# CSC420: Intro to Image Understanding Introduction

Ahmed Ashraf

September 7, 2017
(Content inherited from Sanja Fidler)

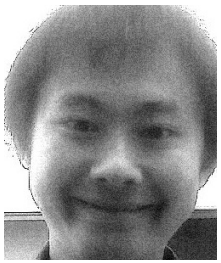
UNIVERSITY OF TORONTO

# The Team

- **Instructor**:

Ahmed Ashraf (aashraf@alumni.cmu.edu)

- **Office**: BA 3219
- **Office hours**: Mondays and Fridays 4-5pm
- **TAs**:



Jake Snell (jsnell@cs.toronto.edu)

Mohammad Kianpisheh(mkian1368@gmail.com)

Hang Chu (chuhang1122@cs.toronto.edu)

Andrei Barsan (andrei.ioan.barsan@gmail.com)

# Course Information

- **Class time**: Tuesday and Thursday at 3-4pm
- **Location**: GB248
- **Tutorials**: on demand
- **Class Website**:

  Coming soon

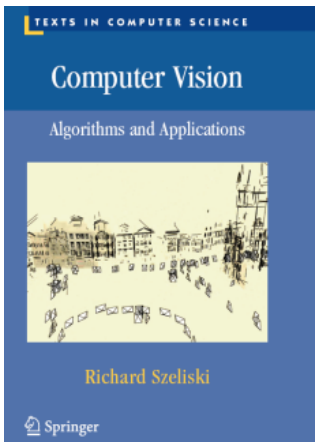
- The class will use Piazza for **announcements** and **discussions**:

  https://piazza.com/utoronto.ca/fall2017/csc420/home

- Your grade will **not depend on your participation on Piazza**. It's just a good way for asking questions, discussing with your instructor, TAs and your peers

# Course Information

- **Textbook**: We won't directly follow any book, but extra reading in this textbook will be useful:




Rick Szeliski

*Computer Vision: Algorithms and Applications*

available free online:

http://szeliski.org/Book/

- Links to other material (papers, code, etc) will be posted on the class webpage

# Course Prerequisites

**Course Prerequisites**:

- Data structures
- Linear Algebra
- Vector calculus

Without this you'll need some serious catching up to do!

**Knowing some basics in this is a plus:**

- Matlab, Python, C++
- Machine Learning
- Neural Networks
- Solving assignments sooner rather than later

# Requirements and Grading

- Each student expected to complete 5 assignments and a project
- **Grading**
  - **Assignments**: 60% (12% each)
  - **Project**: 40%
- **Assignments:**
  - Short **theoretical questions** and **programming exercises**
  - Will be given every **two weeks** (starting with second week of class)
  - You will have **a week to hand in the solution** to each assignment
  - You need to solve the assignment **alone**
- **Project:**
  - You will be able to choose from a list of projects or come up with your own project (discussed prior with your instructor)
  - Need to hand in a **report** and do an oral **presentation**
  - Can work in upto teams of **3**

# Term Work Dates

| Term Work | Post Date | Due Date | % of grade |
|---|---|---|---|
| Assignment 1 | Sept 14 | Sept 21 | 12% |
| Assignment 2 | Sept 28 | Oct 5 | 12% |
| Assignment 3 | Oct 12 | Oct 19 | 12% |
| Project Proposal | | Oct 24 | 5% |
| Assignment 4 | Oct 26 | Nov 2 | 12% |
| Assignment 5 | Nov 16 | Nov 23 | 12% |
| Project Report | | Nov 25 | 20% |
| Project Presentation | | Nov 28,30,Dec 5 | 15% |

# Programming Language?

- Your assignments / project can be in Matlab, Python, C++
- As long as it compiles, runs, and you know how to defend it, we're happy
- HOWEVER, most code and examples we will provide during the class will be in Matlab
- Choose wisely

# Lateness

**Deadline** The solutions to the assignments / project should be submitted **by 11.59pm on the date they are due**. Anything from 1 minute late to 24 hours will count as **one late day**.

**Lateness** Each student will be given a total of **3 free late days**. This means that you can hand in three of the assignments one day late, or one assignment three days late. It is up to the you to make a good planning of your work. **After you have used the 3 day budget, the late assignments will not be accepted.**

# Syllabus

**Tentative** syllabus

| Week nb. | Date | Topic |
|---|---|---|
| 1 | Sept 7 | Intro |
| 2 | Sept 12 & Sept 14 | Linear filters, edges |
| 3 | Sept 19 & Sept 21 | Image features |
| 4 | Sept 26 & Sept 28 | Keypoint detection |
| 5 | Oct 3 & Oct 5 | Matching |
| 6 | Oct 10 & Oct 12 | Segmentation |
| 7 | Oct 17 & Oct 19 | Grouping |
| 8 | Oct 24 & Oct 26 | Object recognition |
| 9 | Oct 31 & Nov 2 | Object detection |
| 10 | Nov 14 & Nov 16 | Neural Networks |
| 11 | Nov 21 & Nov 23 | Stereo, multi-view |
| 12 | Nov 25,28 & Dec 5 | Project Presentations |

# Let's begin!

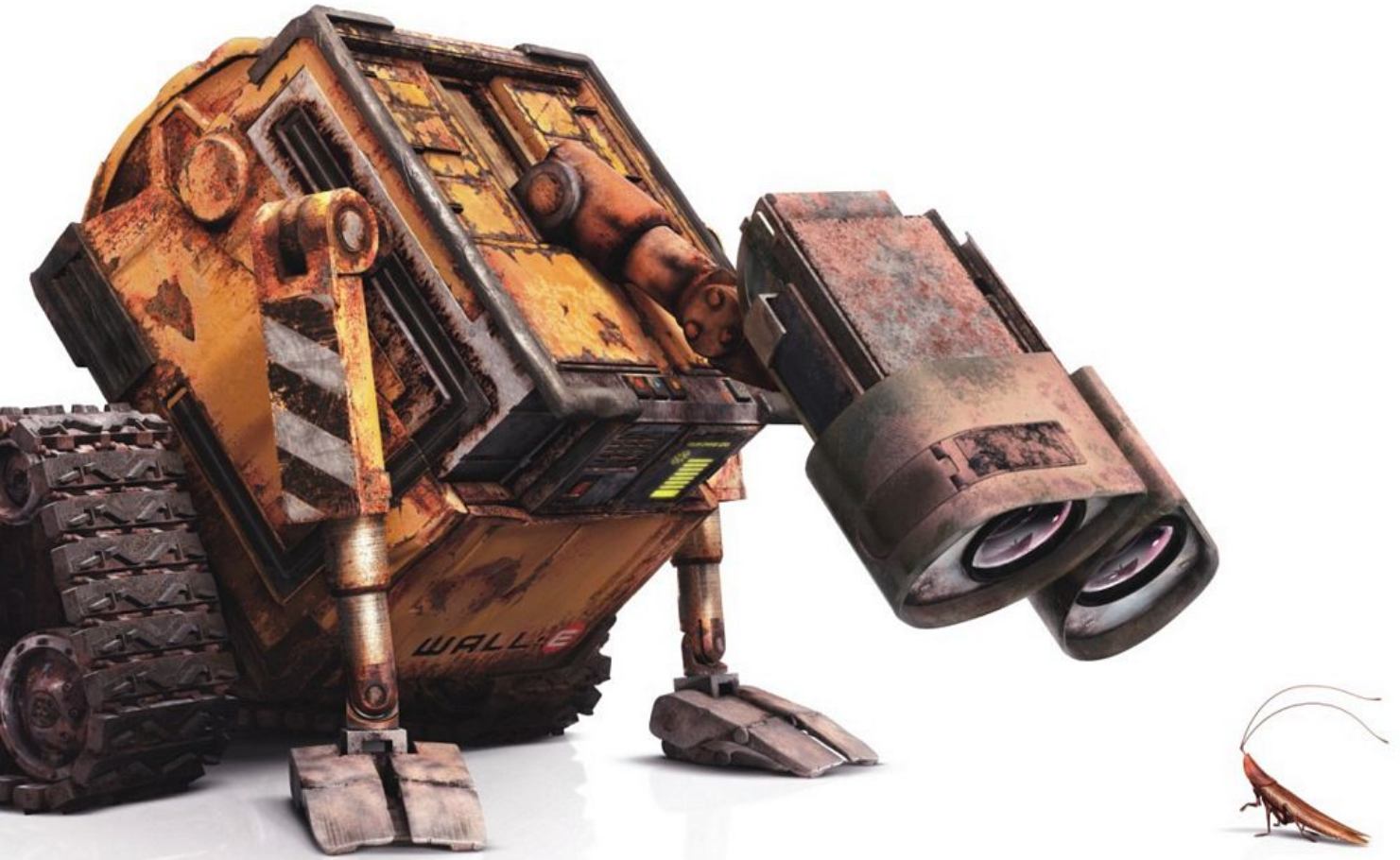
Introduction to Intro to Image Understanding

- What is Computer Vision?
- Why study Computer Vision?
- Which cool applications can we do with it?
- Is vision a hard problem?

# What is Computer Vision?

# What is Computer Vision?

- A field trying to develop automatic algorithms that would **“see”**






Disney · PIXAR
WALL·E

# What is Computer Vision?

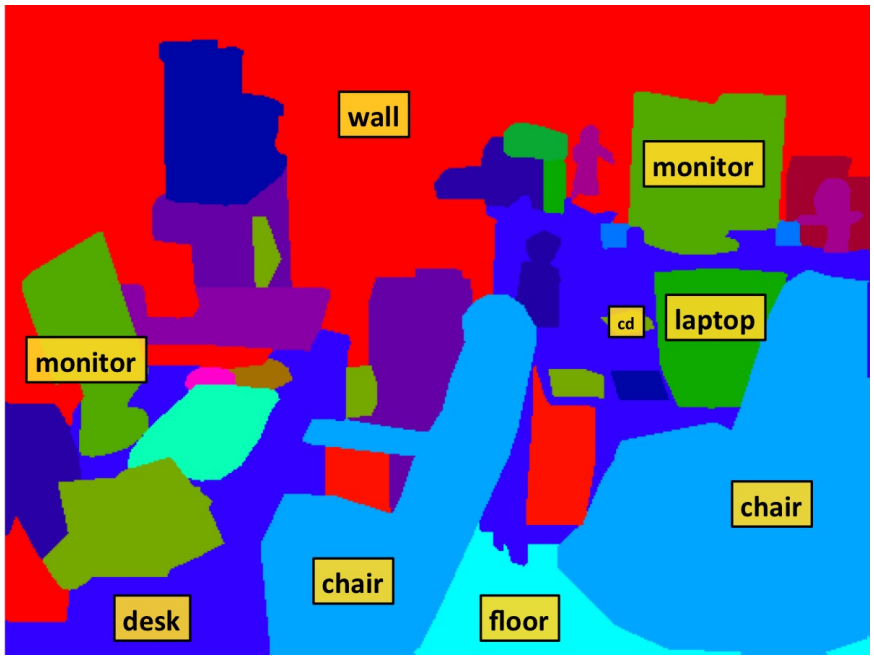
[text adopted from A. Torralba]

- What does it mean to see?
  - To know what is where by looking – Marr, 1982

# What is Computer Vision?

[text adopted from A. Torralba]

- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world

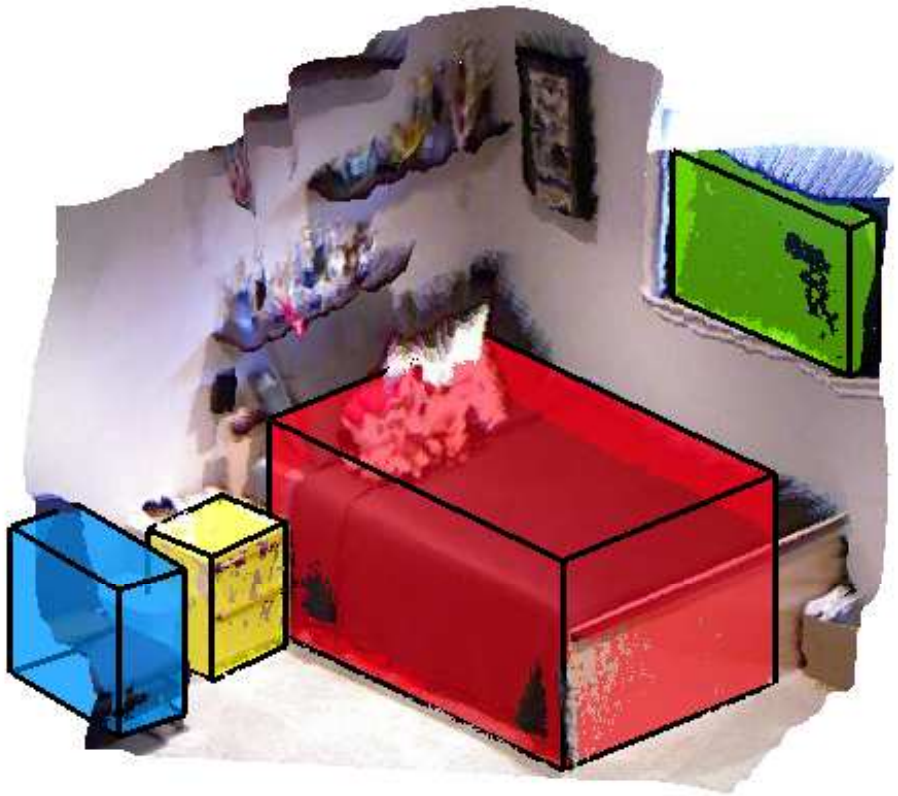


# What is Computer Vision?

[text adopted from A. Torralba]

