CSC420: Intro to Image Understanding Introduction

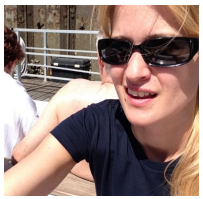
Sanja Fidler

January 8, 2024

UNIVERSITY OF TORONTO

The Team

- **Instructor**:



Sanja Fidler (fidler@cs.toronto.edu)

- **Office**: online
- **Office hours**: Mon 11am-11.30am. Please send an email to schedule outside of these hours.
- **TAs**:

Yun-Chun Chen (yunchun.chen@mail.utoronto.ca)

Parsa Mirdehghan (parsa.mirdehghan@mail.utoronto.ca)

Mohammad Kianpisheh (kian@cs.toronto.edu)

Sina Davari (sina.davari@mail.utoronto.ca)

Arash Rasti Meymandi (arash.rasti@mail.utoronto.ca)

Course Information

- **Class time**: Monday at 9-11am
- **Location**: online
- **Tutorials**: TUT0101 on Monday 1-2pm, TUT0102 on Monday 2-3pm. Tutorials will consist of demos and Q&A. Tutorials are online.
- **Class Website**:

  http://www.cs.toronto.edu/~fidler/teaching/2024/CSC420.html

- The class will use Quercus for **announcements** and **discussions**

Course Information

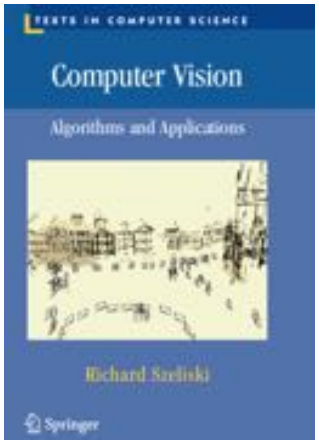
- **Class time**: Monday at 9-11am
- **Location**: online
- **Tutorials**: TUT0101 on Monday 1-2pm, TUT0102 on Monday 2-3pm. Tutorials will consist of demos and Q&A. Tutorials are online.
- **Class Website**:

  http://www.cs.toronto.edu/~fidler/teaching/2024/CSC420.html

- The class will use Quercus for **announcements** and **discussions**

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
- Numerical Analysis

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

Knowing some basics in this is a plus:

- Python, Matlab, 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 + oral exam**: 40%($project report + oral\ exam$)
- For the project you will need to do
  - Short project proposal
  - Project report
  - Project presentation (oral)
- Oral exam: During the project presentation, you will be asked questions about the class material

Term Work Dates

Term Work	Post Date	Due Date
Assignment 1	Jan 21	Jan 28
Assignment 2	Feb 4	Feb 11
Assignment 3	March 3	March 10
Assignment 4	March 17	March 25
Project Report		April 15
Project Presentation		TBD

- All dates are for 2024
- Dates are approximate (depend on what material we cover in class)

Programming Language?

- Your assignments / project can be implemented either in Python, Matlab, or C++. Python is preferred, but not a requirement.
- Most code and examples we will provide during the class will be in Python and Matlab.
- Choose wisely

Lateness

Deadline The solutions to 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

Intro
Linear filters, edges
Image features
Keypoint detection
Matching
Stereo, multi-view
Stereo, multi-view
Object recognition
Object detection
Neural Networks
Segmentation

We will have invited lectures on state of the art foundation models.

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"**

Embodied Agents

- Understand the scene in order to take actions: perception, prediction, planning, reasoning

Figure: How do I make dinner in this household?

Embodied Agents

- Understand the scene in order to take actions: perception, prediction, planning, reasoning

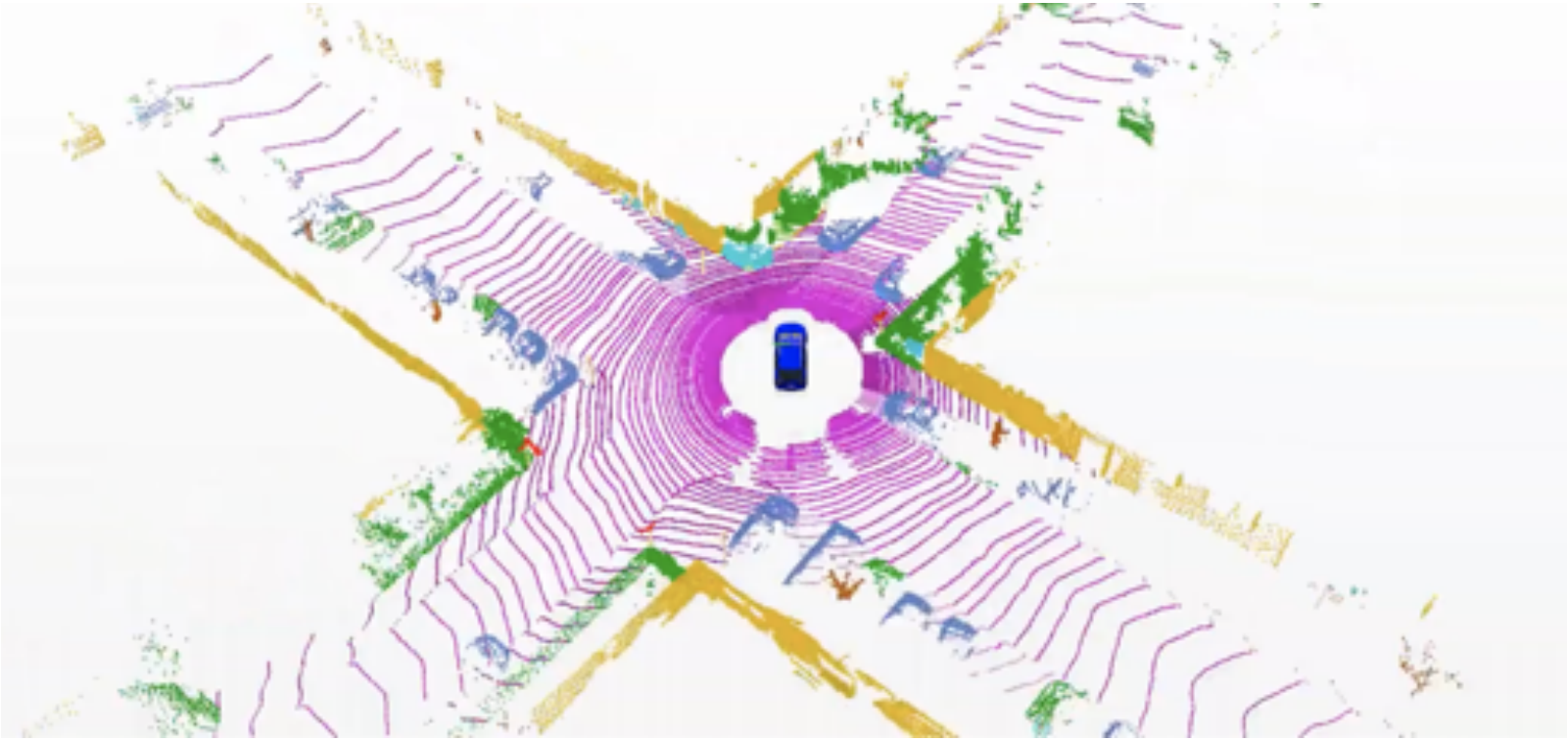


Figure: Autonomous driving

What is Computer Vision?

- What does it mean to see?

