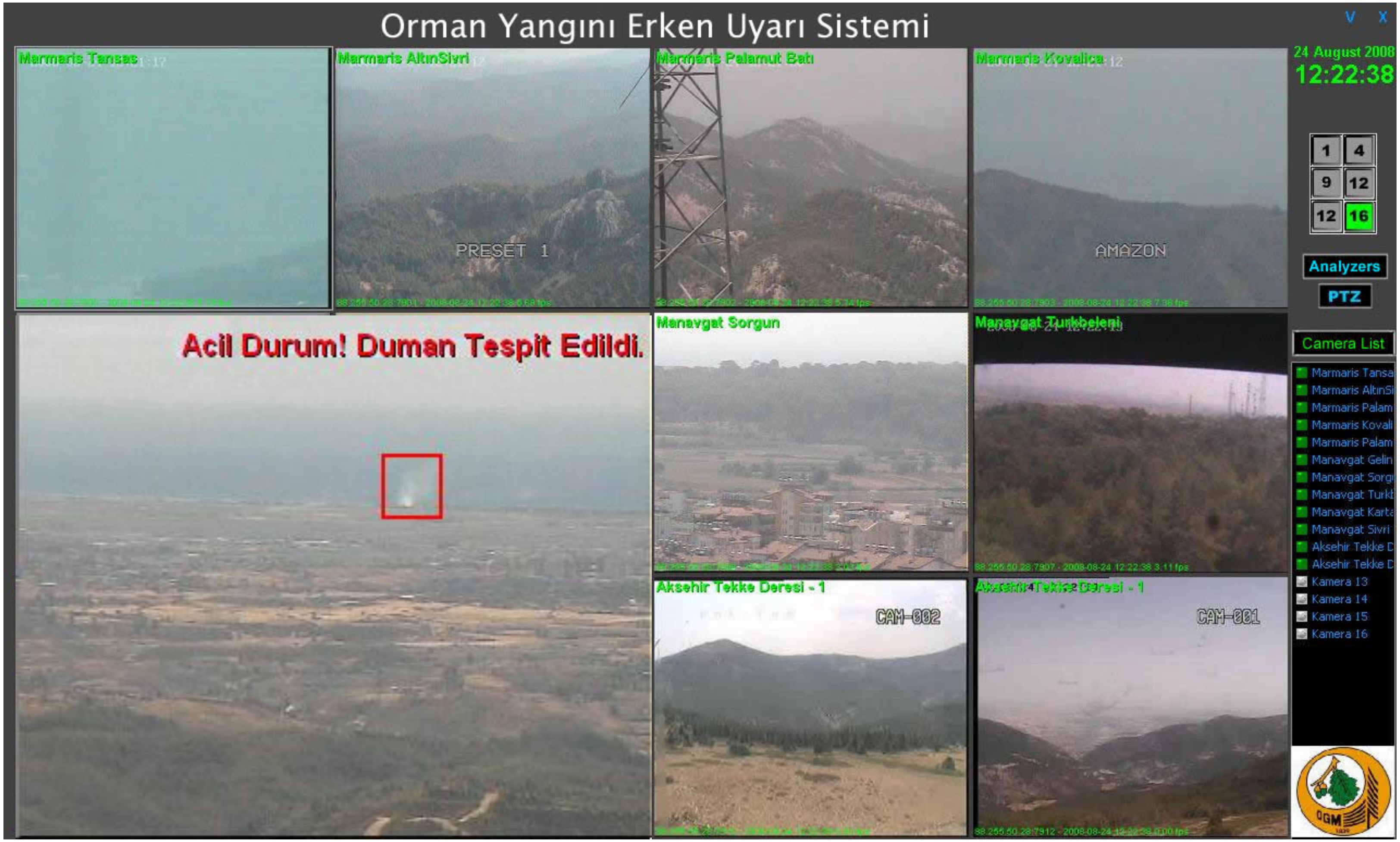
Autonomous navigation



Michigan State University

General Dynamics Robotics Systems
http://www.gdrs.com

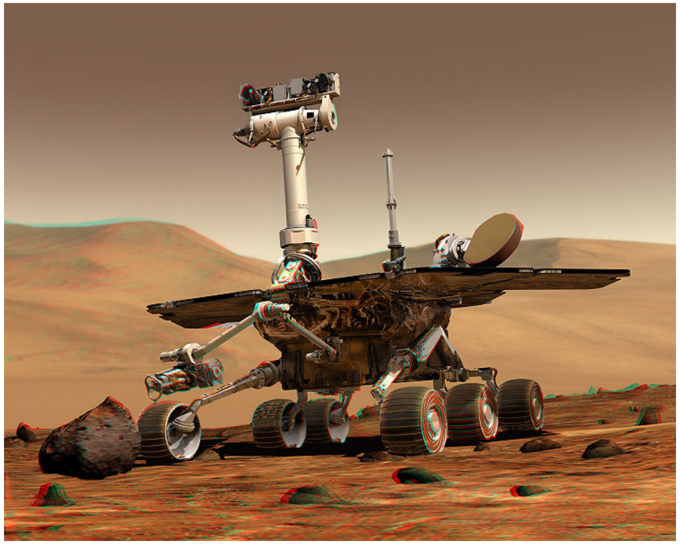
Forest fire monitoring system



Early warning of forest fires

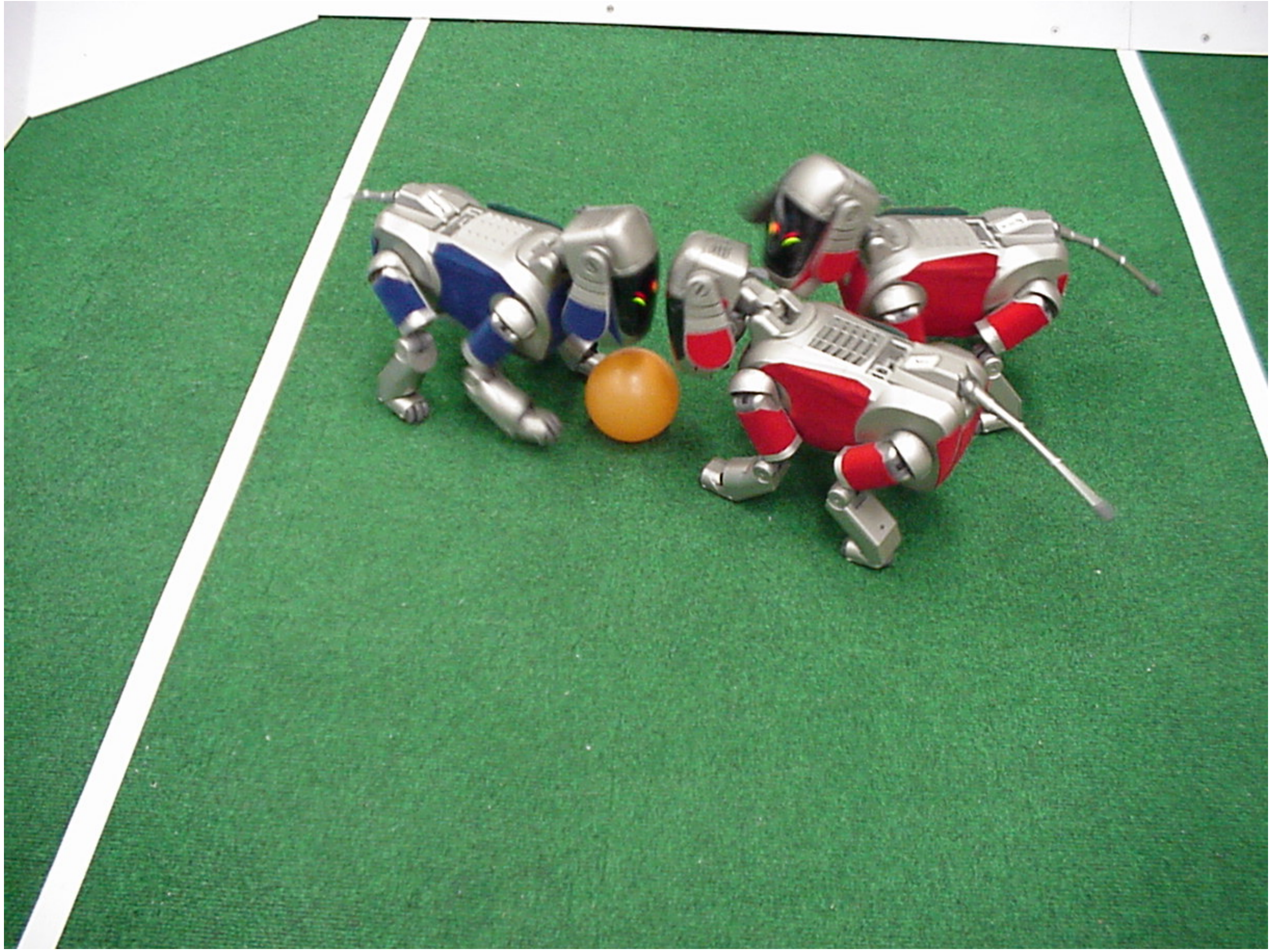
Adapted from Enis Cetin, Bilkent University

