CSC420: Intro to Image Understanding Introduction

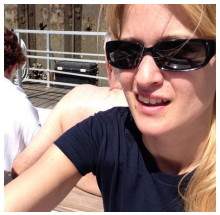
Sanja Fidler

January 8, 2018

UNIVERSITY OF TORONTO

The Team

- **Instructor**:



Sanja Fidler (fidler@cs.toronto.edu)

- **Office**: DH 3094
- **Office hours**: Monday 3-4pm, or by appointment
- **TAs**:

Amlan Kar (amlan@cs.toronto.edu)

Hang Chu (chuhang1122@gmail.com)

Course Information

- **Class time**: Monday at 1-3pm
- **Location**: CC 3150
- **Tutorials**: TUT0101 on Monday 4-5pm (IB 340), TUT0102 on Monday 5-6pm (IB 320), demos and Q&A, we'll do it on demand
- **Class Website**:

  http://www.cs.toronto.edu/~fidler/teaching/2018/CSC420.html

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

  https://piazza.com/utoronto.ca/winter2018/csc420

- 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

- **Class time**: Monday at 1-3pm
- **Location**: CC 3150
- **Tutorials**: TUT0101 on Monday 4-5pm (IB 340), TUT0102 on Monday 5-6pm (IB 320), demos and Q&A, we'll do it on demand
- **Class Website**:

  http://www.cs.toronto.edu/~fidler/teaching/2018/CSC420.html

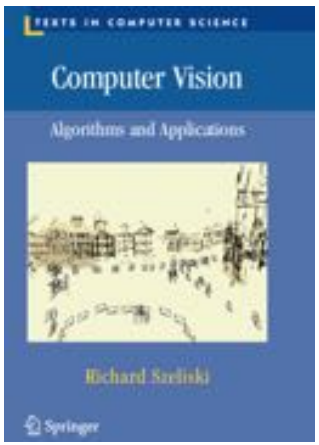
- The class will use Piazza for **announcements** and **discussions**:

  https://piazza.com/utoronto.ca/winter2018/csc420

- 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

- Each student expected to complete 4 assignments and a project
- **Assignments:**
  - Short **theoretical questions** and **programming exercises**
  - Will be given roughly every **two weeks** (starting second week of class)
  - You will have **a week to hand in the solution** to each assignment
  - You need to solve the assignment **alone**

Requirements

- Each student expected to complete 4 assignments and a project
- **Assignments:**
  - Short **theoretical questions** and **programming exercises**
  - Will be given roughly every **two weeks** (starting 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 **individually** or in **pairs**

Requirements

- Each student expected to complete 4 assignments and a project
- **Assignments:**
  - Short **theoretical questions** and **programming exercises**
  - Will be given roughly every **two weeks** (starting 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 **individually** or in **pairs**

Grading

- **Grade breakdown**
  - **Assignments**: 60% (15% each)
  - **Project**: 40%
- For the project you will need to hand in a:
  - Short project proposal
  - Project report
  - Project presentation (oral)
- I will be asking questions about relevant part of the material during project presentations

Term Work Dates

Term Work	Post Date	Due Date
Assignment 1	Jan 17	Jan 24
Assignment 2	Jan 31	Feb 7
Assignment 3	Feb 14	Feb 21
Assignment 4	Mar 7	Mar 14
Project Report		First week of April
Project Presentation		First week of April

- All dates are for 2018. ;)
- Dates are approximate

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 and Python
- 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	Jan 8	Intro
2	Jan 15	Linear filters, edges
3	Jan 22	Image features
4	Jan 29	Keypoint detection
5	Feb 5	Matching
6	Feb 12	Grouping
7	Feb 19	Stereo, multi-view
8	Feb 26	Stereo, multi-view
9	March 5	Object recognition
10	March 12	Object detection
11	March 19	Neural Networks
12	March 26	Segmentation
13	April ?	Project Presentations

Introduction

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"**

What is Computer Vision?

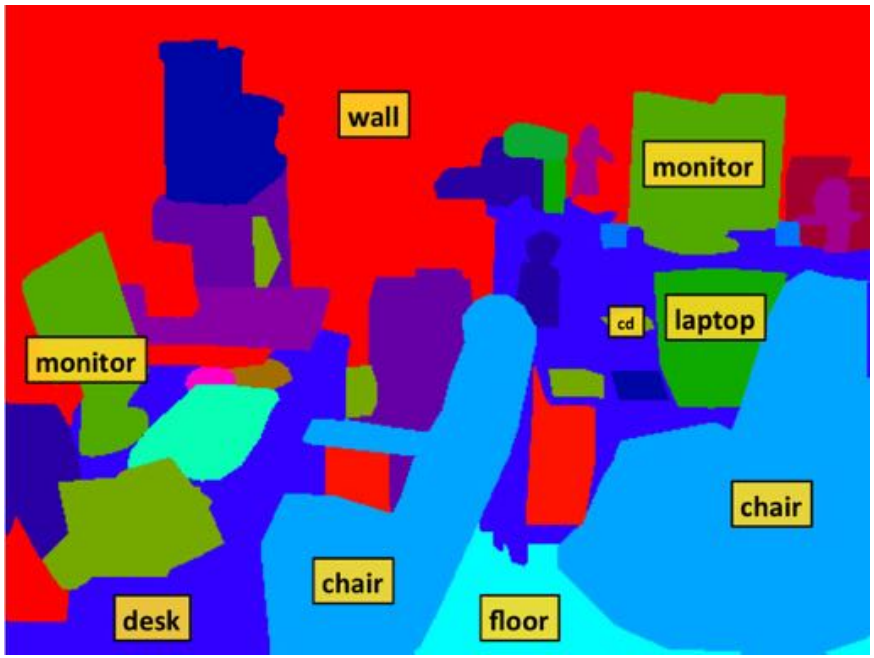
[text adopted from A. Torralba]

- 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

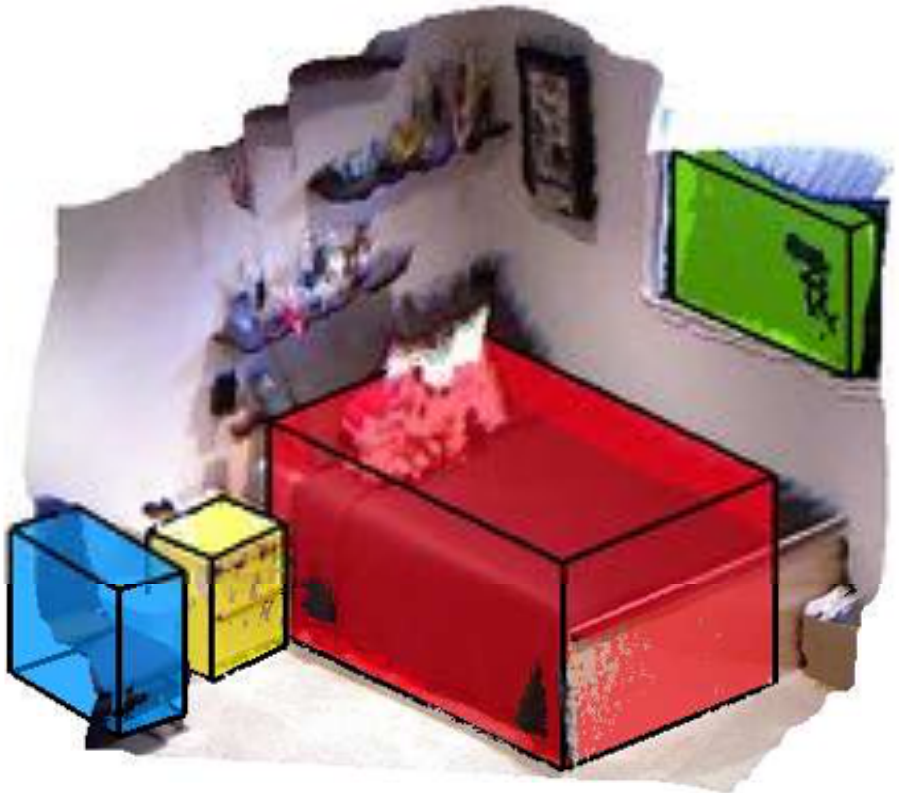


What is Computer Vision?

