# CAP 5415: Computer Vision

Dr. Sedat Ozer

sedat@crcv.ucf.edu

# 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,
  - 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. Examples are:
  - Robotics,
  - Medical field,
  - Surveillance,
  - Entertainment, etc.

# Logistics:

**Exams:**

- See Syllabus for details.
- 1 Midterm

**Hands-on Experience:**

- Final Project
- Programming assignments (PA)

**Presentation Experience:**

- Research paper presentation
- Final project presentation

**Office hours:**

- Tuesday/Thursday 4:30pm-5:30pm
  - (after the class)
- Office: HEC 214

**Notes:**

- Follow the slides

Special thanks to:

- Dr. Mubarak Shah
- Dr. Ulas Bagci
- Dr. Yogesh Singh Rawat

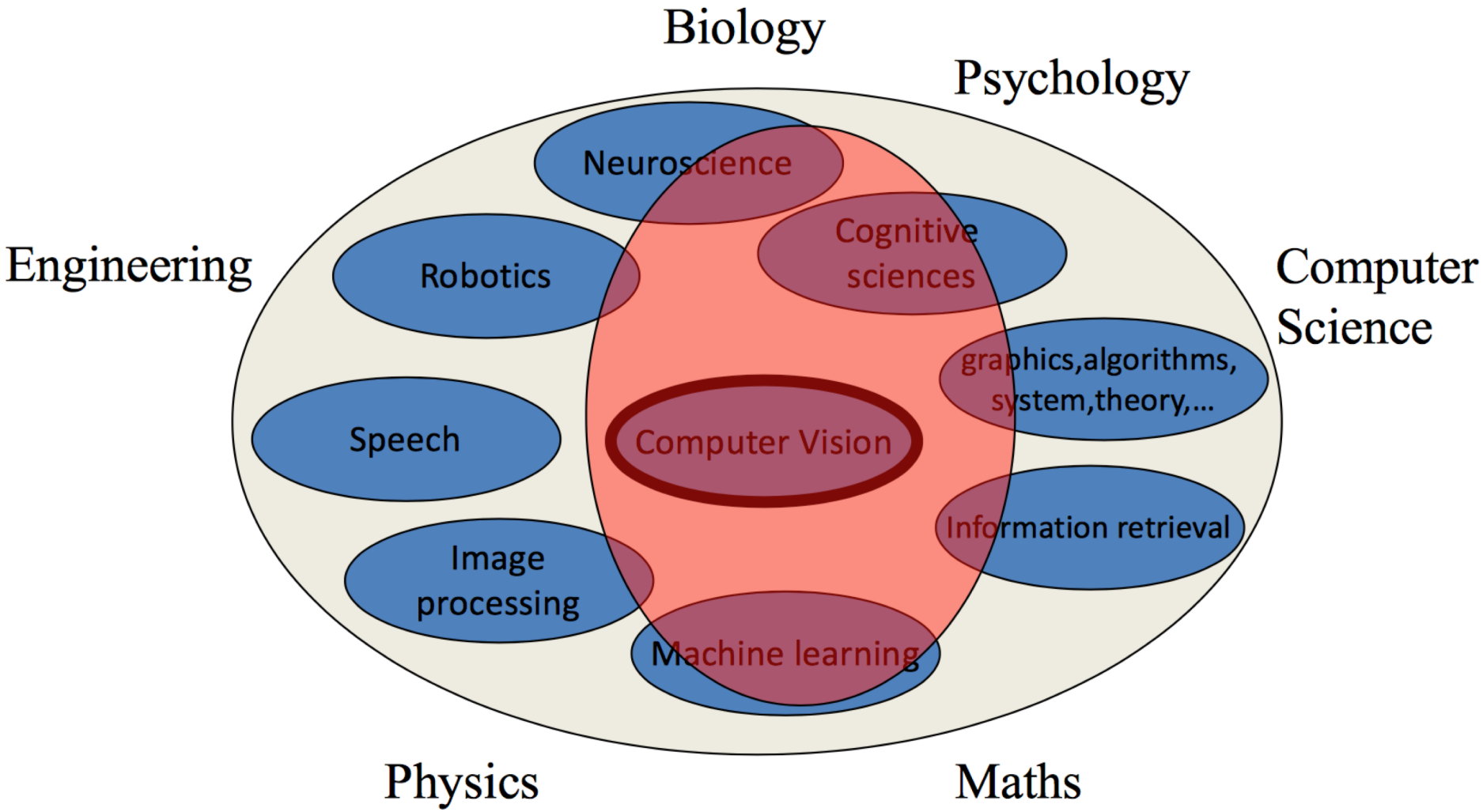
**Teaching Assistants:**

- Google
- Bing
- TBA

- No text book is required!
  - There are many online resources available.
- Attend the class (mandatory), follow the lecture notes and in-class material.
- Python is the required language for PA.

