Lecture 1:
Introduction to Computer Vision

Laura Sevilla-Lara

Who am I?
Dr. Laura Sevilla

- Reader (Associate Professor) in Computer Vision
- Doing research for over 15 years.


BROWN

University of Massachusetts Amherst

Meta

Masters 2009 **PhD 2014** **Researcher 2014-2017** **Researcher 2017-2019** **2019 - Present**

- Understanding the world in motion (video understanding, video and language, generation, etc).
- Always looking for excellent PhD students
- laurasevilla.me

Teaching Team

Laura Sevilla

Reader (Associate Professor)
Computer Vision

Eleonora D’Arnese

Lecturer (Assistant Professor)
Biomedical Science

Changjian Li

Lecturer (Assistant Professor)
Computer Graphics

+ TAs (help in Piazza), Lab Tutors (help with coursework)

Plan for today

- What is Computer Vision
- Why is it hard
- How have people tried to solve it in the past
- What will we learn in this course
- Admin

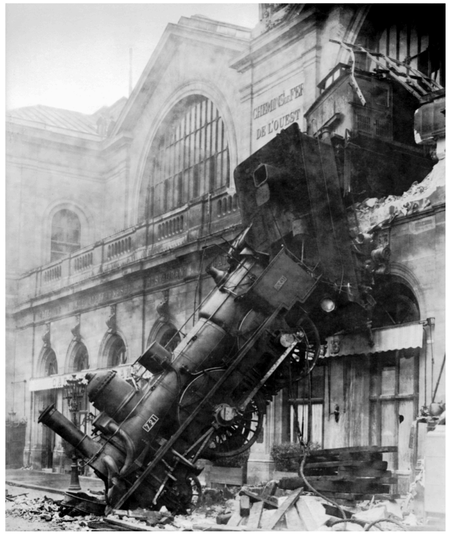
Plan for today

- What is Computer Vision
- Why is it hard
- How have people tried to solve it in the past
- What will we learn in this course
- Admin

What is Computer Vision?

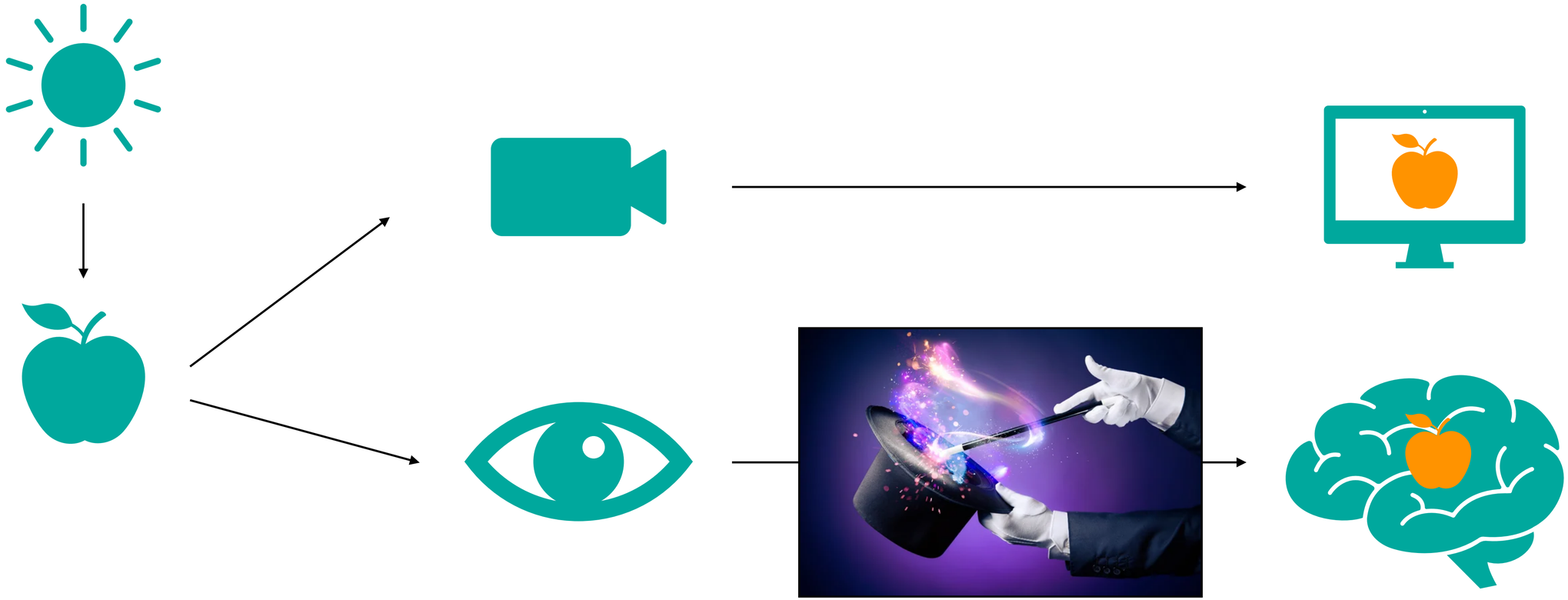
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What the computer “sees”



What you “see”

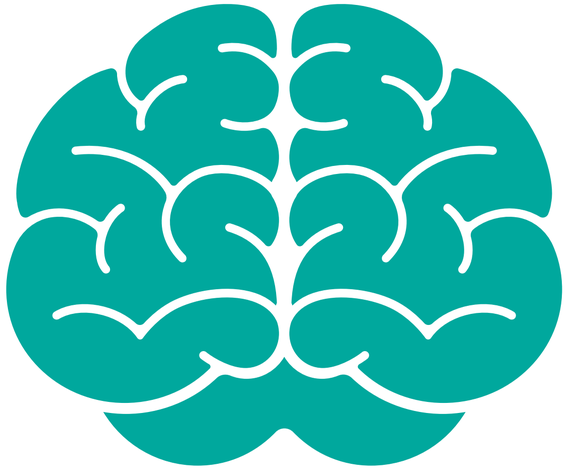
What is ~~Computer~~ Vision?