Robotics



Adapted from CSE 455, U of Washington

Robotics



Adapted from Steven Seitz, U of Washington

Face detection

Adapted from CSE 455, U of Washington

Face recognition

Adapted from CSE 455, U of Washington

Industrial automation

Automatic fruit sorting

Color Vision Systems
http://www.cvs.com.au

Industrial automation

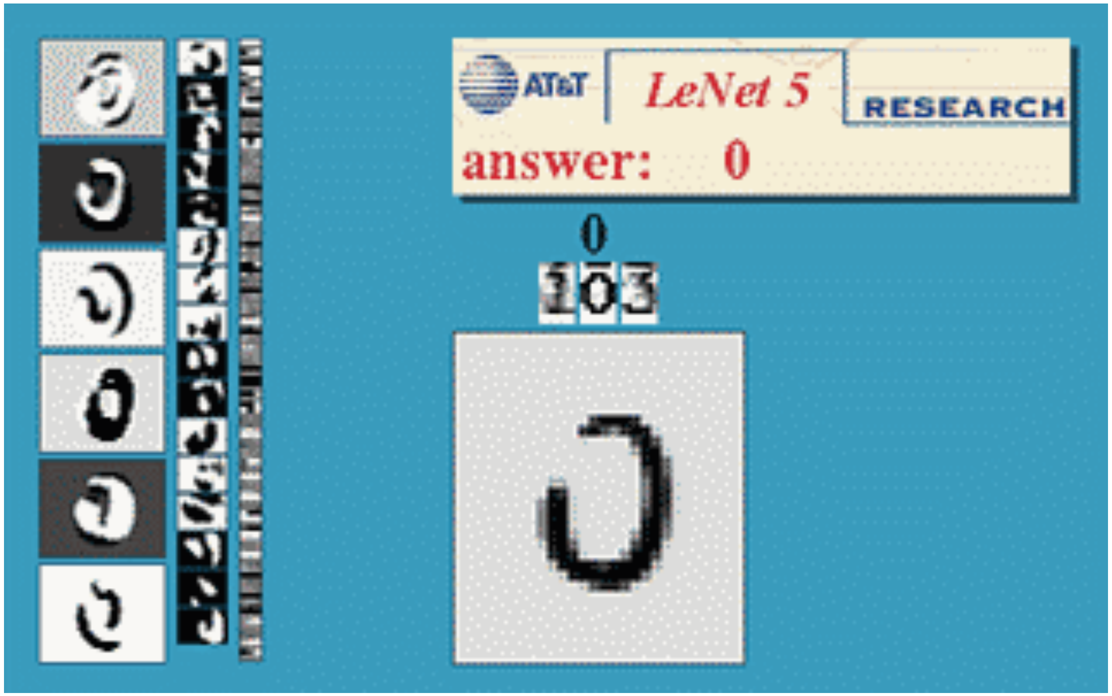
Industrial robotics;
bin picking

http://www.braintech.com

Postal service automation

General Dynamics Robotics Systems
http://www.gdrs.com

Optical character recognition

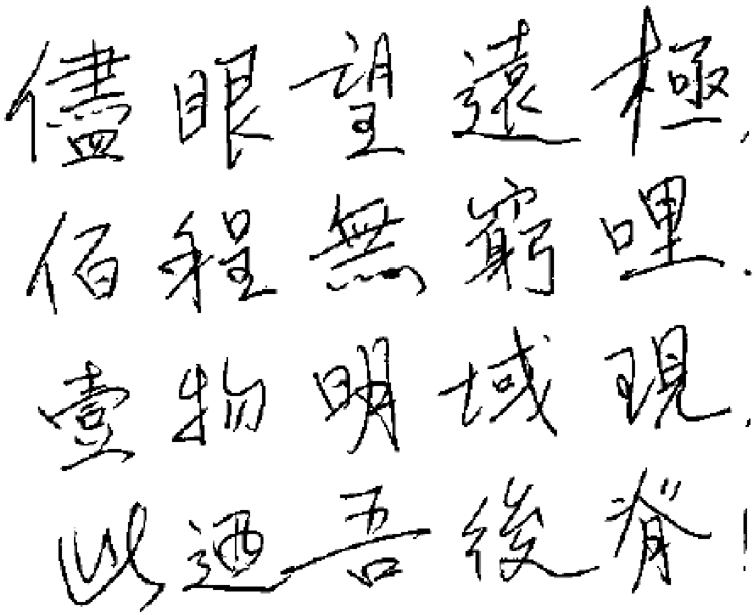


Digit recognition, AT&T labs
http://www.research.att.com/~yann

License place recognition

Adapted from Steven Seitz, U of Washington

Document analysis

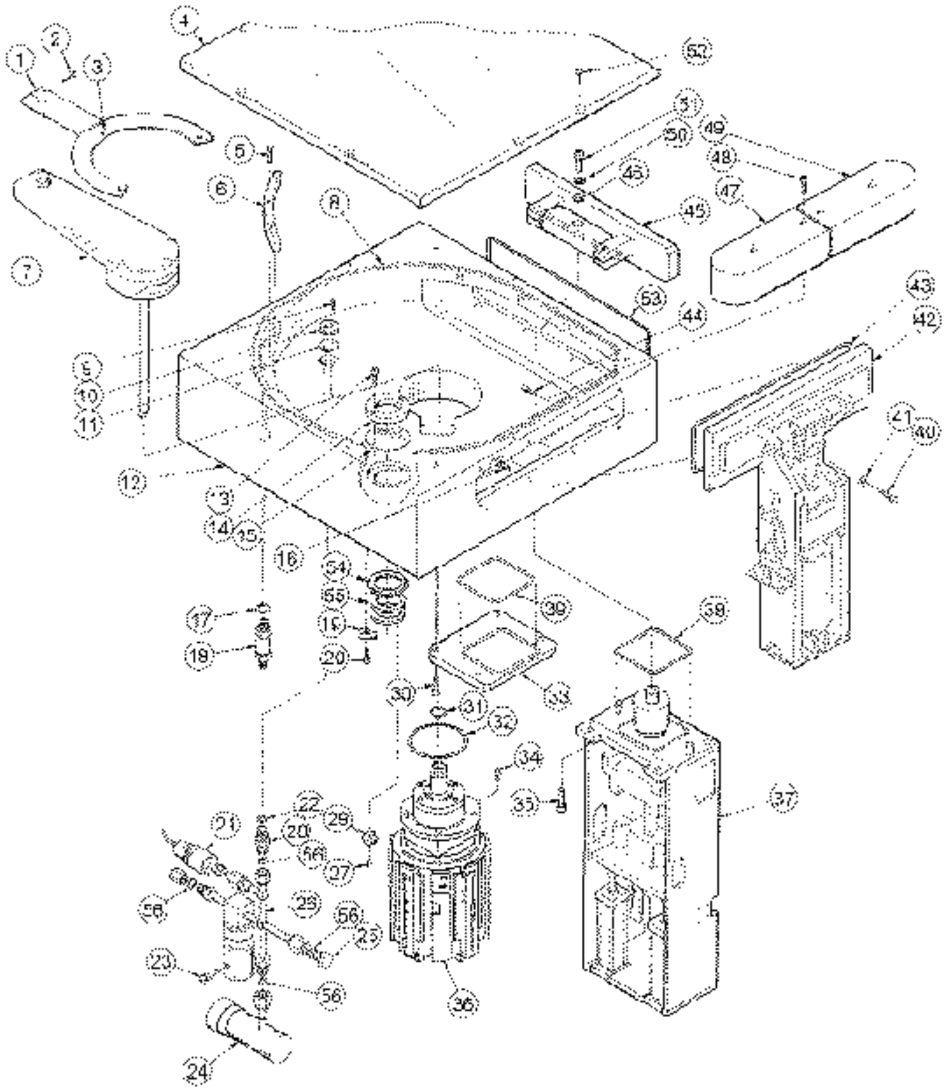


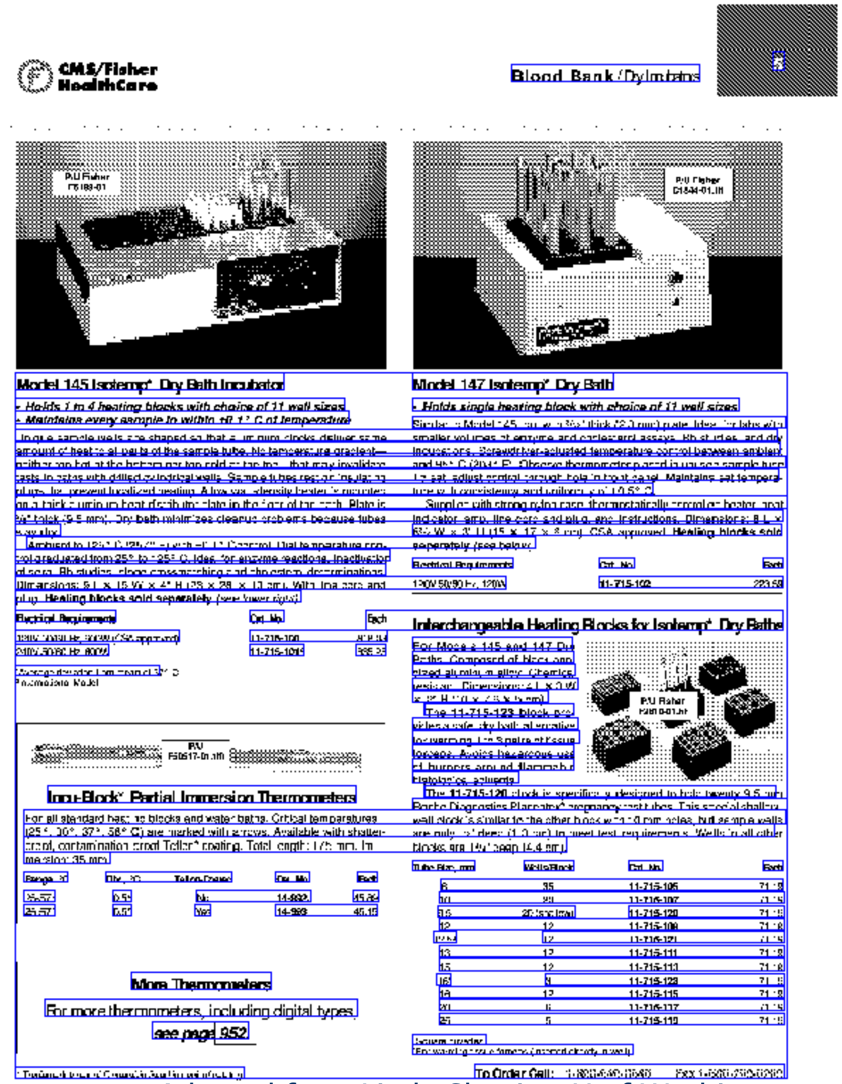
I looked as hard as I could see,
beyond 100 plus infinity