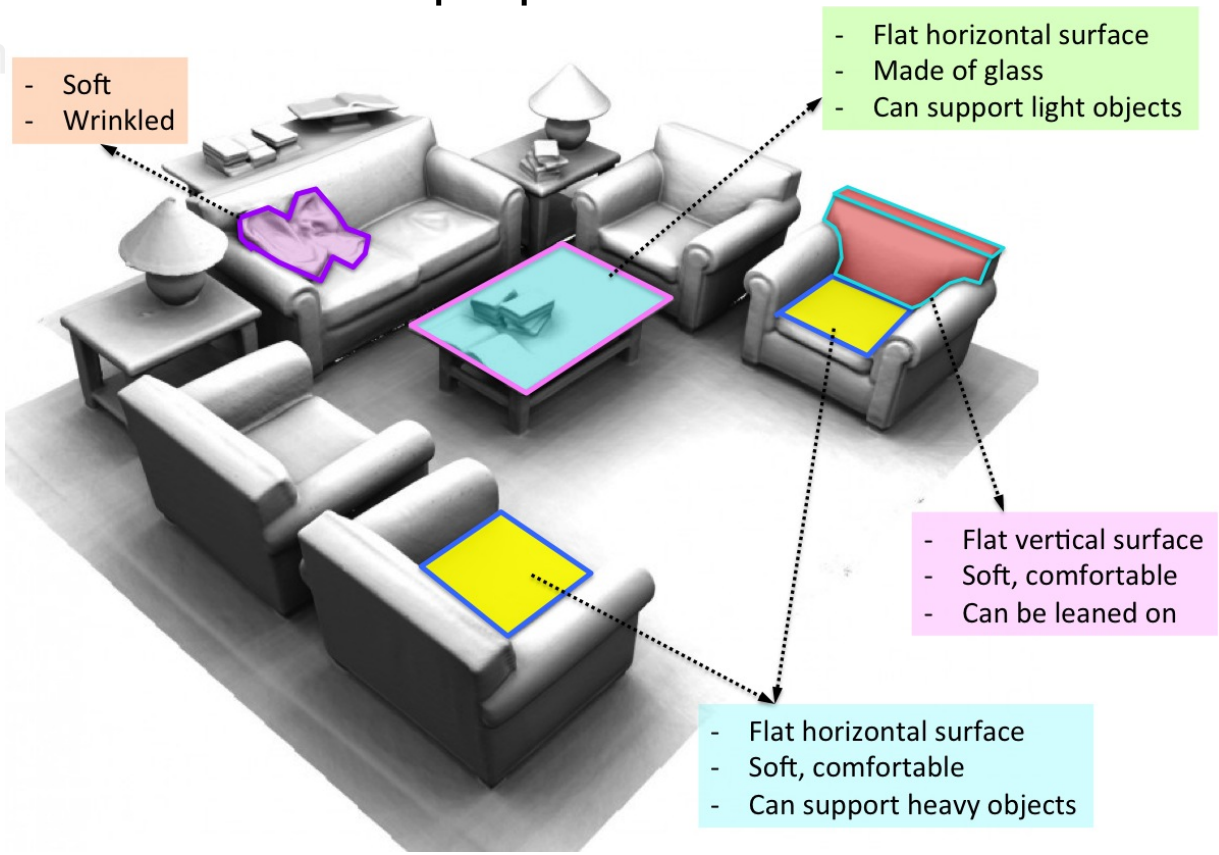
- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D properties?

**image**



# What is Computer Vision?

- What does it mean to see? [text adopted from A. Torralba]
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D properties?
  - Wh




Depth pic from http://vladlen.info

# What is Computer Vision?

[text adopted from A. Torralba]

- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D properties?
  - What actions are taking place?


Pic from www.cobblehillpuzzles.com

# “Full” Image Understanding?

- Full understanding of an image?

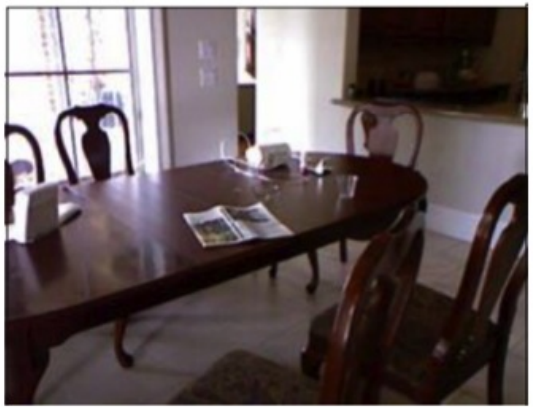
# “Full” Image Understanding?

- Full understanding of an image? **You can answer any question about it**

[M. Malinowski, M. Fritz, A Multi-World Approach to Question Answering about Real-World Scenes based on Uncertain Input, *NIPS*, 2014]

# “Full” Image Understanding?

- Full understanding of an image? **You can answer any question about it**



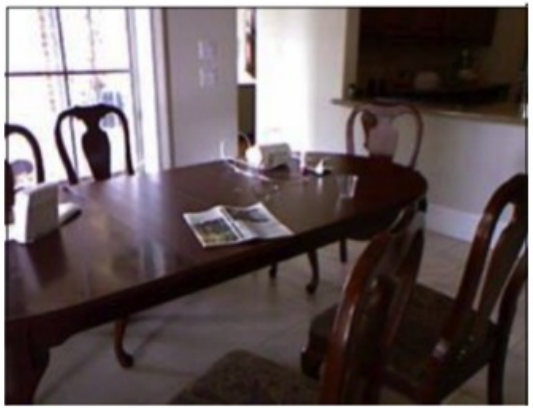
Q: What is behind the table?
A: window

Q: What is in front of the toilet?
A: door

Q: What is on the counter in the corner? A: microwave

# "Full" Image Understanding?

- Full understanding of an image? **You can answer any question about it**



Q: What is behind the table?
A: window

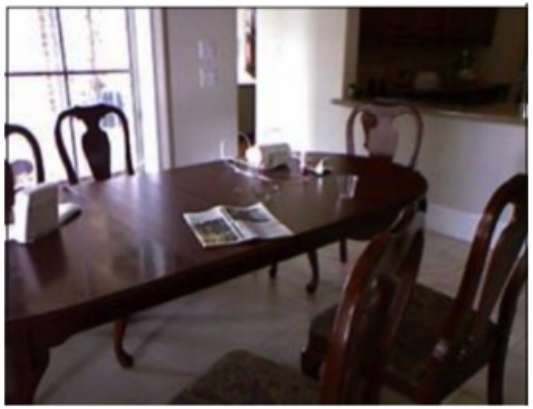
Q: What is in front of the toilet?
A: door

Q: What is on the counter in the corner? A: microwave

Q: What is the shape of the green chair? A: horse shaped

# "Full" Image Understanding?

- Full understanding of an image? **You can answer any question about it**



Q: What is behind the table?
A: window

Q: What is in front of the toilet?
A: door

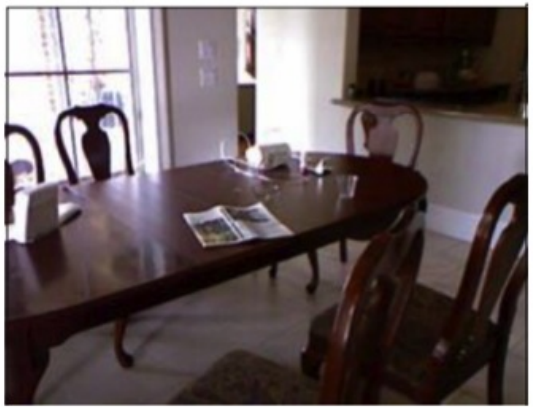
Q: What is on the counter in the corner? A: microwave

Q: What is the shape of the green chair? A: horse shaped

Q: Where is the oven?
A: on the right side of the fridge

# "Full" Image Understanding?

- Full understanding of an image? **You can answer any question about it**



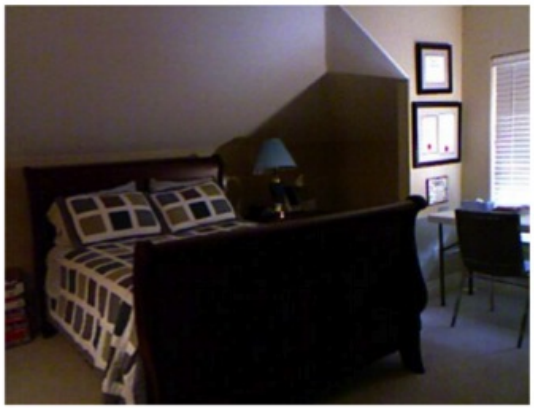
Q: What is behind the table?
A: window

Q: What is in front of the toilet?
A: door

Q: What is on the counter in the corner? A: microwave

Q: What is the shape of the green chair? A: horse shaped

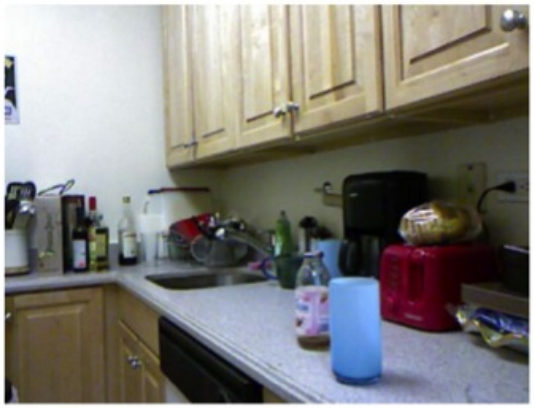
Q: Where is the oven?
A: on the right side of the fridge



Q: What is the largest object?
A: bed

# “Full” Image Understanding?

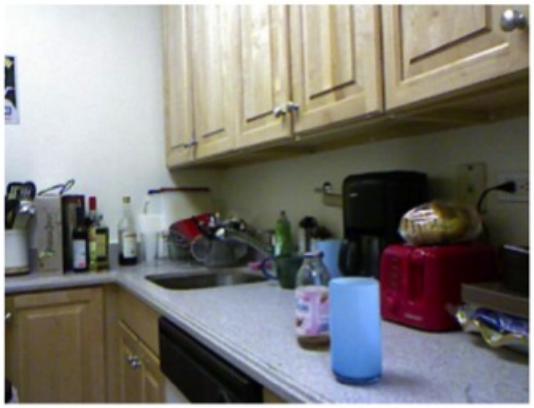
- Full understanding of an image? **You can answer any question about it**



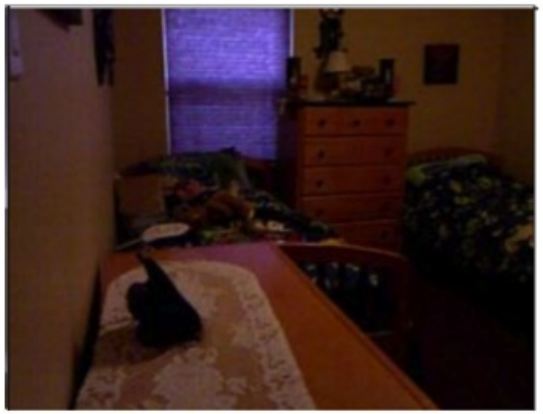
Q: Which object is red?
A: toaster

# "Full" Image Understanding?

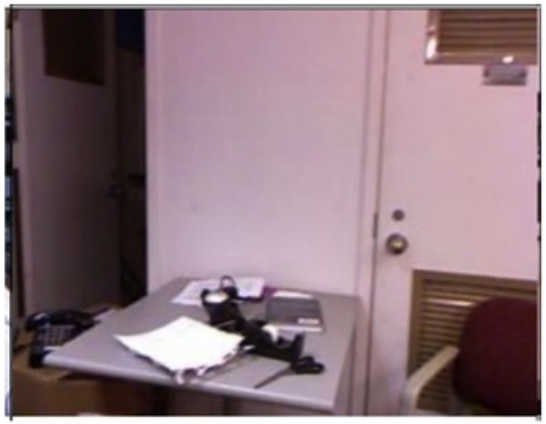
- Full understanding of an image? **You can answer any question about it**



Q: Which object is red?
A: toaster



Q: How many drawers are there?
A: 6

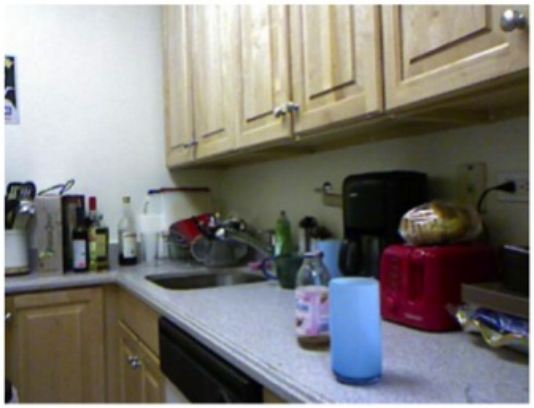


Q: How many doors are open
A: 1

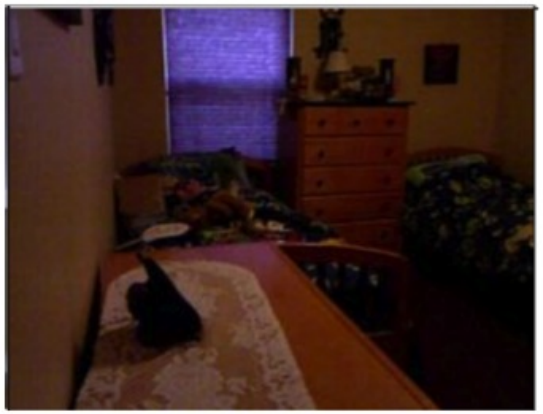
Q: How many lights are on?
A: 6

# "Full" Image Understanding?

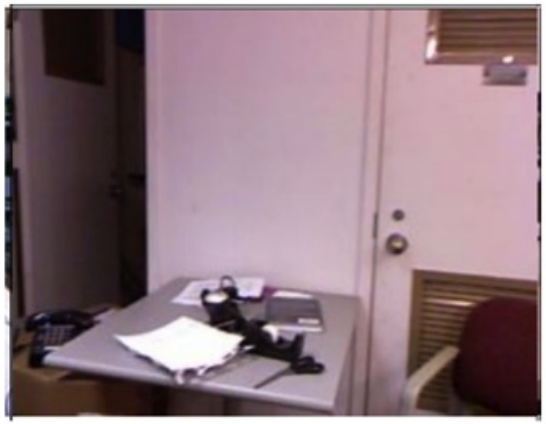
- Full understanding of an image? **You can answer any question about it**