[text adopted from A. Torralba]

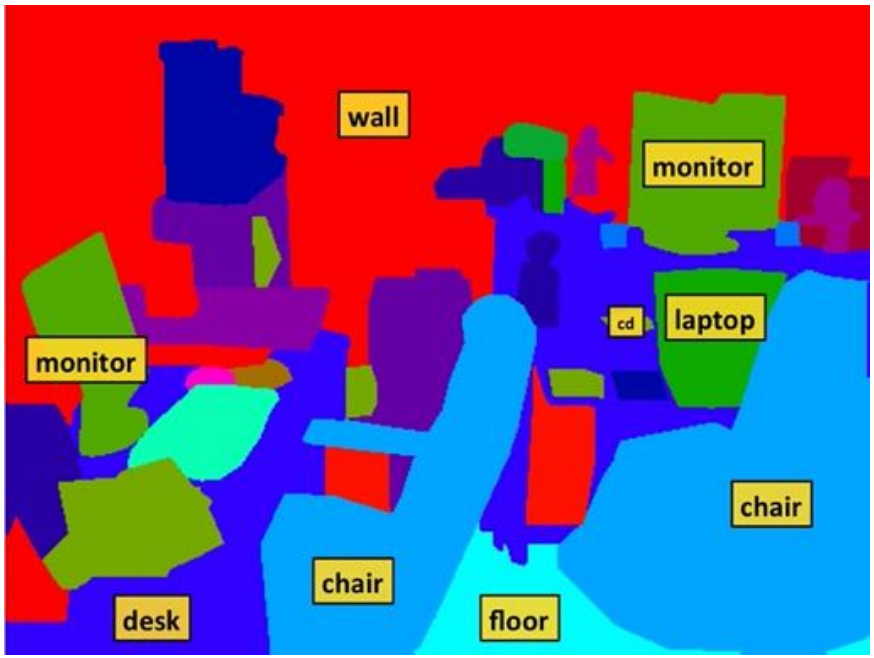
  - To know what is where by looking – Marr, 1982

What is Computer Vision?

- What does it mean to see?

[text adopted from A. Torralba]

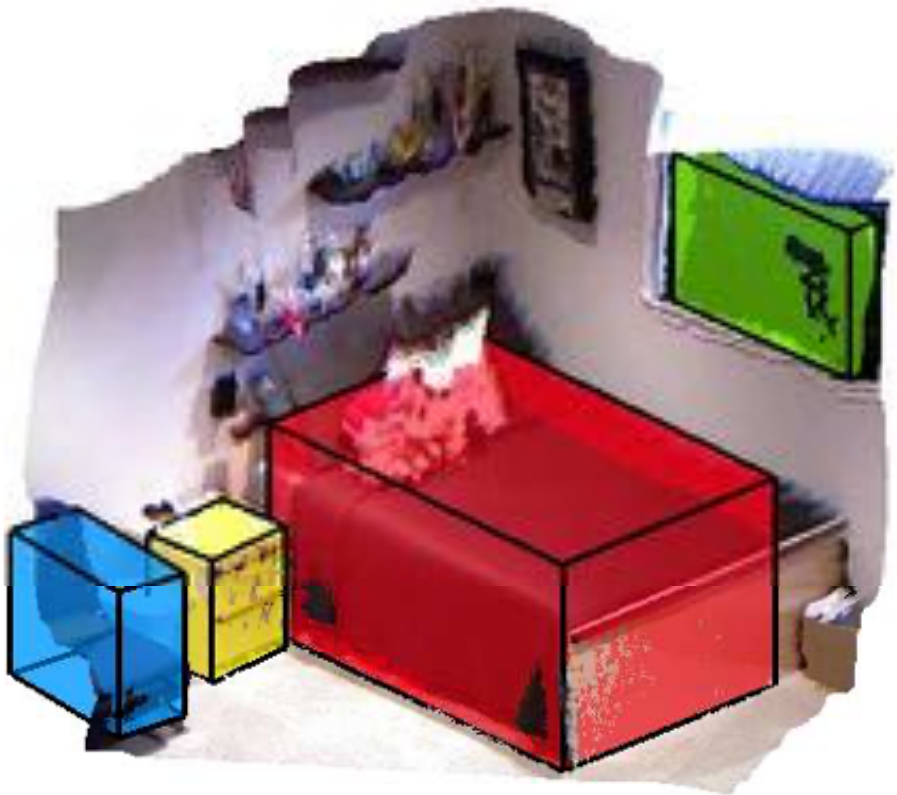
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world



What is Computer Vision?

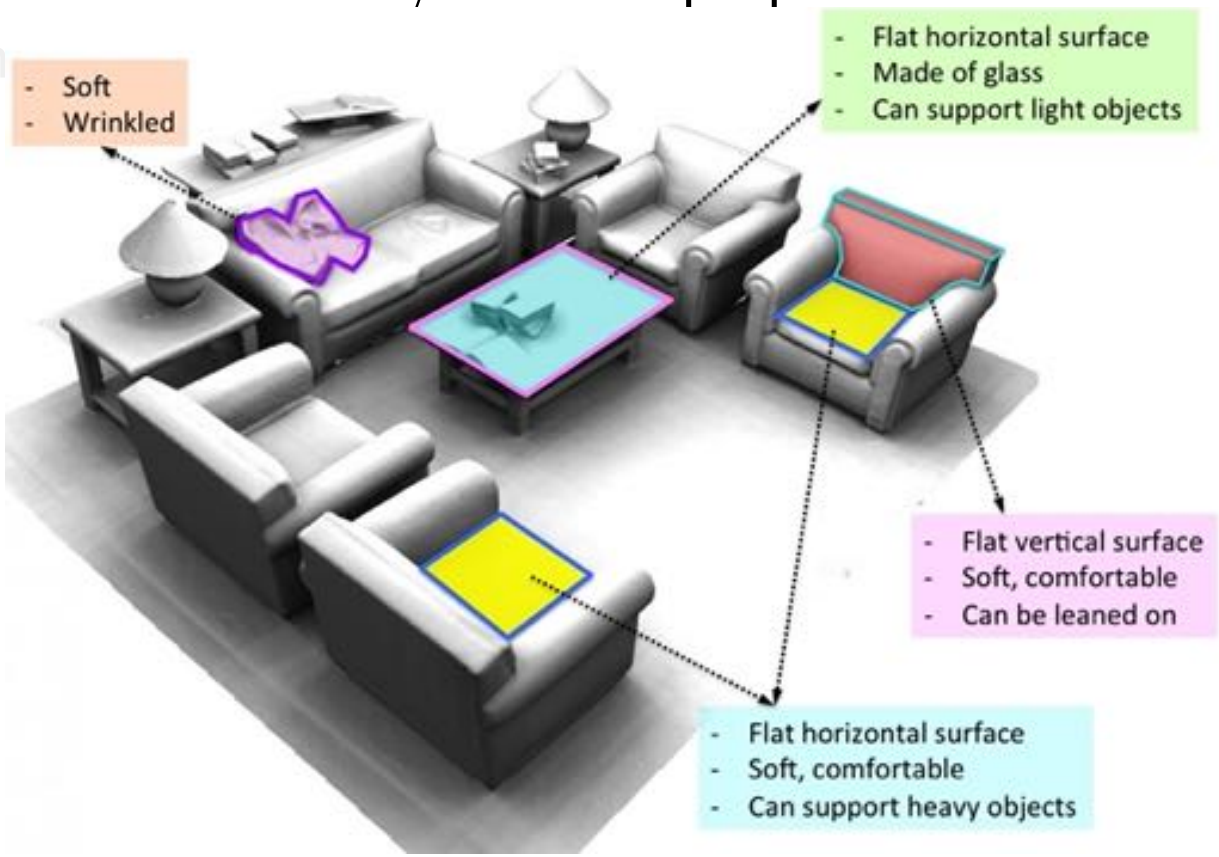
- What does it mean to see? [text adopted from A. Torralba]
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D/material properties?

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?

- 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?
  - What actions are taking place?

Pic from www.cobblehillpuzzles.com

"Full" Image Understanding?

- Full understanding of an image? To answer any question about it. To perform any task on it.

Demo: https://llava.hliu.cc/

[Haotian Liu, Chunyuan Li, Qingyang Wu, Yong Jae Lee, LLaVA: Large Language and Vision Assistant, *NeurIPS*, 2023]

Why study Computer Vision?