an object of bright intensity
– it was the back of me!

Figure 1.5: (Left) Chinese characters and (right) English equivalent. Is it possible that a machine could automatically translate one into the other? Chinese characters and poem courtesy of John Weng.

Adapted from Shapiro and Stockman

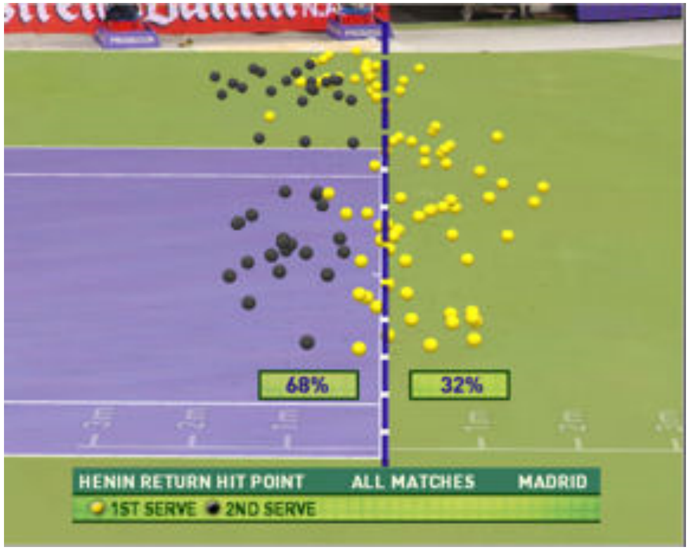
Document analysis

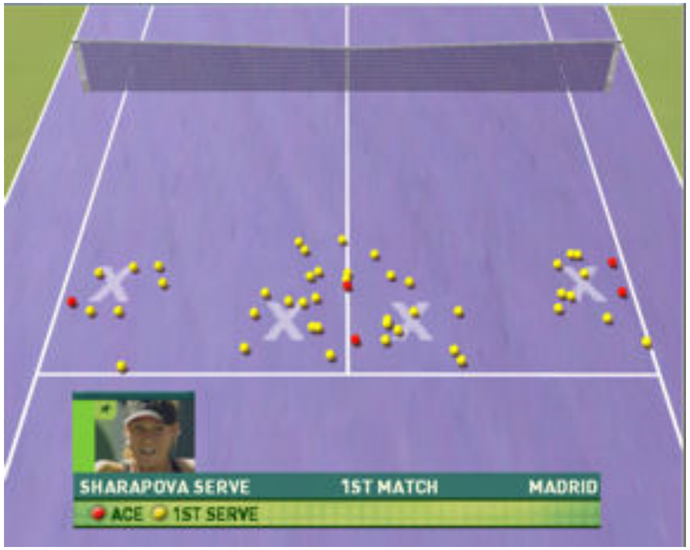


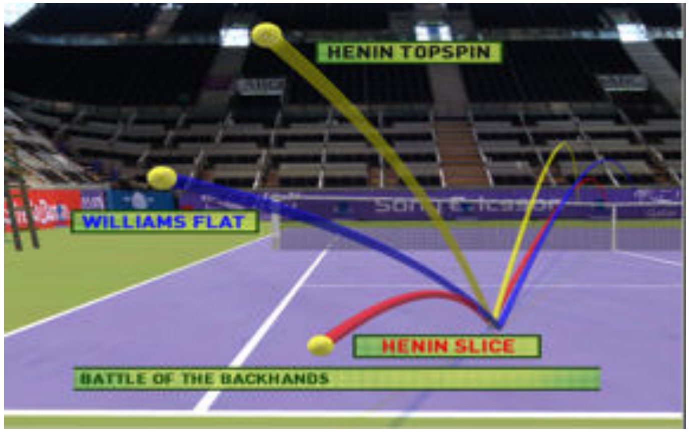


Adapted from Linda Shapiro, U of Washington

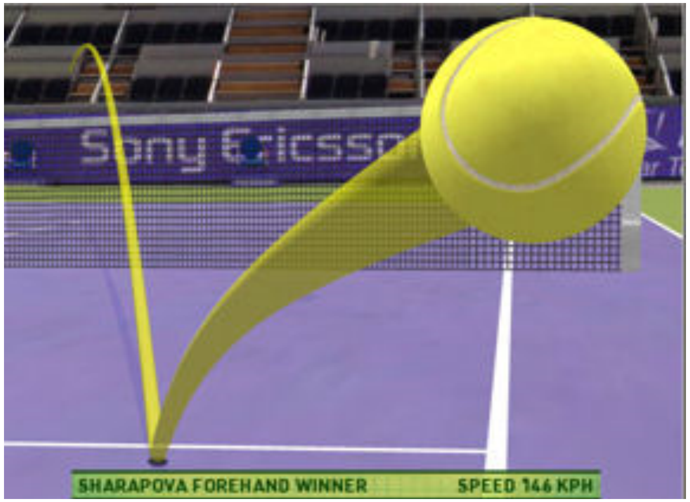
Sports video analysis







Tennis review system



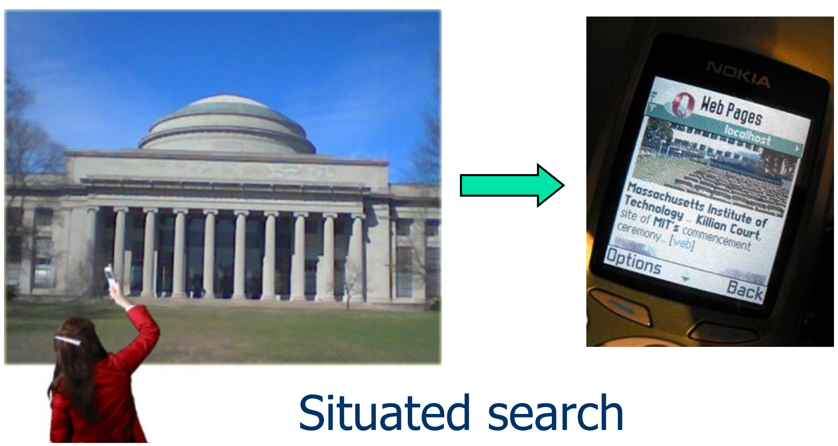
http://www.hawkeyeinnovations.co.uk

Scene classification

Object recognition

Lincoln, Microsoft Research

kooaba



Situated search
Yeh et al., MIT

Google Goggles
Bing Vision

Image captioning

"woman is holding bunch of bananas."

"black cat is sitting on top of suitcase."