Q: Which object is red?
A: toaster

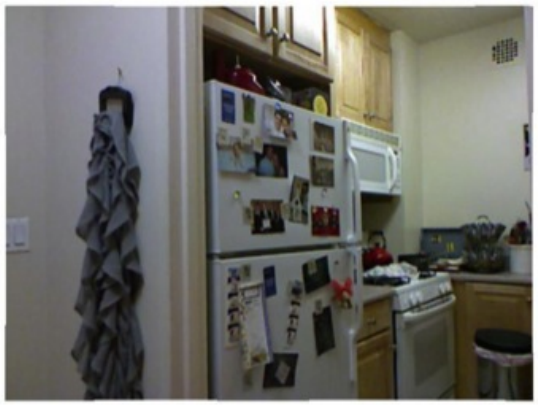


Q: How many drawers are there?
A: 6



Q: How many doors are open
A: 1

Q: How many lights are on?
A: 6



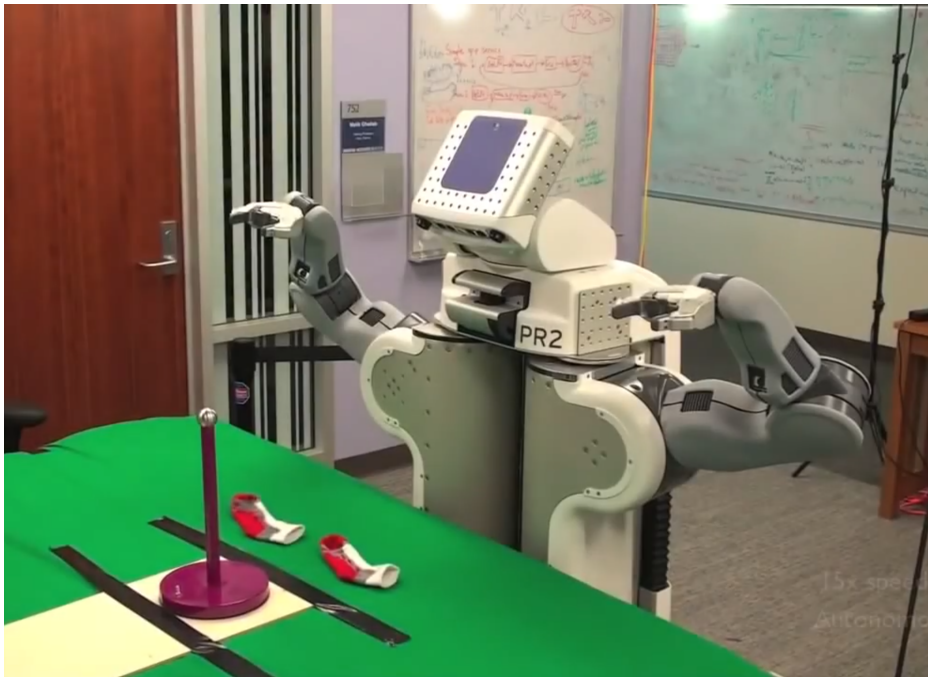
Q: Can you make pizza in this room?
A: yes

Q: Where can you sit?
A: chairs, table, floor

# Why study Computer Vision?

# Why study Computer Vision?

- Because you want your robot to fold your laundry



# Why study Computer Vision?

- ... and drive you to work (video)


Amnon Shashua's Mobileye autonomous driving system

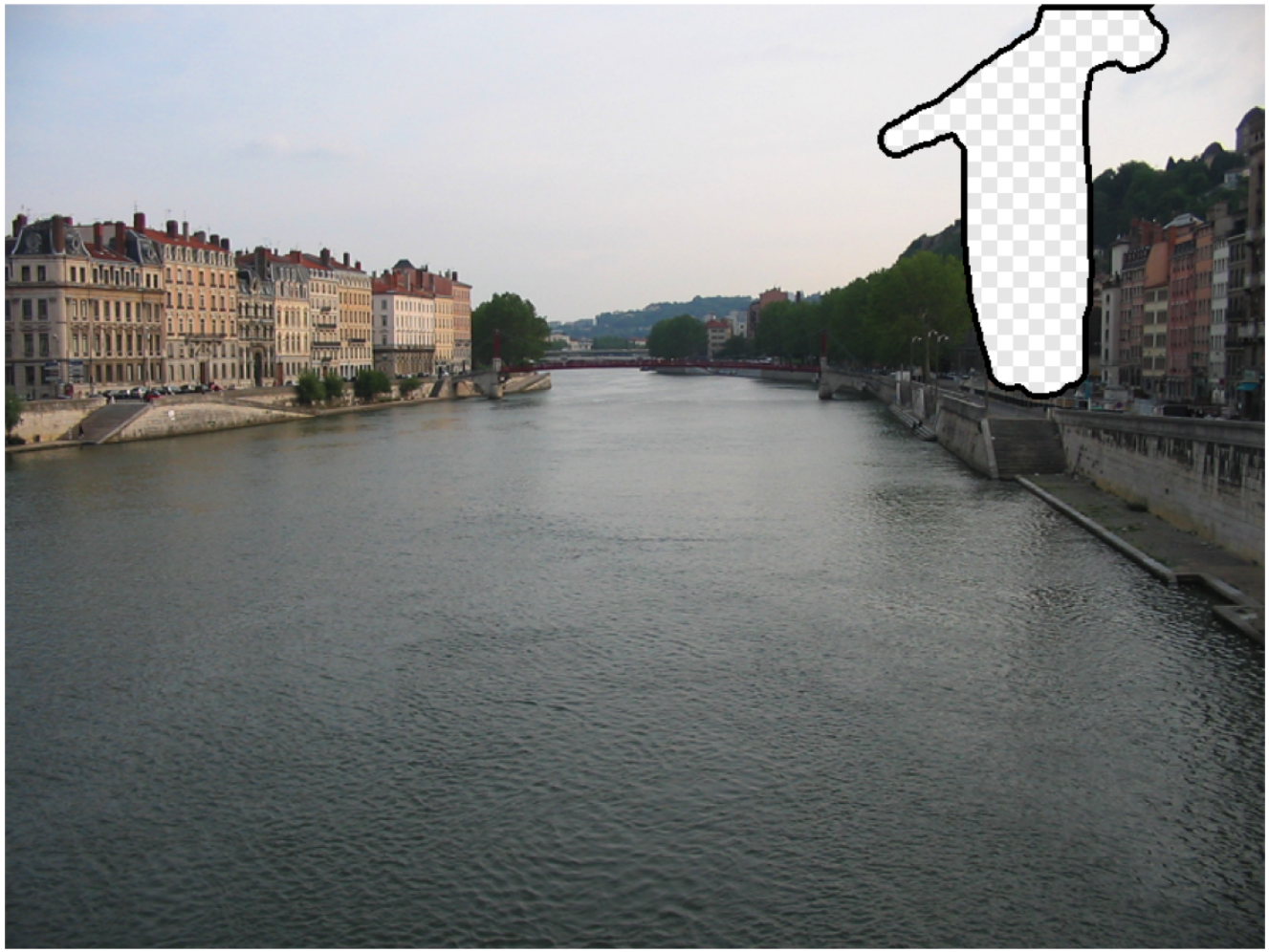
# Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007

# Why study Computer Vision?

- Allows you to manipulate your images



*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007

# Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007

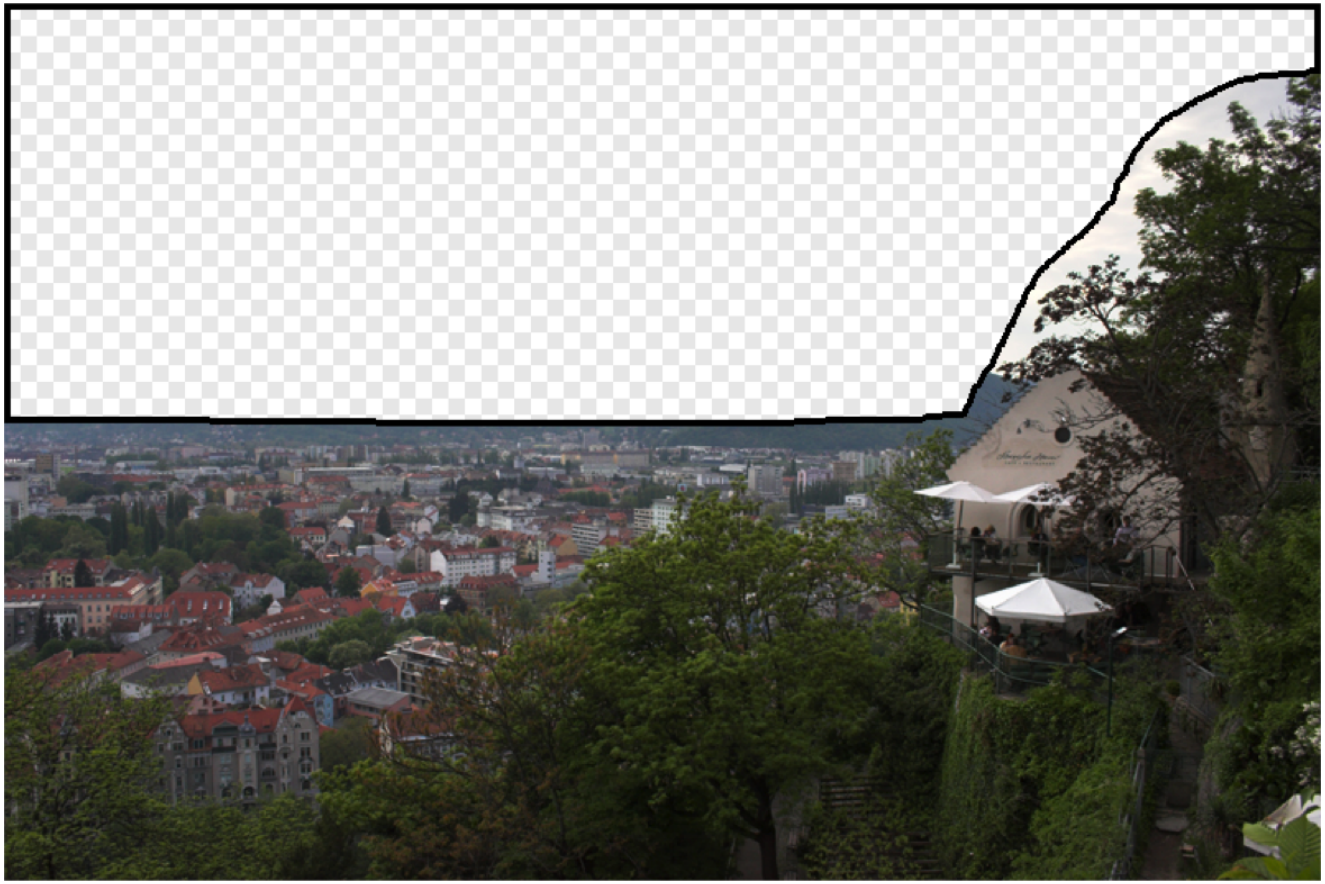
# Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007

# Why study Computer Vision?

- Allows you to manipulate your images



*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007

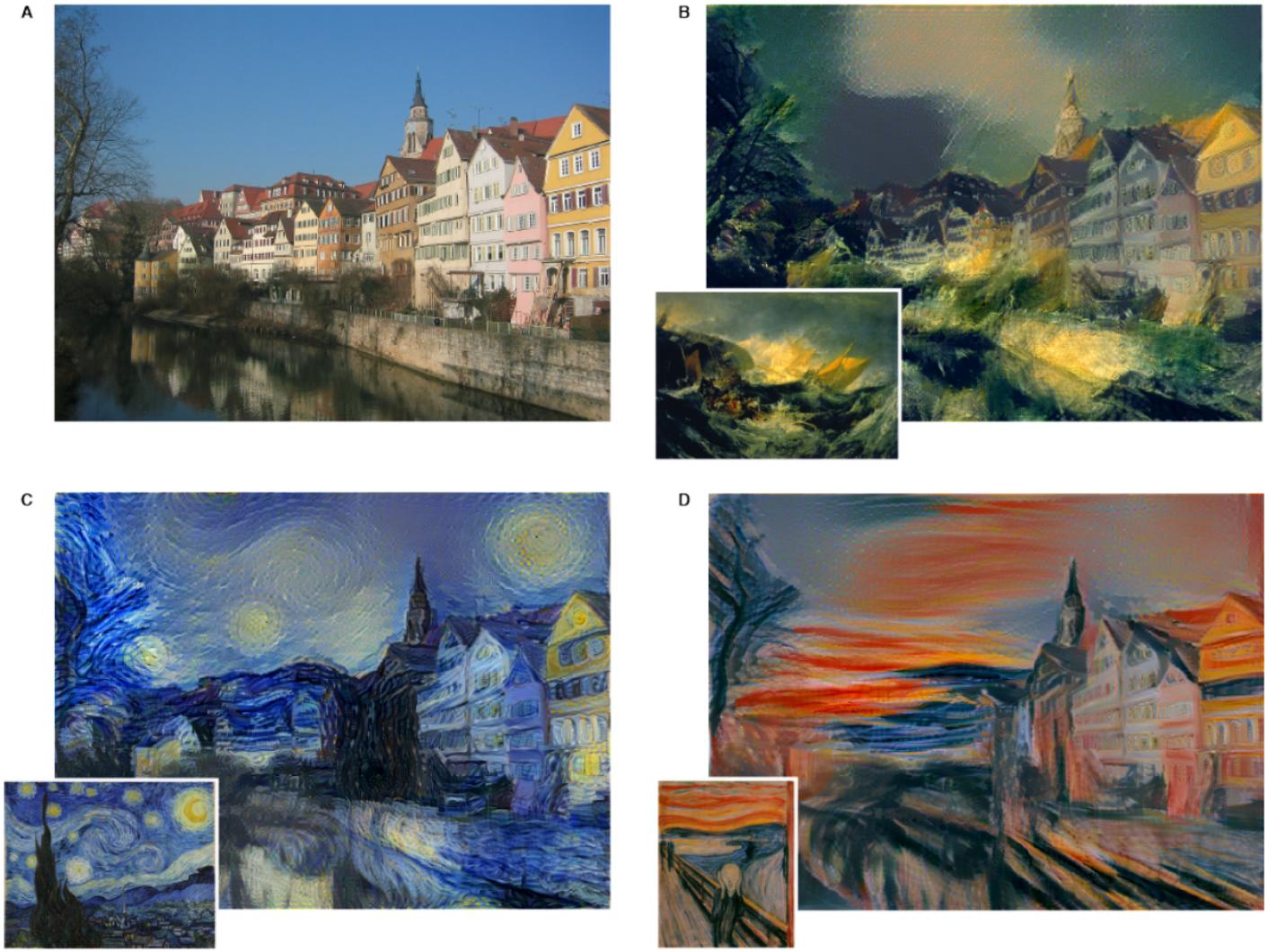
# Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007

# Why study Computer Vision?

- Change style of images




[Gatys, Ecker, Bethge. A Neural Algorithm of Artistic Style. Arxiv'15.]