# Computer Vision

## What is it related to?



# Computer vision: introduction

# Visual recognition: Classification

**Problem:**
Are there any human in this image?

**Answer: Yes** / No

**Problem:**
Are there any cars in this image?

**Answer:** Yes / **No**

⋮

# Visual recognition: Detection (localization)

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

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

# Visual recognition: 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

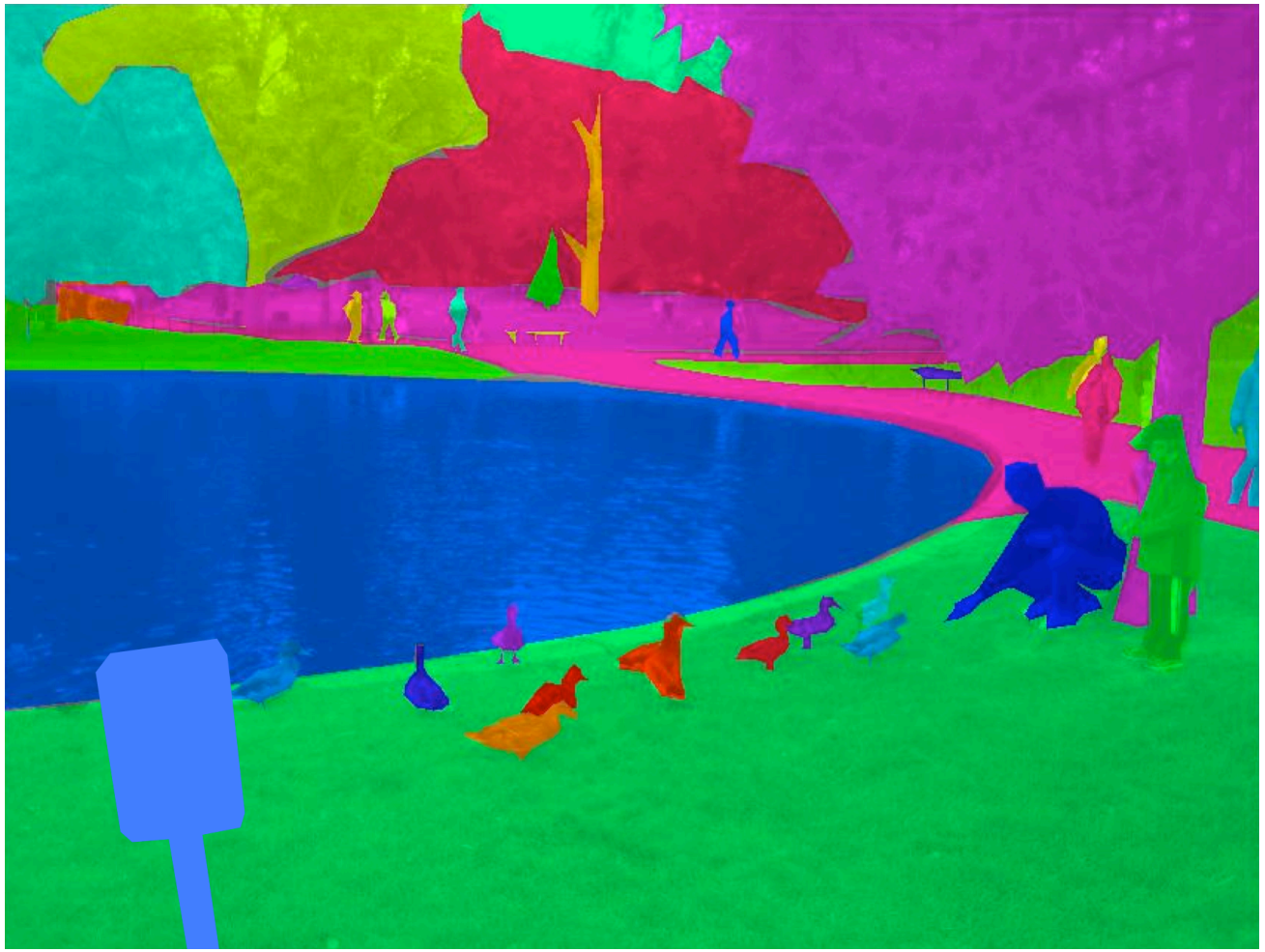


Image source: Antonio Torralba

# Segmentation

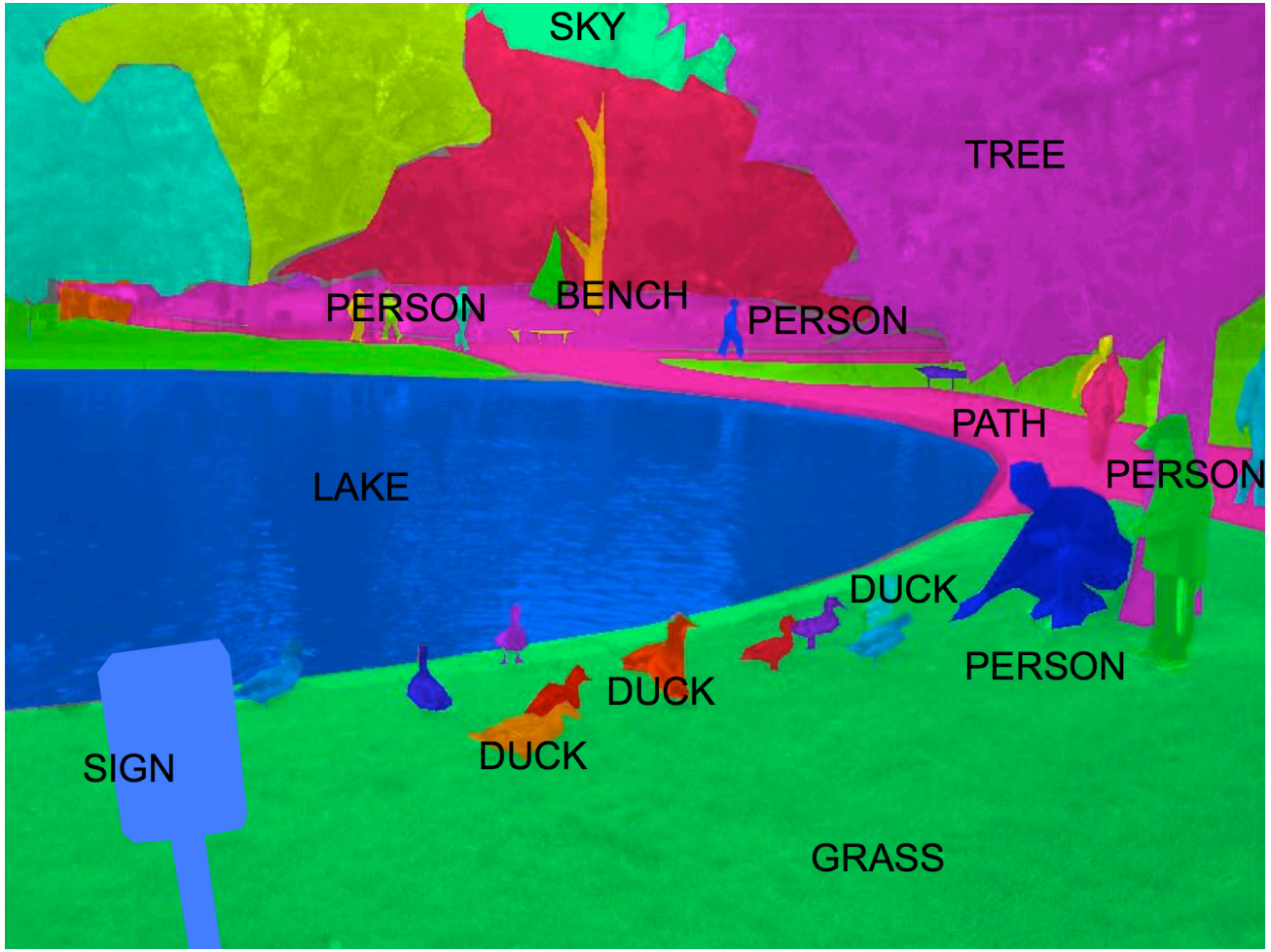
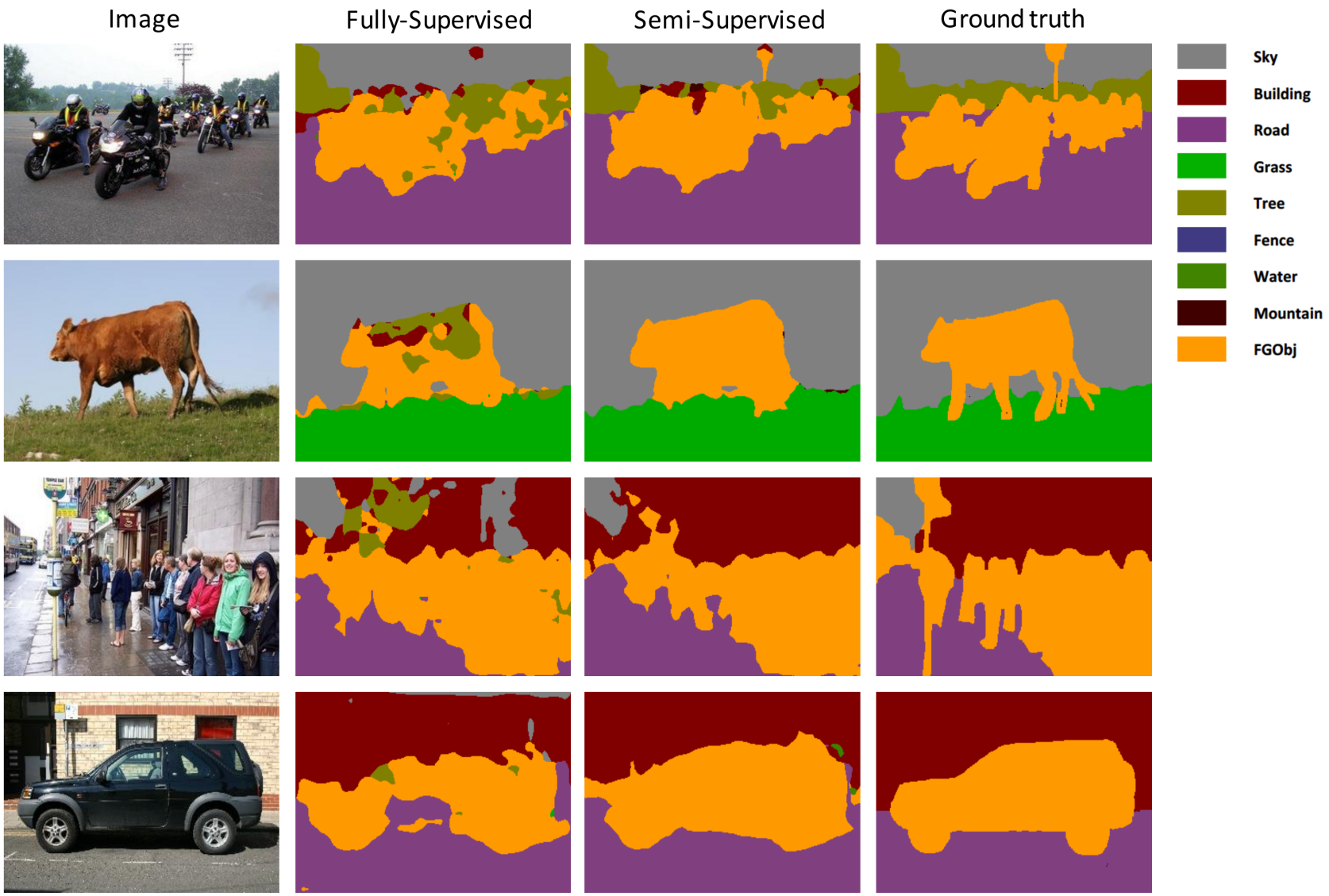


Image source: Antonio Torralba

# Segmentation

Image source: Antonio Torralba

# Semantic Segmentation: Results




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

# Computer Vision Applications

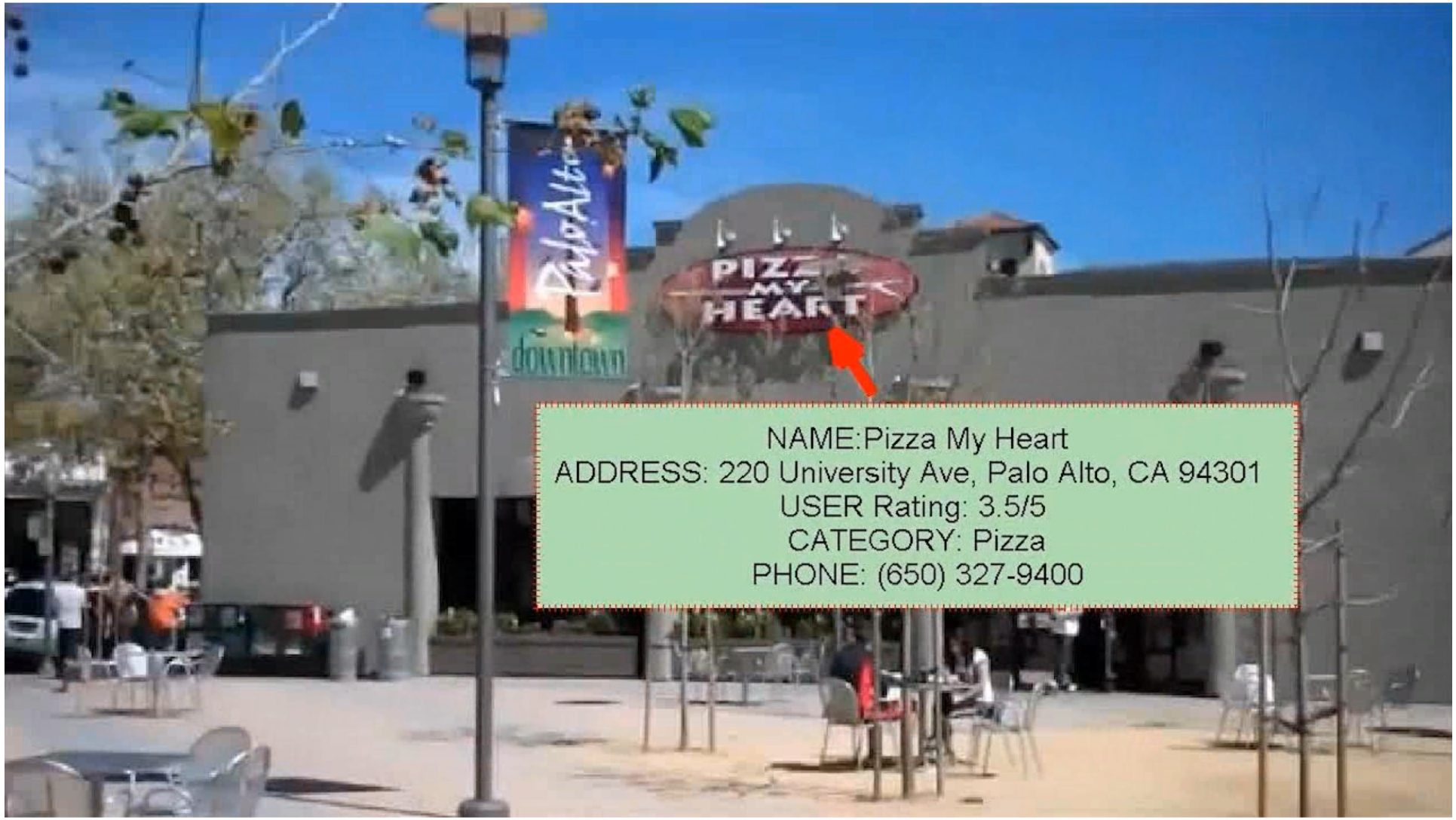
# Counting in Extremely Dense Crowd Images