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 drive you to work

Amnon Shashua's Mobileye autonomous driving system

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

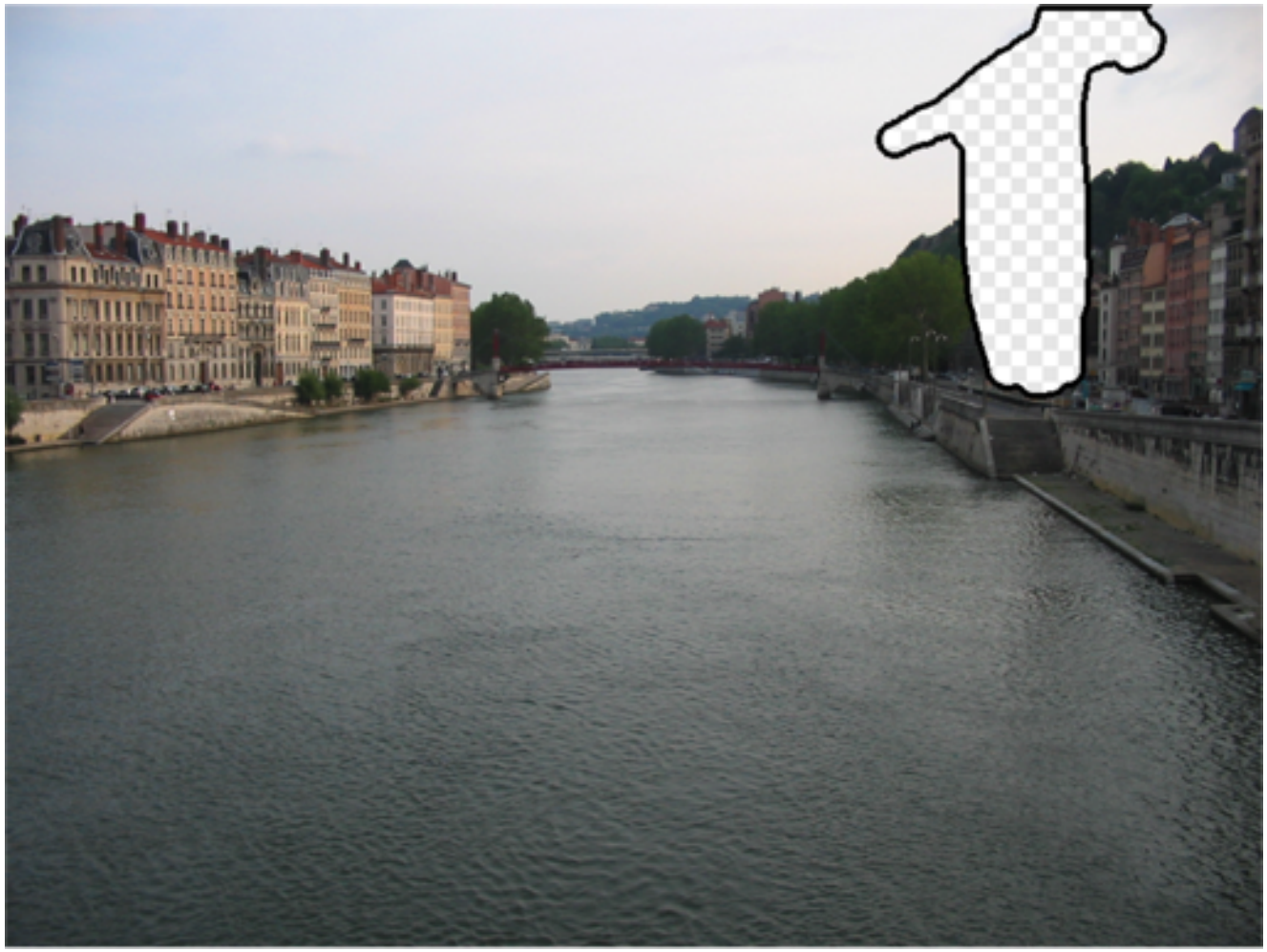
Why study Computer Vision?

- 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

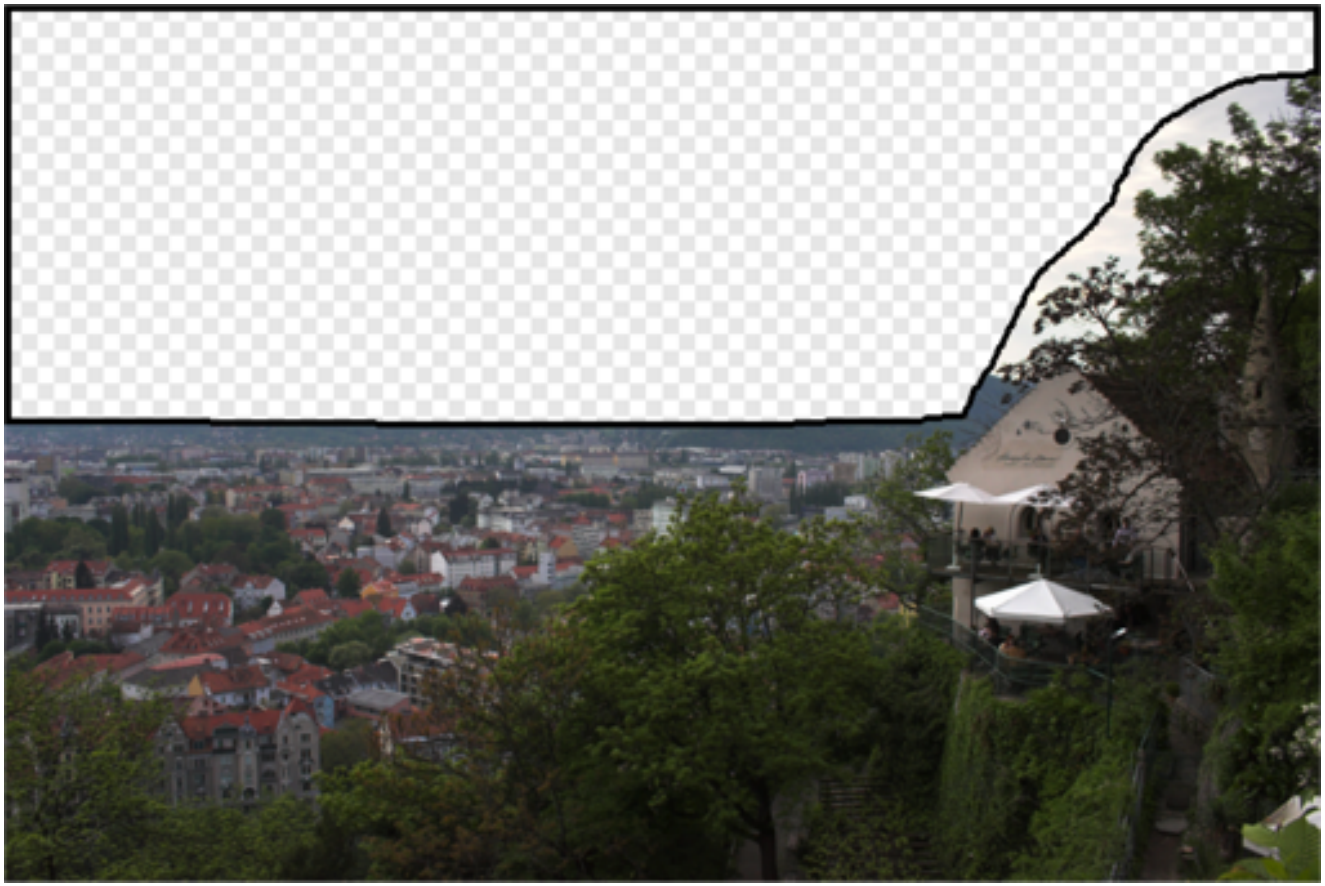
Why study Computer Vision?

- 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?

- Allows you to manipulate your images

https://www.youtube.com/watch?v=p5U4NgVGAwg

GauGan, Ming-Yu Liu et al., http://nvidia-research-mingyuliu.com/gaugan/]

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?