[text adopted from A. Torralba]

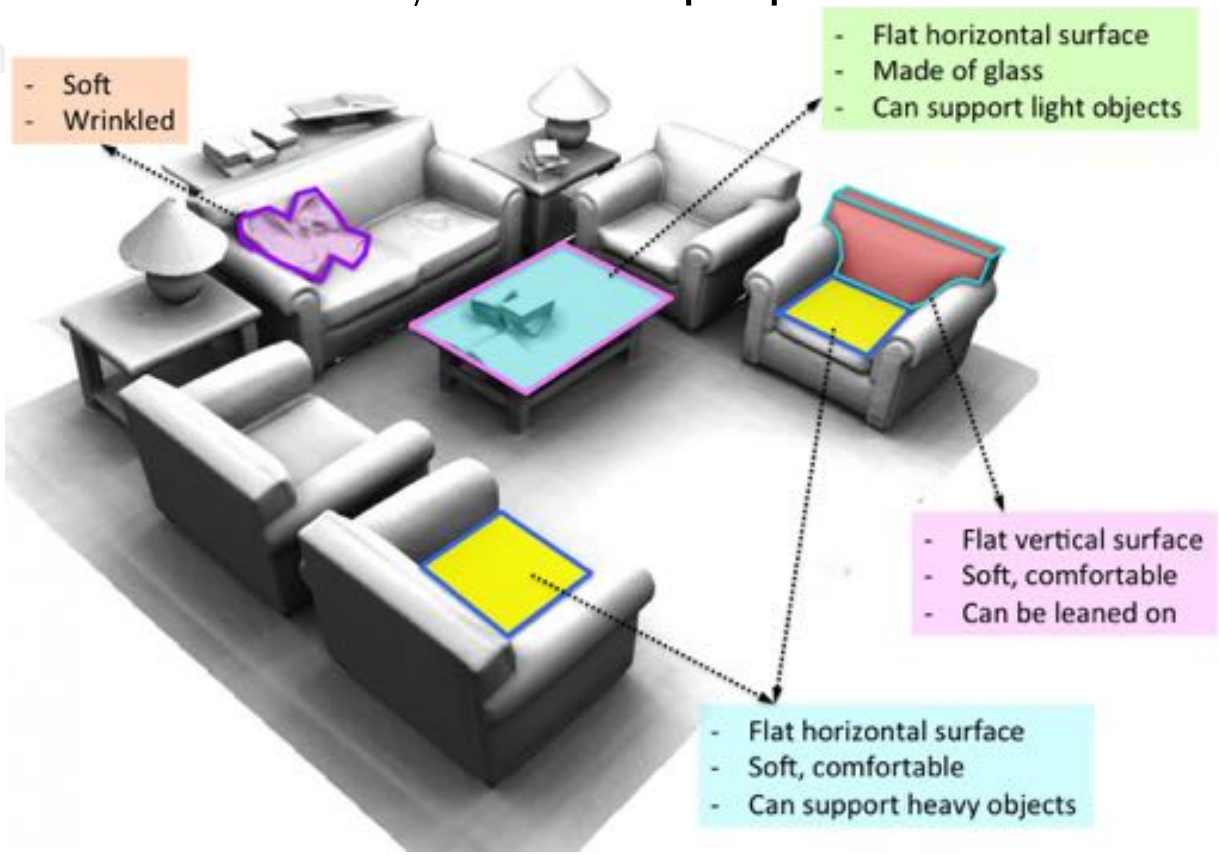
- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D/material properties?

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/material 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/material 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 it is challenging and fun

Jialiang Wang's (4th undergraduate year, UofT) video about his summer research in computer vision

Video: http://www.cs.toronto.edu/~fidler/vcompvideos/Research_Video_JW.mp4

Why study Computer Vision?

- You are curious how to one day make the robot walk your dog

http://www.cs.toronto.edu/~fidler/videos/robotsmovies.mov

Why study Computer Vision?

- ... and fold your laundry

https://www.youtube.com/watch?v=gy5g33S0Gzo

https://www.youtube.com/watch?v=KKUaVzf3Oqw

Why study Computer Vision?

- ... and drive you to work

Amnon Shashua's Mobileye autonomous driving system

https://www.youtube.com/watch?v=4fxFDypHZLs

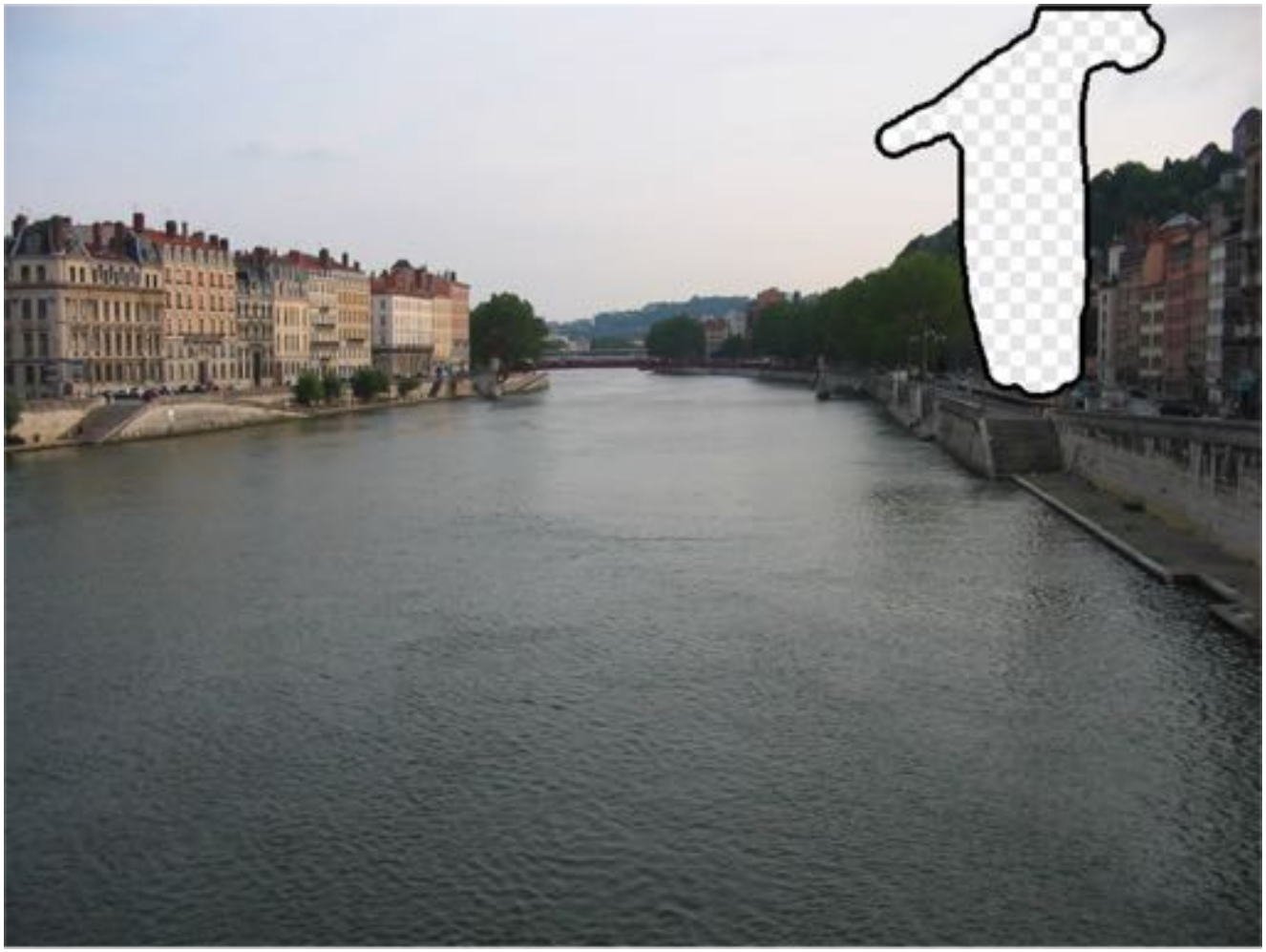
Why study Computer Vision?