Demo: http://gradcam.cloudcv.org/captioning

http://cs.stanford.edu/people/karpathy/deepimagesent/

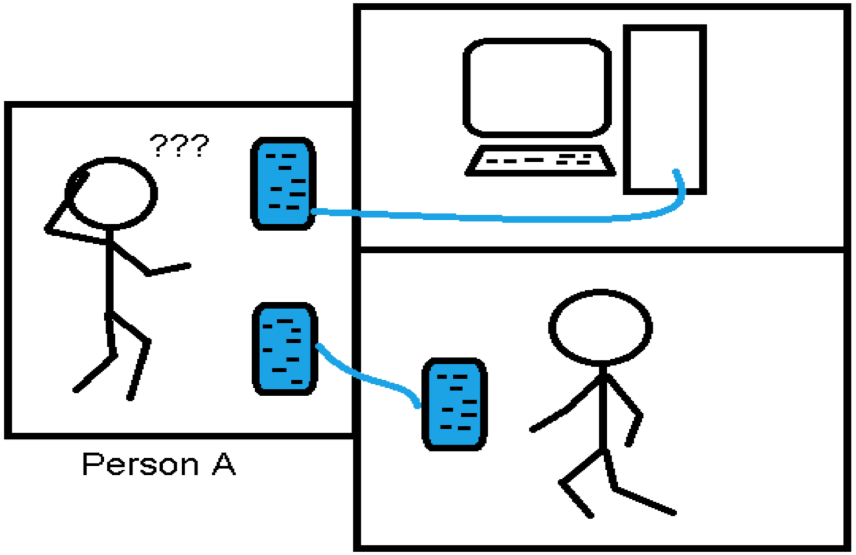
Visual QA

Q: How many slices of pizza are there?