It seems trivial!

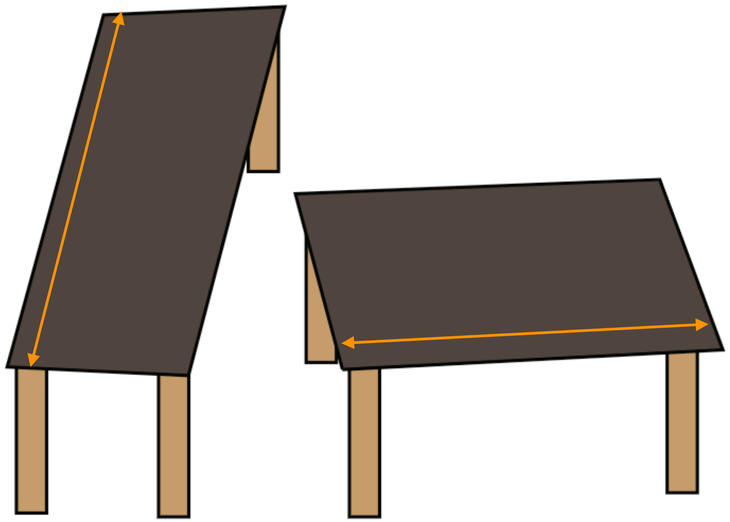
Humans do it “effortlessly”

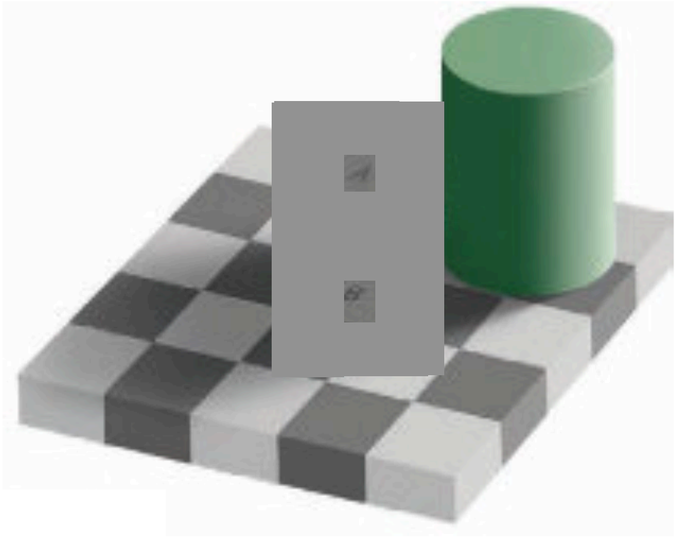
2/3s of electrical activity is dedicated to vision. That is 2 of 3 billion firings per second



30-50% of cortex dedicated to vision (compared to 8% touch or 3% hearing)