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

Ground truth=1428 Proposed Method=1468

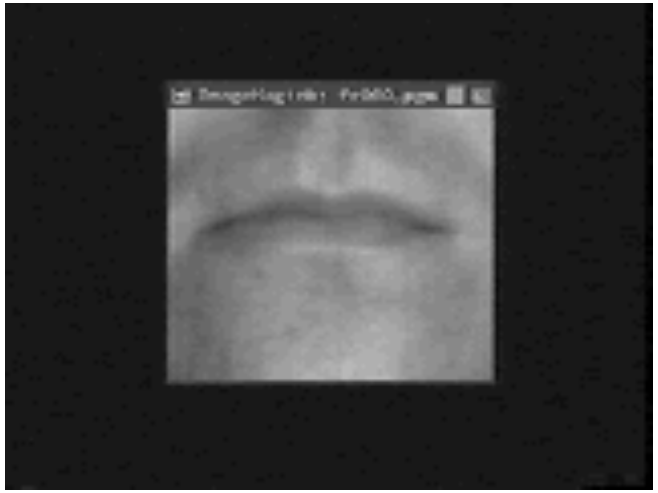
Ground truth=2319 Proposed Method=2496

# Visual Business Recognition

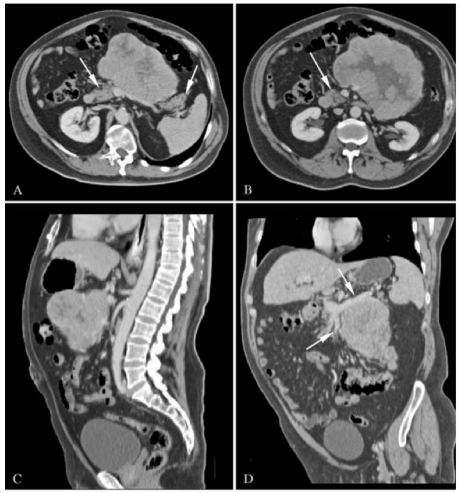


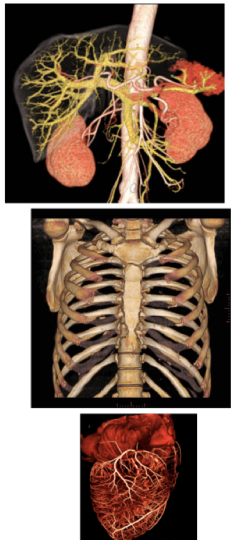
# So many other applications…

- Face identification,
- Lip reading,
- Emotion detection,
- Medical Computer Vision / Analysis,



Surprised Happy

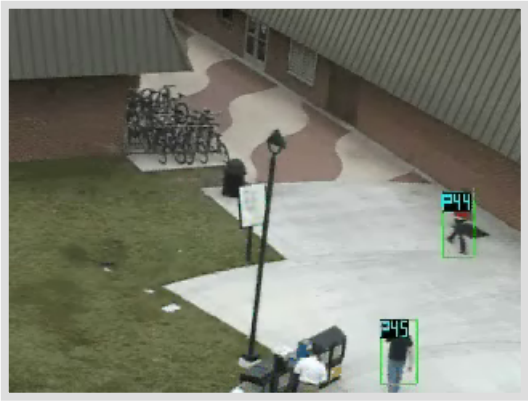


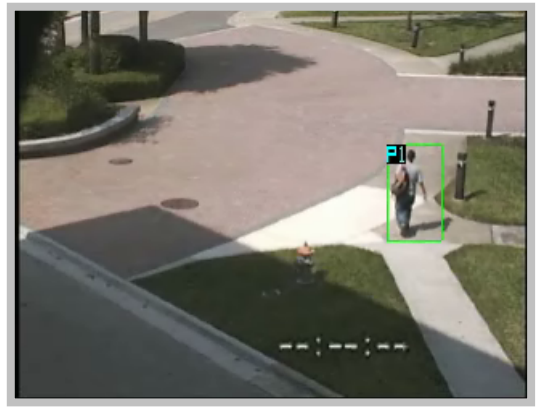


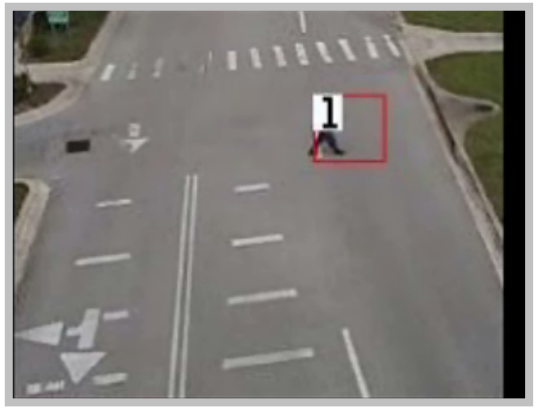
# Video Analysis

Video: image sequence (over time)

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

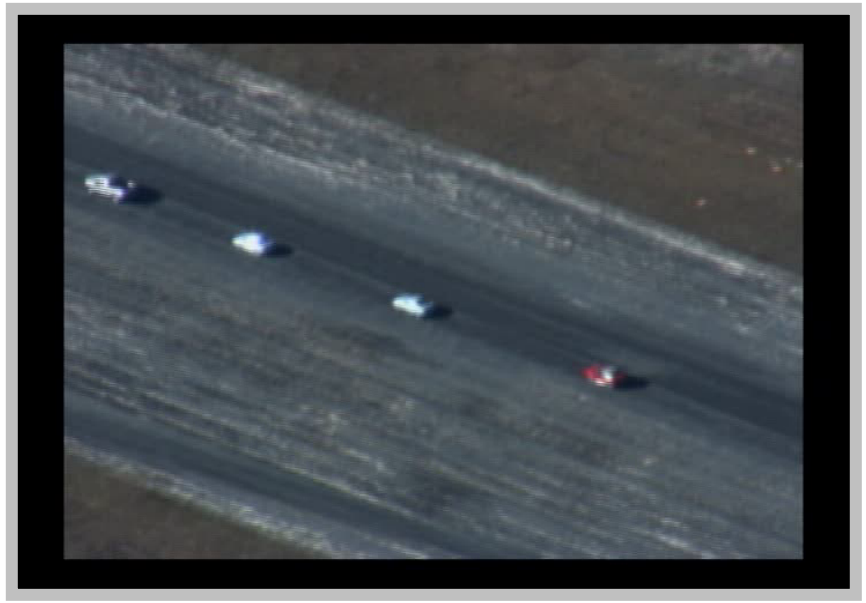


# Object detection in videos

# (Object) Tracking



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

# Human Action Recognition

Action: A spatial and temporal pattern

9 actions, 142 videos (UCF database, Shah).