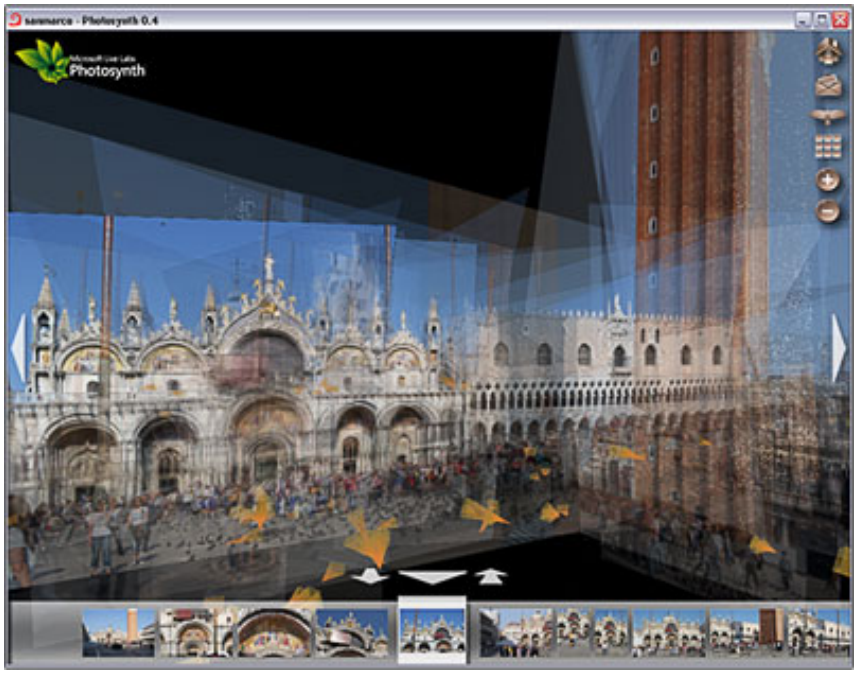
# Why study Computer Vision?

- ... and make cool videos using a single image

*3D Object Manipulation in a Single Photograph using Stock 3D Models*, Kholgade, Simon, Efros, Sheikh, SIGGRAPH 2014

# Why study Computer Vision?

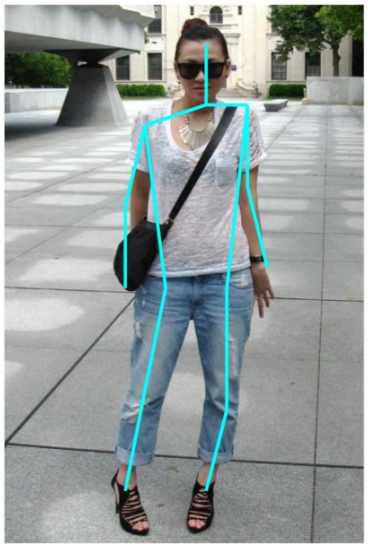
- Reconstruct the world in 3D from online photos!

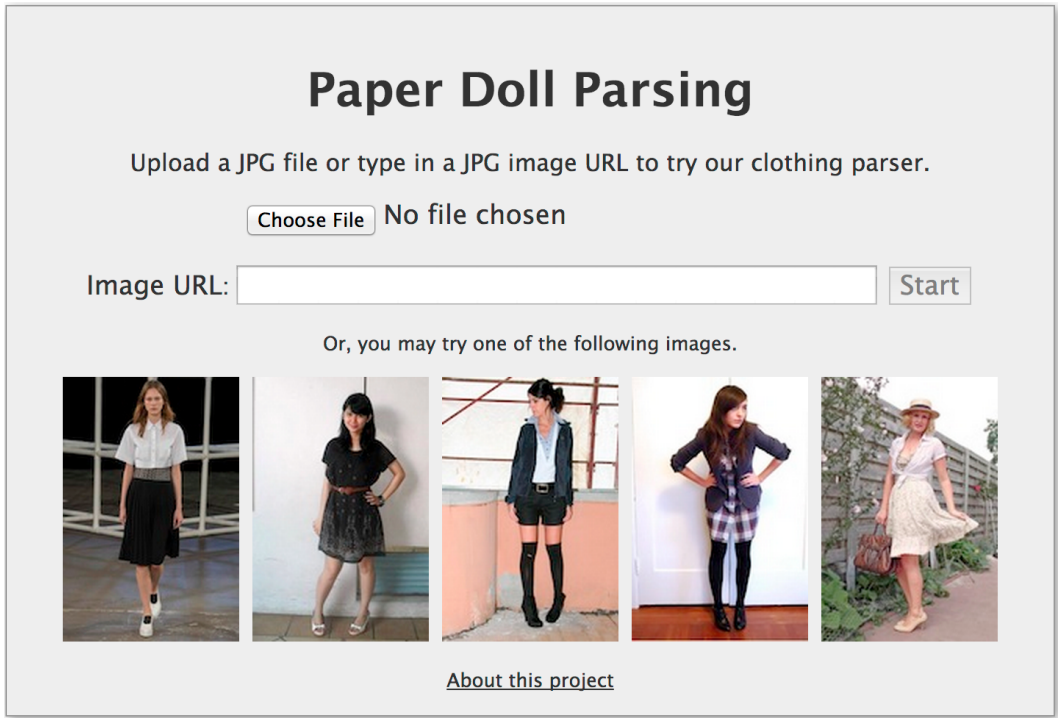



Photosynth, `https://photosynth.net/` (try it!)

# Why study Computer Vision?

- Figure out what people are wearing




`http://clothingparsing.com` (try it!)

# Why study Computer Vision?

- Crazy media attention!!!

New Scientist

Quartz

Tech Times

Wired, UK

Mashable

AOL News (video)

Huffington Post, UK (video)

Huffington Post, Canada

MSN, Canada

Protein

Yahoo, Canada

Science Daily

Daily Mail, UK

PSFK

Toronto Star

Gizmag

TheRecord.com

iDigitalTimes

Scientific Computing

Bustle

Harper's Bazaar

Glamour

Elle

Cosmopolitan, UK

Marie Claire

Fashion Magazine

Yahoo style

Red Magazine, UK

The Pool, UK

FashionNotes

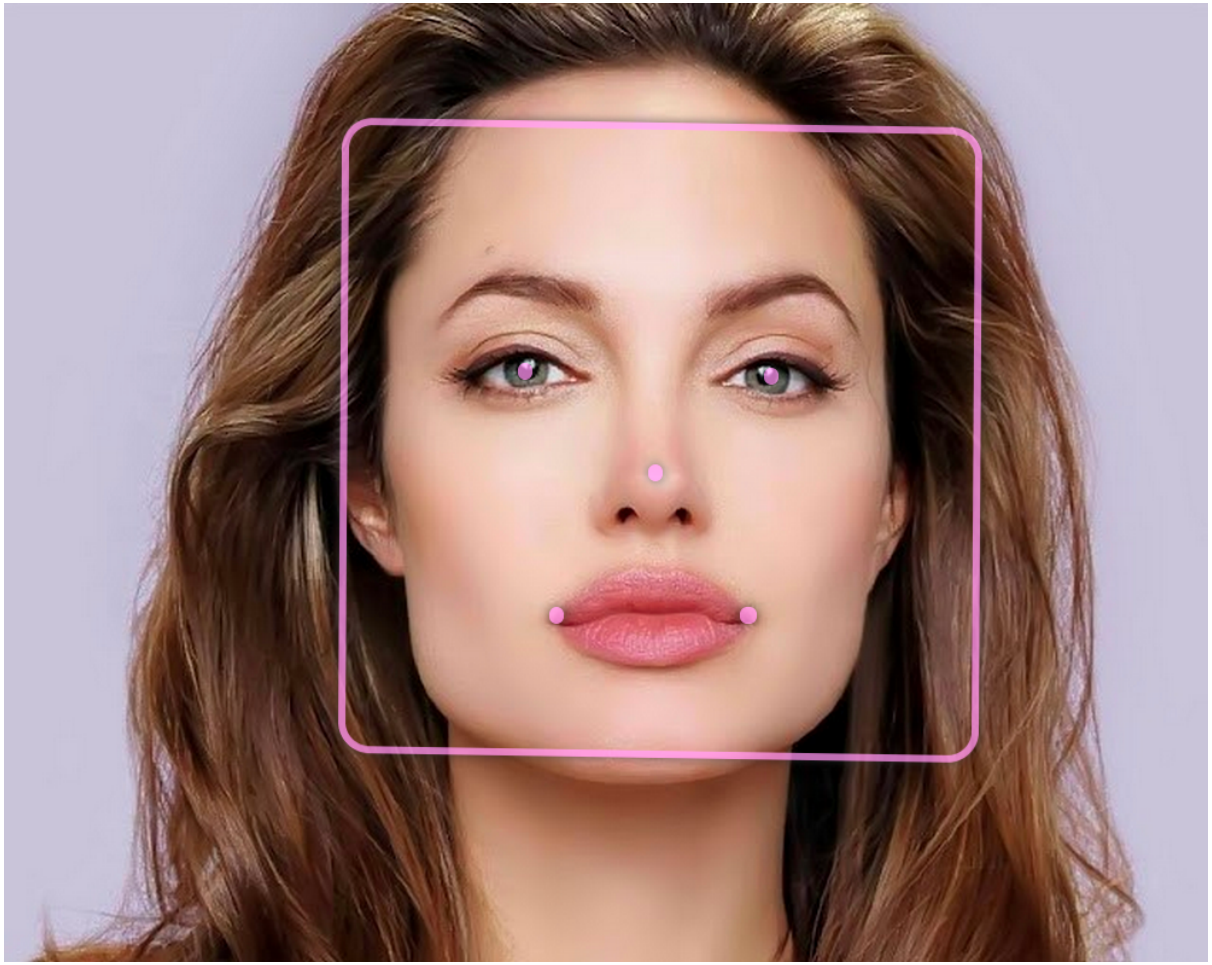
# Why study Computer Vision?

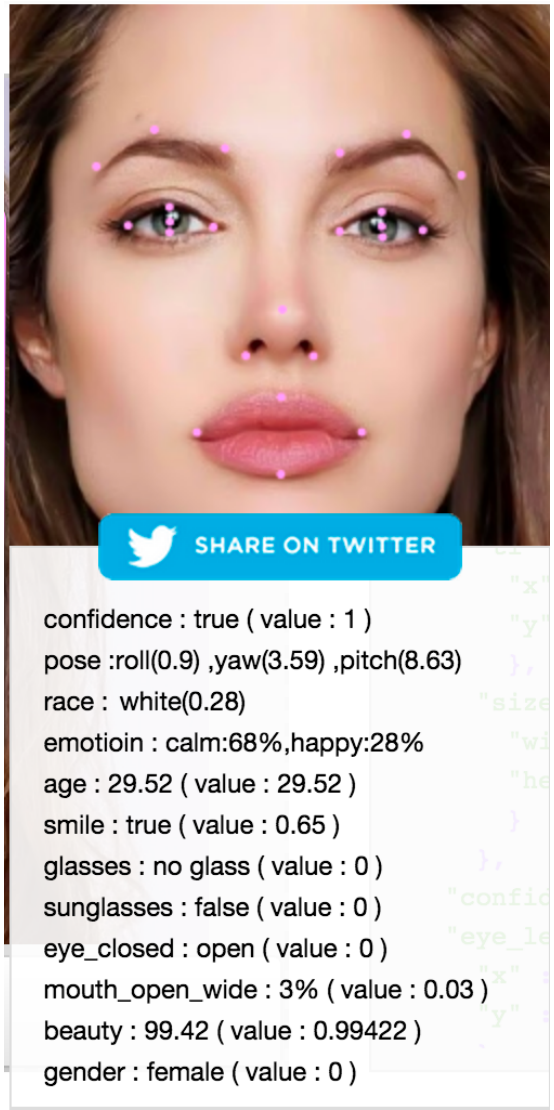
- Crazy media attention!!!


From Cosmopolitan: *The technology scores your facial attributes (this just keeps getting better, doesn't it) from your looks, to your age, and the emotion you're showing, before combining all the information* **using an equation SO complex we won't begin to go into it**.