Some Optical Shortcuts

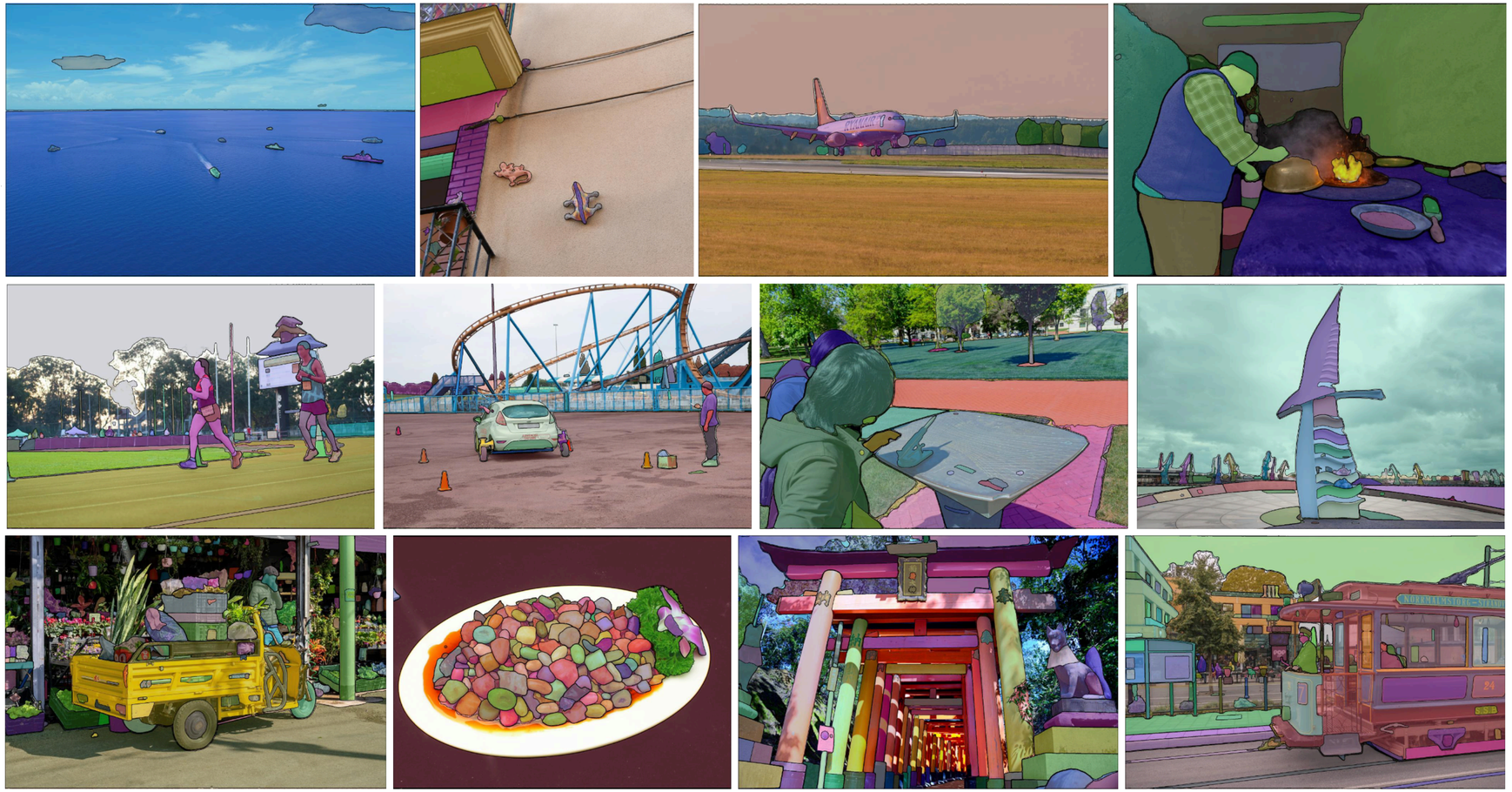




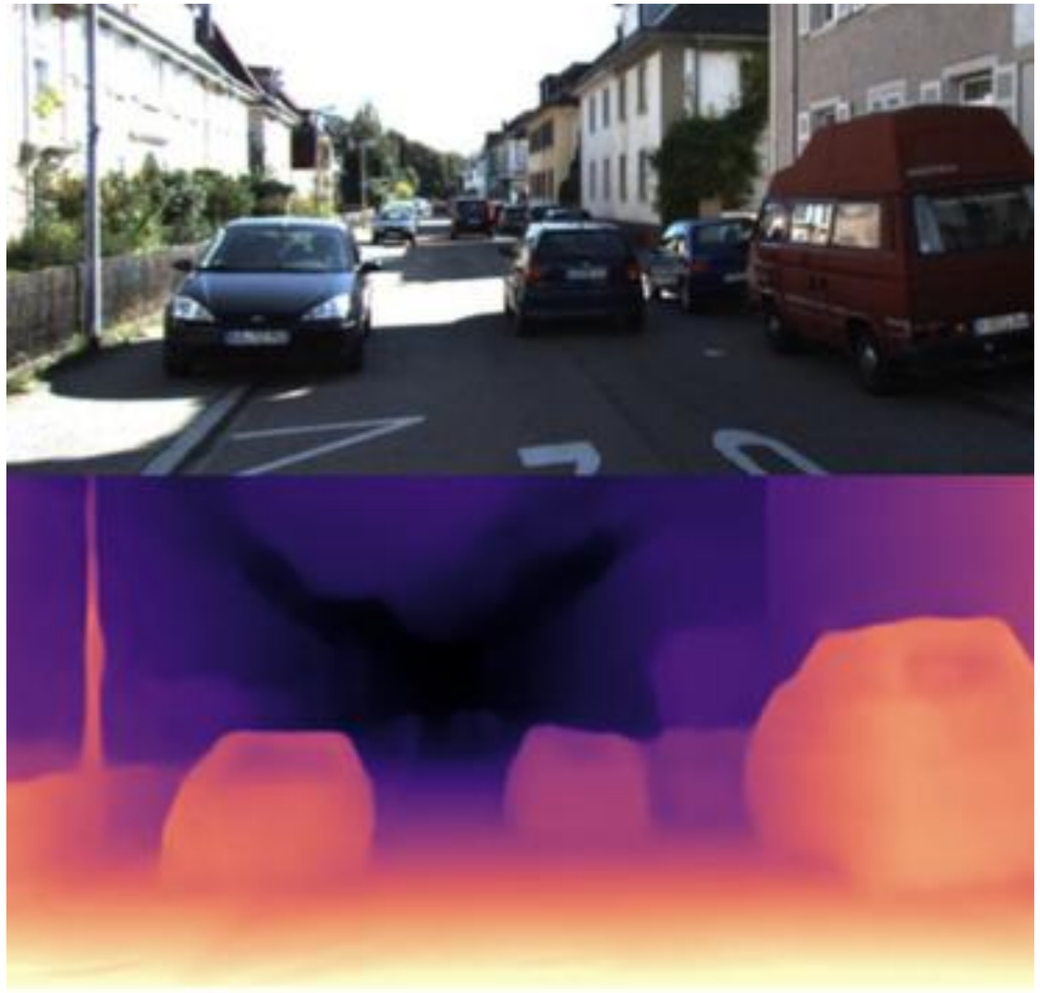
Some Computer Vision Problems

Recognition

Segmentation