- 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

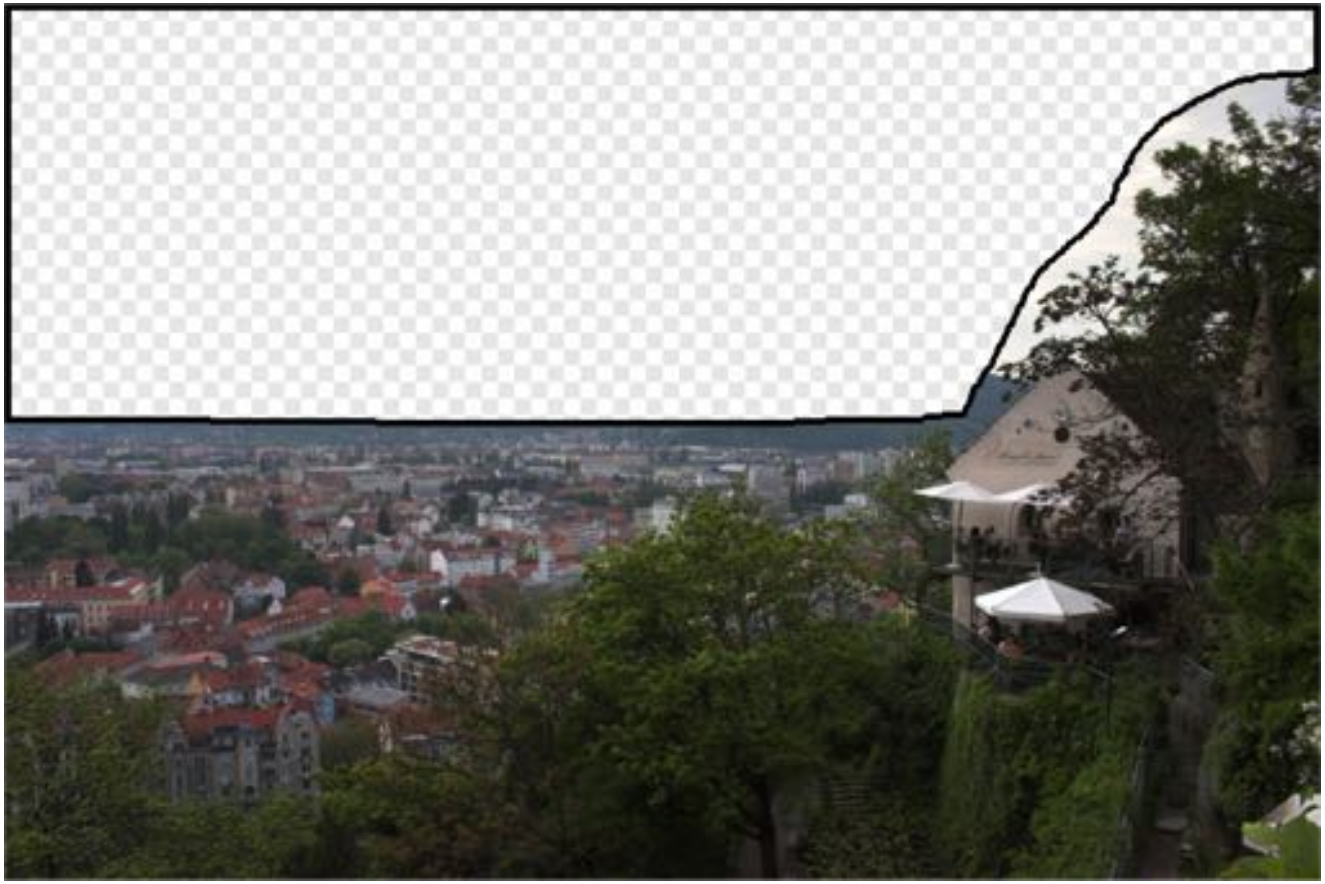
Why study Computer Vision?

- 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?

- Change style of videos

https://www.youtube.com/watch?v=Khuj4ASldmU

[Ruder, Dosovitskiy, Brox. Artistic style transfer for videos, 2016]

Why study Computer Vision?

- Change style of videos

Bringing Impressionism to Life with Neural Style Transfer in *Come Swim*

Bhautik J Joshi*
Research Engineer, Adobe

Kristen Stewart
Director, *Come Swim*

David Shapiro
Producer, Starlight Studios

Figure 1: *Usage of Neural Style Transfer in Come Swim; left: content image, middle: style image, right: upsampled result. Images used with permission, (c) 2017 Starlight Studios LLC & Kristen Stewart.*

Abstract

Neural Style Transfer is a striking, recently-developed technique that uses neural networks to artistically redraw an image in the style of a source style image. This paper explores the use of this technique in a production setting, applying Neural Style Transfer to redraw key scenes in *Come Swim* in the style of the impressionistic painting that inspired the film. We document how the technique can be driven within the framework of an iterative creative process to achieve a desired look, and propose a mapping of the broad parameter space to a key set of creative controls. We hope that this mapping can provide insights into priorities for future research.

execute efficiently and predictably. In a production setting, however, a great deal of creative control is needed to tune the result, and a rigid set of algorithmic constraints run counter to the need for this creative exploration. While early investigations to better map the low-level neural net evaluations to stylistic effects are underway [Li et al. 2017], in our paper we focused on examining the higher-level parameter space for Neural Style Transfer and found a set of working shortcuts to map them to a reduced but meaningful set of creative controls.

2 Realizing Directorial Intent

https://arxiv.org/pdf/1701.04928.pdf

Why study Computer Vision?

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

http://www.cs.cmu.edu/~om3d/

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

Why study Computer Vision?

- Fancy visualization and game analysis in sports

Why study Computer Vision?

- Fancy visualization and special effects in movies

[Source: http://cvfxbook.com and http://vimeo.com/100095868]

Why study Computer Vision?

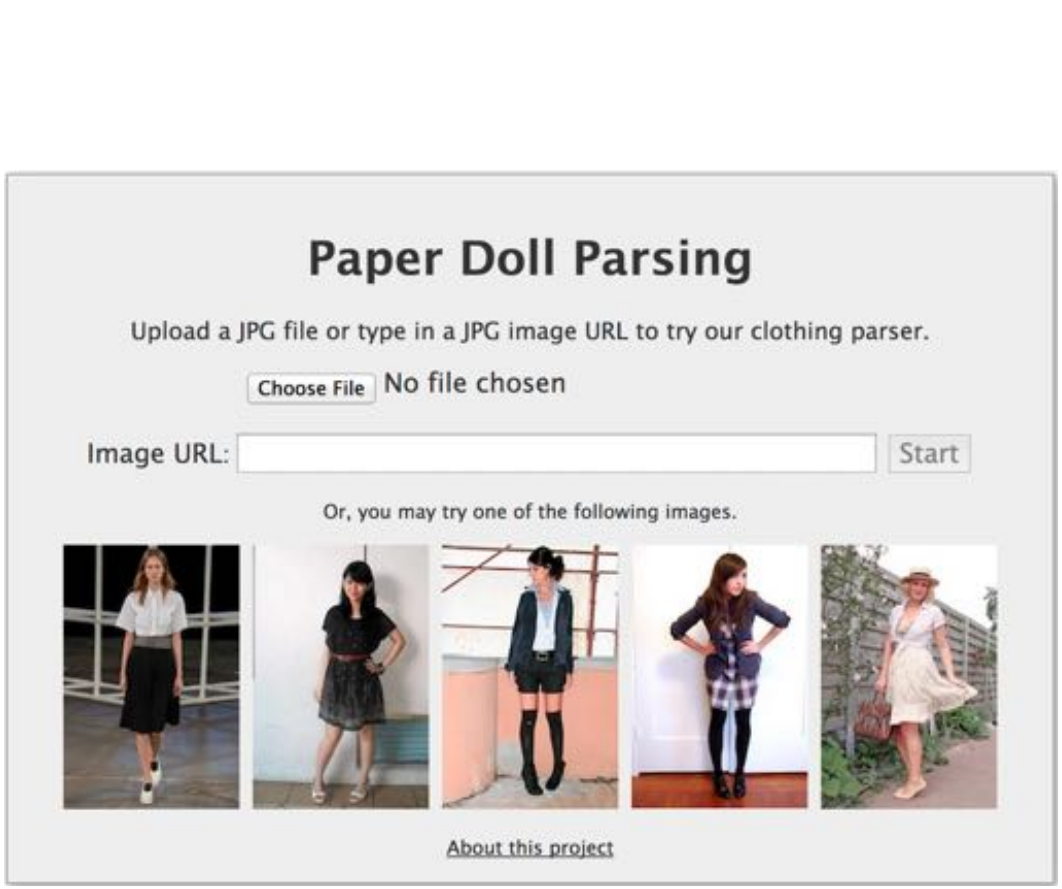
- Reconstruct the world in 3D from online photos!

https://www.youtube.com/watch?v=IgBQCoEfiMs

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?

- How Fashionable Are You?

LOS ANGELES, CA

466 FANS
288 VOTES
62 FAVOURITES

TAGS
CHIC
EVERDAY
FALL

COLOURS
WHITE-BOOTS

NOVEMBER 10, 2014

GARMENTS
White Cheap Monday Boots
Chilli Beans Sunglasses
Missguided Romper
Daniel Wellington Watch

COMMENTS
Nice!!
Love the top!
cute
...

Figure: An example of a post on `http://www.chictopia.com`. We crawled the site for 180K posts.

Why study Computer Vision?

- How Fashionable Can You Become?

Figure: Examples of recommendations provided by our model. The parenthesis we show the fashionability scores.

[E. Simo-Serra, S. Fidler, F. Moreno, R. Urtasun. CVPR'15.]

Why study Computer Vision?

[S. Zhu, C.C Loy, D. Lin, R. Urtasun, S. Fidler. In submission.]

The lady's upper-clothes contain the pattern of flowers

Why study Computer Vision?

[S. Zhu, C.C Loy, D. Lin, R. Urtasun, S. Fidler. In submission.]

The woman is wearing a blue short-sleeved T-shirt and blue jeans

Why study Computer Vision?

[S. Zhu, C.C Loy, D. Lin, R. Urtasun, S. Fidler. In submission.]

A woman wearing a black overcoat and white shorts

Why study Computer Vision?