**Bench Swing**

**Dive**

**Swing**

**Run**

**Kick**

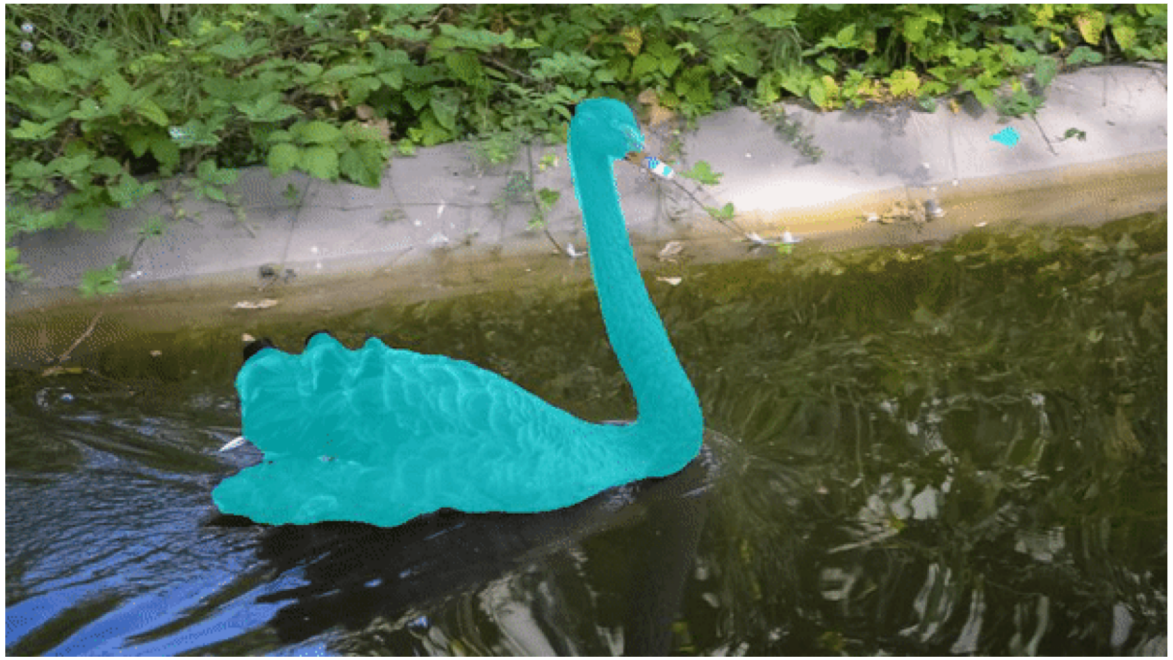
**Lift**

**Ride**

**Golf Swing**

**Skate**

# Actor – action segmentation



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

# Test: what do you see in this image?

Image source: Antonio Torralba

# What do you see in this image?

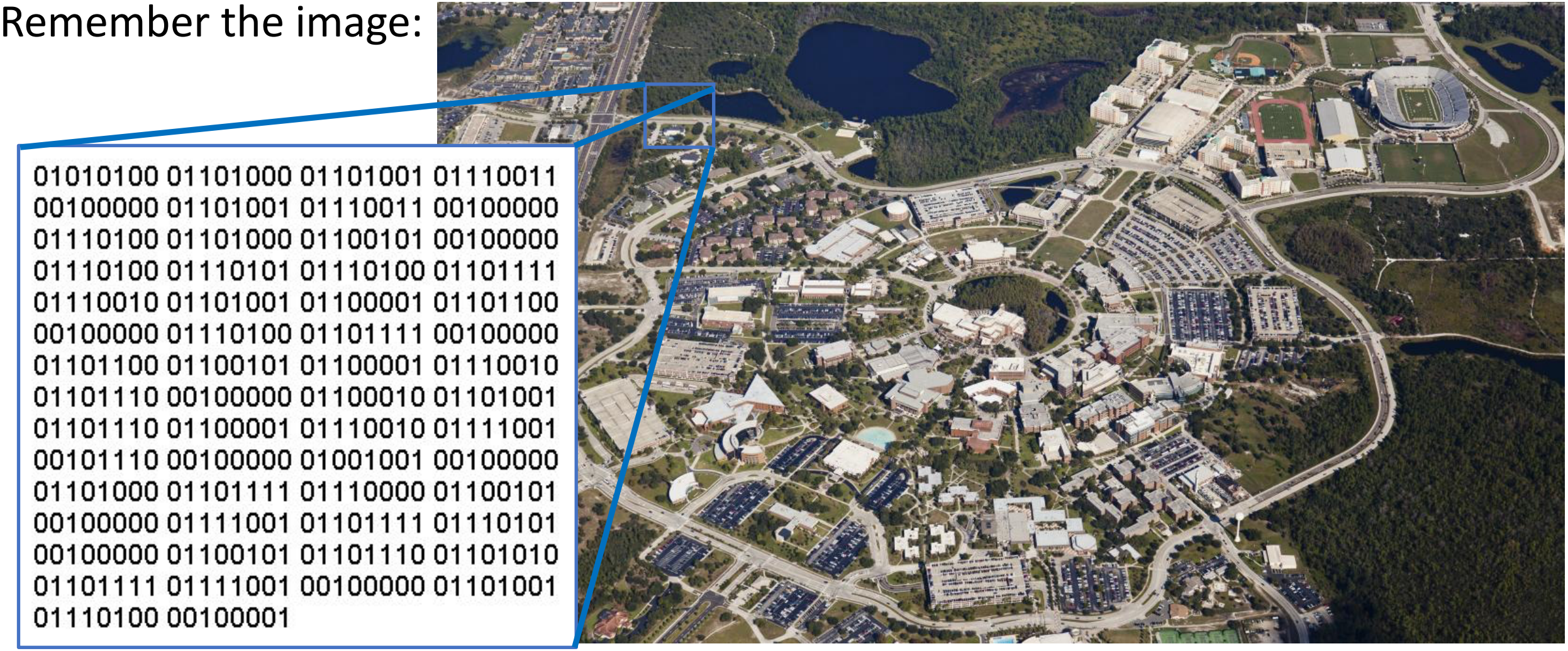
# Now can you see it?

*Credit: Thompson, Basic Vision, Oxford Press, 2012.*

# Question:

How would you model (represent) an object in visual recognition? Ideas?

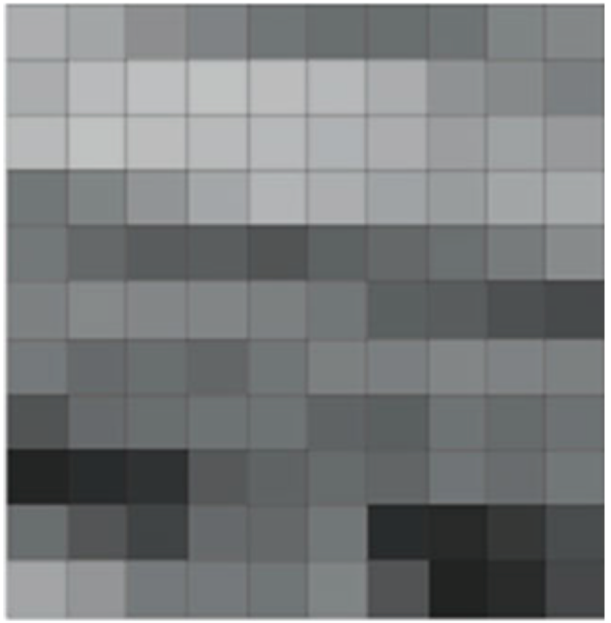
Remember the image:




This is what computers get (see?)!

This is what you see…

# Our common units: Pixels

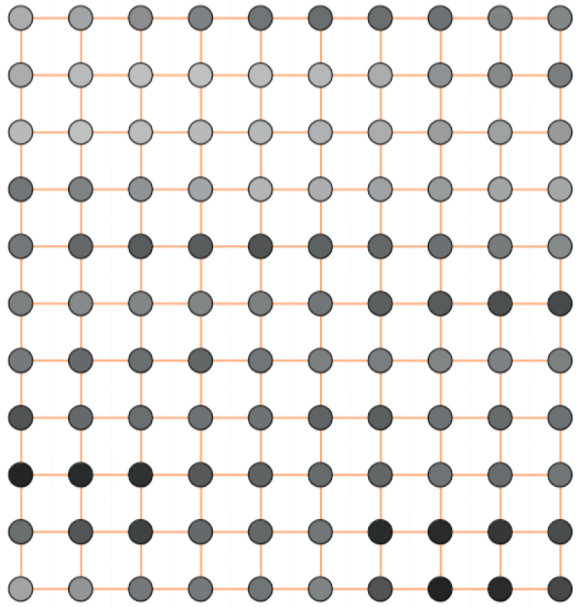


01010100 01101000 01101001 01110011
00100000 01101001 01110011 00100000
01110100 01101000 01100101 00100000
01110100 01110101 01110100 01101111
01110010 01101001 01100001 01101100
00100000 01110100 01101111 00100000
01101100 01100101 01100001 01110010
01101110 00100000 01100010 01101001
01101110 01100001 01110010 01111001
00101110 00100000 01001001 00100000
01101000 01101111 01110000 01100101
00100000 01111001 01101111 01110101
00100000 01100101 01101110 01101010
01101111 01111001 00100000 01101001
01110100 00100001

Binary

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 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 |

Decimal - grid



Grayscale image

# Image Types: (Gray)Scalar and Binary

- A scalar image has integer values
  $u \in \{0, 1, ..., 2^a - 1\}$

a: level (bit)