Depth

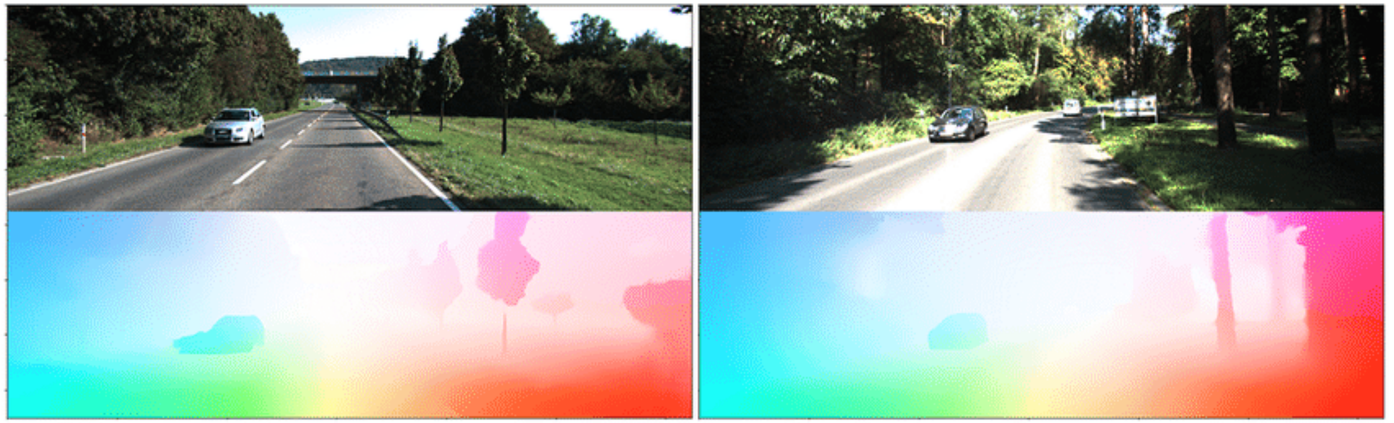


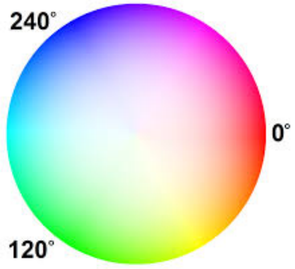
Optical Flow

MPI-Sintel

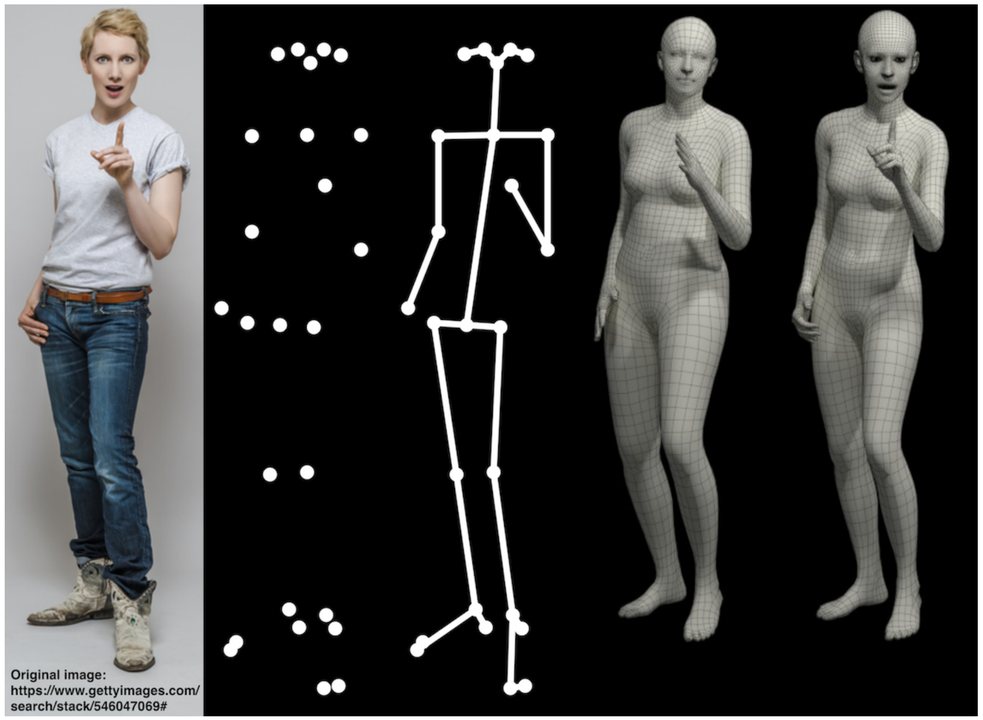
(a) (b) (c) (d)