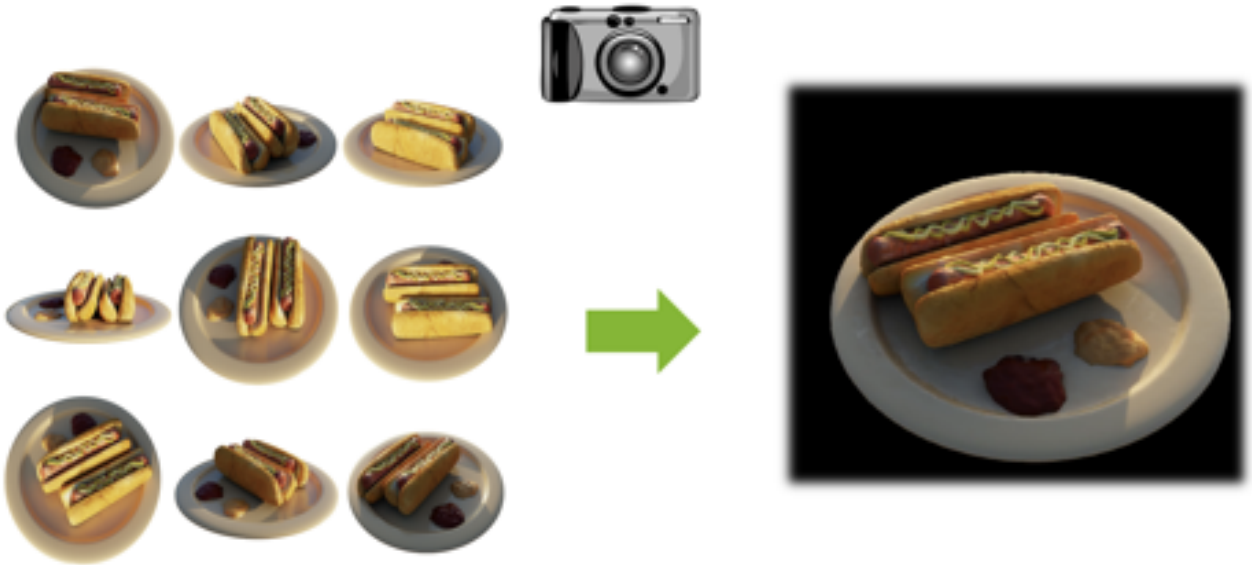
- ... and make cool videos using a single image

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?

- Reconstruct the world in 3D from captured photos!

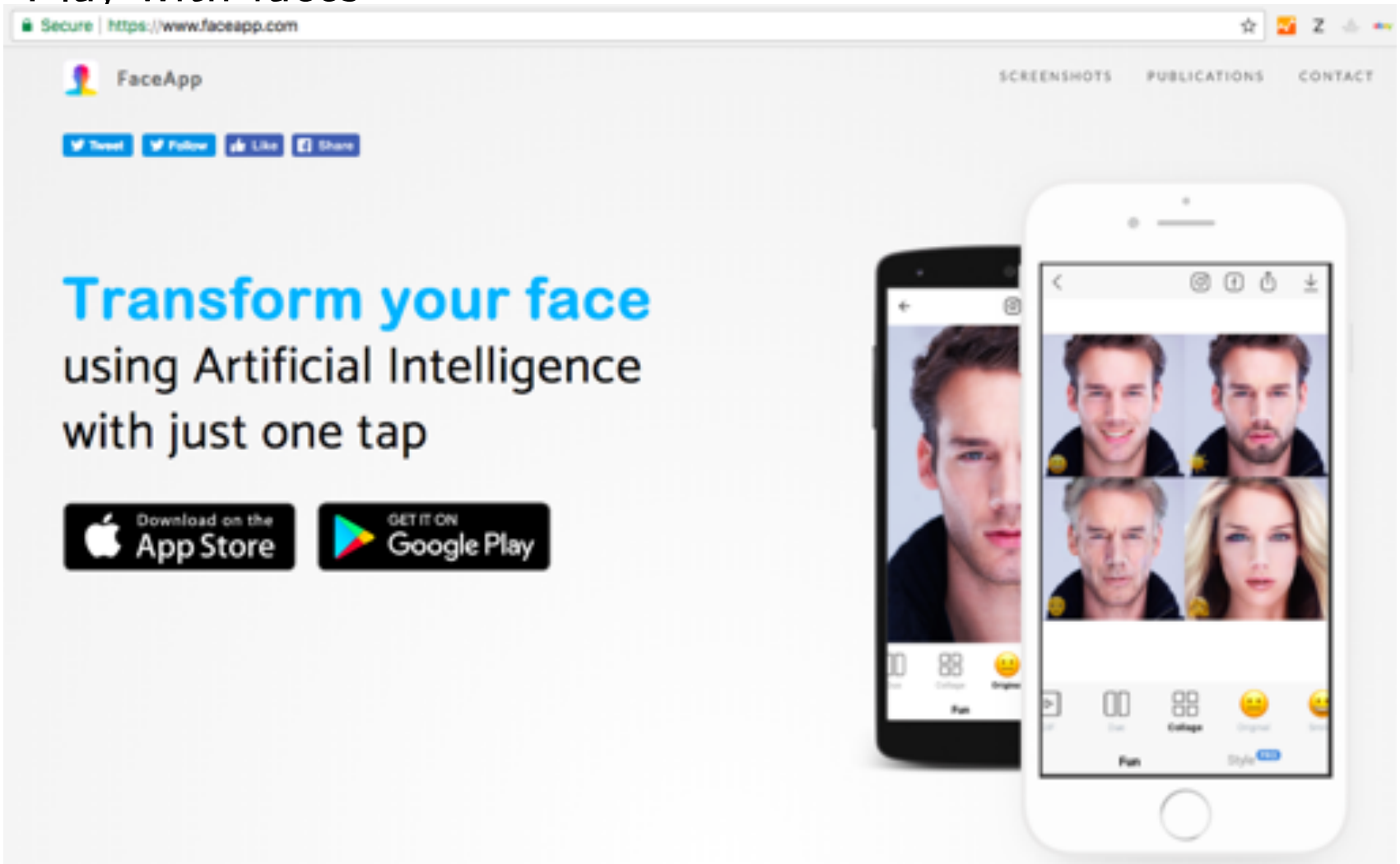


https://www.youtube.com/watch?v=DJ2hcC1orc4

Photosynth: https://photosynth.net/
Nerf: https://www.youtube.com/watch?v=yPKIxoN2Vf0

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 new faces

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

StyleGAN, Tero Karras et al., https://github.com/NVlabs/stylegan]

Why study Computer Vision?

- Generate image descriptions automatically

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

Why study Computer Vision?

- Generate images from descriptions automatically

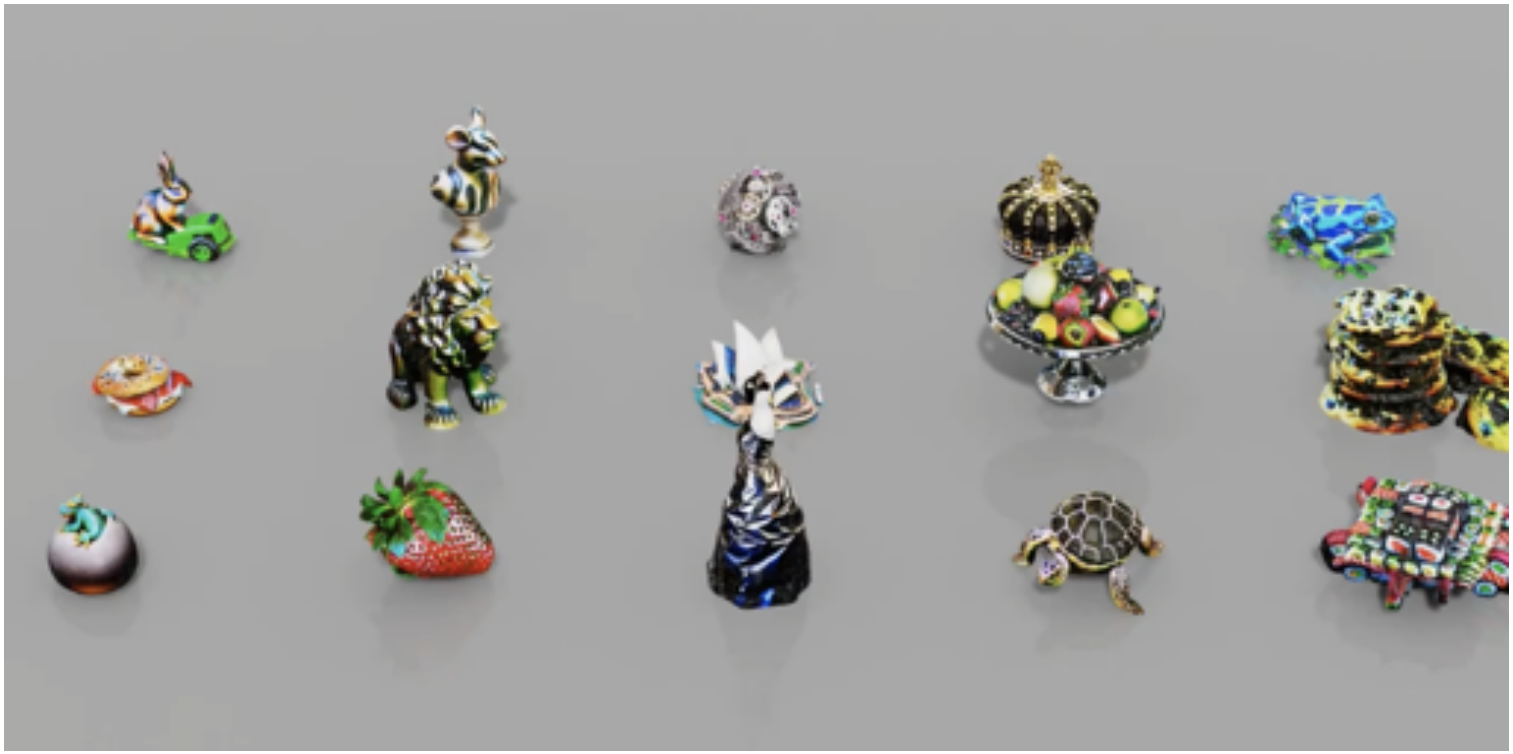
DALL-E:

Imagen:

[DALL-E: `https://openai.com/blog/dall-e/`, Imagen: `https://imagen.research.google/`, Imagen-video: `https://imagen.research.google/video/`, ediffi: `https://deepimagination.cc/eDiff-I/`]

Why study Computer Vision?

- Generate 3D models from descriptions automatically



Magic3D: Text-to-3D generation

[Dreamfusion: `https://dreamfusionpaper.github.io/`, Magic3D: `https://deepimagination.cc/Magic3D/`, GET3D: `https://nv-tlabs.github.io/GET3D/`]

Why study Computer Vision?

- Generate animated 3D models from descriptions automatically

Align Your Gaussians: Text-to-4D generation

[Make-a-Video3D https://make-a-video3d.github.io/,
Align Your Gaussians:: https://research.nvidia.com/labs/toronto-ai/AlignYourGaussians/]

Why study Computer Vision?

- Have a computer do math for you

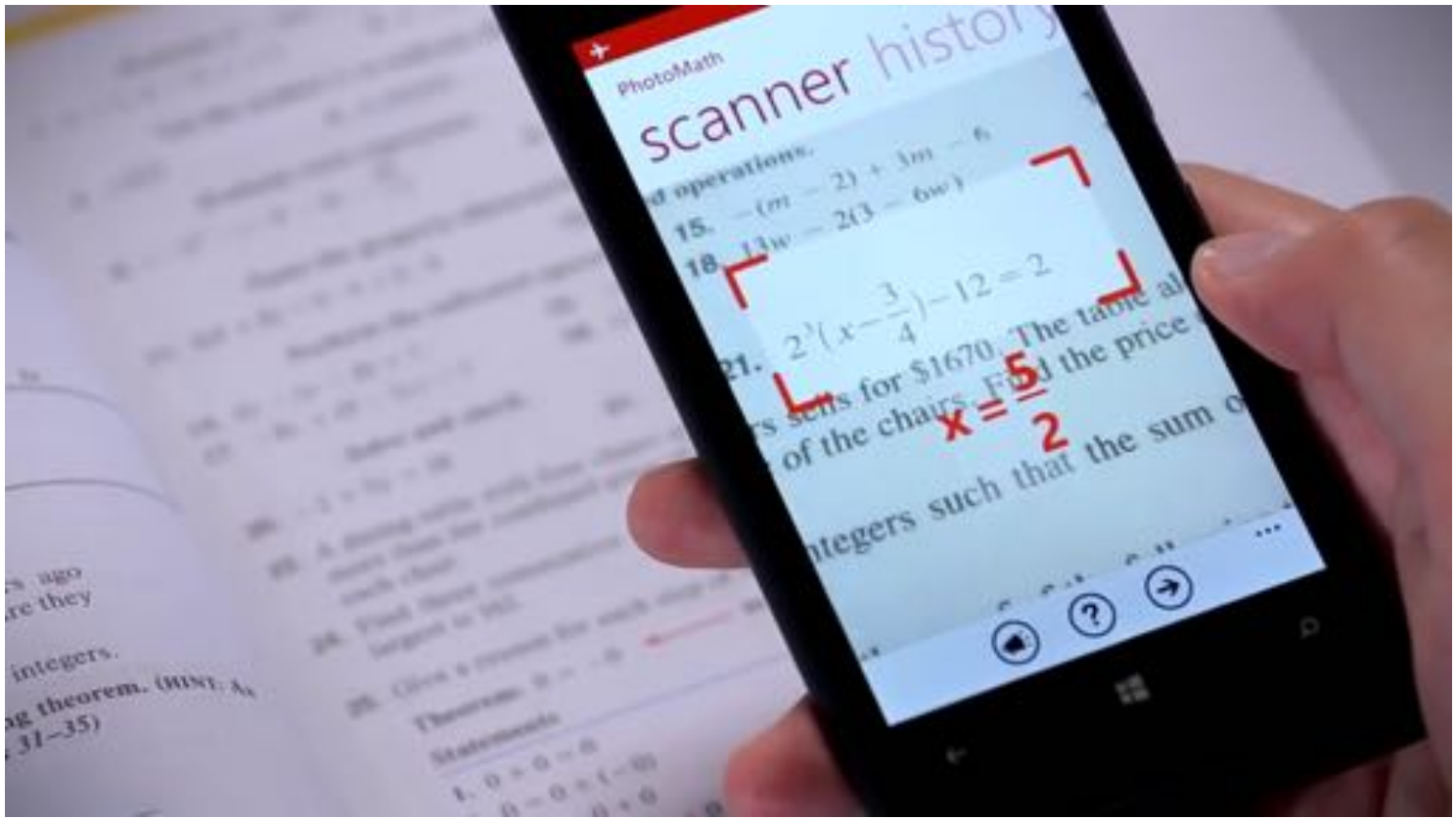


Figure: Photomath: https://photomath.net/, http://www.youtube.com/watch?v=XlbVB50mIh4

Why study Computer Vision?

- You can do movie-like Forensics

Figure: Source: Nayar and Nishino, "Eyes for Relighting"

[Source: N. Snavely]

Why study Computer Vision?

[Source: N. Snavely]

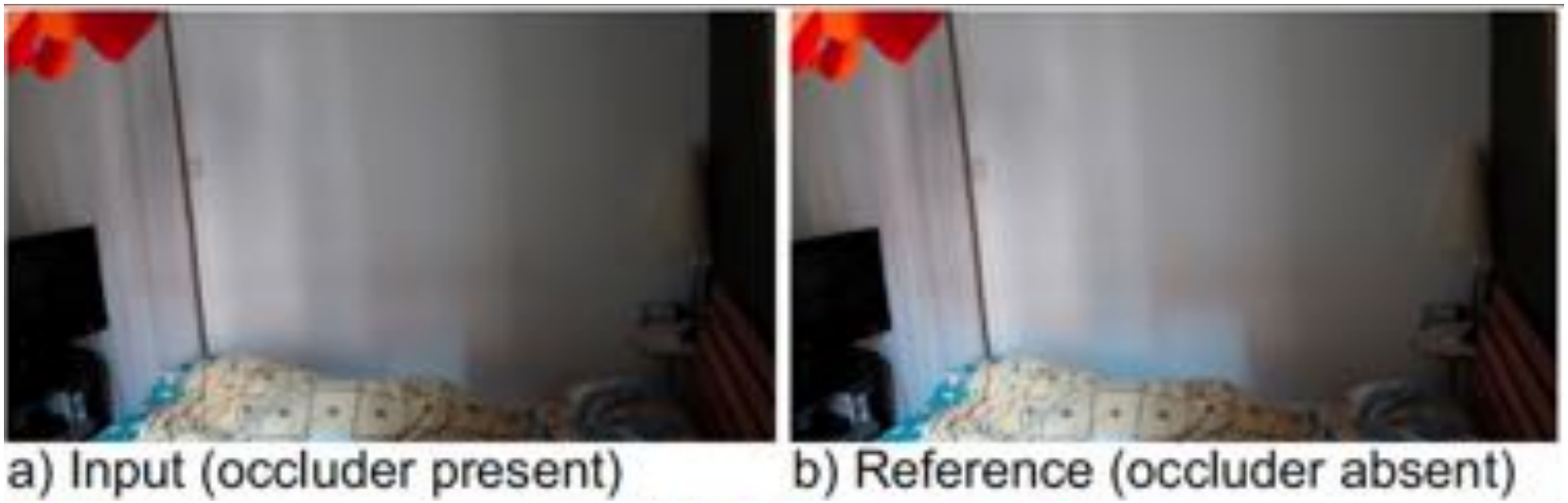
Why study Computer Vision?