# Why study Computer Vision?

- Detect and analyze faces

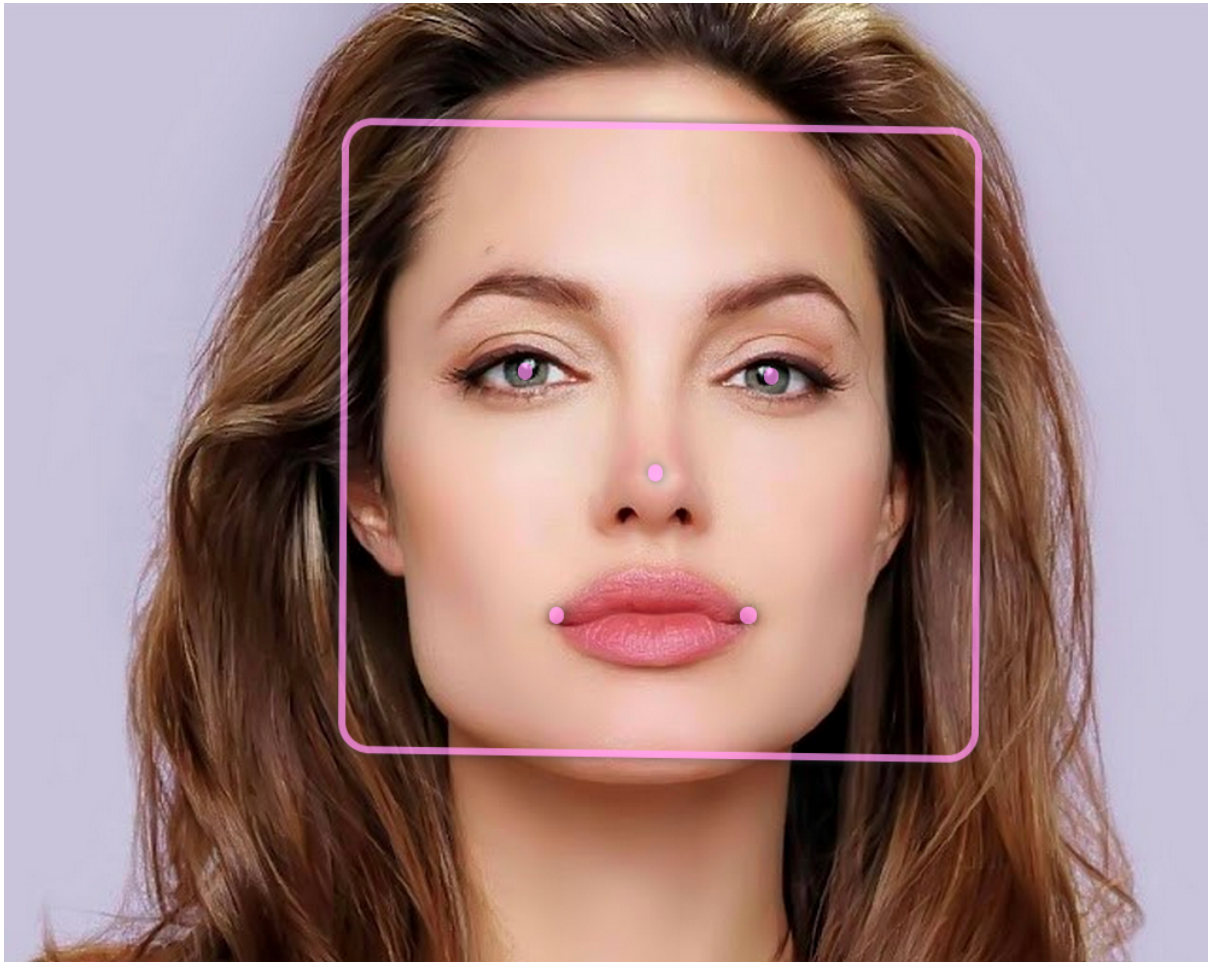


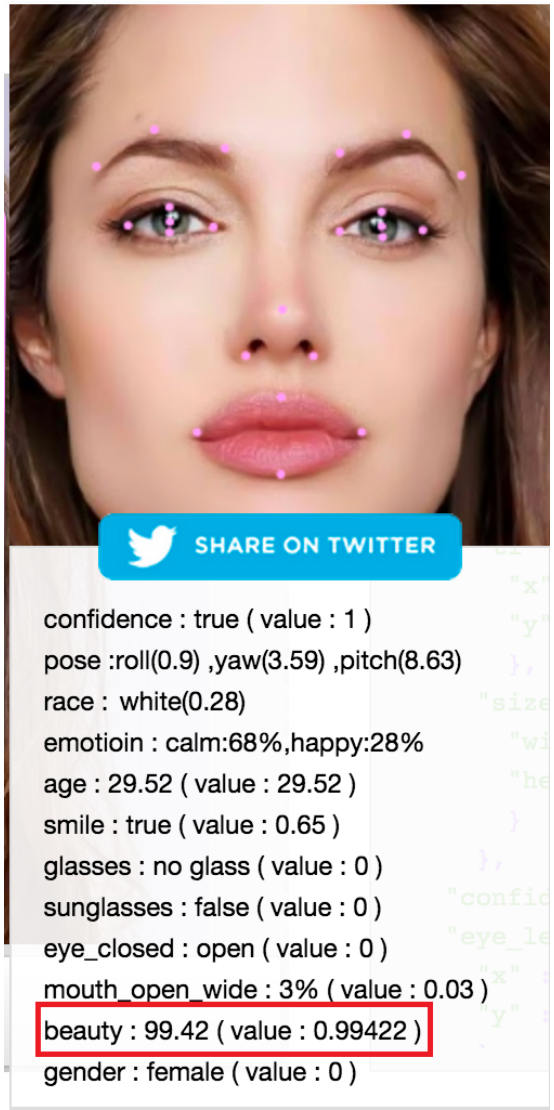



`http://www.rekognition.com` (try it!)

# Why study Computer Vision?

- Detect and analyze faces

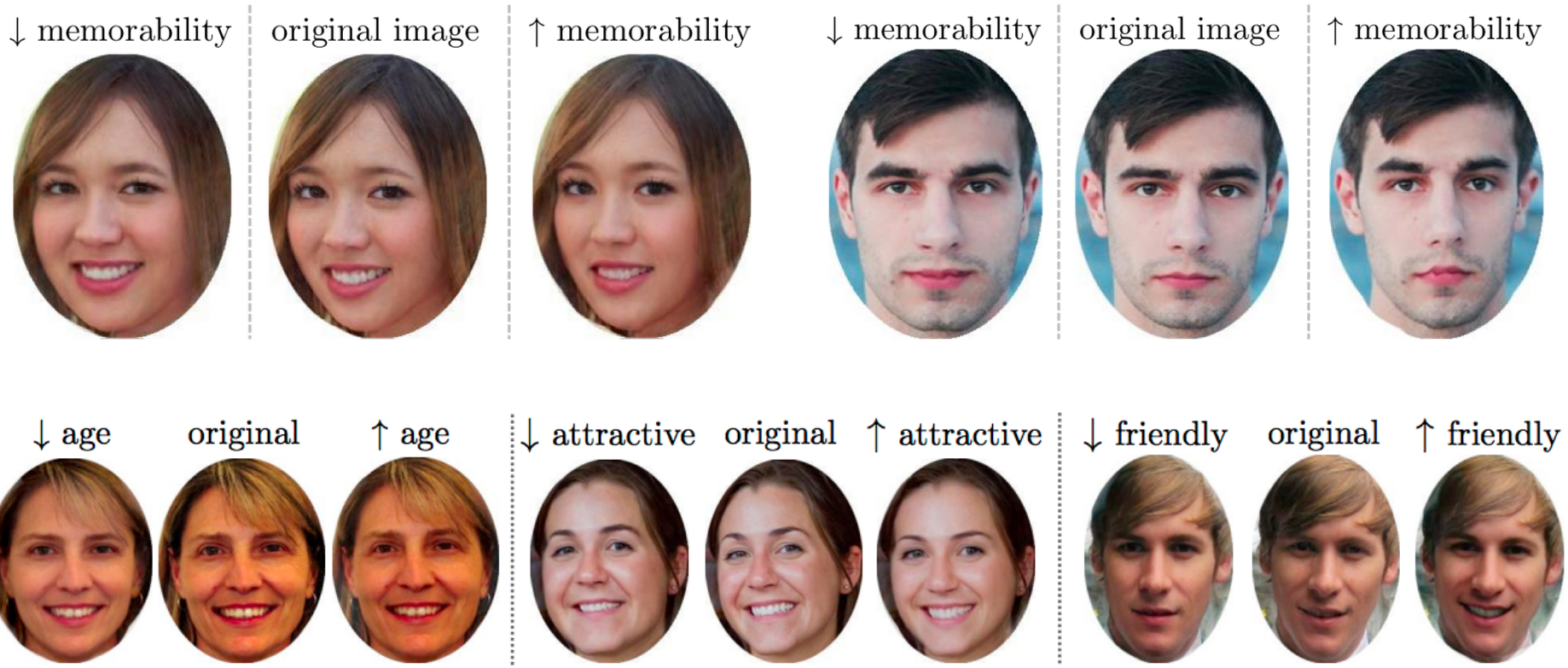





`http://www.rekognition.com` (try it!)

# Why study Computer Vision?

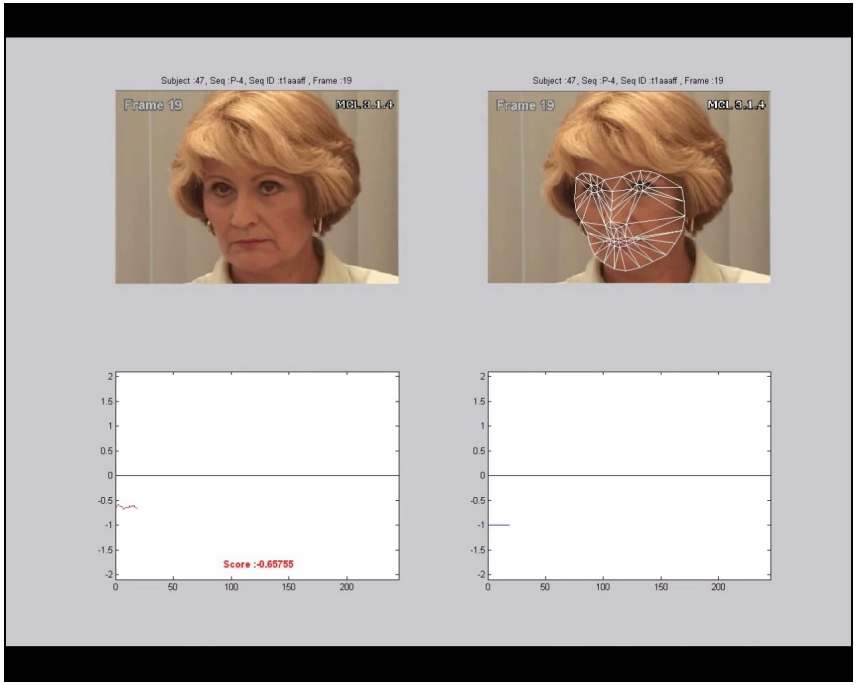
- You can make yourself look better (and competitors worse)




[Khosla, Bainbridge, Oliva, Torralba, Modifying the Memorability of Face Photographs, ICCV 2013]

# Why study Computer Vision?

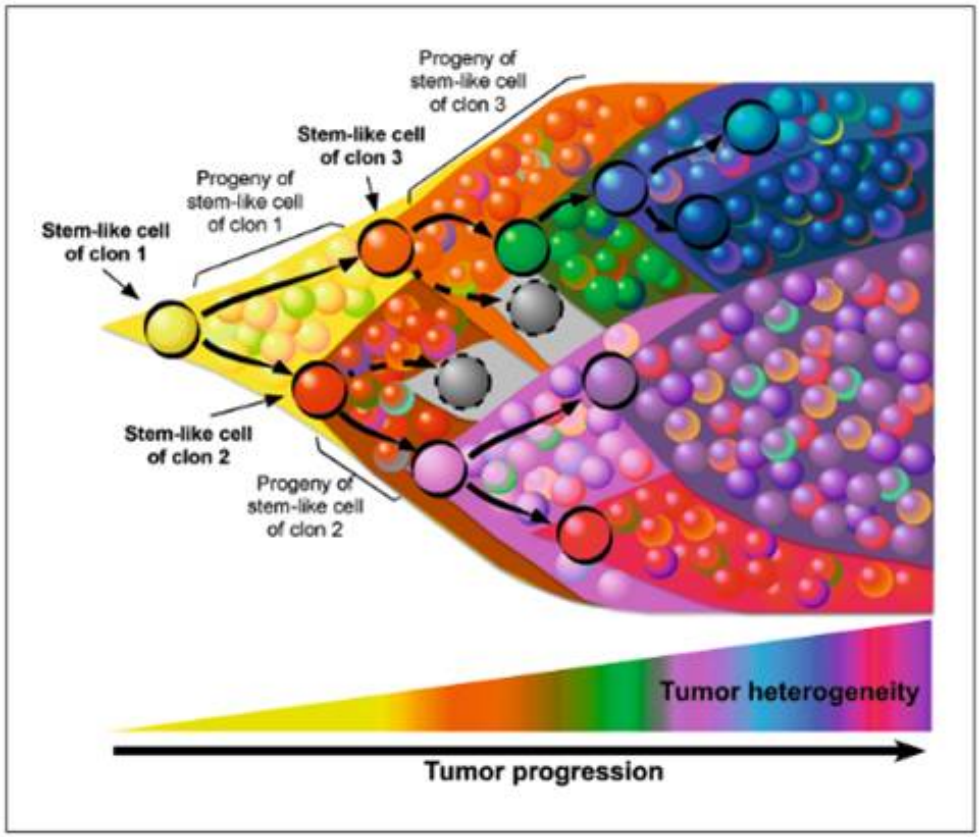
- An Eye for Pain (video)




Ashraf et al's Pain Detection System

# Why study Computer Vision?

- Beyond Selective Tissue Sampling in Breast Cancer




Ashraf et al's (2012) Identifying Intrinsic Cancer Imaging Phenotypes

# Why study Computer Vision?

- Generate image captions automatically


[Source: L. Zitnick, NIPS'14 Workshop on Learning Semantics]

# Why study Computer Vision?

- Generate image captions automatically

A man with a colorful umbrella walking down a street.