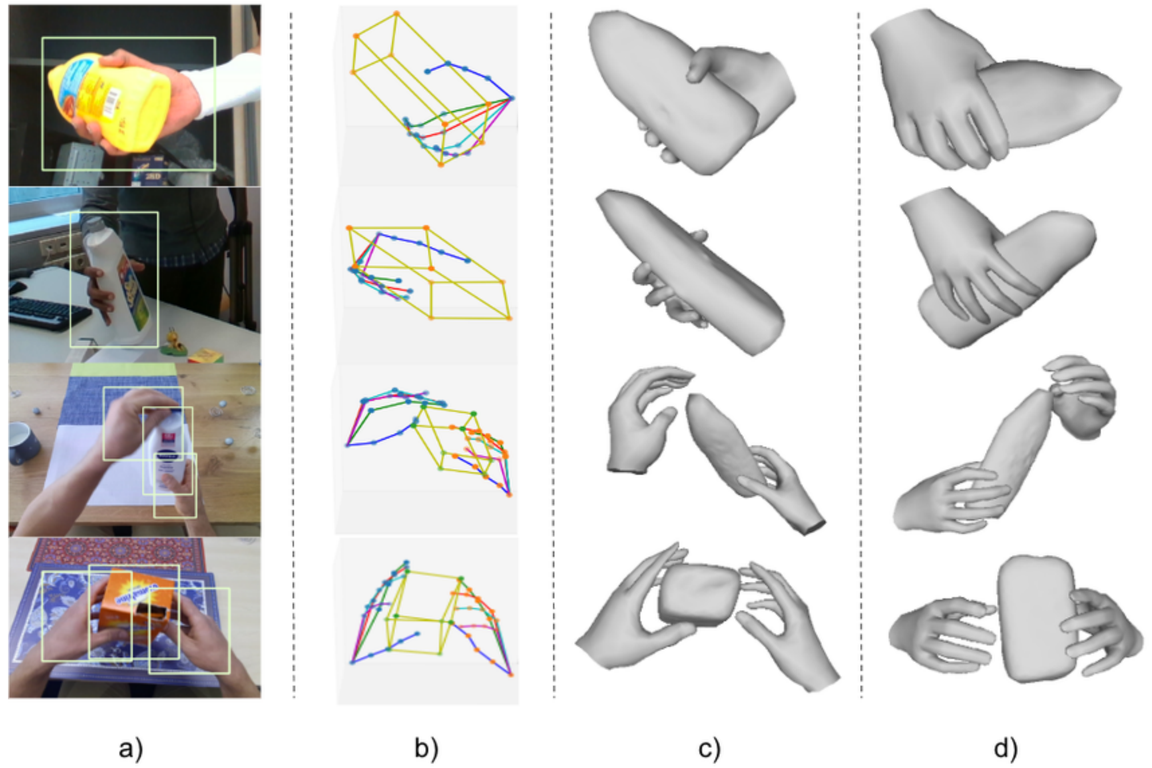
KITTI 2015





3D estimation



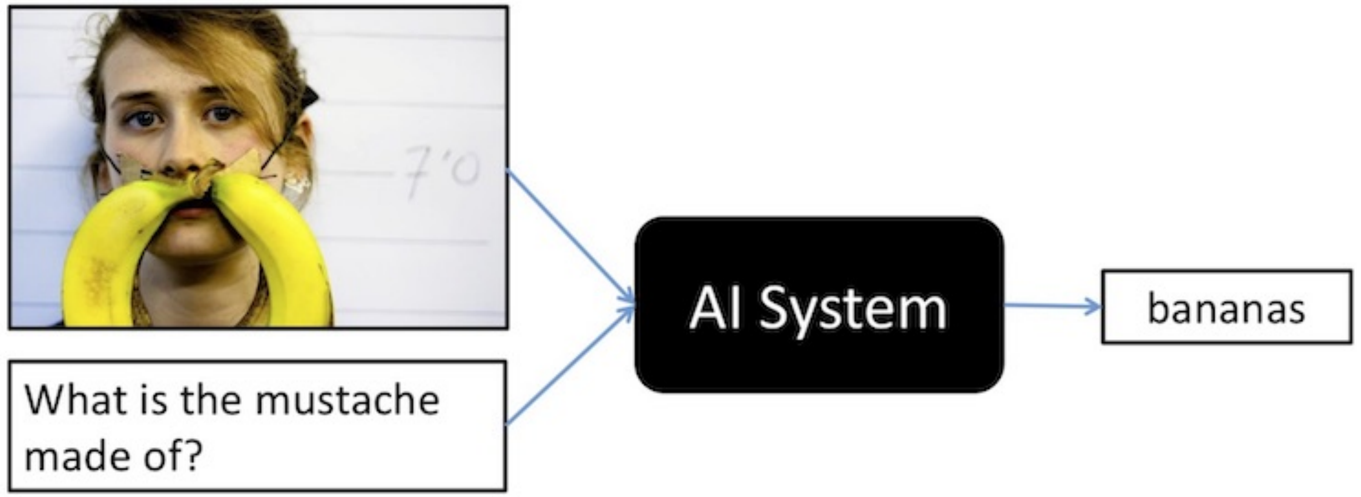


Captioning

SM (ours): This image shows an inviting dining space with plenty of natural light.

SM (ours): People gather under a blossoming cherry tree, enjoying the beauty of nature together.

Visual Q&A



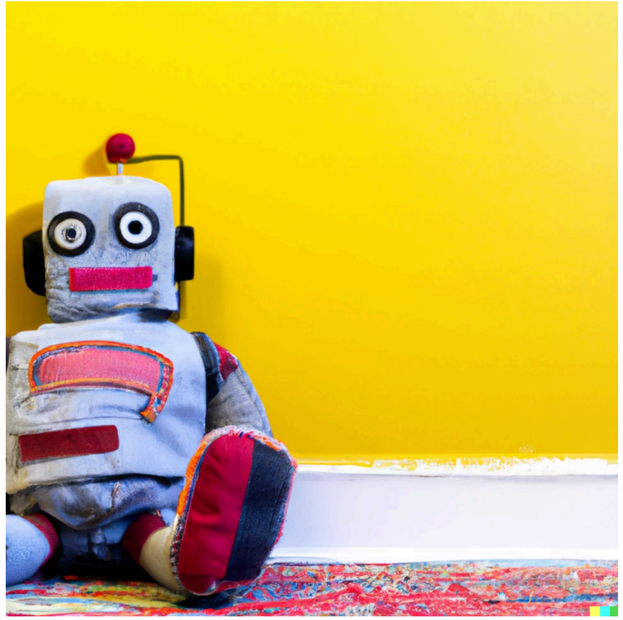
Multi-modal analysis

original video
(before separation)

visual predictions:
acoustic guitar & harmonica

Generation

3D render of a cute tropical fish in an aquarium on a dark blue background, digital art



A plush toy robot sitting against a yellow wall

Plan for today

- What is Computer Vision
- Why is it hard
- How have people tried to solve it in the past
- What will we learn in this course
- Admin

Why is Computer Vision hard?