Ex. If 8 bit (a=8), image spans from 0 to 255

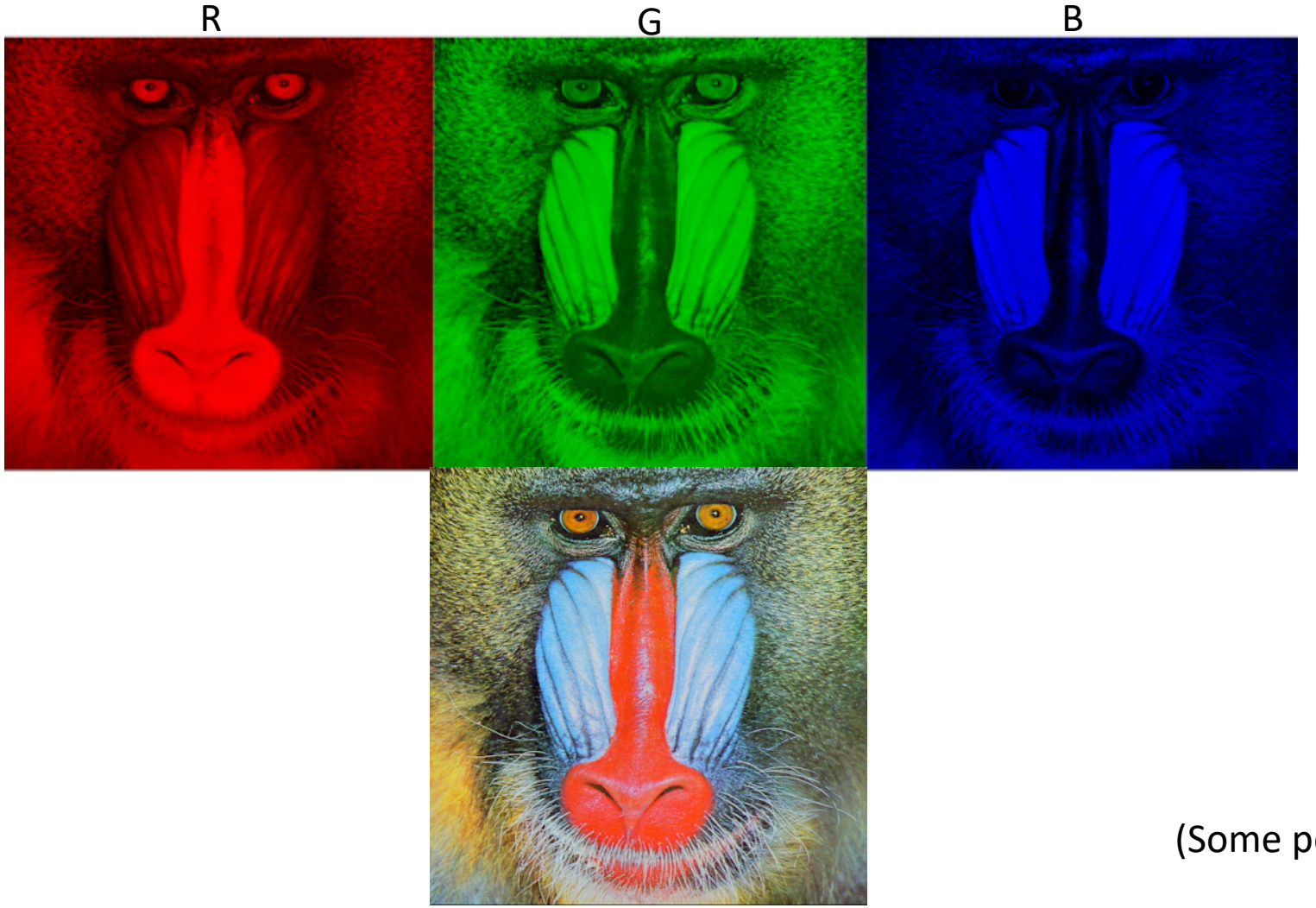
0 black

255 white

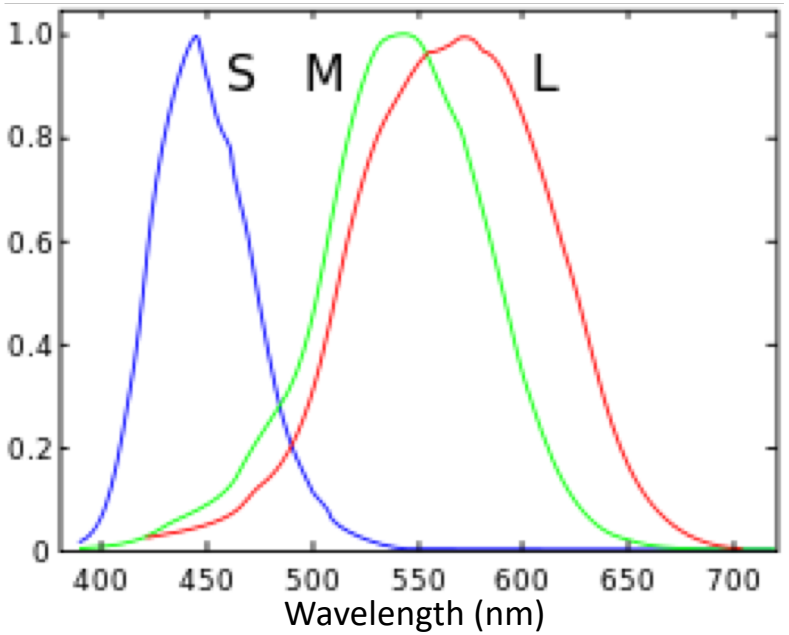
Ex. If 1 bit (a=1), it is binary image, 0 and 1 only.

# Image Type: RGB (red, green, blue)

- Image has three channels (bands), each channel spans a-bit values.




Human Cone-cells (normalized) responsivity spectra

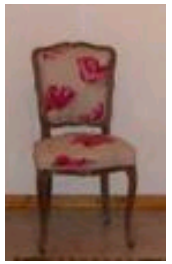



(Some people might have 4 cone-types!)

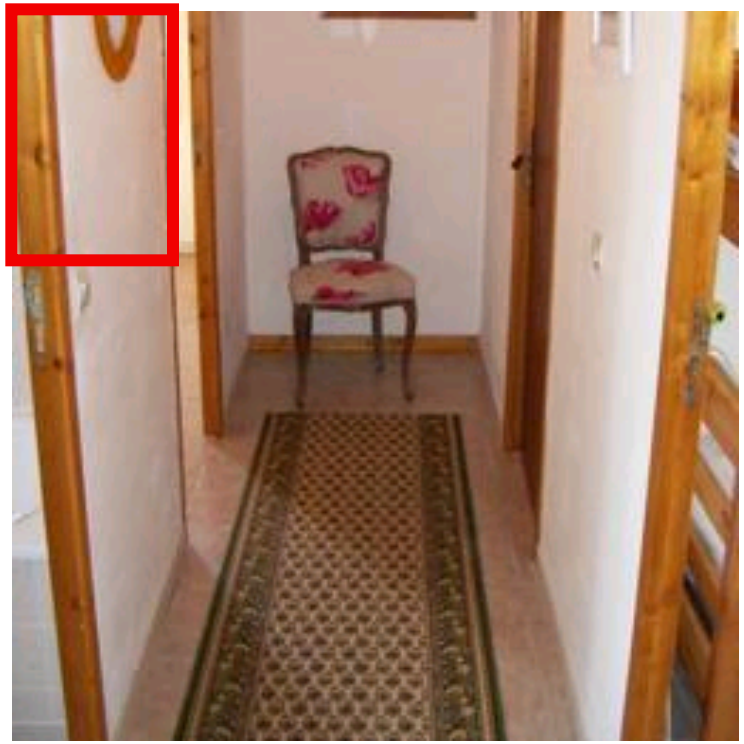
# So…. How do we detect an object in an image?

# Naïve approach: Template Matching

This is a chair



Find the chair in this image



Output of correlation

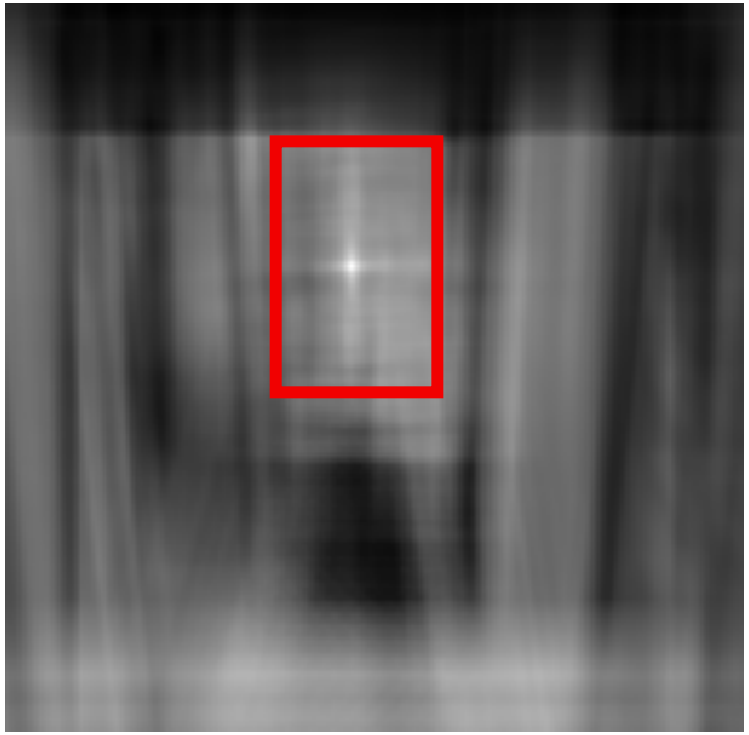
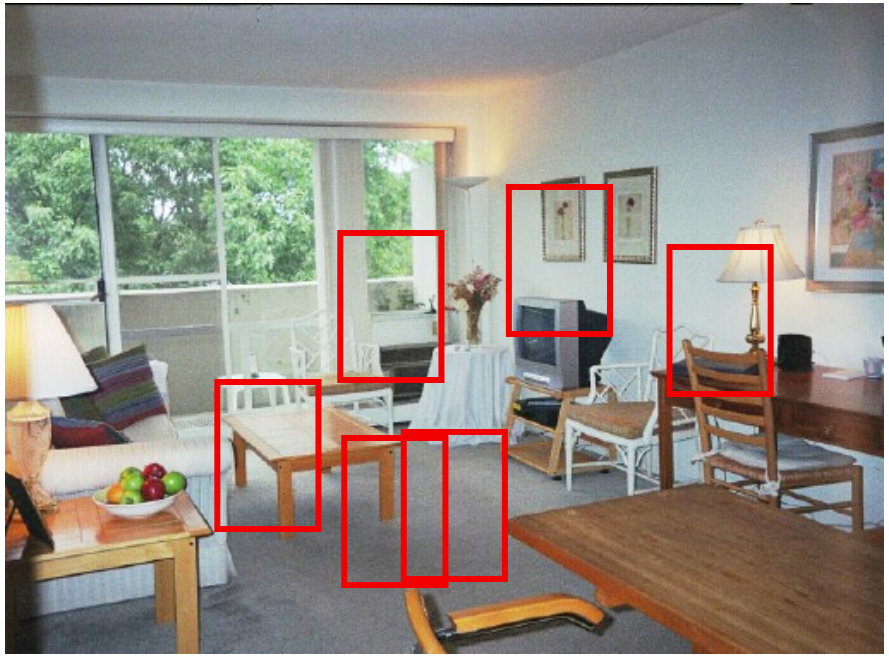
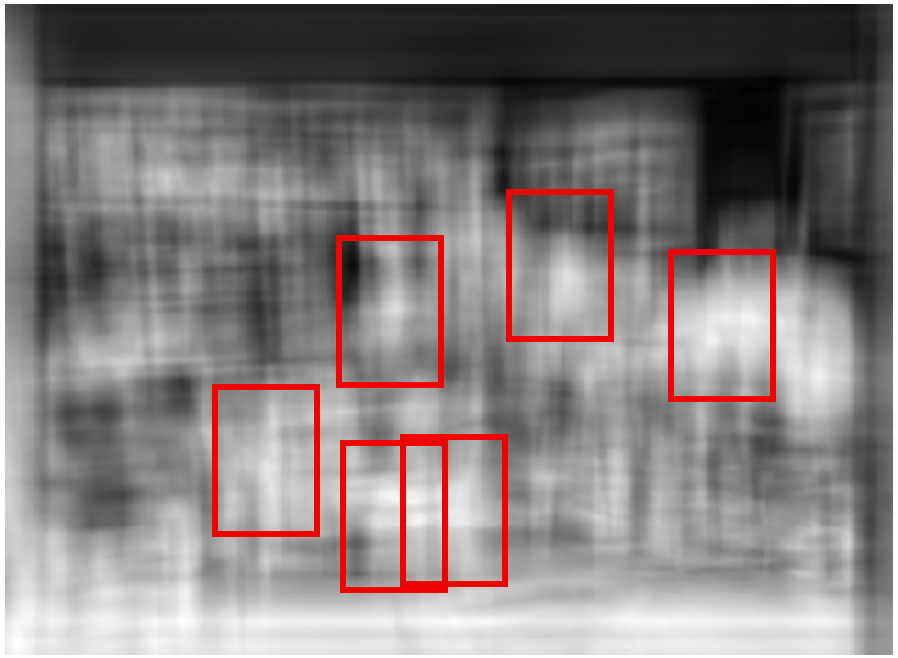