x

y

A: 6



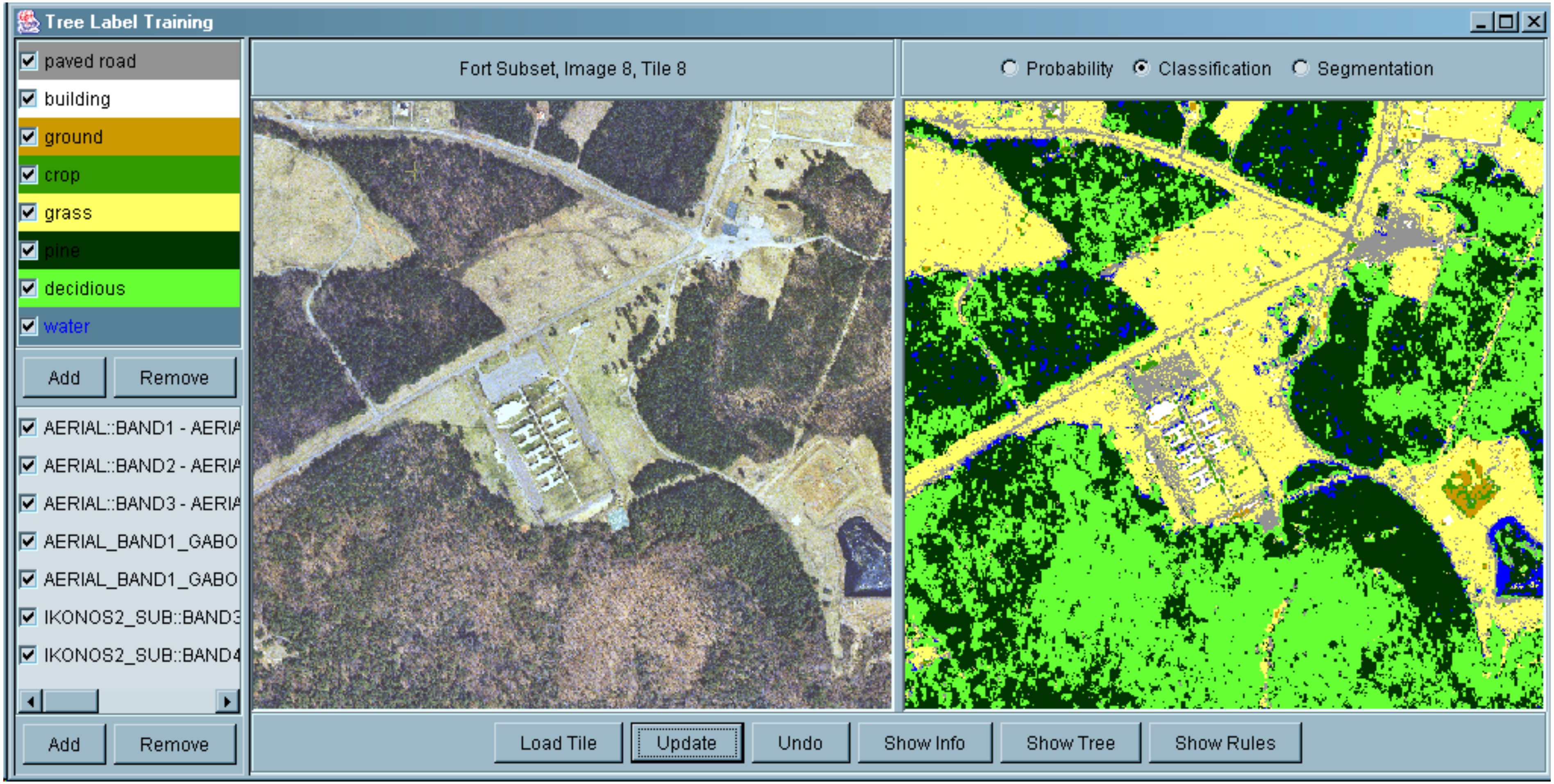
Demo: http://gradcam.cloudcv.org/vqa

Slide by Dhruv Batra

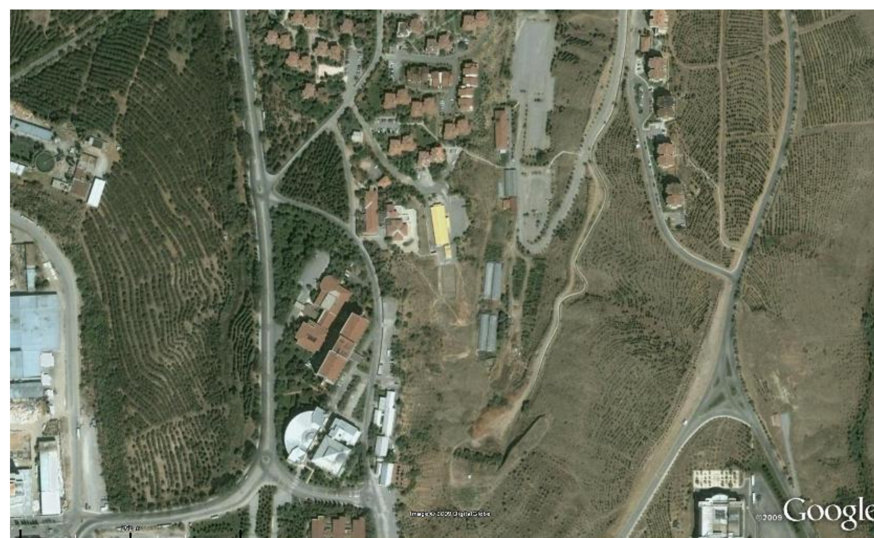
Augmented reality

Adapted from CSE 455, U of Washington

Land cover classification



Land cover classification

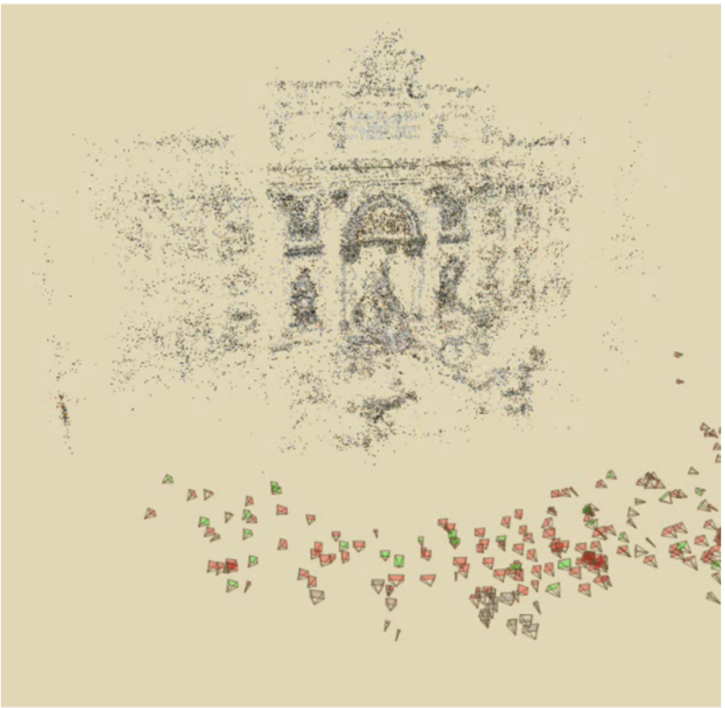


Object recognition

Recognition of buildings and building groups

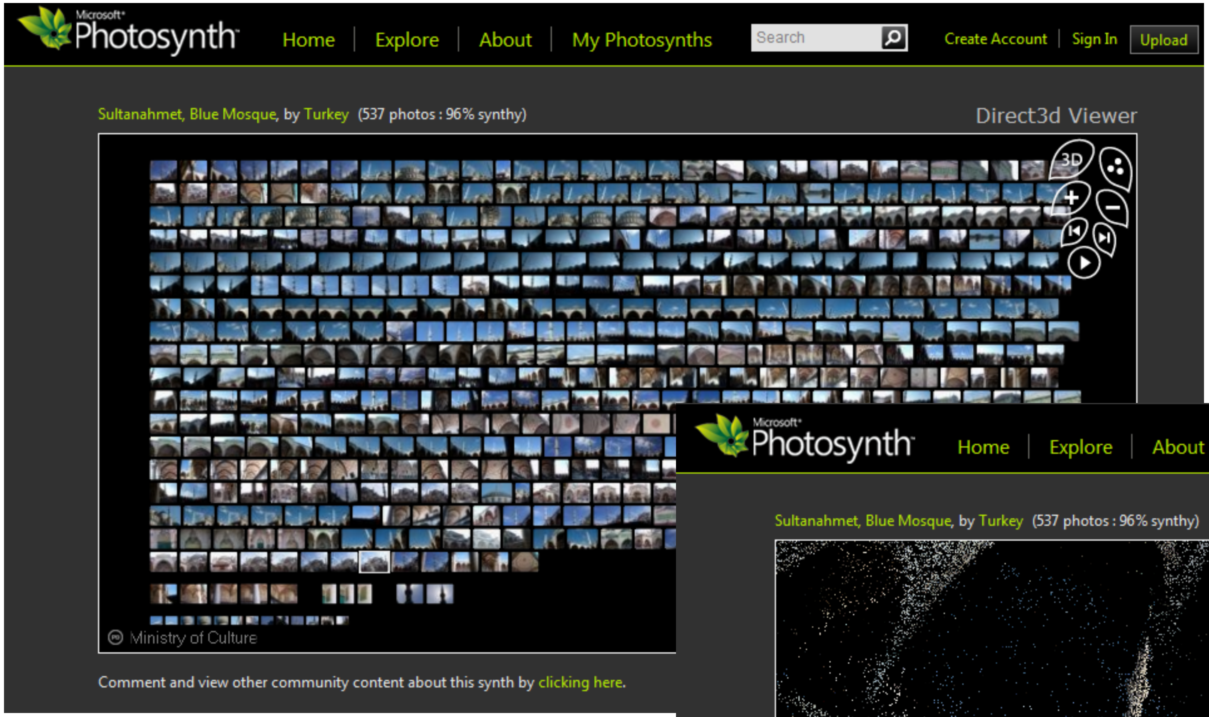
Photo tourism: exploring photo collections



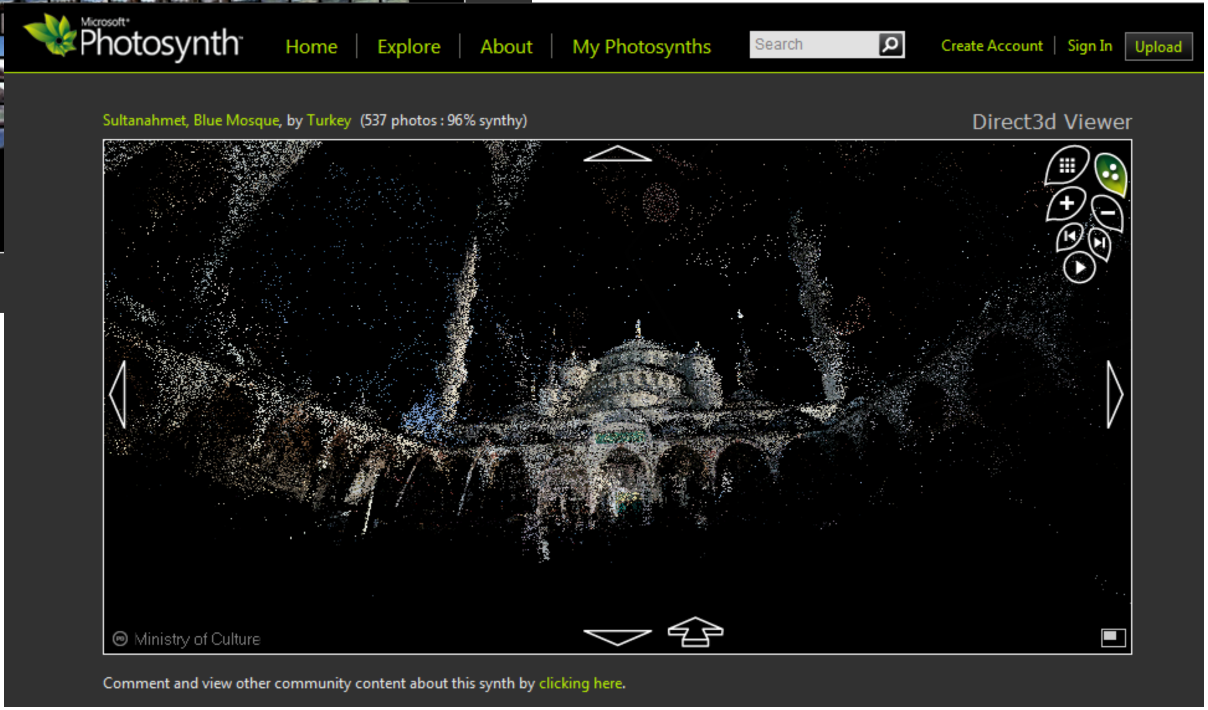
Building 3D scene models from individual photos

Adapted from Steven Seitz, U of Washington

Photosynth

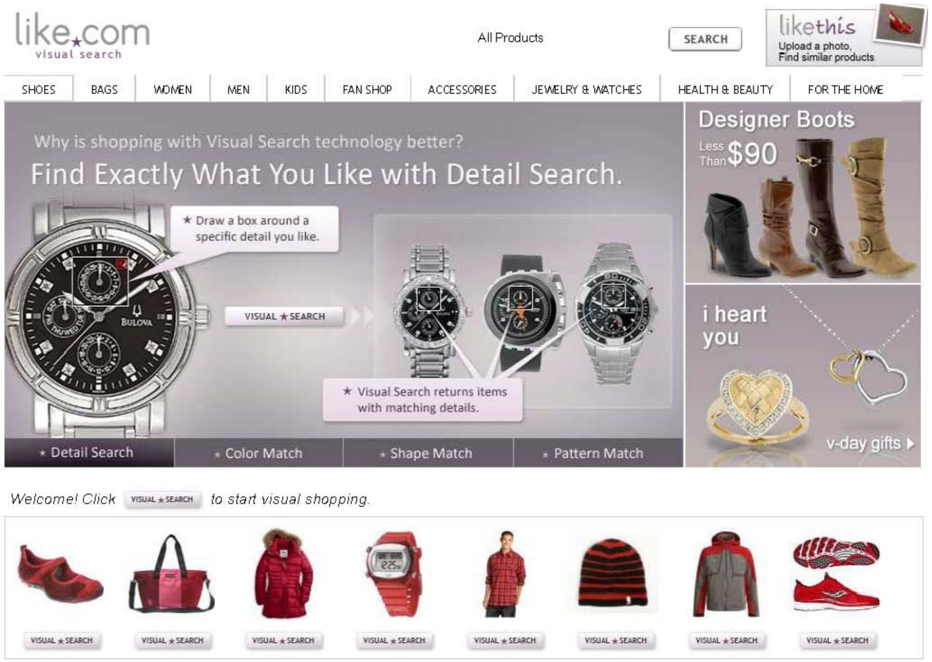


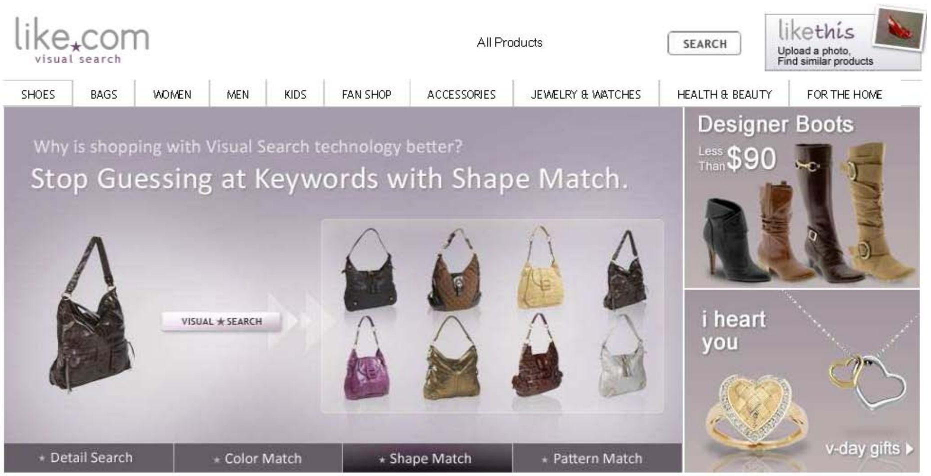
Microsoft
Photosynth
Home
Explore
About
My Photosynths
Search
Create Account
Sign In
Upload
Sultanahmet, Blue Mosque, by Turkey (537 photos : 96% synthy)
Direct3d Viewer
3D
Ministry of Culture
Comment and view other community content about this synth by clicking here.