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

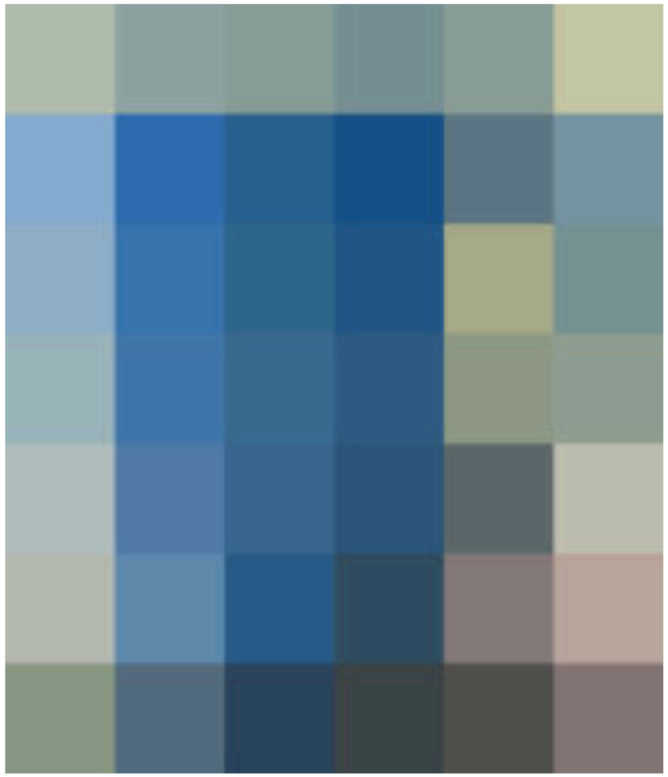
THE SUMMER VISION PROJECT

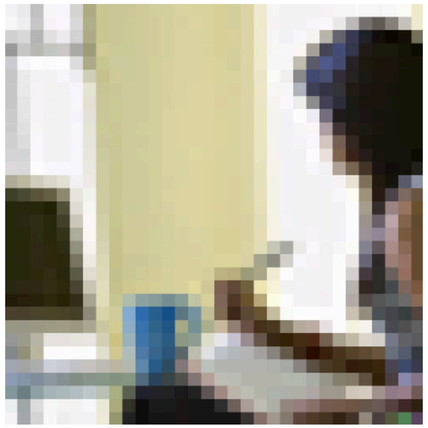
Seymour Papert

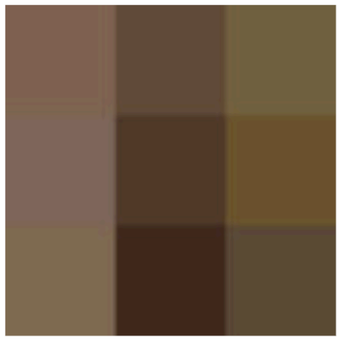
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

[Credit: Fei-Fei Li]

Context







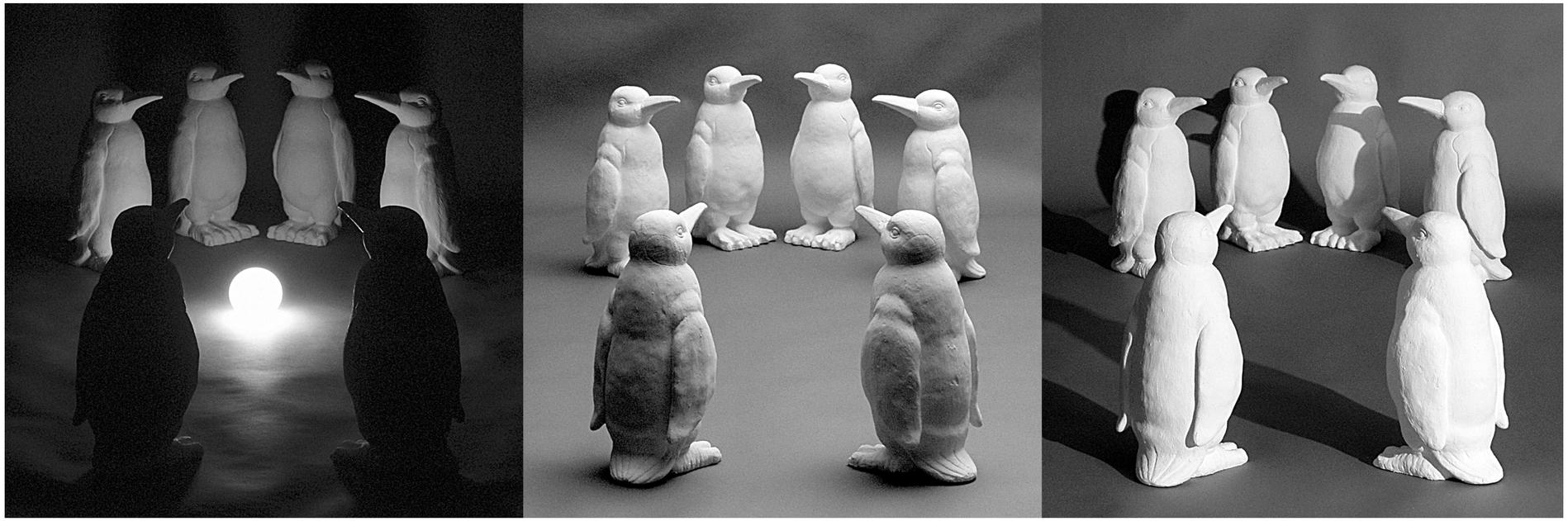
Viewpoint Variation

Michelangelo 1475-1564

[Credit: Fei-Fei, Fergus & Torralba]