Figure: Source: Nayar and Nishino, "Eyes for Relighting"

[Source: N. Snavely]

Why study Computer Vision?

- Some more CSI

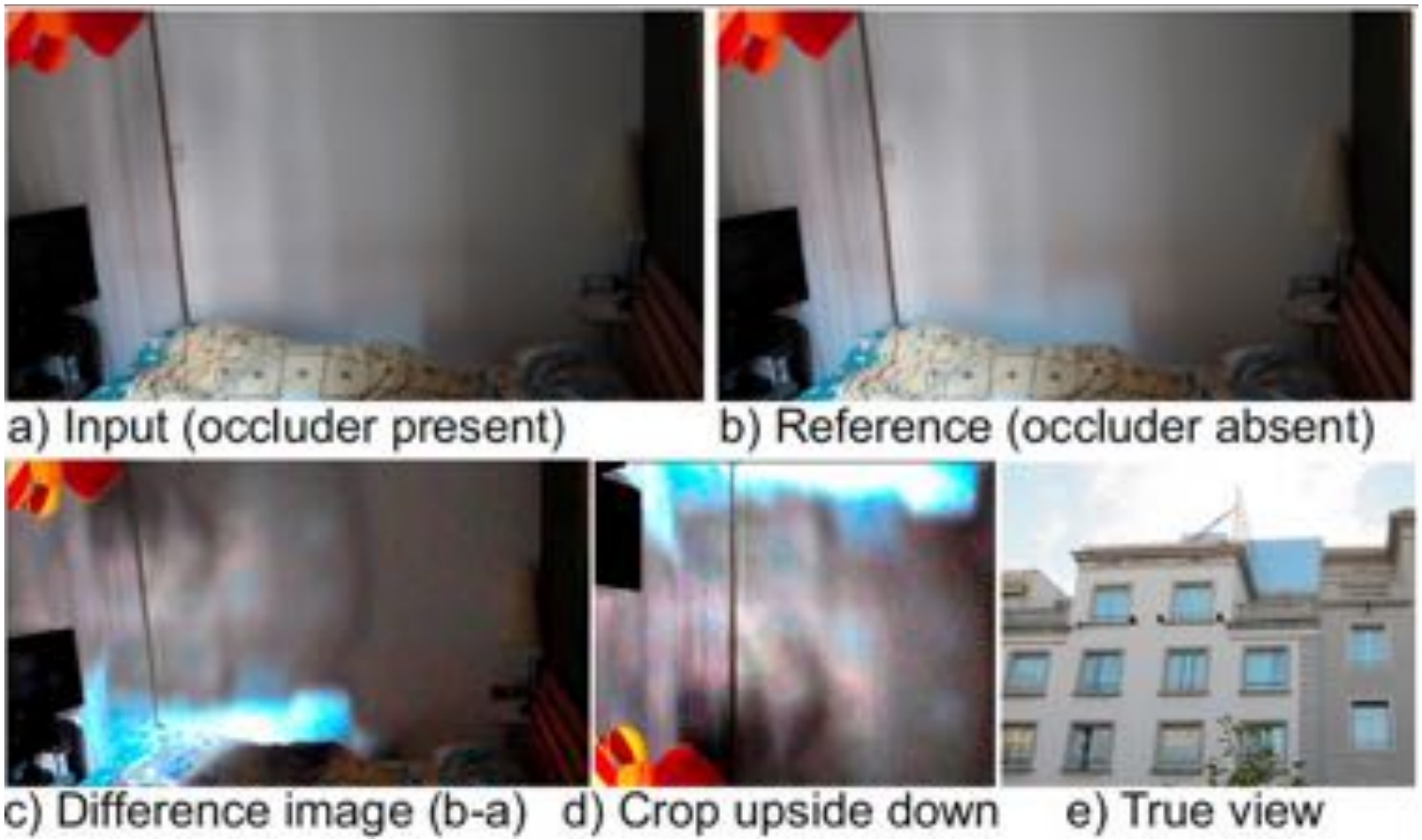


- Can you see something on the wall?

Torralba & Freeman, CVPR'12

Why study Computer Vision?

- Some more CSI



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

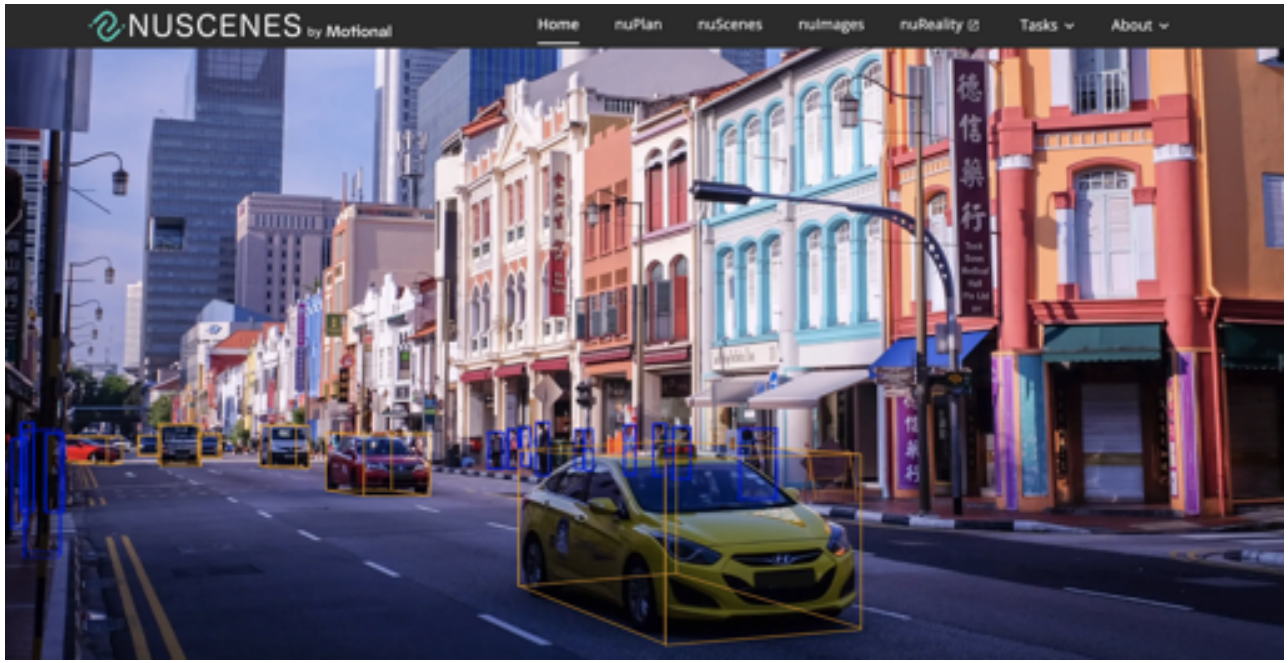
Seymour Papert

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: KITTI, Waymo Open, nuScenes, ImageNet, PASCAL, Cityscapes, MS-COCO