[Source: L. Zitnick, NIPS'14 Workshop on Learning Semantics]

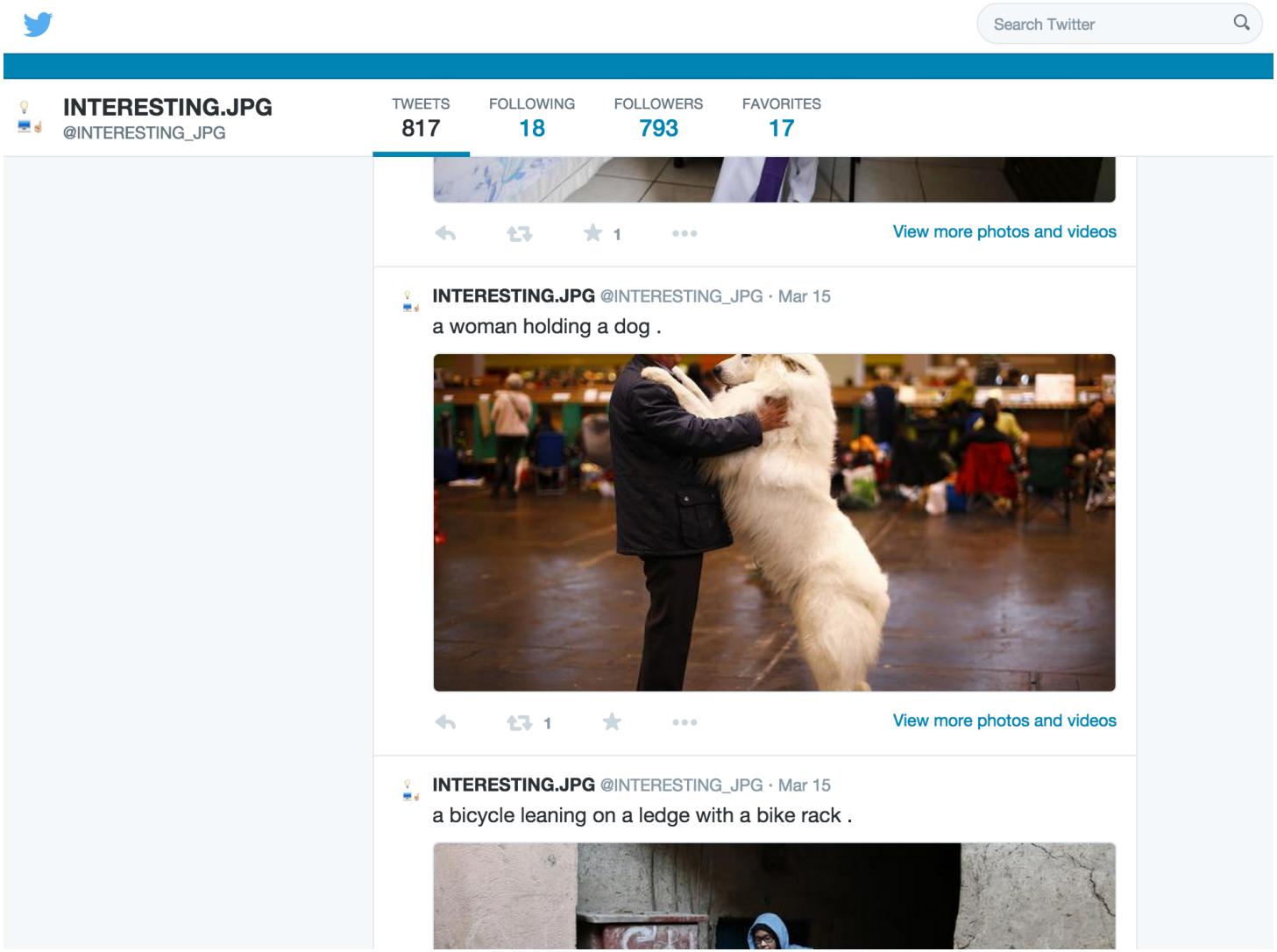
# Why study Computer Vision?

- Generate image captions automatically


[Source: L. Zitnick, NIPS'14 Workshop on Learning Semantics]

# Why study Computer Vision?

- Generate image captions automatically




[Kiros, Salakhutdinov, Zemel. Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models. 2014]

# Why study Computer Vision?

- Have a computer do math for you

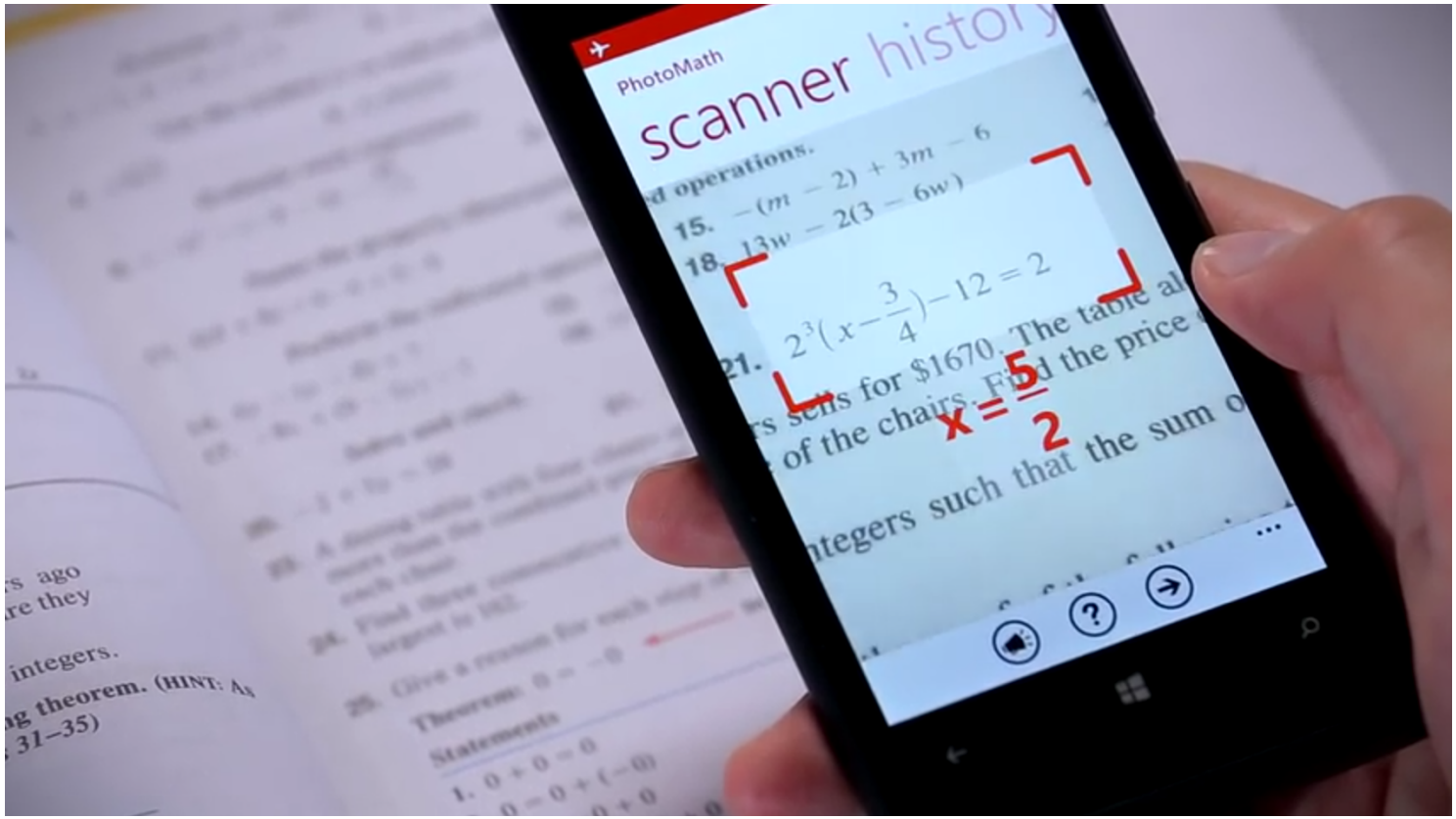
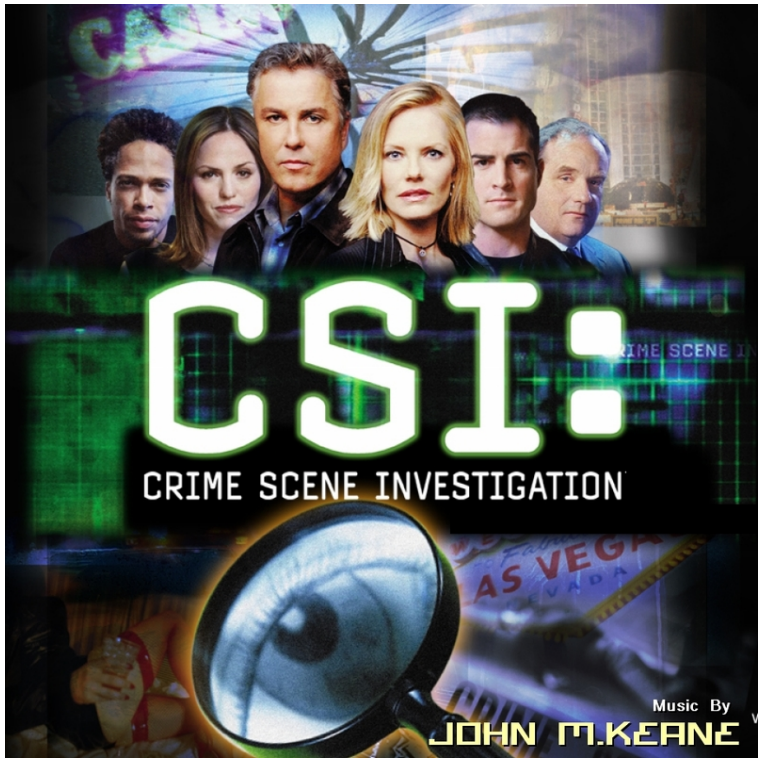
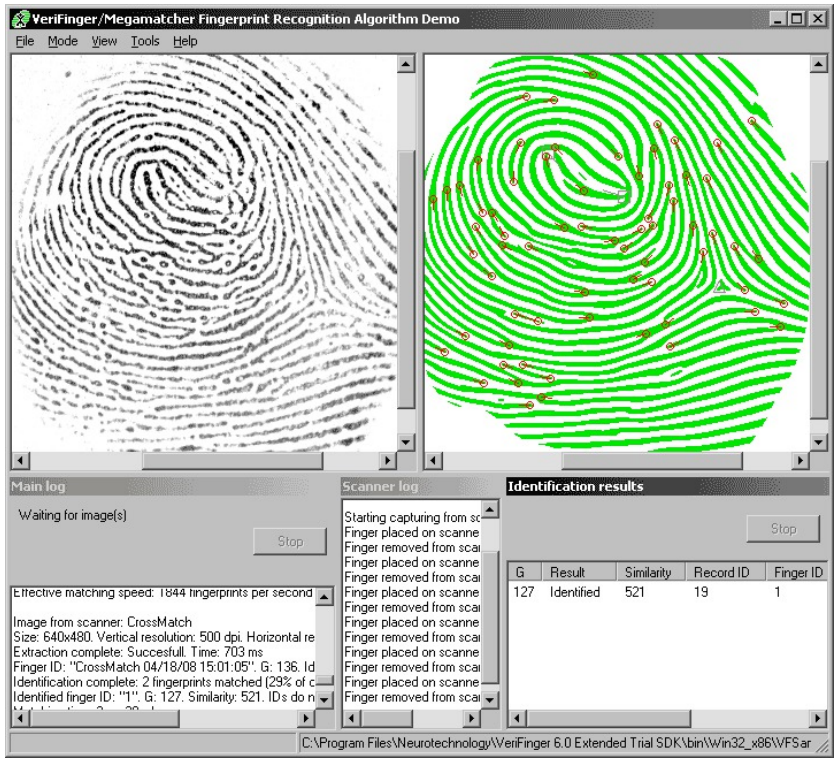



Figure: Photomath: https://photomath.net/

# Why study Computer Vision?

- Fingerprint recognition




[Source: S. Lazebnik]

# Why study Computer Vision?

- You can do some movie-like Forensics

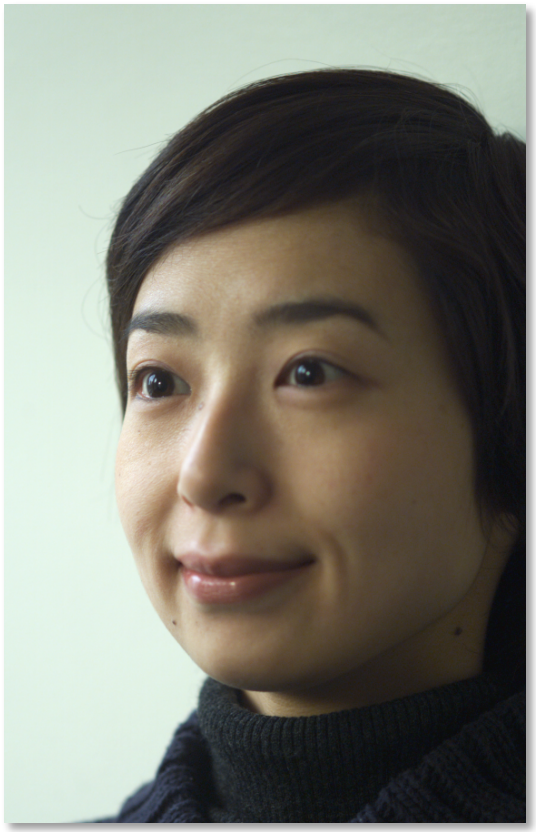


Figure: Source: Nayar and Nishino, Eyes for Relighting

[Source: N. Snavely]

# Why study Computer Vision?


[Source: N. Snavely]