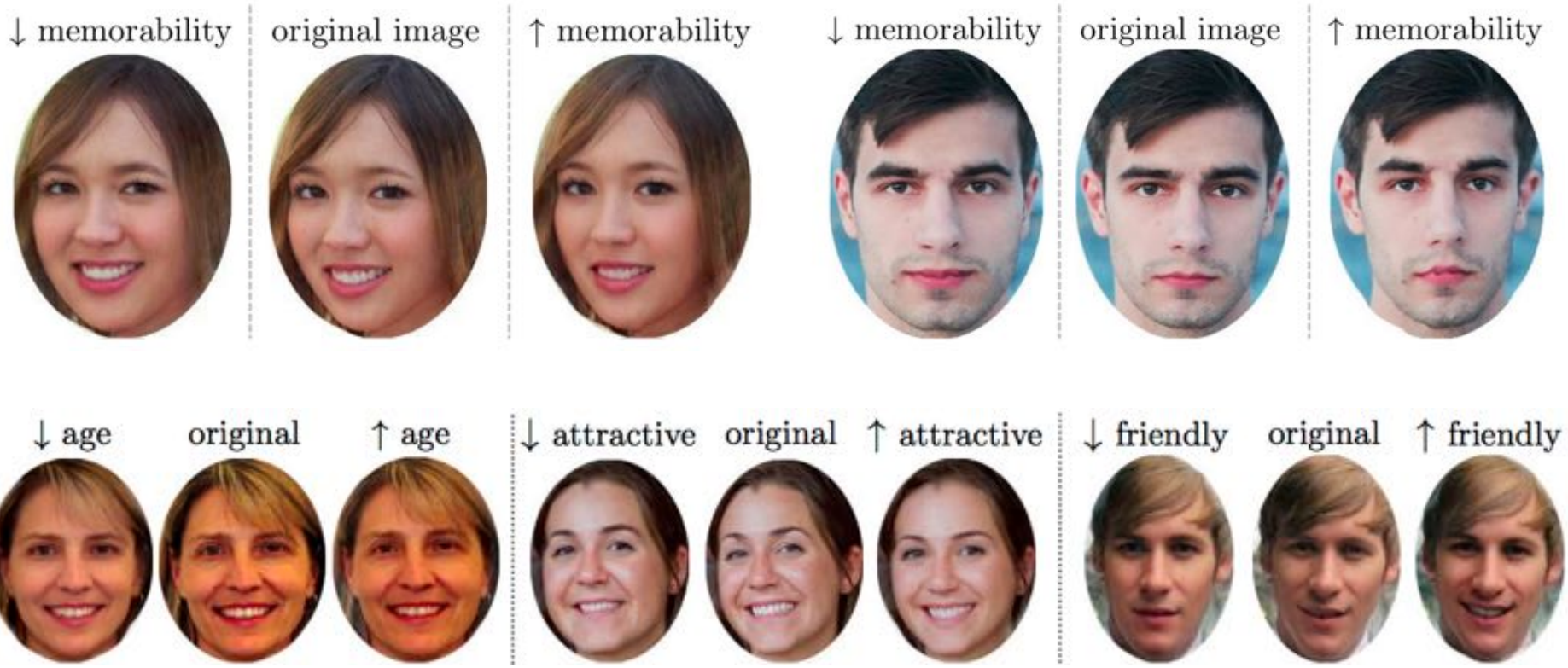
[S. Zhu, C.C Loy, D. Lin, R. Urtasun, S. Fidler. In submission.]

Why study Computer Vision?

- Crazy media attention!!!

Why study Computer Vision?

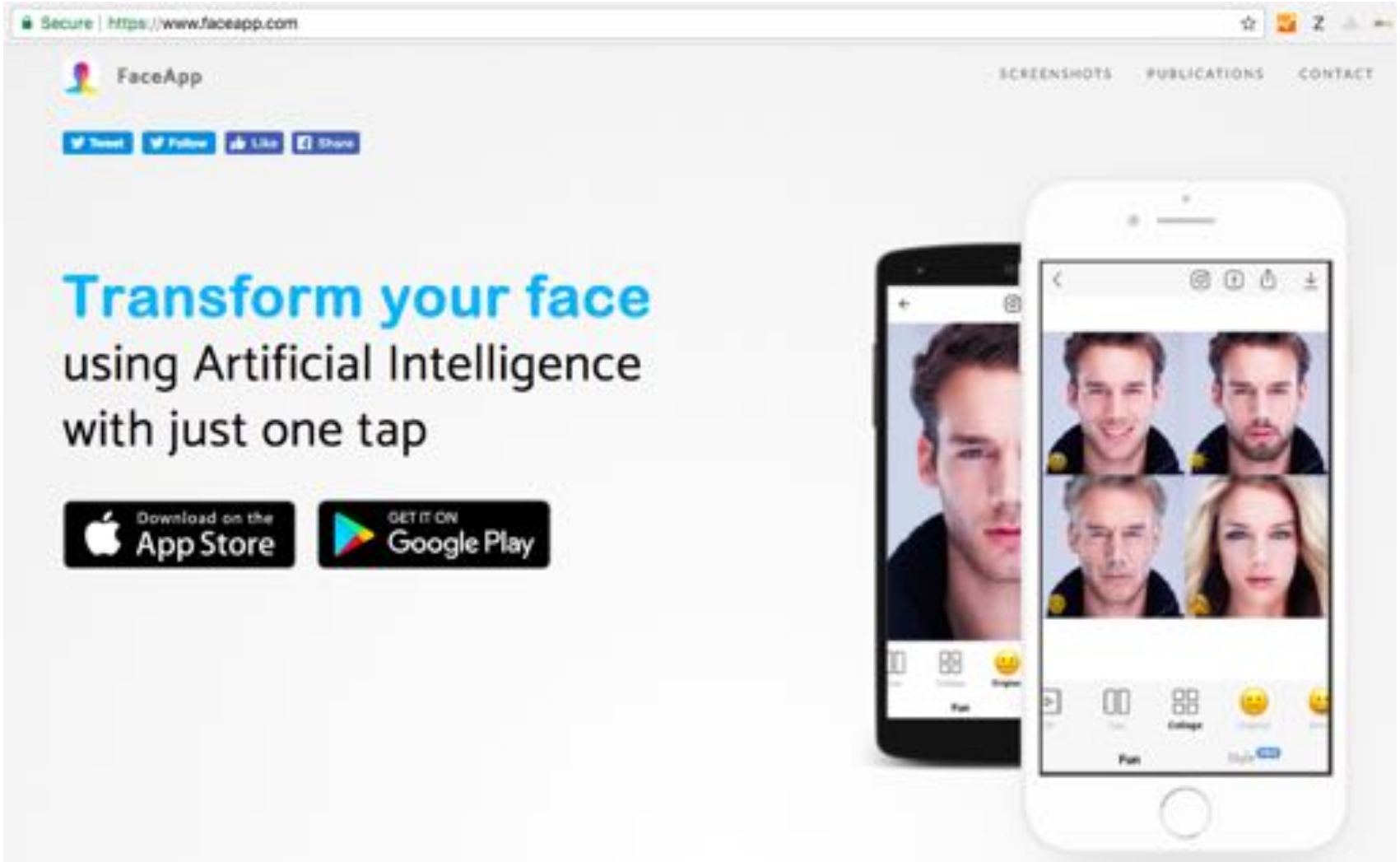
- You can make yourself look better (and others worse)



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

Why study Computer Vision?

- Play with faces



Why study Computer Vision?

- Play with faces

Why study Computer Vision?

- Play with faces

Why study Computer Vision?