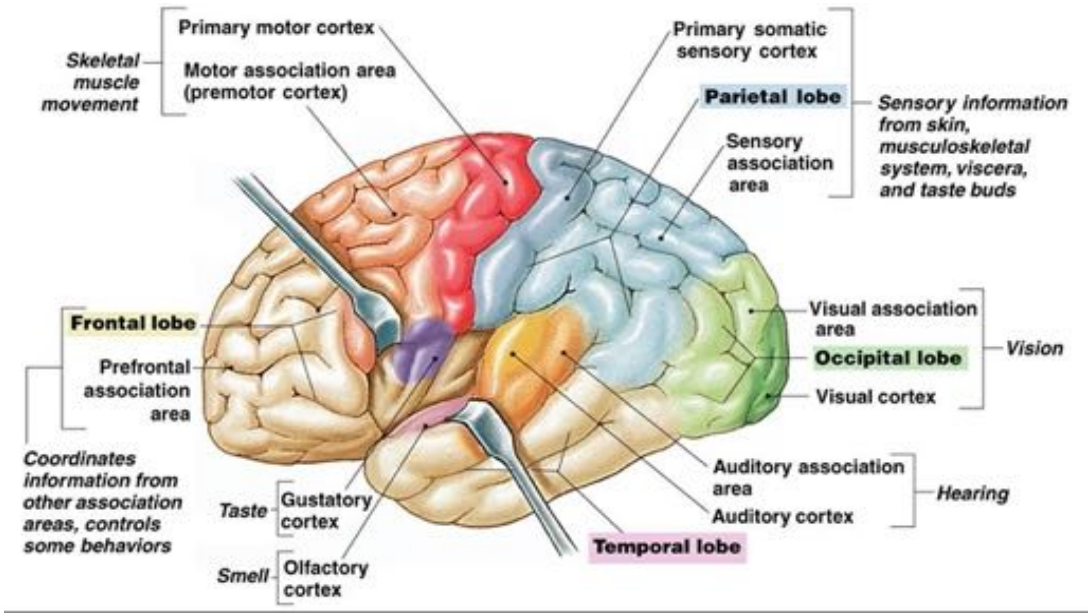


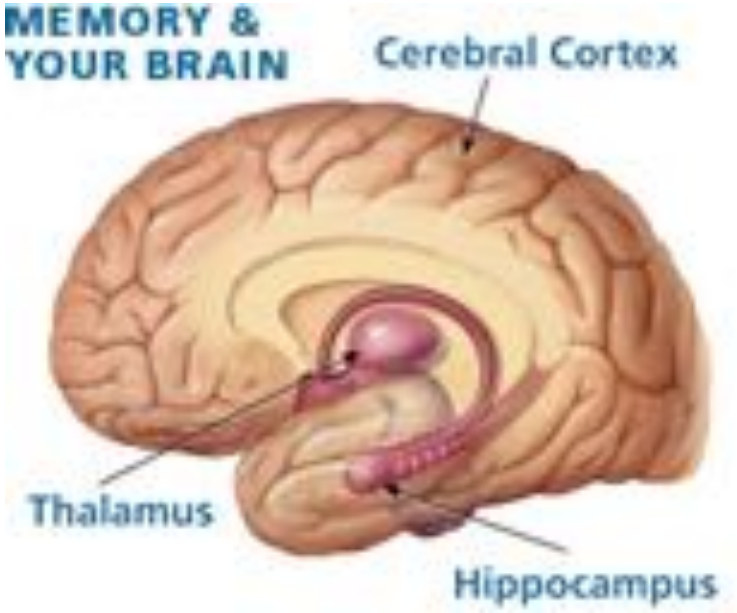
Method					Metrics									
Date	Name	Modalities	Map data	External data	mAP	mATE (m)	mASE (1-IOU)	mAOE (rad)	mAVE (m/s)	mAAE (1-acc)	NDS	PKL *	FPS (Hz)	Stats
		Any	All	All										
2023-08-22	EA-LSS	Camera, Lidar	no	no	0.766	0.234	0.228	0.278	0.204	0.124	0.776	0.505	n/a	
2023-03-29	IEI-BEVFusion++	Camera, Lidar	no	no	0.757	0.236	0.235	0.283	0.143	0.126	0.776	0.535	n/a	
2023-03-25	BEVFusion4D-e	Camera, Lidar	no	no	0.768	0.229	0.229	0.302	0.225	0.135	0.772	0.506	n/a	
2022-11-21	MMFusion-e	Camera, Lidar, Rad.	no	no	0.750	0.220	0.218	0.278	0.192	0.132	0.771	0.512	n/a	
2022-10-17	MegFusion	Camera, Lidar	no	no	0.753	0.233	0.220	0.271	0.212	0.127	0.770	0.516	n/a	

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?

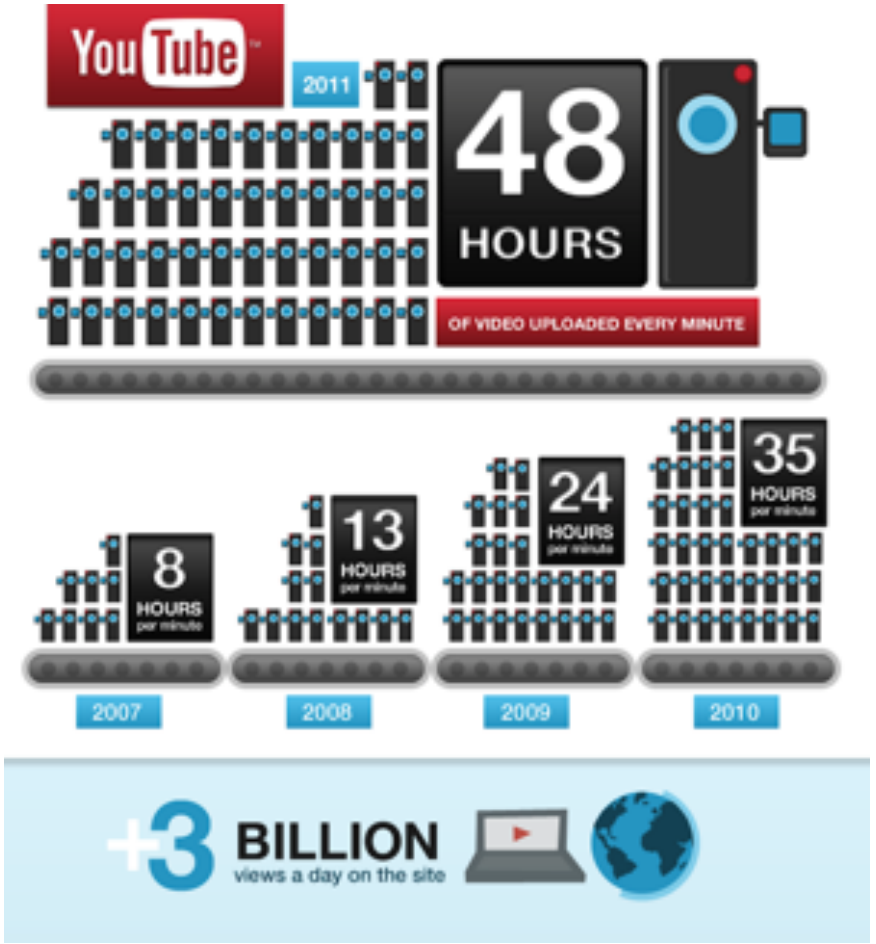
Biederman, 1987

[slide credit: R. Urtasun]

Why is vision hard?

Lots of data to process:

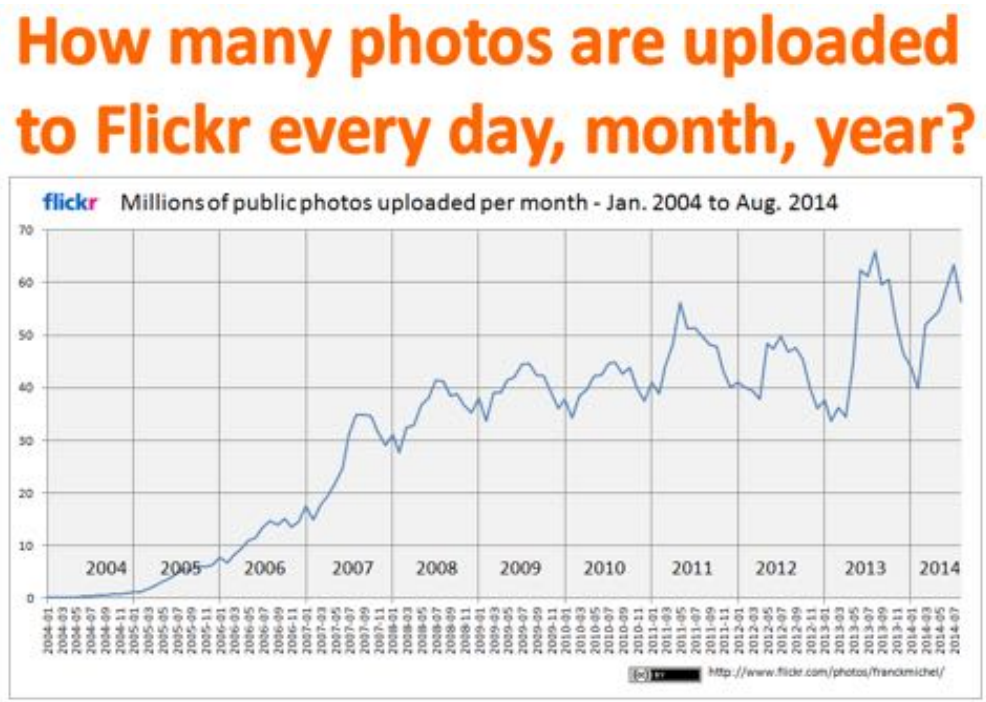
- Thousands to millions of pixels in an image
- 100 hours of video added to YouTube per minute [source: YouTube]
- Over 6 billion hours of video are watched each month on YouTube – almost an hour for every person on Earth [source: YouTube]