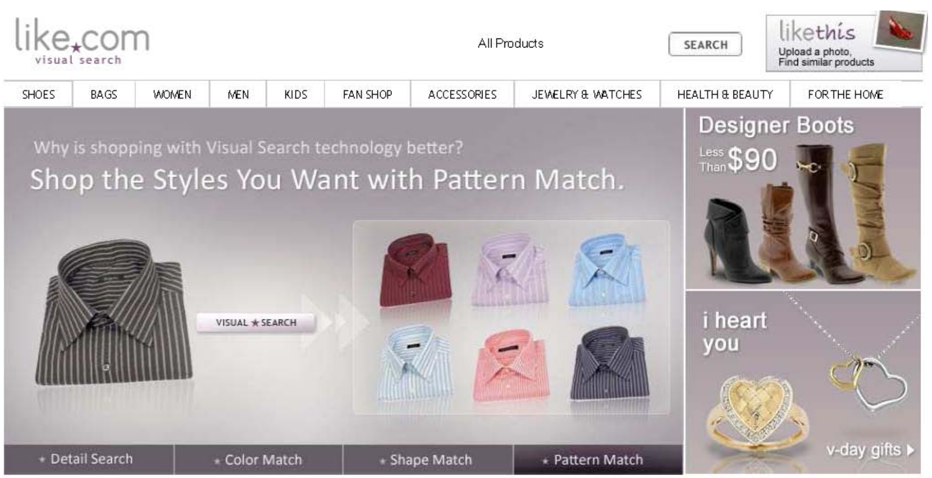


Microsoft
Photosynth
Home
Explore
About
My Photosynths
Search
Create Account
Sign In
Upload
Sultanahmet, Blue Mosque, by Turkey (537 photos : 96% synthy)
Direct3d Viewer
Ministry of Culture
Comment and view other community content about this synth by clicking here.