Non-rigid Deformation

Illumination



[Image credit: J. Koenderink]

Clutter

[Credit: V. Ferrari]

Occlusion

[Credit: S. Lazebnik, National Geographic]

Motion

[Credit: S. Lazebnik]

Intra-class variation

[Slide credit: Fei-Fei, Fergus & Torralba]

Scale

Humans Recognise: ~**30K** basic categories

Millions of “Fine-Grained” categories

Ambiguity

- Many different 3D scenes can give rise to a particular 2D image

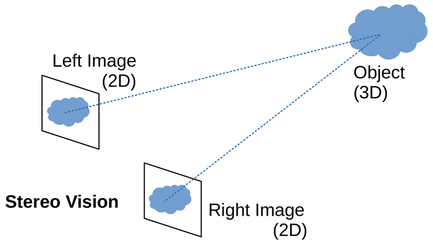
Plan for today

- What is Computer Vision
- Why is it hard
- How have people tried to solve it in the past
- What will we learn in this course
- Admin

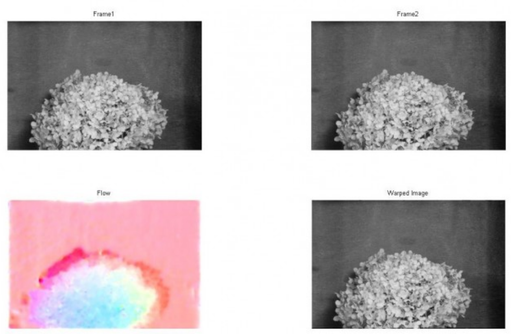
A Little Bit of History

Classic Vision (from 60s to 2012)

Low-level problems

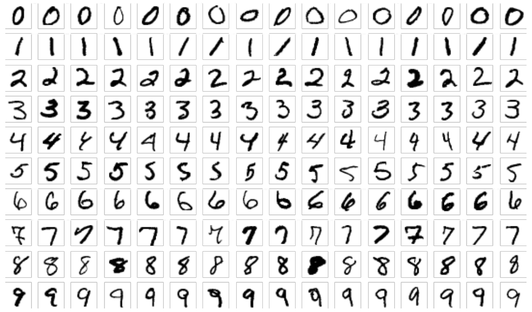


Stereo



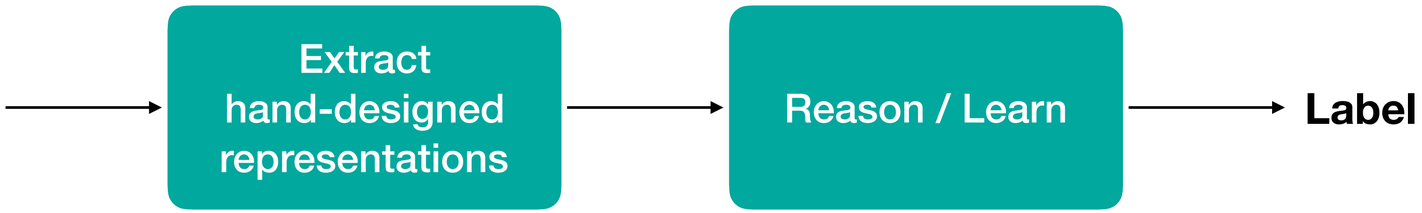
Flow

Simple high-level problems



A Little Bit of History

Classic Vision
(from 60s to 2012)



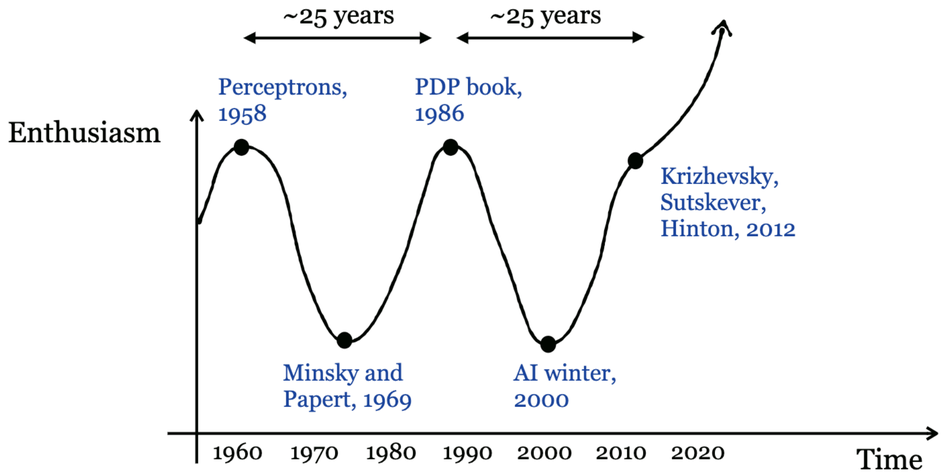
A Little Bit of History

“Learning” era, CNNs
(2012-2017)

+

Faster compute

AlexNet (Krizhevsky, 2012)



A little bit of History

“Learning” era, Transformers (2017-now)