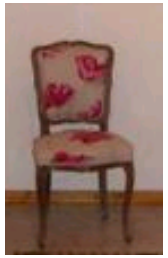
Image source: Antonio Torralba

# Template Matching

Find the chair in this image







Epic fail!
Simple template matching is not going to make it

Image source: Antonio Torralba

# Idea:

- Instead of comparing raw image pixels:
  - first map those pixels into another (more robust) form,
  - and then compare those mapped forms.
  - Finally, select the closest image map (how do you define “closest”? Metrics).
- Features
  - Examples: compute edges, compute color histograms, Gradients, HOG, SIFT, …

# “Bag-of-Words” Representation

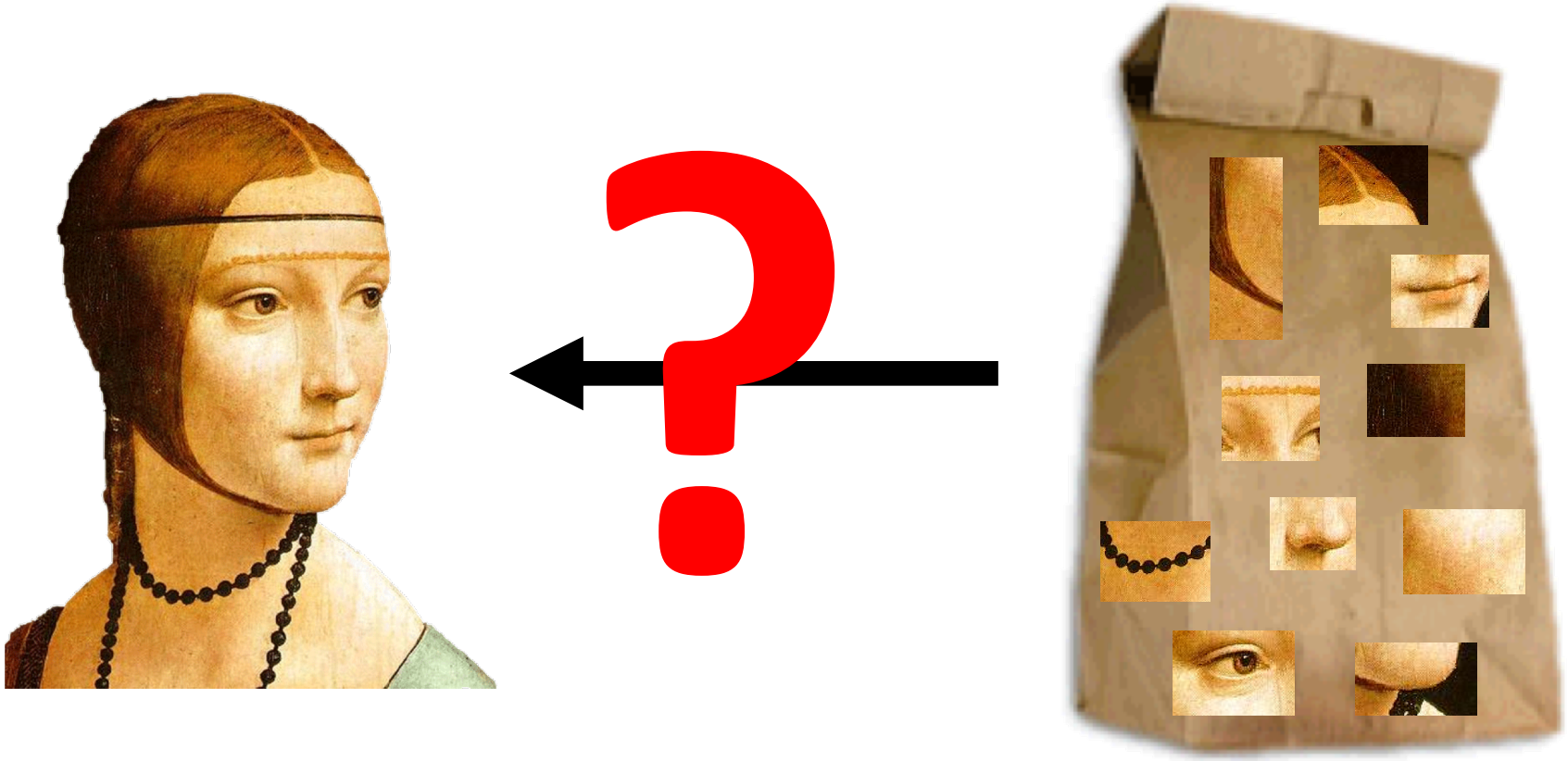
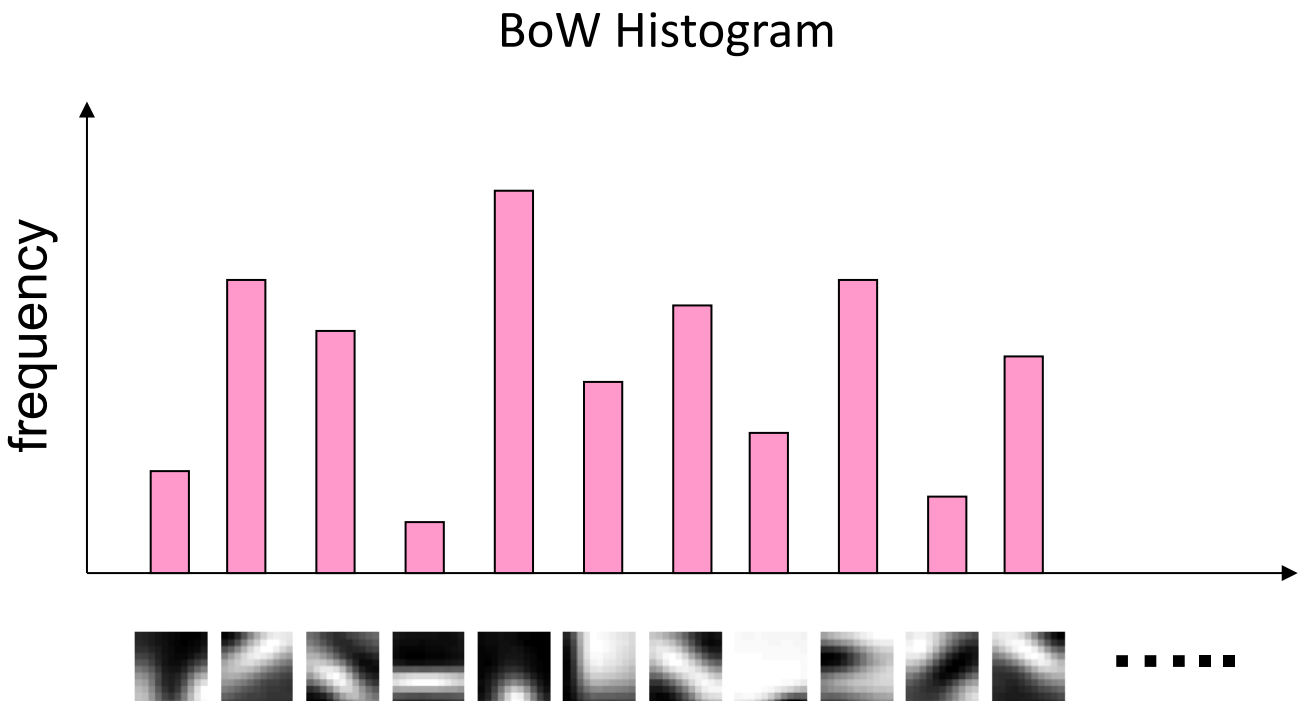
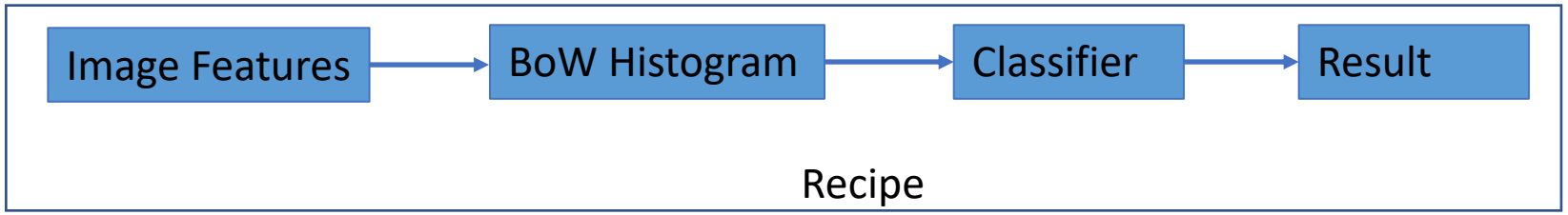