- Play with faces

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

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

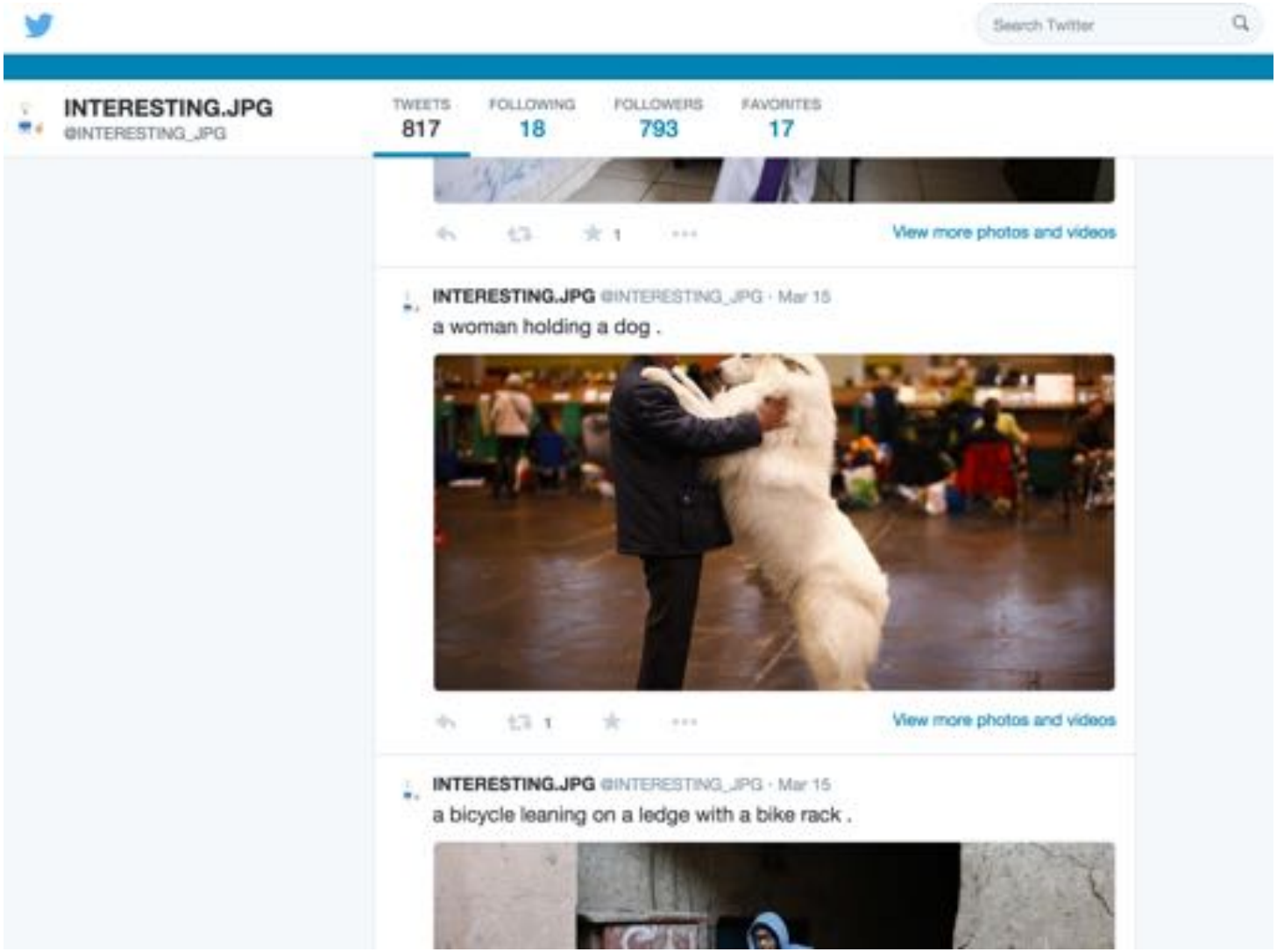
Why study Computer Vision?

- 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

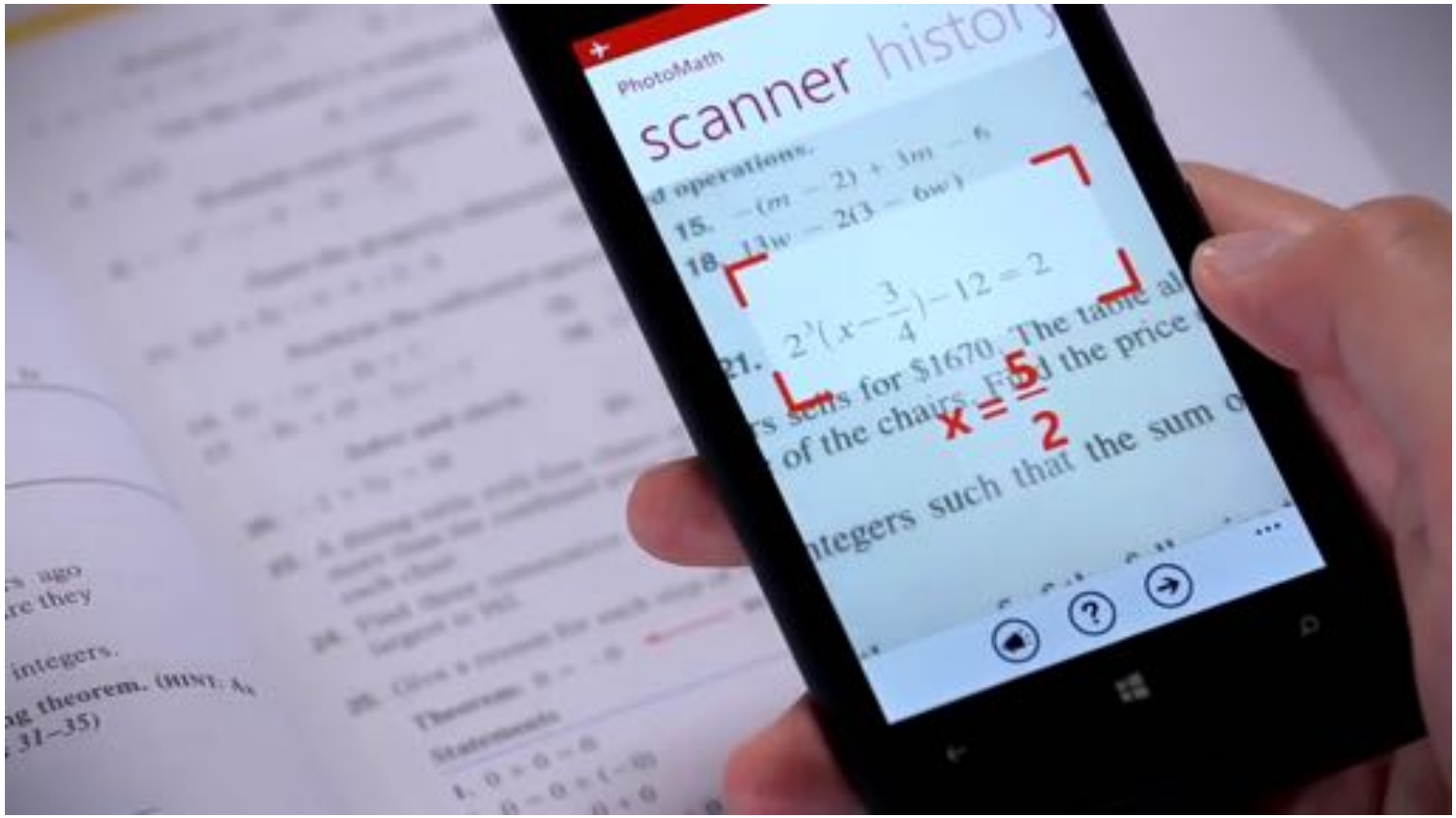
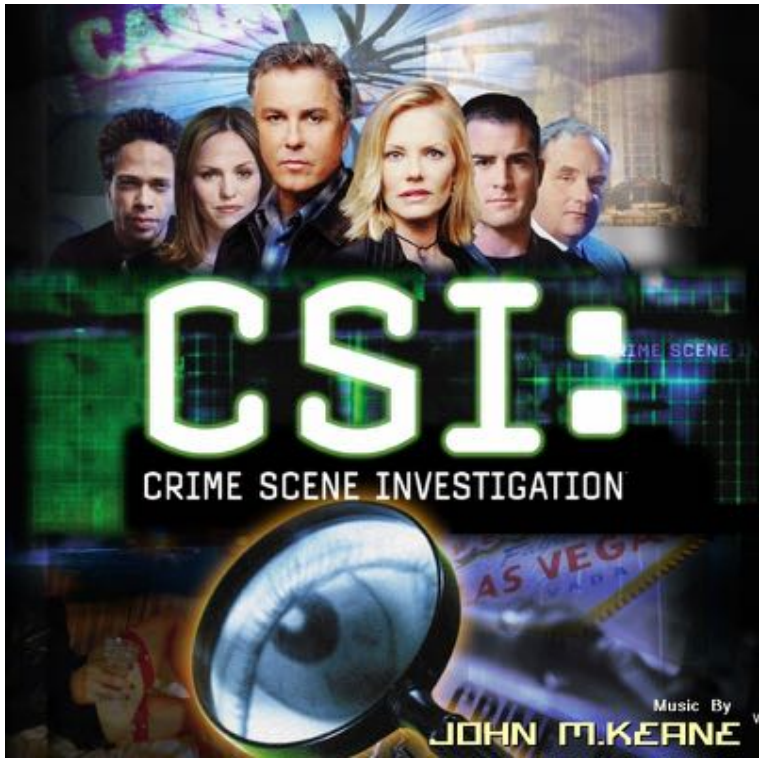
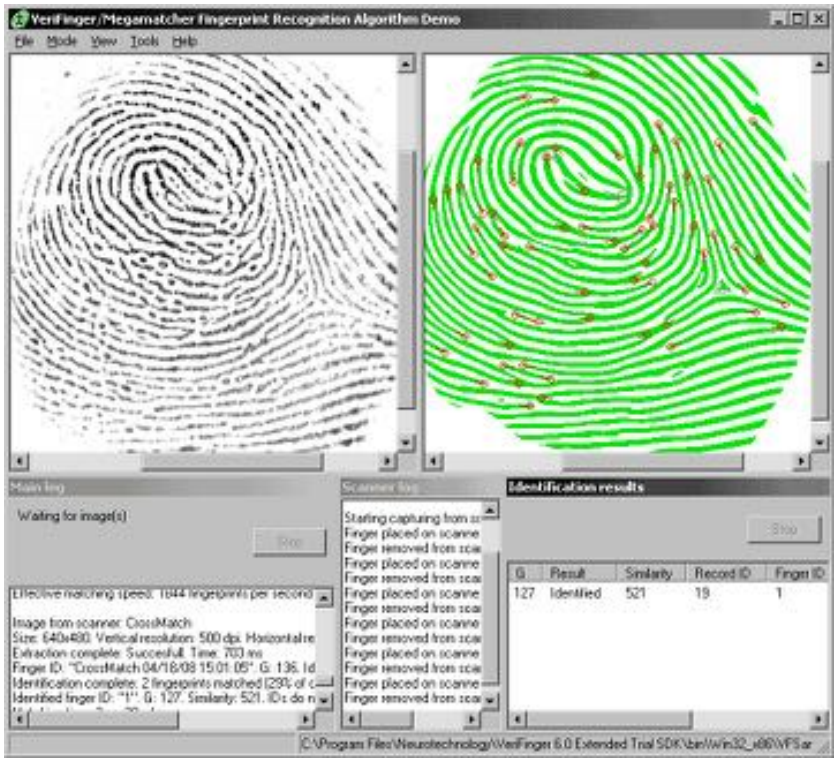