# Why study Computer Vision?

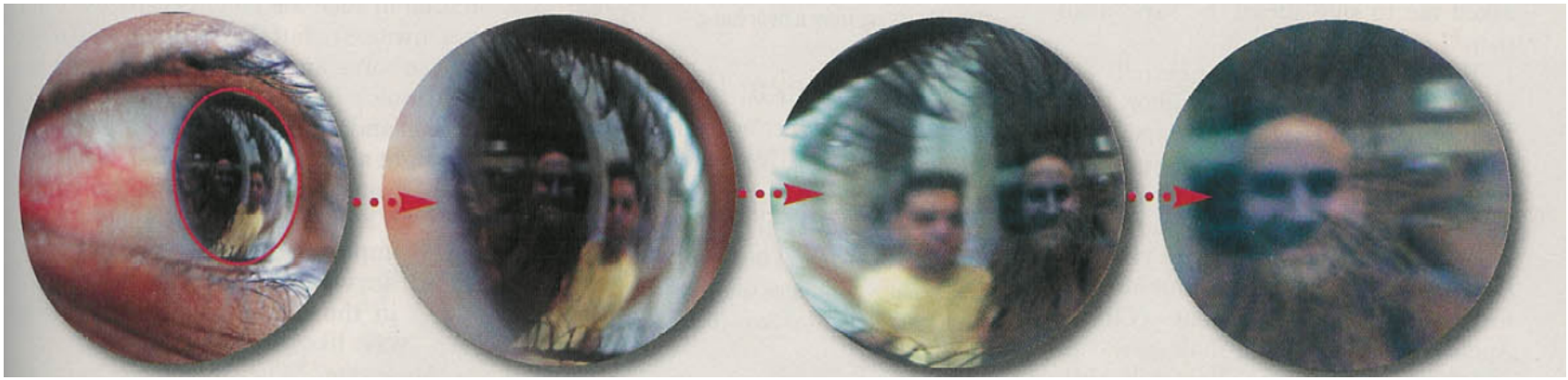


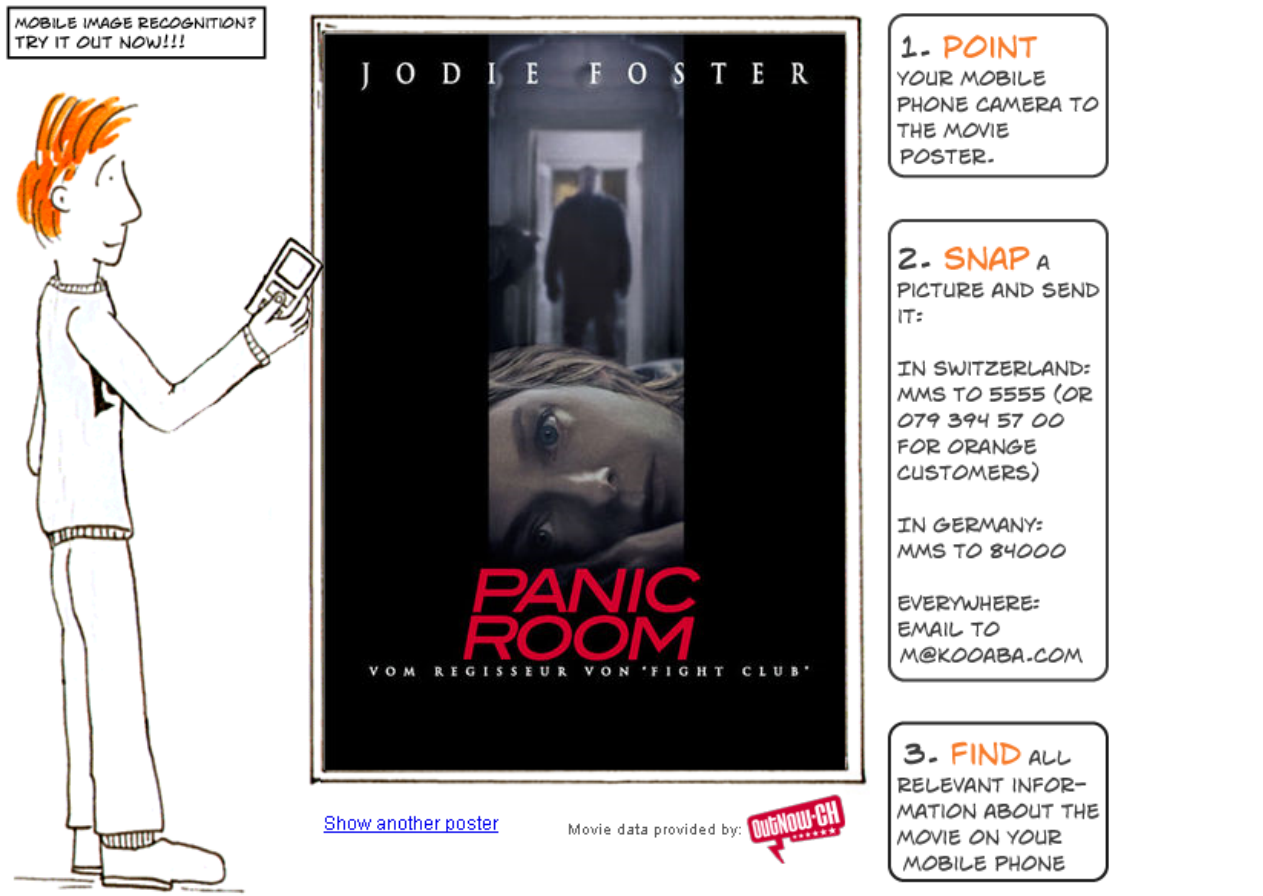
Figure: Source: Nayar and Nishino, Eyes for Relighting

[Source: N. Snavely]

# Why study Computer Vision?

- Recognizing movie posters (in mobile phones)

iPhone Apps: kooaba (www.kooaba.com)




**Source: S. Lazebnik**

# Why study Computer Vision?

- Games, games & games: 3D Pose Estimation with Depth Sensors


[Source: Microsoft Kinect]

# How It All Began...

# How It All Began...

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".

[Slide credit: A. Torralba]

# 50 years and thousands of PhDs later...

**Popular benchmarks**:

The KITTI Vision Benchmark Suite
A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

## Car

| | Method | Setting | Code | Moderate | Easy | Hard | Runtime | Environment | Compare |
|---|---|---|---|---|---|---|---|---|---|
| 1 | DenseBox2 | | | 89.32 % | 93.94 % | 79.81 % | 5 s | GPU @ 2.5 Ghz (C/C++) | |
| 2 | DJML | | | 88.79 % | 91.31 % | 77.73 % | x s | GPU @ 1.5 Ghz (Matlab + C/C++) | |
| 3 | 3DOP | | | 88.64 % | 93.04 % | 79.10 % | 3s | GPU @ 2.5 Ghz (Matlab + C/C++) | |

X. Chen, K. Kundu, Y. Zhu, A. Berneshawi, H. Ma, S. Fidler and R. Urtasun: 3D Object Proposals for Accurate Object Class Detection. NIPS 2015.

## Cyclist

| | Method | Setting | Code | Moderate | Easy | Hard | Runtime | Environment | Compare |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3DOP | | | 68.94 % | 78.39 % | 61.37 % | 3s | GPU @ 2.5 Ghz (Matlab + C/C++) | |

X. Chen, K. Kundu, Y. Zhu, A. Berneshawi, H. Ma, S. Fidler and R. Urtasun: 3D Object Proposals for Accurate Object Class Detection. NIPS 2015.

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Regionlets | | | 58.72 % | 70.41 % | 51.83 % | 1 s | >8 cores @ 2.5 Ghz (C/C++) | |

X. Wang, M. Yang, S. Zhu and Y. Lin: Regionlets for Generic Object Detection. International Conference on Computer Vision 2013.
C. Long, X. Wang, G. Hua, M. Yang and Y. Lin: Accurate Object Detection with Location Relaxation and Regionlets Relocalization. Asian Conference on Computer Vision 2014.

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | MV-RGBD-RF | | | 42.61 % | 52.97 % | 37.42 % | 4 s | 4 cores @ 2.5 Ghz (C/C++) | |

A. Gonzalez, G. Villalonga, J. Xu, D. Vazquez, J. Amores and A. Lopez: Multiview Random Forest of Local Experts Combining RGB and LIDAR data for Pedestrian Detection. IEEE Intelligent Vehicles Symposium (IV) 2015.

| | mean | aero plane | bicycle | bird | boat | bottle | bus | car | cat | chair | cow | dining table | dog | horse | motor bike | person | potted plant | sheep | sofa | train | tv/ monitor | submission date |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fast R-CNN + YOLO [?] | 70.8 | 82.7 | 77.7 | 74.3 | 59.1 | 47.1 | 78.0 | 73.1 | 89.2 | 49.6 | 74.3 | 55.9 | 87.4 | 79.8 | 82.2 | 75.3 | 43.1 | 71.4 | 67.8 | 81.9 | 65.6 | 05-Jun-2015 |
| Fast R-CNN VGG16 extra data [?] | 68.8 | 82.0 | 77.8 | 71.6 | 55.3 | 42.4 | 77.3 | 71.7 | 89.3 | 44.5 | 72.1 | 53.7 | 87.7 | 80.0 | 82.5 | 72.7 | 36.6 | 68.7 | 65.4 | 81.1 | 62.7 | 18-Apr-2015 |
| segDeepM [?] | 67.2 | 82.3 | 75.2 | 67.1 | 50.7 | 49.8 | 71.1 | 69.6 | 88.2 | 42.5 | 71.2 | 50.0 | 85.7 | 76.6 | 81.8 | 69.3 | 41.5 | 71.9 | 62.2 | 73.2 | 64.6 | 29-Jan-2015 |
| BabyLearning [?] | 63.8 | 77.7 | 73.8 | 62.3 | 48.8 | 45.4 | 67.3 | 67.0 | 80.3 | 41.3 | 70.8 | 49.7 | 79.5 | 74.7 | 78.6 | 64.5 | 36.0 | 69.9 | 55.7 | 70.4 | 61.7 | 12-Nov-2014 |

# 50 years and thousands of PhDs later...

- Algorithms work **pretty** well
- Still some embarrassing mistakes...
- The general vision problem is not yet solved

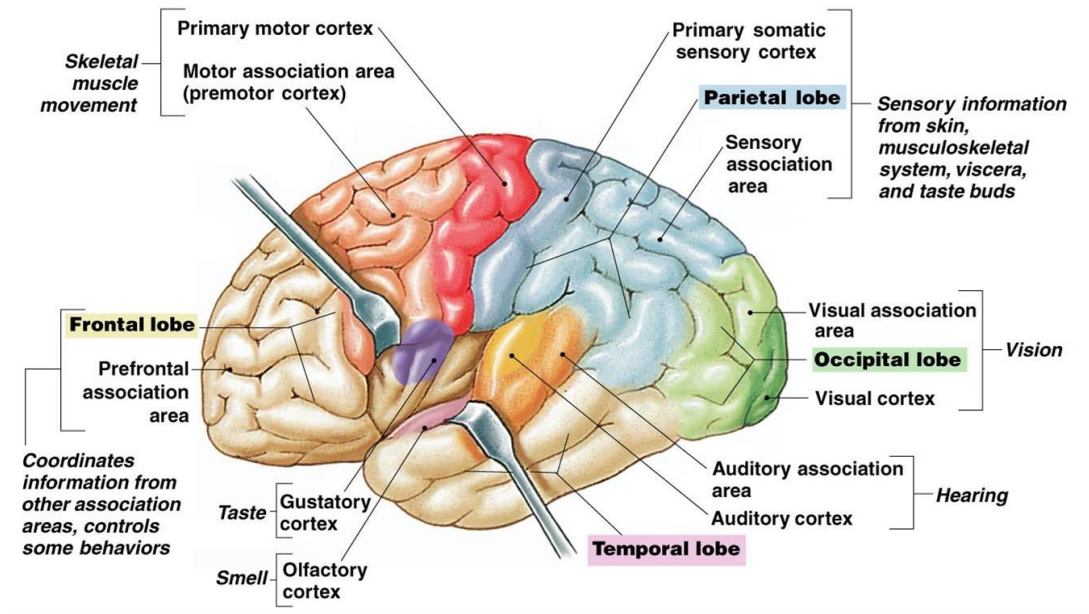
Where pink means “person”

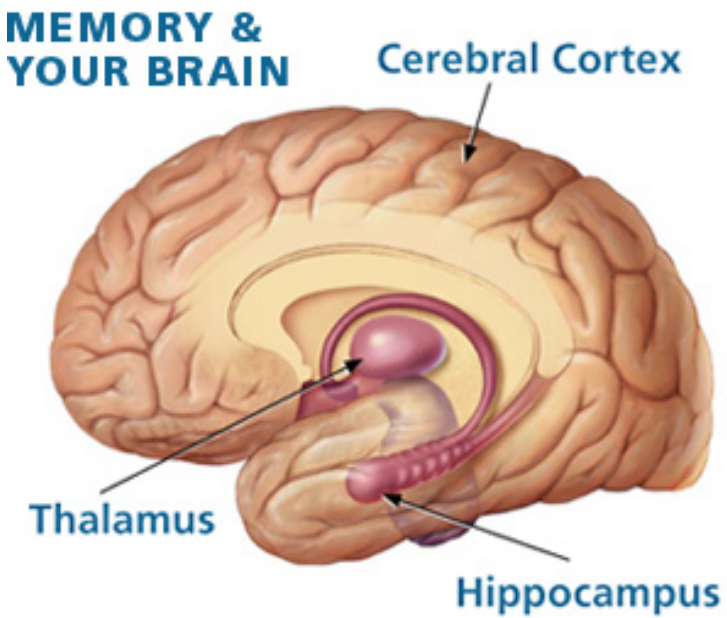
[This pic is from 2014]

# Why is vision hard?

# Why is vision hard?

- Half of the cerebral cortex in primates is devoted to processing visual information. This is a lot. Means that vision has to be pretty hard!

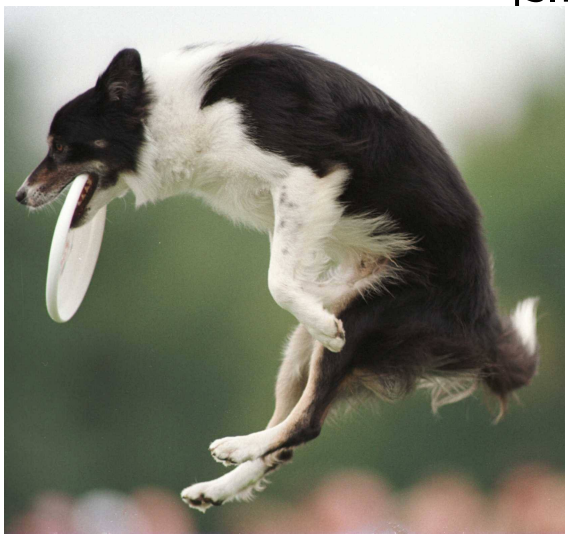


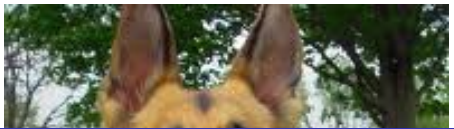



# Why is vision hard?

All this is dog...