Attention is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

→ **Language** → **Vision**

Plan for today

- What is Computer Vision
- Why is it hard
- How have people tried to solve it in the past
- What will we learn in this course
- Admin

Goals if this course

- Learn principles of computer vision
- Be able to reason about what works and why
- Be able to design some computer vision models
- Not: about tweaking parameters without understanding. From the book:

> “ Computer vision is not just an engineering discipline that tries to build systems that see. Instead computer vision is another aspect of the interdisciplinary scientific quest that has focused on understanding how natural intelligence and perception works. ”

Plan for today

- What is Computer Vision
- Why is it hard
- How have people tried to solve it in the past
- What will we learn in this course
- Admin

Schedule

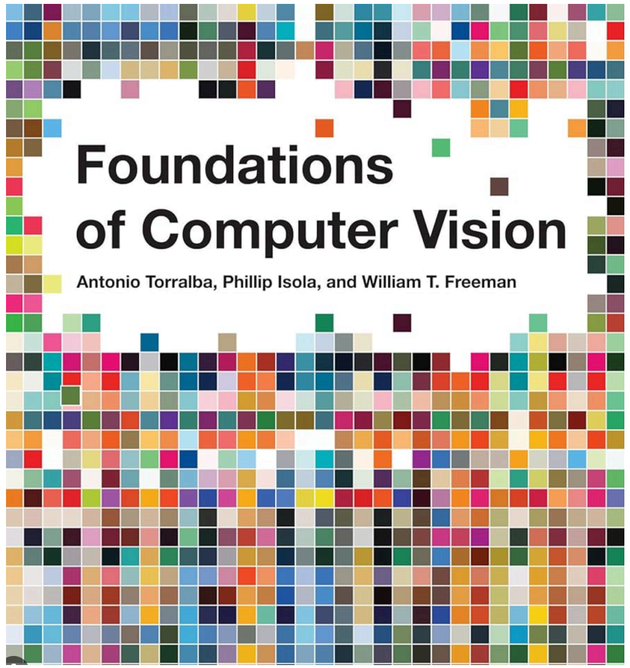
Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11
Intro	Image Processing	Classic Vision	Deep Learning (I)	Deep Learning (II)	Generative Models	Vision & Language	3D	Video	Ethics	Limited Data
Image Formation		Tutorial		Tutorial		Tutorial		Tutorial		Review
			Lab1		Lab2		Lab3		Lab4	

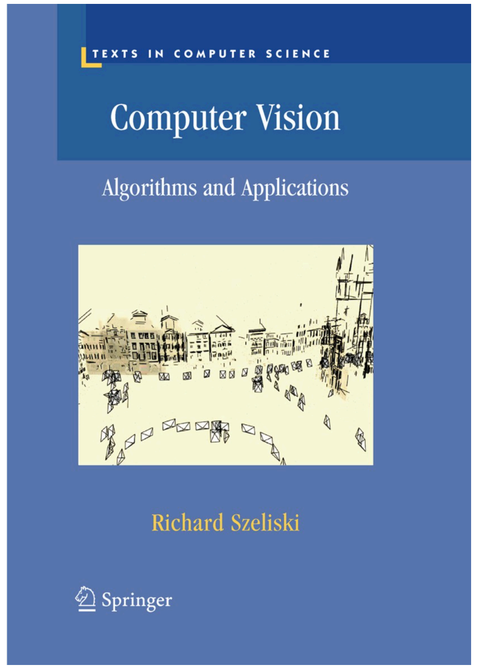
Marks

- Theoretical Content (Exam) : 50%
- Practical Content (Mini-project): 50%

- Average mark tends to be around 60-65%, according to the Common Marking Scheme of the University.

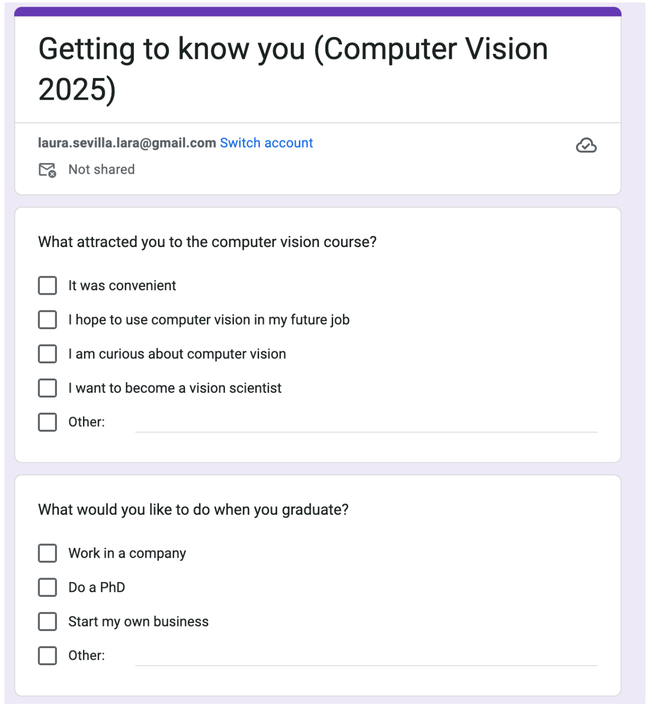
Books


Foundations
of Computer Vision
Antonio Torralba, Phillip Isola, and William T. Freeman


TEXTS IN COMPUTER SCIENCE
Computer Vision
Algorithms and Applications
Richard Szeliski
Springer

Getting to Know You

Please fill in this ANONYMOUS form:



- https://forms.gle/Rp3zt6JKAiuk394t8