Image source: Antonio Torralba

# “Bag-of-Words” (BoW) Histograms

Image



# BoW Representation

- All have equal probability for bag-of-words methods,
- Location (spatial) information is important but lost.



Image source: Antonio Torralba

# 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

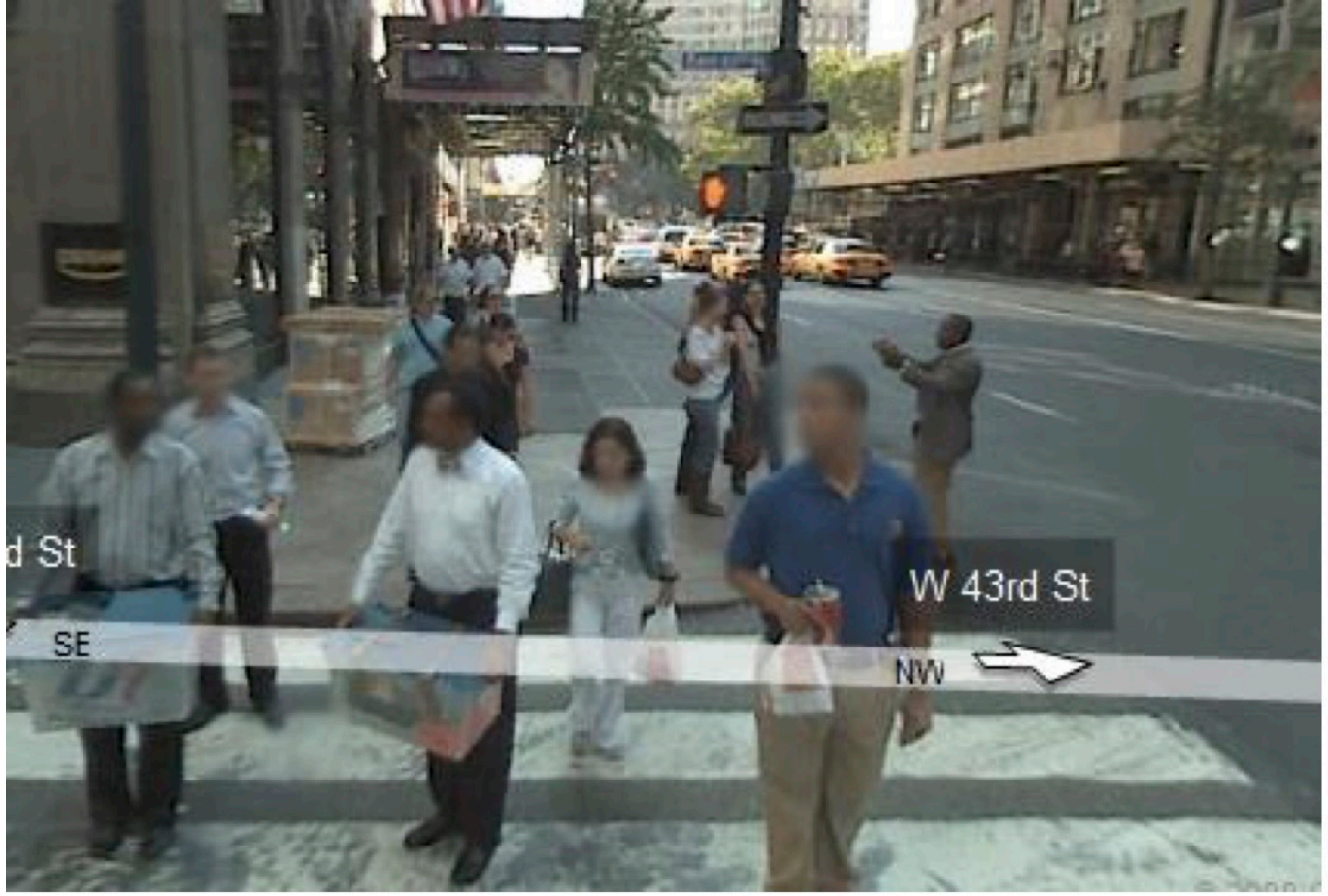
Seymour Papert

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

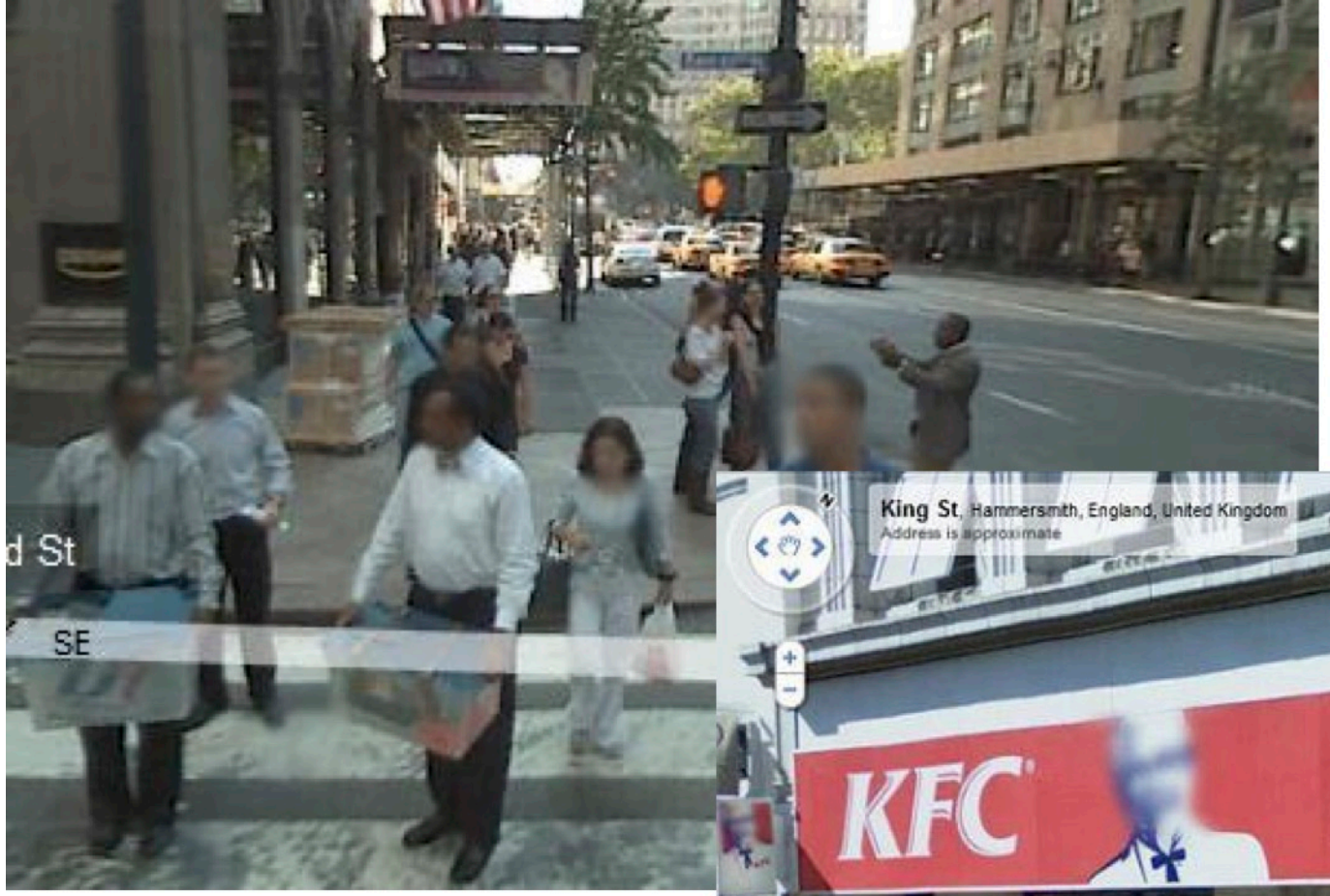
# 25 Years ago

# 15 Years ago




• 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)
•….