Figure: Photomath: https://photomath.net/, http://www.youtube.com/watch?v=XlbVB50mIh4

Why study Computer Vision?

- Fingerprint recognition





[Source: S. Lazebnik]

Why study Computer Vision?

- You can do some movie-like Forensics

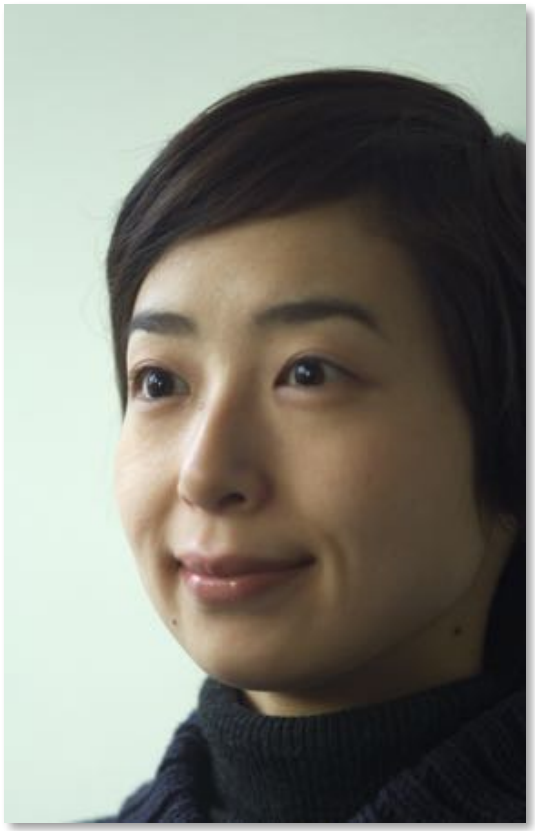


Figure: Source: Nayar and Nishino, Eyes for Relighting

[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?

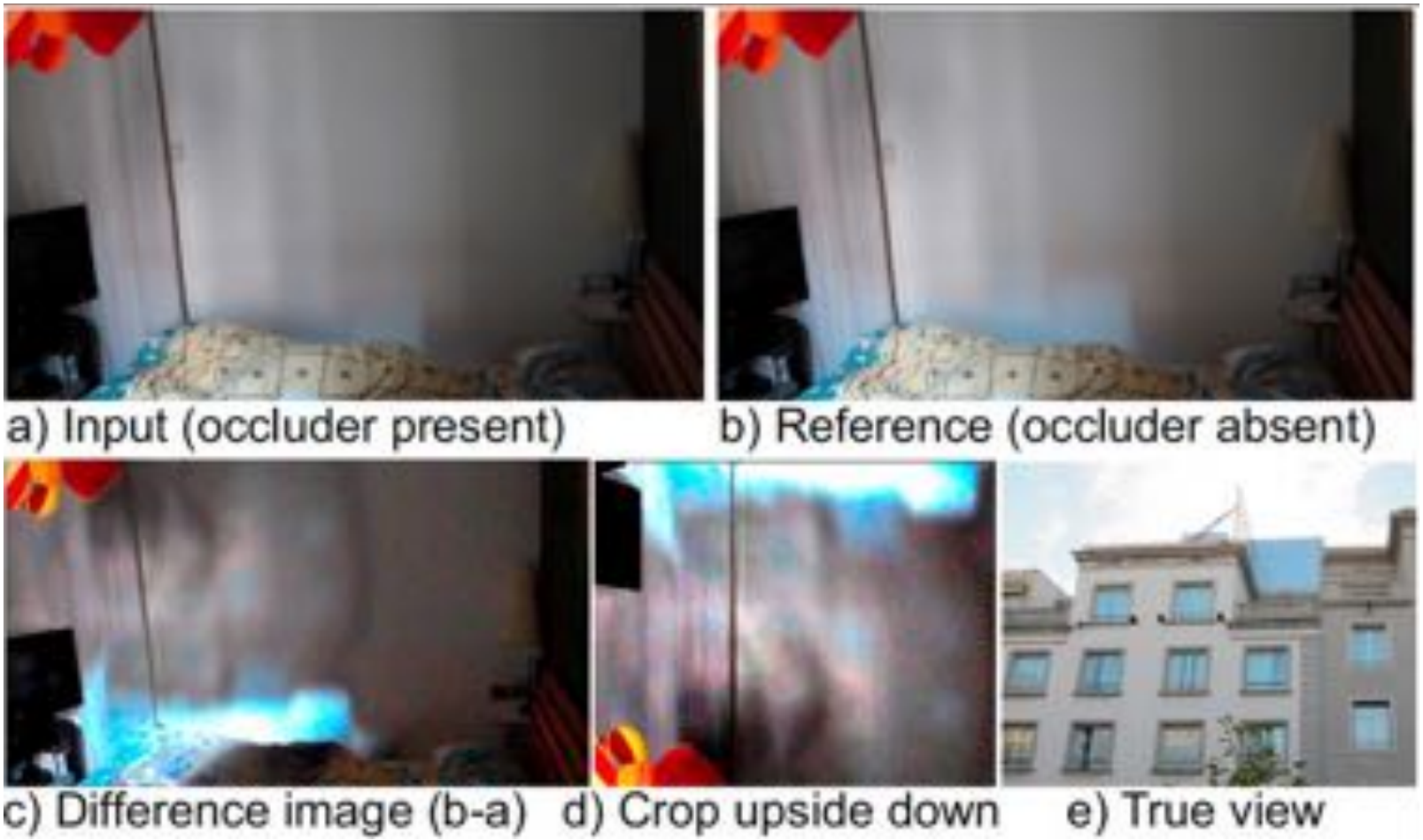
- Some more CSI

- Can you see something on the wall?

Torralba & Freeman, CVPR'12

Why study Computer Vision?

- Some more CSI



Why study Computer Vision?

- Object recognition (in mobile phones)

[Source: S. Seitz]

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	DuEye			92.65 %	91.43 %	86.18 %	4 s	GPU @ 2.5 Ghz (C/C++)	
2	RV-CNN			91.67 %	91.28 %	85.43 %	3.5 s	GPU @ 2.5 Ghz (Python + C/C++)	
3	eagle			91.28 %	91.06 %	85.66 %	4 s	GPU @ 2.5 Ghz (C/C++)	
4	Genome			90.63 %	90.85 %	85.82 %	4 s	GPU @ 2.5 Ghz (C/C++)	

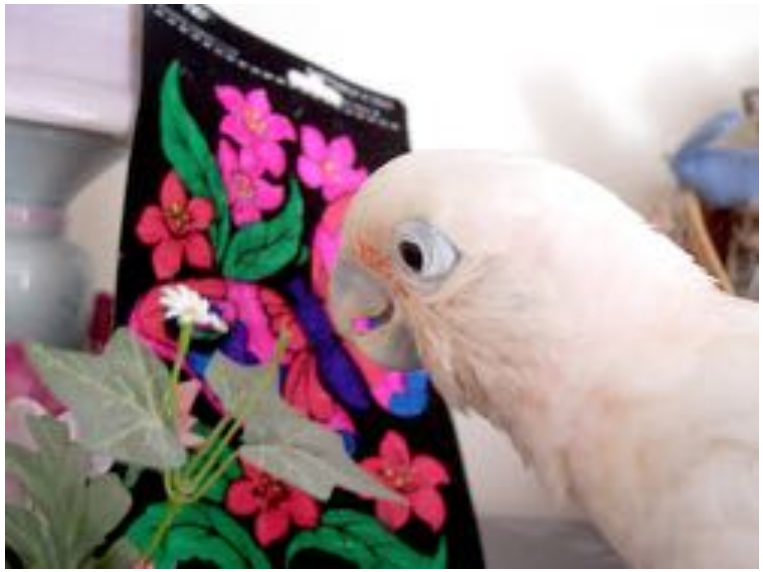
Cyclist

	Method	Setting	Code	Moderate	Easy	Hard	Runtime	Environment	Compare
1	Pie			76.25 %	84.62 %	67.57 %	1.2 s	1 core @ 2.5 Ghz (C/C++)	
2	SAIT			76.13 %	83.88 %	66.60 %	0.15 s	GPU @ >3.5 Ghz (Python + C/C++)	
3	TICNN			75.83 %	84.28 %	66.50 %	0.5 s	GPU @ 2.5 Ghz (Matlab + C/C++)	
4	TuSimple			75.59 %	84.15 %	66.35 %	1.6 s	GPU @ 2.5 Ghz (Python + C/C++)	