[slide adopted from: R. Urtasun]

# Why is vision hard?

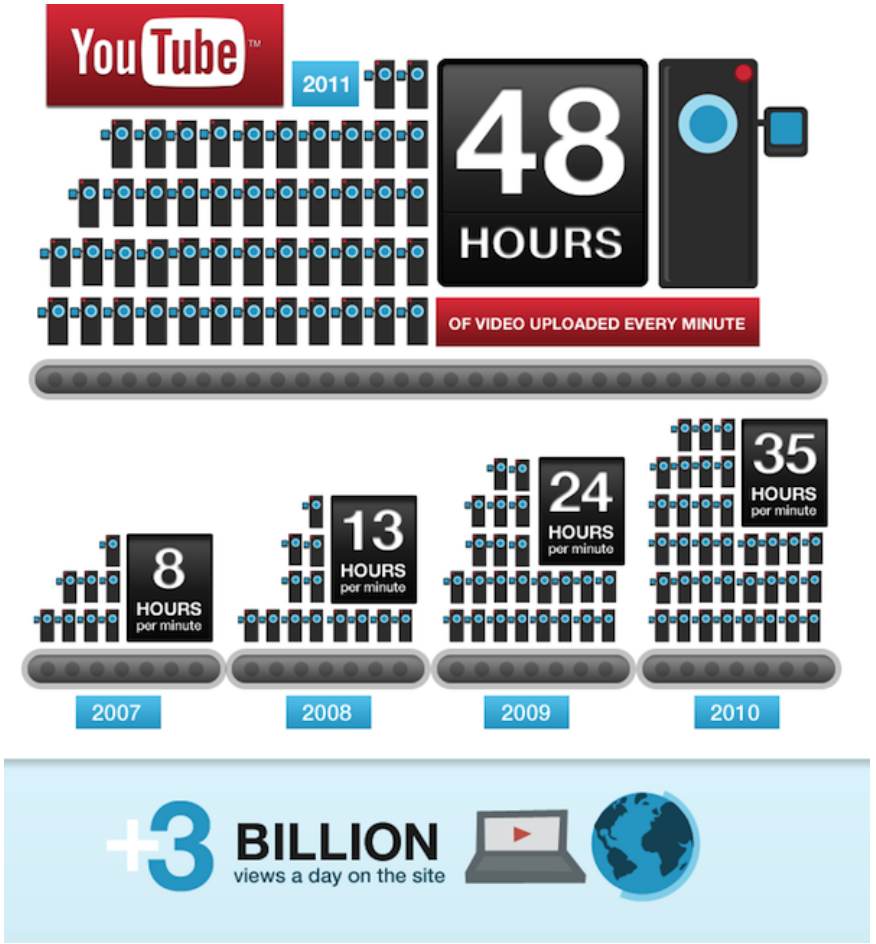

Biederman, 1987

[slide credit: R. Urtasun]

# Why is vision hard?

Lots of data to process:

- Thousands to millions of pixels in an image
- 100 hours of video added to YouTube per minute [source: YouTube]
- Over 6 billion hours of video are watched each month on YouTube – almost an hour for every person on Earth [source: YouTube]

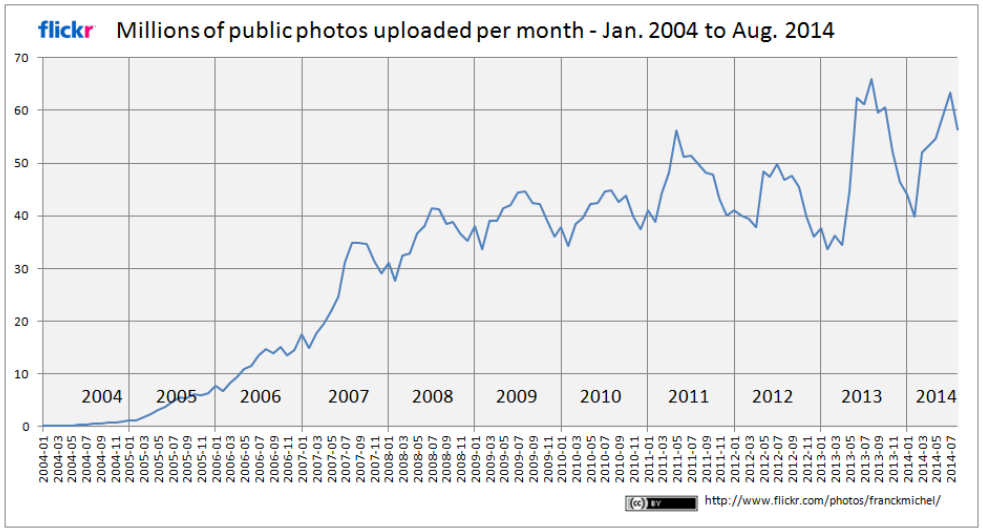


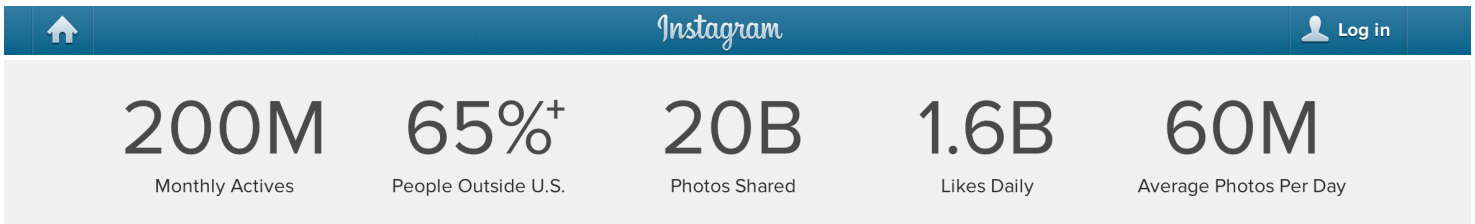
# Why is vision hard?

Lots of data to process:

- $\sim$ 5000 new tagged photos added to Flickr per minute (7M per day)
- $\sim$ 60M photos uploaded to Instagram every day [source: Instagram]

**How many photos are uploaded to Flickr every day, month, year?**

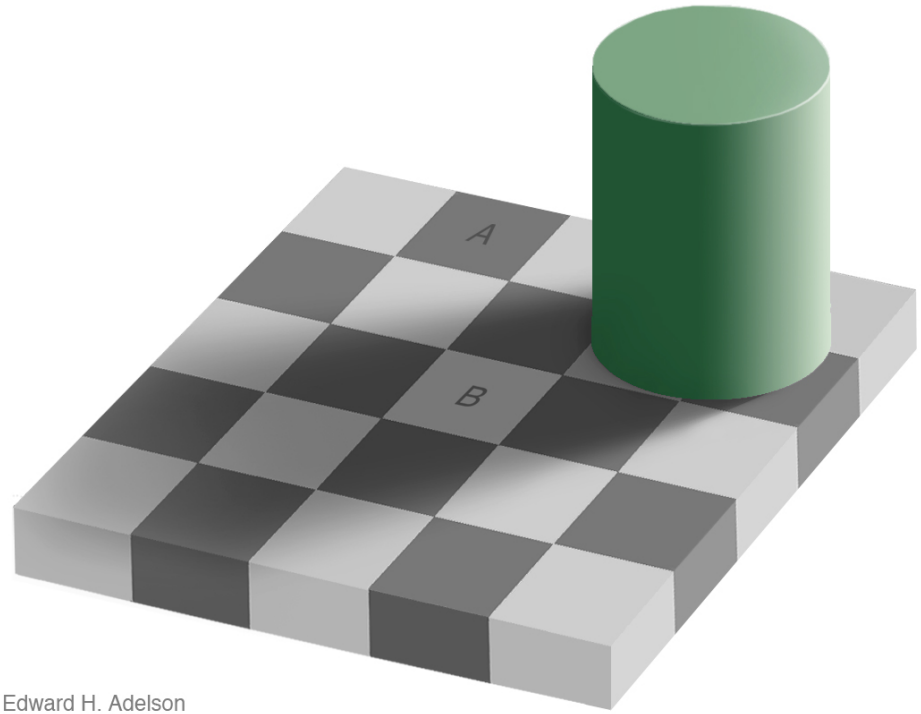


# Why is vision hard?

- Human vision seems to work quite well.
- How well does it really work?
- Let's play some games!

# How good are humans?

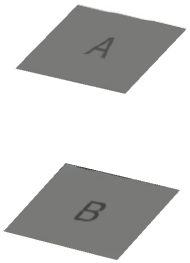



Edward H. Adelson

- Which square is lighter, A or B?

[Slide credit: A. Torralba]

# How good are humans?




Edward H. Adelson

- Which square is lighter, A or B?

[Slide credit: A. Torralba]

# How good are humans?

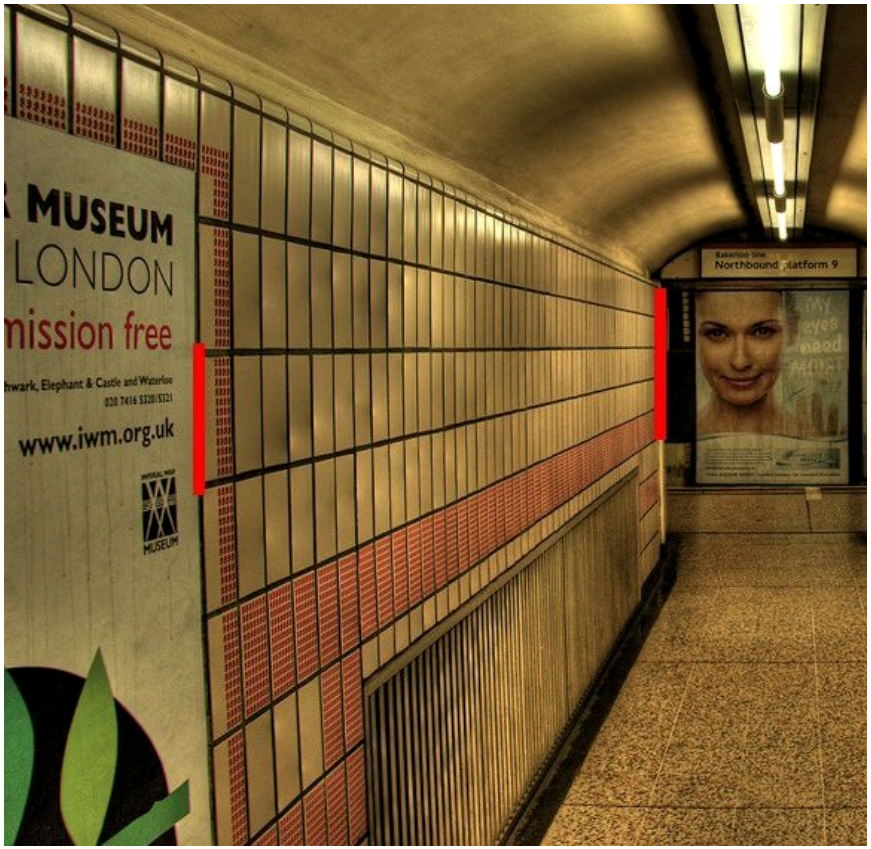



Figure: 2006 Walt Anthony

- Which red line is longer?

[Slide credit: A. Torralba]

# How good are humans?

Figure: 2006 Walt Anthony

- Which red line is longer?

[Slide credit: A. Torralba]

# How good are humans?

Figure: Chabris & Simons

- Count the number of times the white team pass the ball
- Concentrate, it's difficult!

# How good are humans?

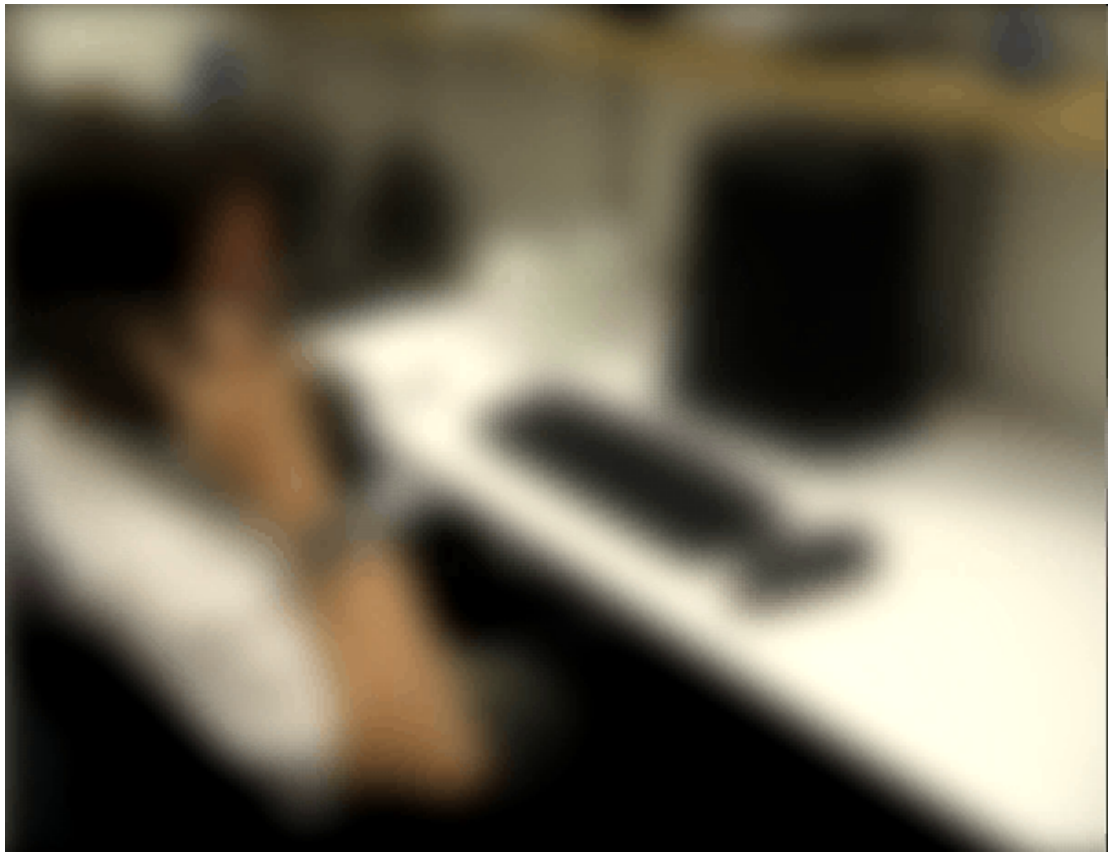


Figure: Torralba et al.

- Can you describe what's going on in the video?

# How good are humans?

Figure: Torralba et al.

- Can you describe what's going on in the video?

# What do I need...

What do I need to become a good Computer Vision researcher?

- Some math knowledge
- Good programming skills
- Imagination
- Even better intuition
- Lots of persistence
- Some luck always helps