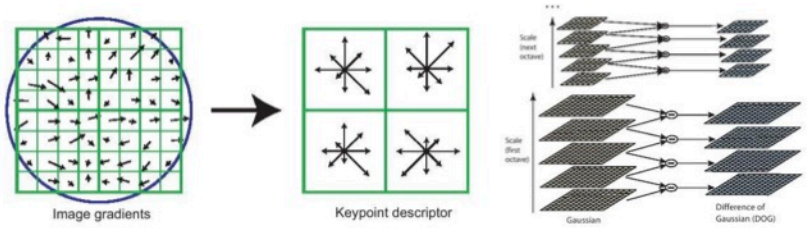
Image source: Antonio Torralba



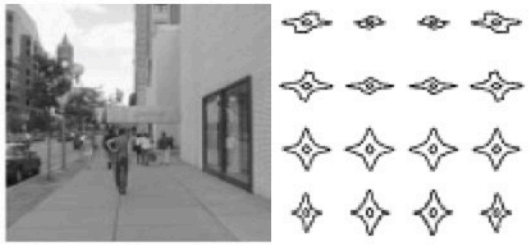
- The representation and matching of pictorial structures Fischler, Elschlager (1973).
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time 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

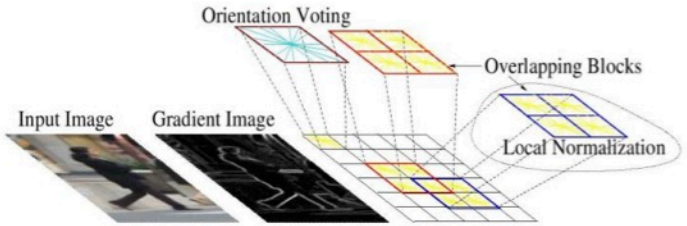
# Advances in computer vision



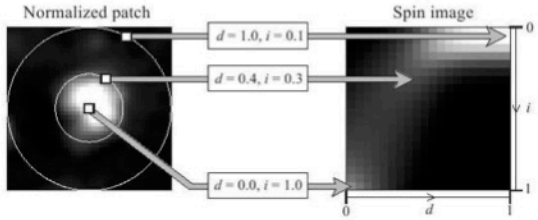

SIFT



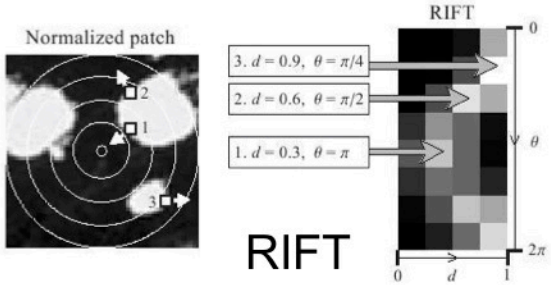
GIST



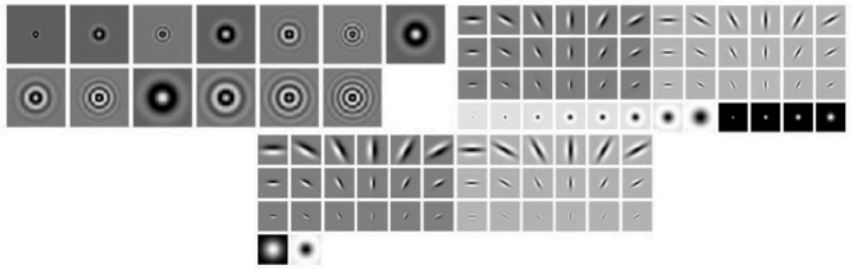

HoG



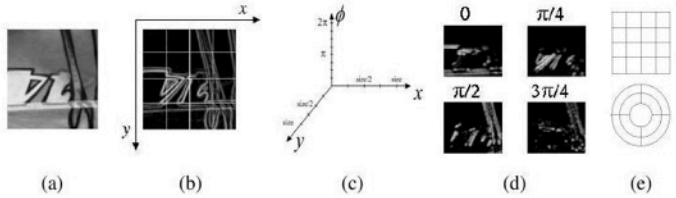

Spin image




RIFT



Textons




GLOH



Image source: Antonio Torralba

# Big data

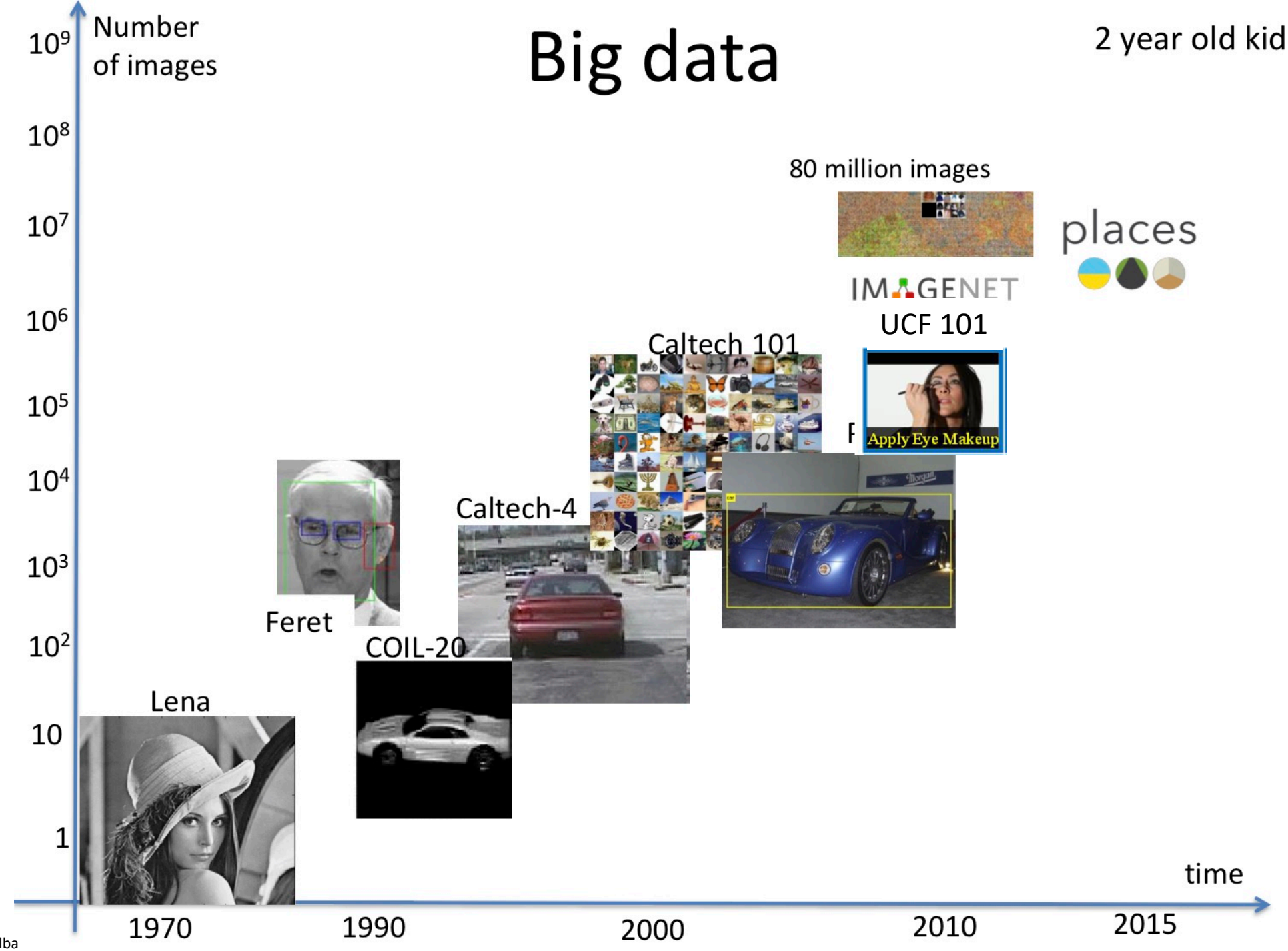



Image source: Antonio Torralba

# The time of big data

Image source: Antonio Torralba

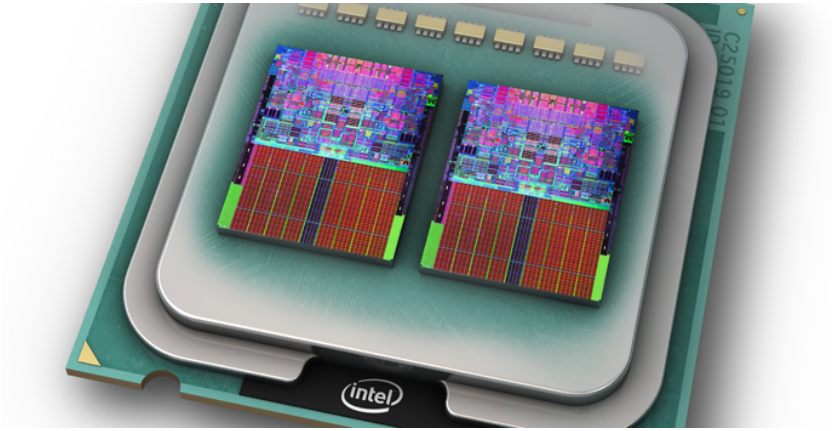
# Distributed / GPU / Cloud Computing

amazon web services™

Google Cloud

Microsoft Azure

NVIDIA GPU CLOUD



# QUESTIONS?

This lecture contains content from:

- Antonio Torralba (MIT),
- Ulas Bagci (UCF),