Content-based retrieval



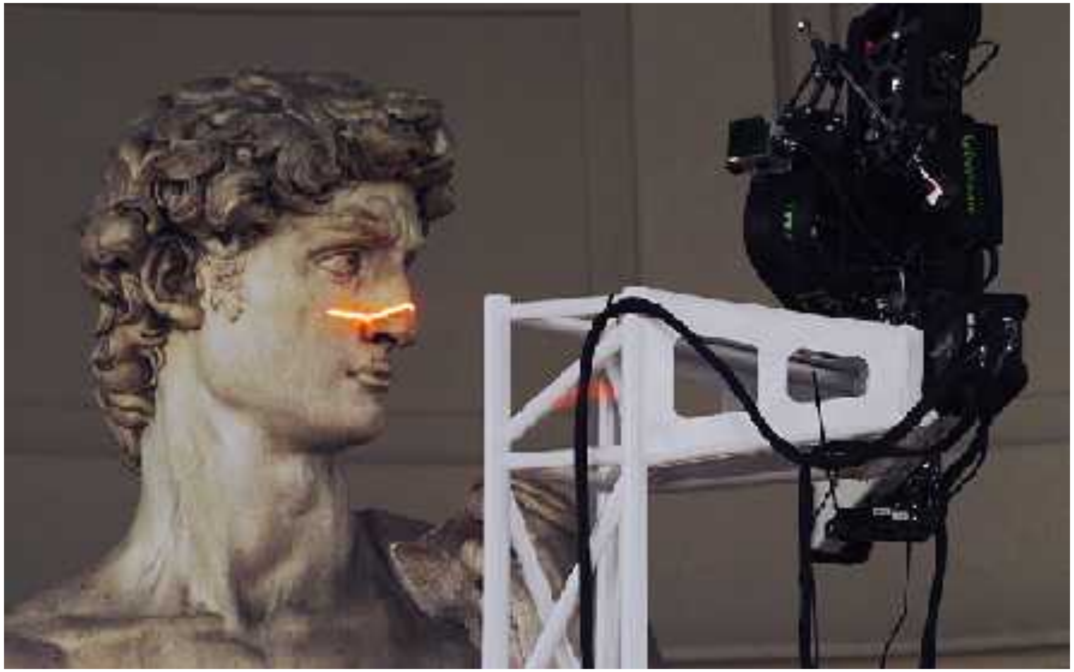




Online shopping catalog search

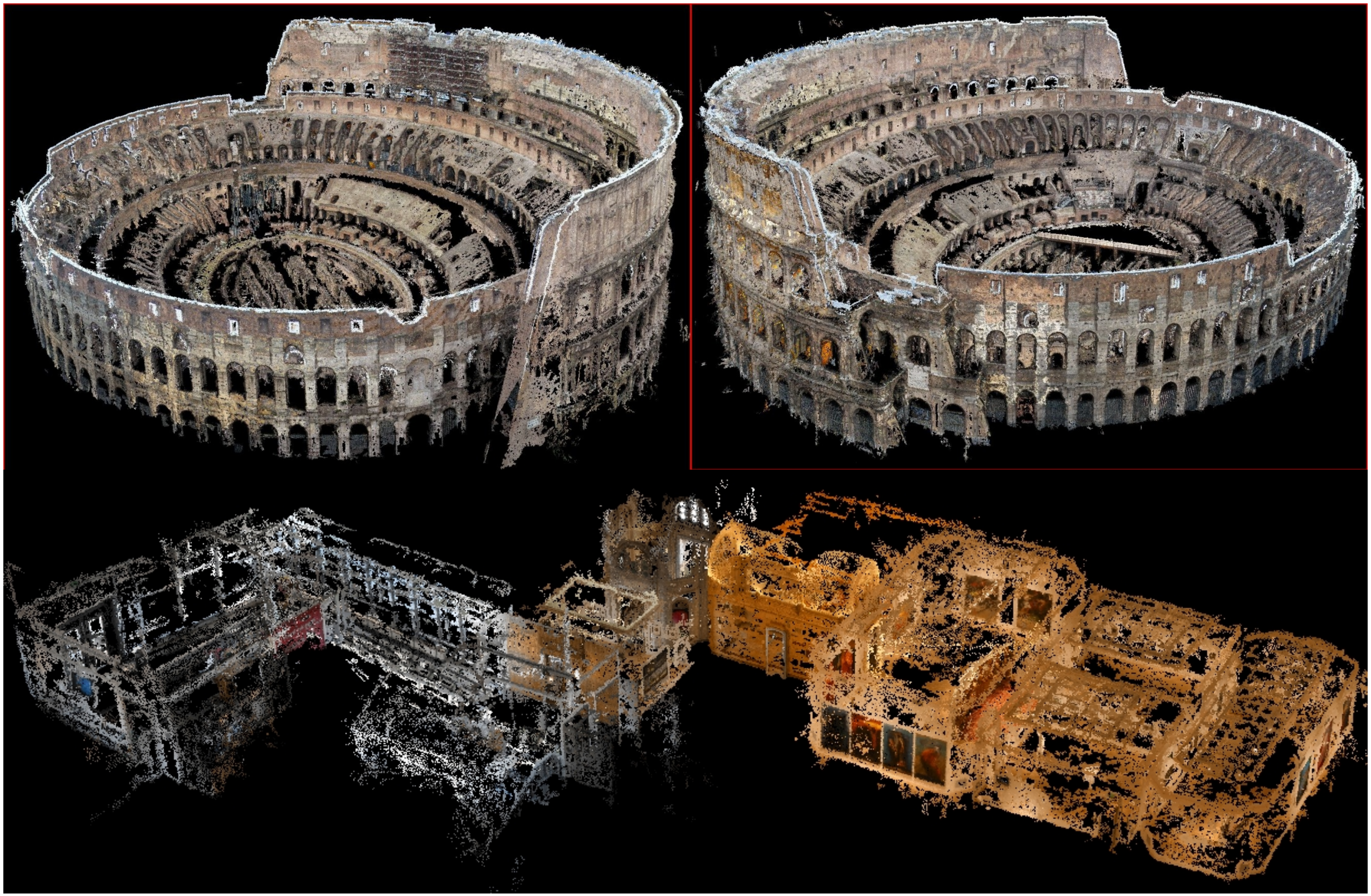
http://www.like.com

3D scanning and reconstruction

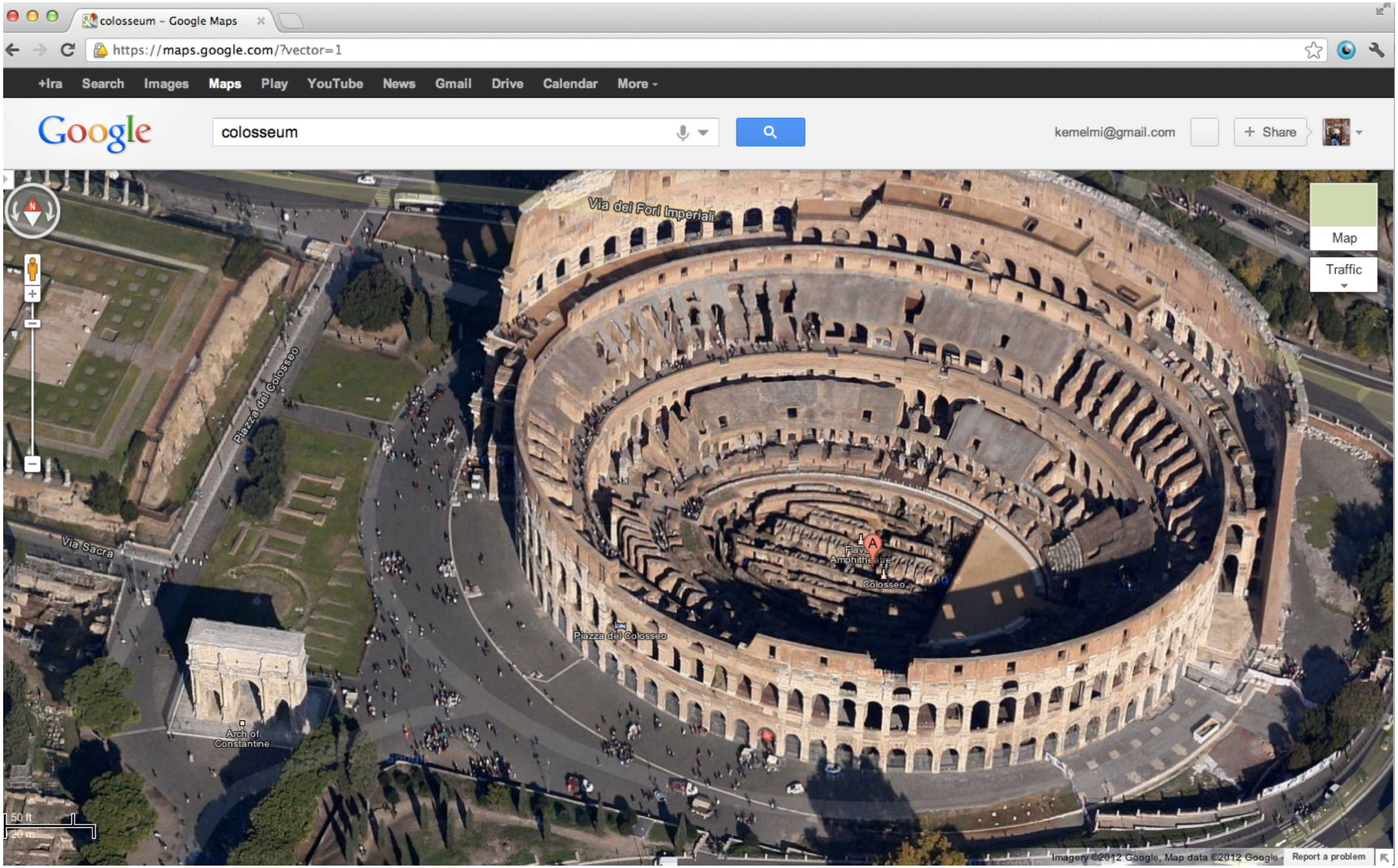


Adapted from Linda Shapiro, U of Washington

3D modeling



Earth viewers



Motion capture

Adapted from Linda Shapiro, U of Washington

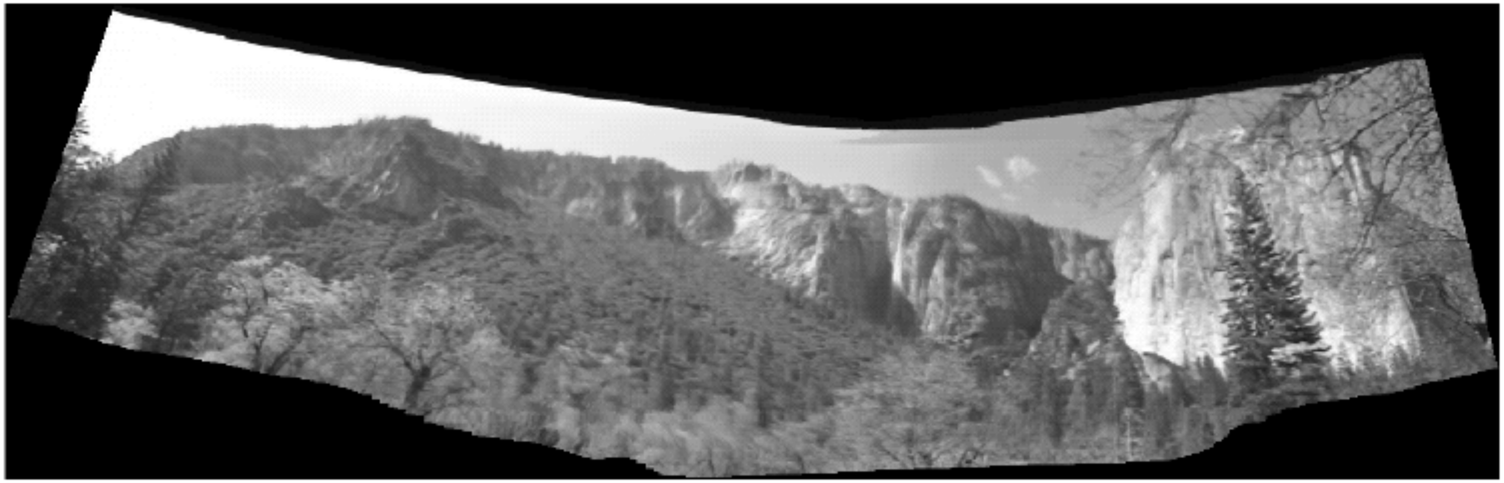
Visual effects

Adapted from CSE 455, U of Washington

Motion capture

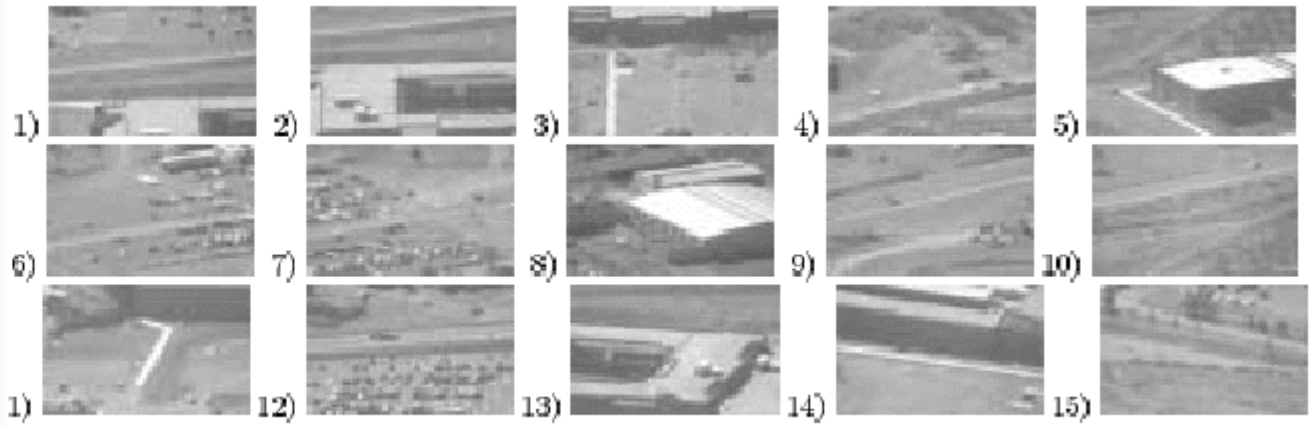
Microsoft's XBox Kinect

Mozaic



Adapted from David Forsyth, UC Berkeley

Mozaic



Adapted from David Forsyth, UC Berkeley

Even More Applications

Counting in Extremely Dense Crowd Images

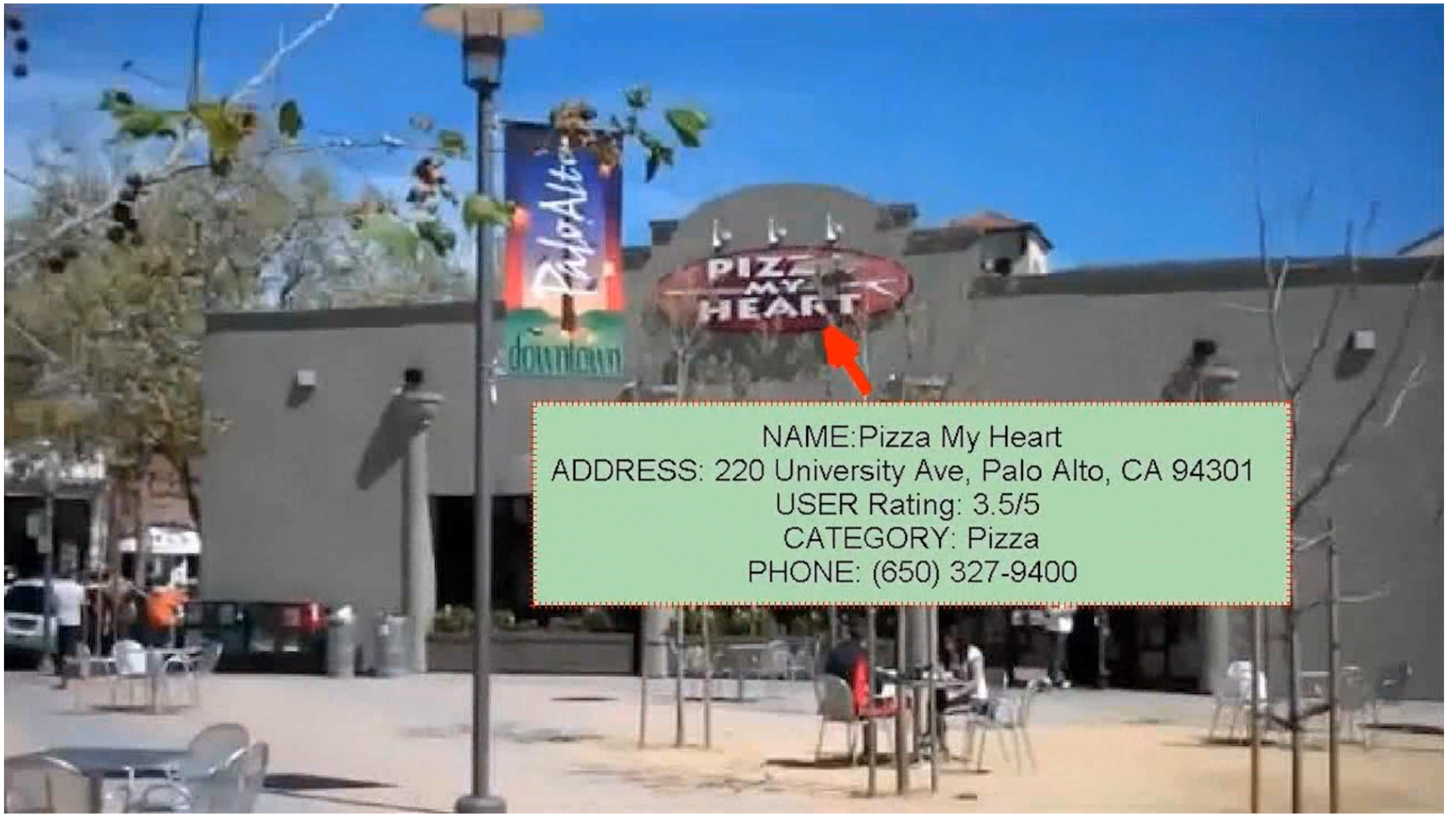
Ground truth=634 Proposed Method by Idrees and Shah=640

Ground truth=1428

Proposed Method=1468

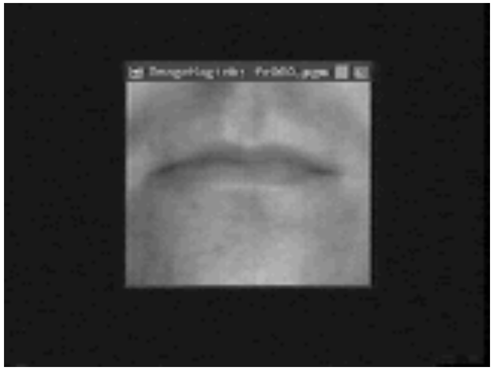
Ground truth=2319 Proposed Method=2496