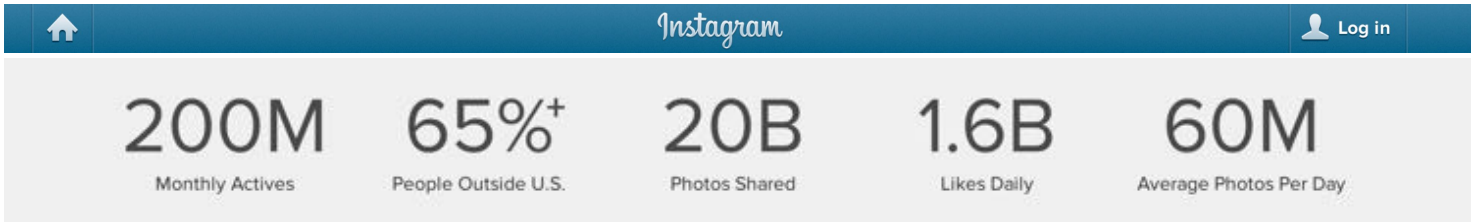


Why is vision hard?

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!

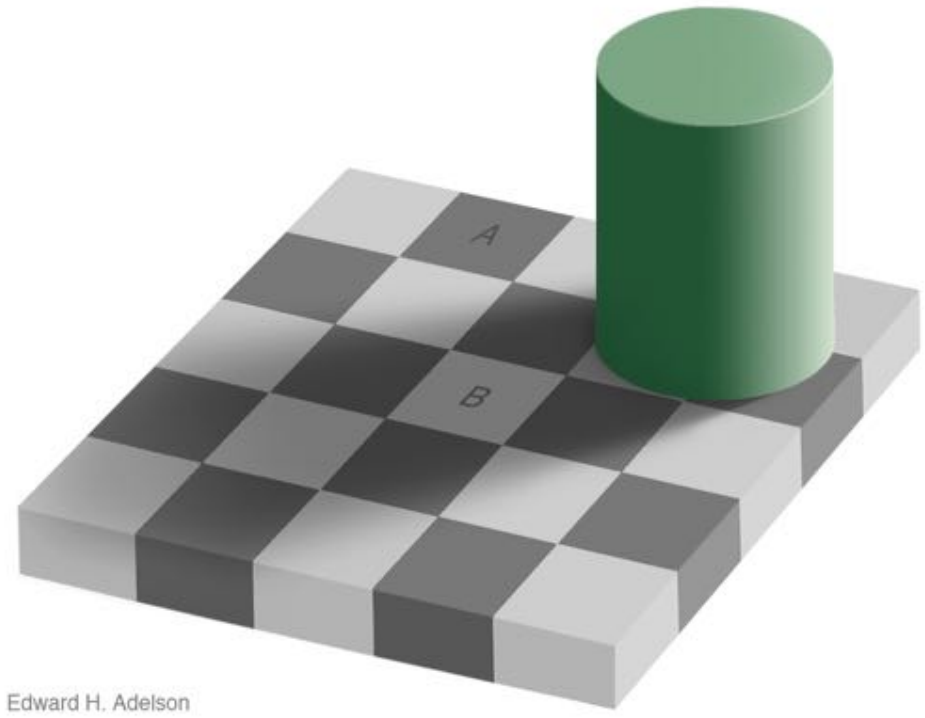
Figure: Vemodalen: The Fear That Everything Has Already Been Done, https://www.youtube.com/watch?v=8ftDjebw8aA

[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!

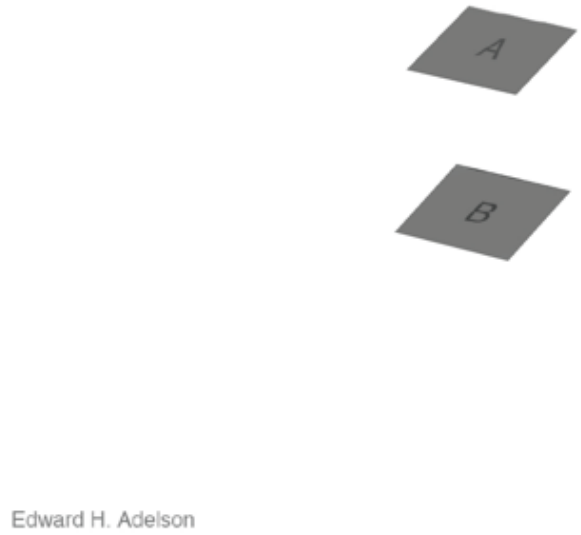
How good are humans?



- 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?

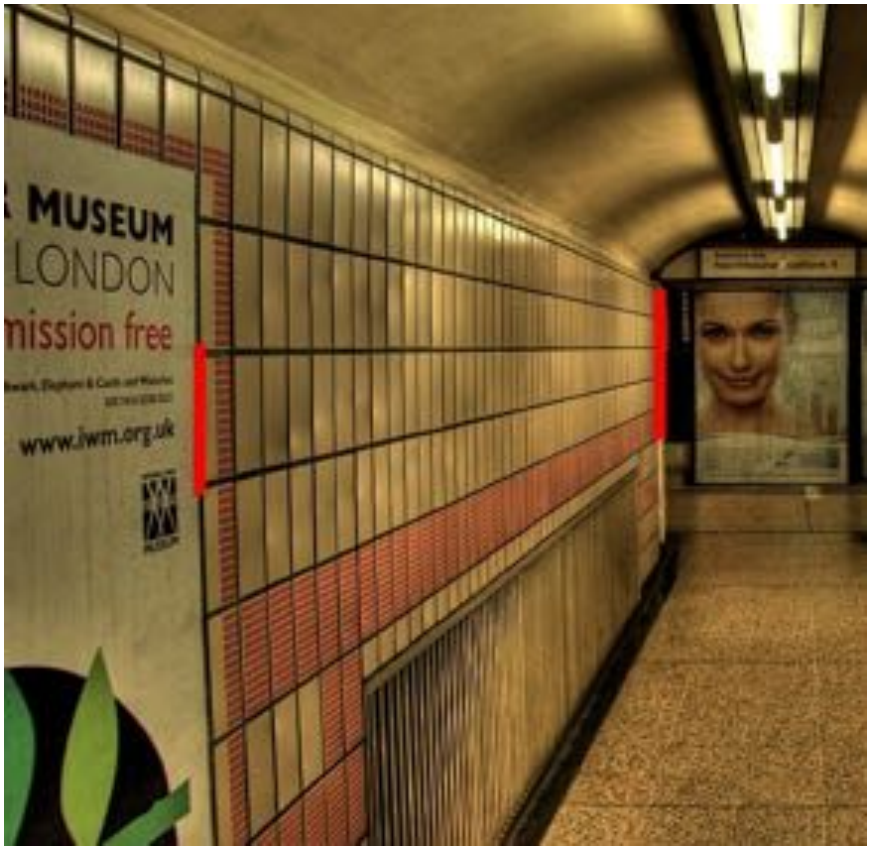


Figure: 2006 Walt Anthony

- 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!

https://www.youtube.com/watch?v=vJG698U2Mvo

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 mathematical foundations
- Good programming skills
- Crativity
- Good intuition (can be obtained with experience)
- Persistence