		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 mistakes...
- The general vision problem is not yet solved



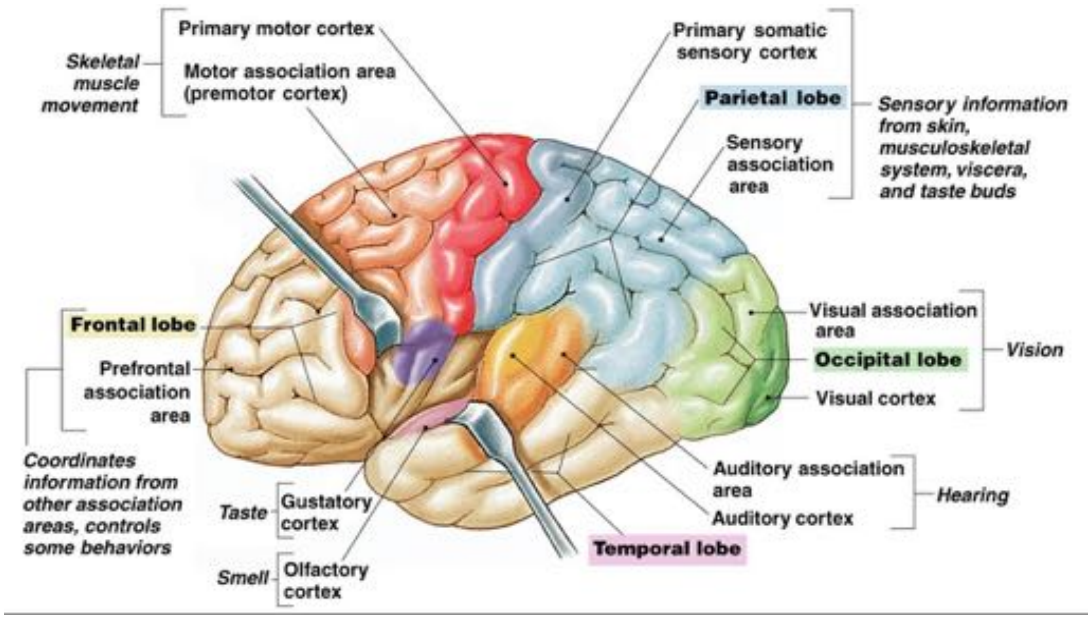
Where pink means “person”

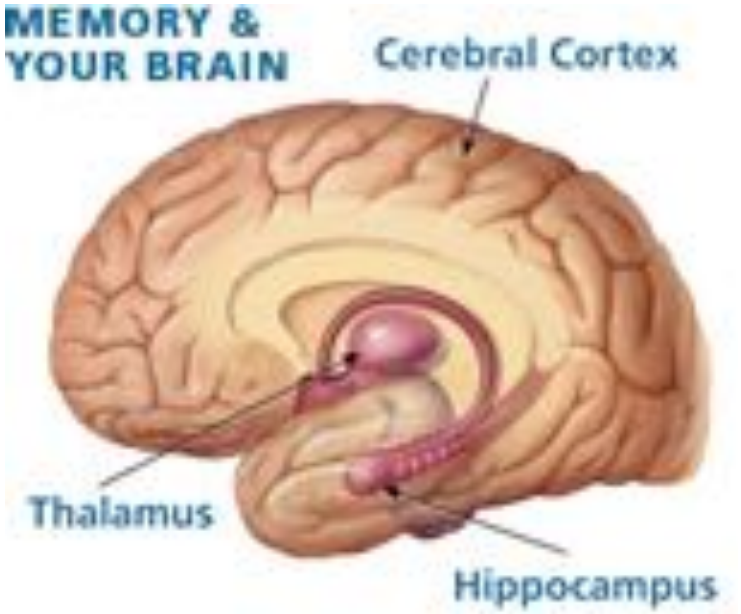
[This pic is from 2014]

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?

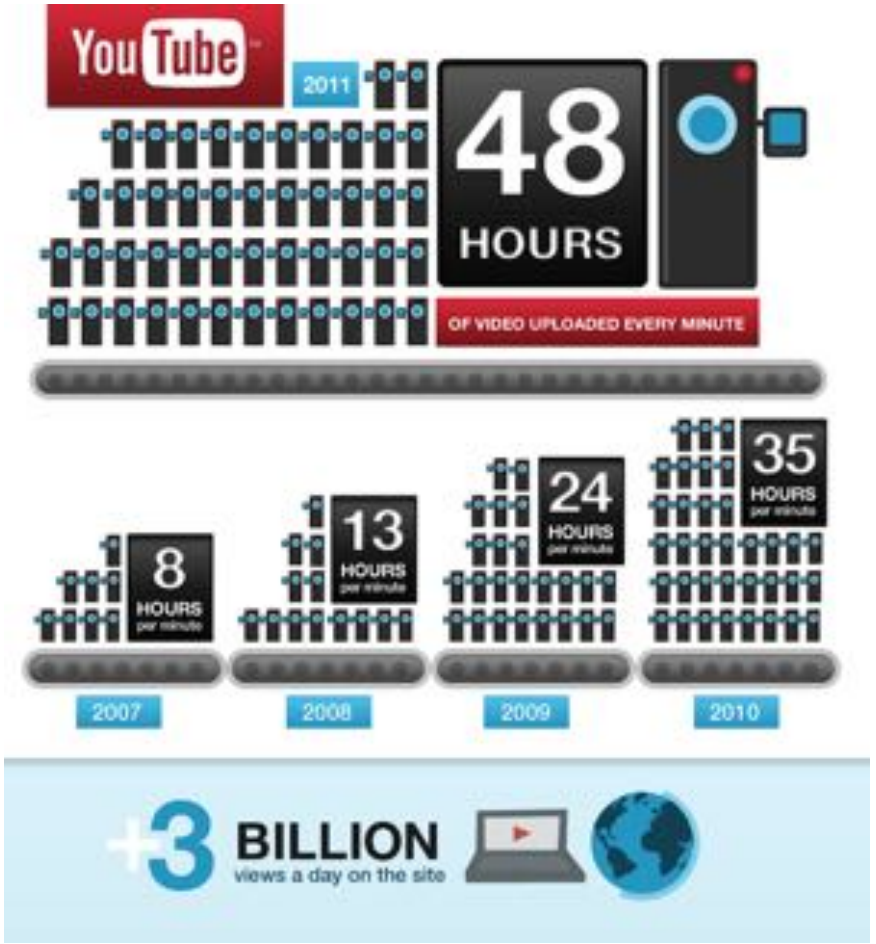
Biederman, 1987

[slide credit: R. Urtasun]

Why is vision hard?

Lots of data to process:

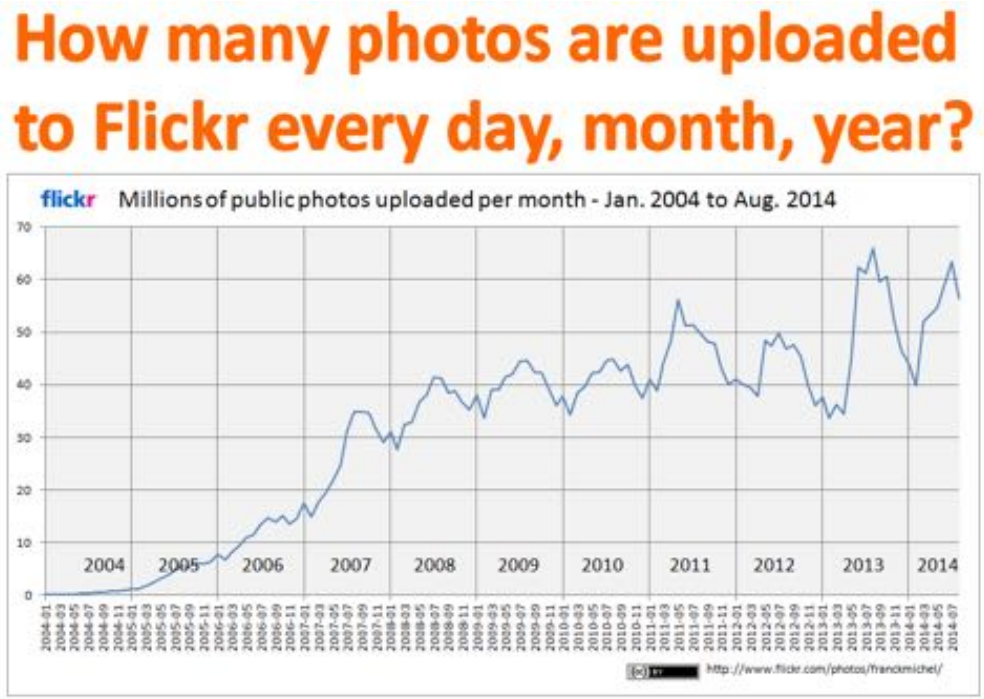
- Thousands to millions of pixels in an image
- 100 hours of video added to YouTube per minute [source: YouTube]
- Over 6 billion hours of video are watched each month on YouTube – almost an hour for every person on Earth [source: YouTube]