Visual Business Recognition



So many other applications…

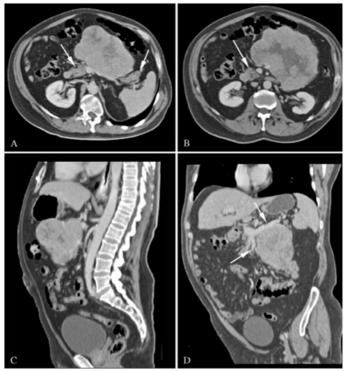
- Face identification,
- Lip reading,

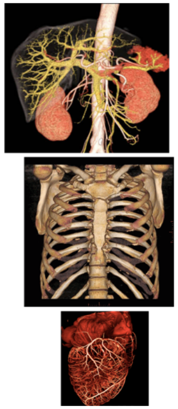


- Emotion detection,

Surprised Happy

- Medical Computer Vision / Analysis,

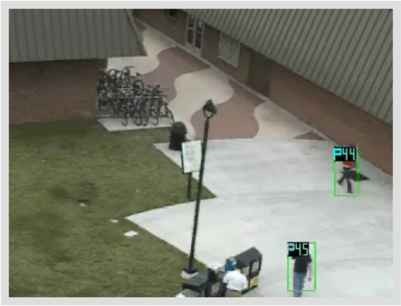


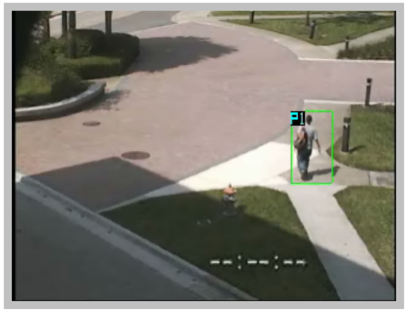


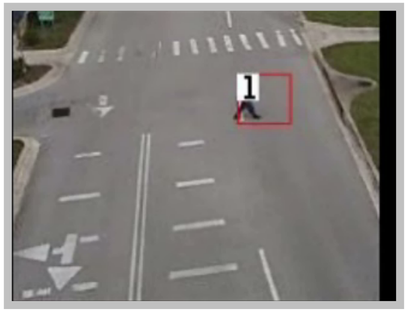
Video Analysis

Video: image sequence (over time)

Object detection → Object tracking → Event, Action or Activity Recognition

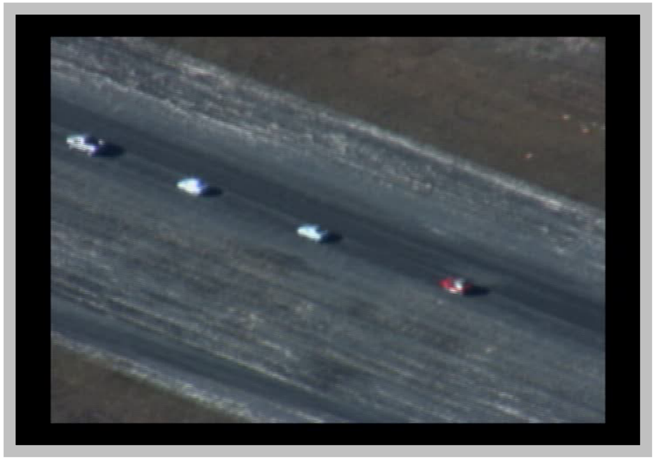






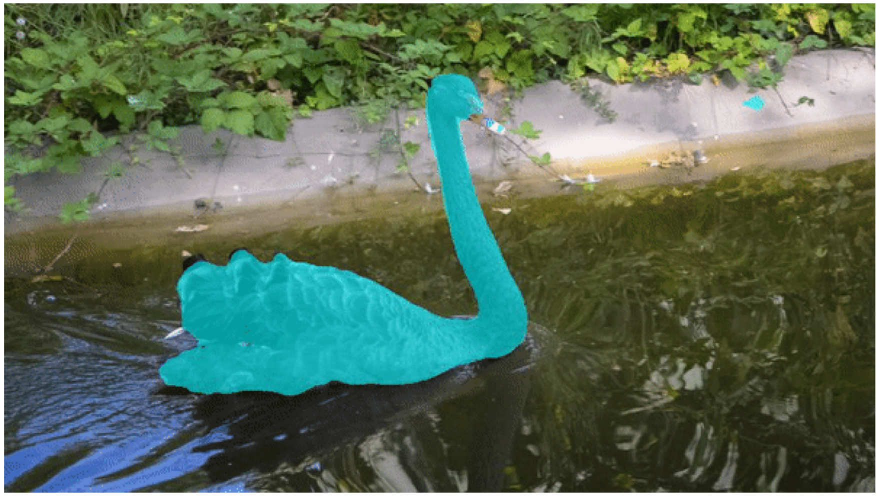
Object detection in videos

(Object) Tracking



- Deals with data association problem
  - (Correspondence problem)
- Correlates objects over time

Actor – action segmentation



R. Hou, C. Chen, and M. Shah, 2017

Acknowledgement

- Some slides are taken / adopted from:
  - Dr. Selim Aksoy
  - Dr. Antonio Torralba