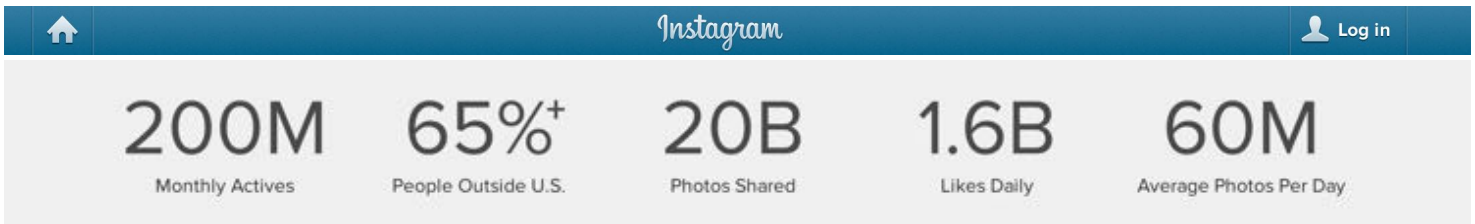


Why is vision hard?

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]





Exploit so Much Data!

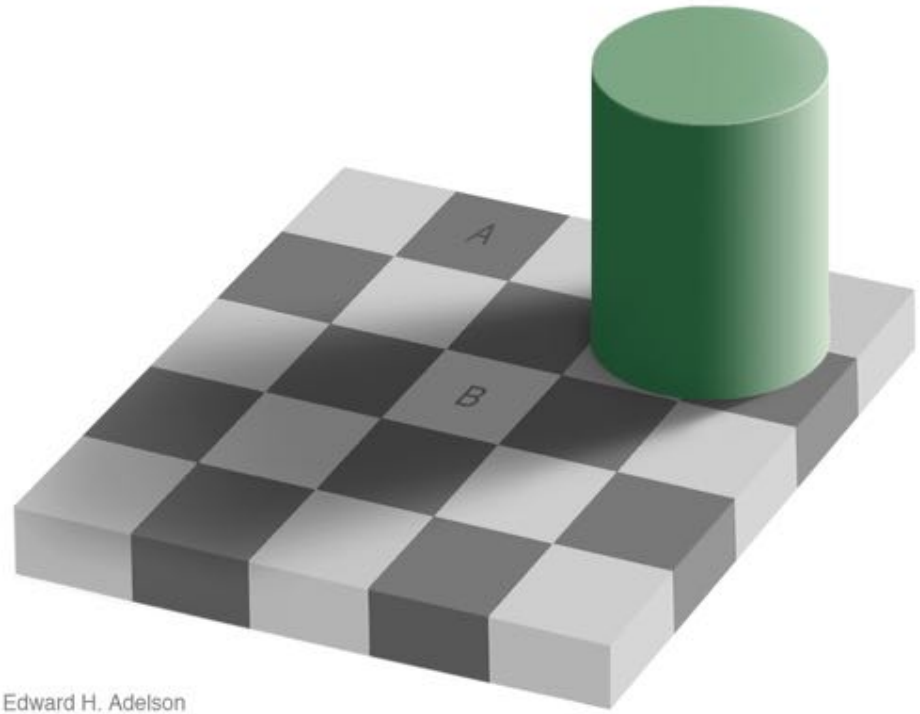
Figure: Vemodalen: The Fear That Everything Has Already Been Done, https://www.youtube.com/watch?v=8ftDjebw8aA

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

Why is vision hard?

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

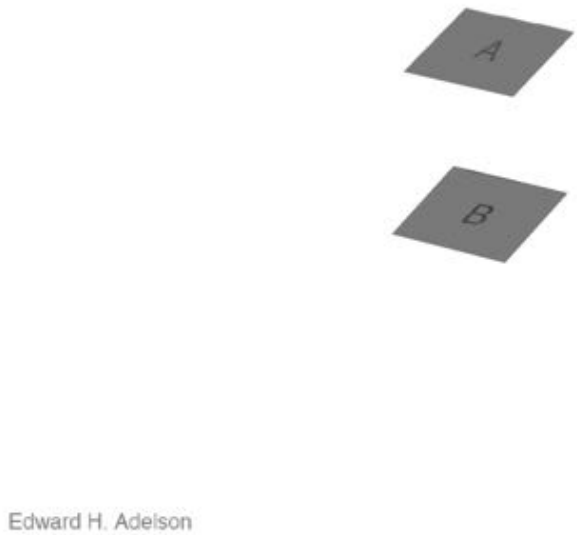
How good are humans?



- Which square is lighter, A or B?

[Slide credit: A. Torralba]

How good are humans?



- Which square is lighter, A or B?

[Slide credit: A. Torralba]

How good are humans?

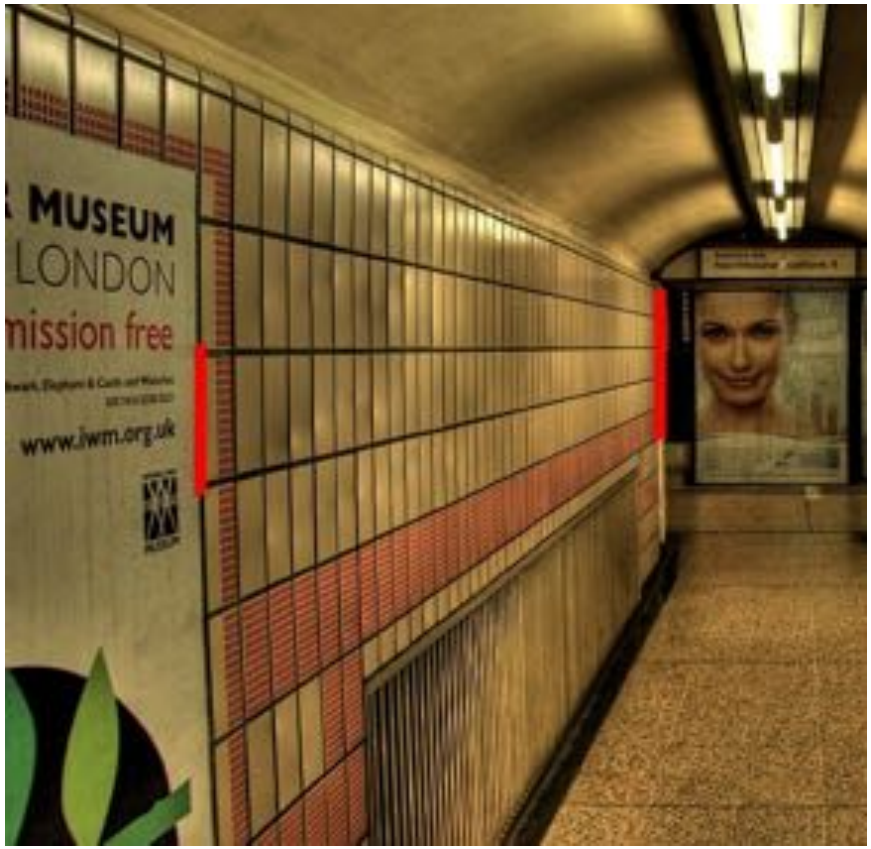


Figure: 2006 Walt Anthony

- 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: Ames room

- Assumptions can be wrong

[Slide credit: A. Torralba]

How good are humans?

Figure: Chabris & Simons, https://www.youtube.com/watch?v=vJG698U2Mvo

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

How good are humans?

Figure: Simons et al., http://www.perceptionweb.com/perception/perc1000/a_d_ex1.mov (more videos here: http://www.perceptionweb.com/misc.cgi?id=p3104)

- Is something happening in the picture?

How good are humans?

Figure: Torralba et al., http://people.csail.mit.edu/torralba/courses/6.870/slides/blur.avi

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

How good are humans?

Figure: Torralba et al., http://people.csail.mit.edu/torralba/courses/6.870/slides/highres.avi

- 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?

- Technical capabilities
- Good programming skills
- Imagination
- Even better intuition
- Lots of persistence
- Some luck always helps